



HAL
open science

Évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Antonello de Galizia

► **To cite this version:**

Antonello de Galizia. Évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel. Automatique / Robotique. Université de Lorraine, 2017. Français. NNT : 2017LORR0018 . tel-01637209

HAL Id: tel-01637209

<https://hal.univ-lorraine.fr/tel-01637209>

Submitted on 17 Nov 2017

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.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



Automatique



Ecole Doctorale IAEM Lorraine - Département de Formation Doctorale en

Thèse

Présentée pour l'obtention du grade de

Docteur de l'Université de Lorraine

Spécialité Automatique, Traitement du Signal et des Images, Génie Informatique

par

Antonello De Galizia

Évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Soutenue publiquement le 28 Février 2017 à l'Université de Lorraine devant le jury composé de :

Rapporteurs	Frédéric VANDERHAEGEN	Professeur à l'UVHC de Valenciennes
	Eric ZAMAI	Maître de Conférences à l'INP de Grenoble
Examineurs	Benoit IUNG	Professeur à l'Université de Lorraine, directeur de thèse
	Philippe WEBER	Professeur à l'Université de Lorraine, co-directeur de thèse
	Christophe SIMON	Maître de Conférences à l'Université de Lorraine, co-encadrant de la thèse
	Carole DUVAL	Ingénieur de recherche expert chez EDF R&D, co-encadrant de la thèse
	Jean-Yves DANTAN	Professeur à l'Ecole Nationale Supérieure d'Arts et Métiers Paristech
	Enrico ZIO	Professeur à l'Ecole Polytechnique de Milan et à École Centrale-Supélec, Paris
Invités	Dounia TAZI	Responsable du pôle expertise et accompagnement de l'ICSI
	Emmanuel SERDET	Ingénieur de recherche chez EDF R&D, co-encadrant de la thèse

A mes parents
A mes frères
A ma chère Vera

Remerciements

La thèse de doctorat représente un travail s'inscrivant dans la durée, et pour cette raison, constitue le fil conducteur d'une tranche de vie de son auteur. Parfois au crépuscule de la candeur étudiante, et souvent à l'aube de la maturité scientifique, elle constitue une aventure collective dont l'aboutissement est possible grâce à tous ceux qui, de près ou de loin, y ont contribué. De nombreuses personnes se retrouvent ainsi, de manière fortuite ou non, entre le doctorant et son doctorat. Ce sont certaines de ces personnes que j'aimerais mettre en avant dans ces remerciements.

Je tiens à exprimer mes plus vifs remerciements à mon directeur, Monsieur Benoît Iung, pour son intérêt et son soutien, sa grande disponibilité et ses nombreux conseils durant la durée de la thèse, jusqu'à la phase finale de la soutenance. Sa compétence, sa rigueur scientifique et sa clairvoyance m'ont permis de prendre confiance en moi et d'assumer mes choix. Elles ont été et resteront des moteurs de mon travail.

Je souhaite également remercier Messieurs Philippe Weber et Christophe Simon, qui ont co-encadré la thèse. Les nombreux échanges que nous avons eus sur les différentes problématiques liées à la résilience et aux distributions de probabilité dans les réseaux bayésiens, leurs disponibilités et leurs aides dans les moments difficiles m'ont beaucoup aidé dans mon apprentissage du métier de chercheur.

Un grand merci à Madame Carole Duval, pour sa capacité à mettre en musique les paroles de ma composition scientifique ! Elle a non seulement été une très bonne encadrante dans le quotidien, et grâce à son implication et sa confiance elle a su me motiver pour faire de ce travail de thèse un parcours de valorisation de mes capacités. Je remercie Monsieur Emmanuel Serdet pour son soutien et pour les discussions que nous avons eues sur les facteurs humains et pour son expertise dans ce domaine.

Je remercie également Monsieur Jean-Yves Dantan d'avoir accepté de présider mon jury de thèse. Je suis particulièrement reconnaissant à Messieurs Frédéric Vanderhaegen et Eric Zamai pour leurs remarques constructives en tant que rapporteurs de ce manuscrit.

Enfin, j'exprime sincèrement tous mes remerciements à Monsieur Enrico Zio, qui a été un professeur formidable pendant ma dernière année à Milan et qui a été toujours présent dans la suite. Il a encadré mon stage de fin d'études à MRI et ensuite il m'a encouragé à poursuivre en thèse. Ce travail n'aurait pu être mené à bien sans sa présence et son support, non seulement scientifique mais aussi humain, durant ces années à Paris. Je le remercie enfin pour l'examen de mon travail et pour sa présence chaleureuse (et son esprit !) lors de la soutenance.

Ce travail a été possible également grâce à la disponibilité et l'accueil chaleureux que m'ont témoigné tous les personnes rencontrées sur mon chemin du début à la fin de la thèse.

Je remercie toutes les personnes formidables du département MRI d'EDF R&D que j'ai rencontrées entre le stage et la thèse. Merci pour votre support et vos encouragements. Je pense particulièrement à Monsieur Philippe Nonclercq pour m'avoir accueilli dans l'équipe T-51 dédiée aux « Etudes probabilistes de sûreté et disponibilité des systèmes ». Je remercie également Roland, Hassane, Mohammed, Richard, Valentin, Gilles, Dominique, Marie, Anne-Marie, Pascal, Alan, Christelle, Tu-Duong, Zakarya et tous les autres agents des équipes T-51 et T-54 avec qui j'ai pu converser et qui m'ont permis de découvrir la diversité des autres disciplines à MRI. Merci aux doctorants et aux stagiaires, notamment Richard et Claudia, qui ont partagé le bureau avec moi. Ils ont été d'extraordinaires copains de voyage dans les moments agréables et aussi ceux difficiles durant cette thèse.

Je remercie également tous les doctorants, maîtres de conférences et professeurs du CRAN, notamment ceux du département ISET au 6^{ème} étage, pour le climat sympathique dans lequel ils m'ont permis de travailler lors de mes venues à Nancy. Les nombreuses conversations que j'ai pu avoir avec eux m'ont beaucoup apporté. Je voudrais exprimer particulièrement toute mon amitié à Mario Lezoche pour sa gentillesse et son soutien durant la durée de la thèse.

Je ne saurais terminer sans remercier toutes ces personnes dans l'ombre dont la contribution à mon travail est non négligeable. Un grand merci aussi à tous les professionnels et les scientifiques rencontrés dans les conférences à travers le monde et dont nos échanges ont enrichi l'expérience de la thèse.

Au terme de ce parcours, je remercie enfin celles et ceux qui me sont chers. Leurs attentions et encouragements m'ont accompagné tout au long de ces années.

Je tiens à exprimer toute mon affection à mes deux meilleurs amis, Marco et Silvia, pour l'intérêt qu'ils ont porté à cette thèse depuis l'Italie.

Je ne serais pas là sans le soutien de ma famille. Malgré mon éloignement depuis plusieurs années, leur intelligence, leur confiance, leur tendresse, leur amour me portent et me guident tous les jours. Merci à mes parents, Dario et Marina, pour m'avoir accompagné dans mes choix. Merci à mes frères, Alessio et Simone, qui sont toujours là quand j'en ai besoin, avec leurs blagues et leurs esprits légers. Je vous aime tous.

Et maintenant, les mots les plus simples étant les plus forts, j'adresse mon dernier remerciement à Vera. C'est toi qui m'as supporté et m'as permis de me lever motivé, le cœur plein des beaux moments passés ensemble durant cette nouvelle vie parisienne. L'esprit vif qui t'appartient m'a nourri depuis le début de ma thèse. Très humblement, je voudrais te dire merci Vera pour ton soutien quotidien pendant mes périodes de doutes et pour tes multiples encouragements répétés. Cette thèse a été aussi notre aventure. Merci d'être à mes côtés à chaque instant.

Table des matières

Chapitre 1 De la hiérarchisation des actions humaines a une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

1 De la hiérarchisation des actions humaines a une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Introduction

1.2 Le management des risques à EDF

1.2.1 Considération « système » des facteurs humains et organisationnels dans les approches multirisques à EDF

1.3 L'analyse intégrée des risques (AiDR)

1.3.1 Intérêt pour les réseaux Bayésiens dans l'AiDR

1.3.2 Limites actuelles de l'AiDR et besoins d'EDF

1.3.3 Expression détaillée des verrous techniques pour l'évolution de l'AiDR

1.3.3.1 Renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel

1.3.3.2 Intégration de la résilience dans l'AiDR

1.3.3.3 Traitement des avis d'experts et estimation des paramètres du modèle probabiliste intégrant la résilience

1.4 Conclusions

Chapitre 2 Des verrous techniques aux verrous scientifiques pour une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

2 Des verrous techniques aux verrous scientifiques pour une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Introduction

2.2 Renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel

2.2.1 Présentation des méthodes d'évaluation probabiliste de la fiabilité humaine

Critère n° 1 – Cadre conceptuel

Critère n° 2 – Approche quantitative

Critère n° 3 – Couverture des PSF

2.2.2 Extrapolation des PSF

2.3 Intégration de la résilience dans l'AiDR

2.3.1 Elaboration des marqueurs des mécanismes résilients

2.3.2 Interaction entre mécanismes résilients et pathogènes dans le cadre probabiliste

2.3.3 Les modèles canoniques probabilistes

2.3.3.1 Les Modèles Canoniques Déterministes (MCD)

2.3.3.2 Les Modèles Canoniques Simples (MCS)

2.3.3.3 Les Modèles Canoniques « ICI »

2.3.4 Le modèle « Leaky » Noisy-OR

2.4 Conclusions

Chapitre 3 Contributions à l'évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

3 Contributions à l'évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Introduction

3.1 Intégration de la résilience dans l'AiDR

- 3.1.1 Élaboration des marqueurs résilients
- 3.1.2 Formalisation de l'interaction entre mécanismes résilients et pathogènes
- 3.1.3 Application de l'interaction entre mécanismes résilients et pathogènes dans les RBs

3.2 Proposition d'un nouveau modèle de barrière humaine

- 3.2.1 Principes génériques de construction du modèle de barrière humaine
 - 3.2.1.1 *Classes de variables*
 - 3.2.1.1 *Espace des possibles*
- 3.2.2 Application des principes génériques pour la construction d'un méta-modèle de barrière humaine dans le cadre de l'AiDR
 - 3.2.1.1 *Niveau I : Segmentation de l'action en trois phases*
 - 3.2.1.1 *Niveau II : Formalisation des PSF pour leur intégration dans la structure du modèle de barrière humaine de l'AiDR*
 - 3.2.1.3 *Niveau III : Généralisation des Facteurs Organisationnels pour la considération des mécanismes résilients*
 - 3.2.1.3 *Niveau III : Adaptation des liens de causalité entre Facteurs Organisationnels et déterminants pour intégrer les mécanismes résilients*

3.3 Principes d'instanciation du méta-modèle de barrière humaine de l'AiDR pour créer un modèle particulier

- 3.3.1 Définition d'un protocole de traitement des avis d'expert
- 3.3.2 Définition d'une approche d'estimation des paramètres pour le modèle intégrant l'interaction entre mécanismes résilients et pathogènes

- 3.3.2.2 *Approche par criblage et création d'arbres décisionnels*
- 3.3.2.1 *Optimisation mathématique*

3.4 Conclusions

Chapitre 4 Validation des propositions à un cas réel d'EDF

4 Validation des propositions à un cas réel d'EDF

Conclusions et perspectives

Conclusions et perspectives

Annexes

I Les méthodes EPFH analysées

Liste des acronymes

Ai	Aides
AiDR	Analyse intégrée Des Risques
ATHEANA	A Technique for Human Event Anal
Cao	Contrôle et atteinte des objectifs
CP	Culture de Production
CQE	Critère de Qualité de l'Estimation
CRAN	Centre de Recherche en Automatique de Nancy
CREAM	Cognitive Reliability and Error Analysis Method
COS	Culture Organisationnelle de Sûreté
CS	Culture de Sûreté
Ctm	Conception du processus de travail, attribution des tâches et management
De	Délégation
EDF	Électricité de France
EPFH	Évaluation Probabiliste de la Fiabilité Humaine
EPS	Études Probabiliste de Sûreté
ER	État de Référence
Ex	Expérience
Fc	Facteurs Contextuels
FH	Facteur Human
Fo	Formation
FO	Facteur Organisationnel
FOP	Facteur Organisationnel Pathogène
Gcdg	Gestion collective et dynamique de groupe
GQS	Gestion Quotidienne de la Sûreté
HEART	Human Error Assessment and Reduction Technique
HEP	Human Error Probability
HRC	Human Cognitive Reliability
ICI	Influences Causales Indépendantes
INERIS	Institut National pour l'Environnement et les Risques
MCD	Modèles Canoniques Déterministes
MCS	Modèles Canoniques Simples
MERMOS	Méthode d'Évaluation de la Réalisation des Missions Opérateur pour la sûreté
MRI	Management des Risques
MT	Mauvais Traitement de la Complexité Organisationnelle
NARA	Nuclear Action Reliability Assessment
NRC	Nuclear Regulatory Commission
OC	Organismes de Contrôle
PSF	Performing Shaping Factors

RB	Réseau Bayésien
REX	Retour d'Expérience (FOP)
Rex	Retour d'expérience (Déterminant)
RHC	Réexamen des Hypothèses de Conception
SAM	System-Action-Management
SLIM-MAUD	Success Likelihood Index Method using Multi-Attribute Utility Decomposition
SPAR-H	Standardized Plant Analysis Risk - Human Reliability Assessment
TCO	Traitement de la Complexité Organisationnelle
THERP	Technique for Human Error Rate Prediction
TPC	Table de Probabilités Conditionnelles

Introduction générale

L'approche dite 'sociotechnique' est née dans les années 50 de la rencontre de trois courants de pensée – la psychologie industrielle, la sociologie du travail et les sciences de l'ingénieur – grâce aux recherches réalisées par Frederick E. Emery et Eric L. Trist au Tavistock Institute de Londres (Emery et Trist, 1978). Emery et Trist ont souligné la corrélation existante entre le système 'technique' et le système 'social' montrant que l'efficacité globale dépend de l'efficacité conjointe de ces deux systèmes – si l'organisation fonctionne mal, les risques de rencontrer des problèmes/dysfonctionnements dans le système technique sont plus importants. Plus récemment, l'expression « systèmes sociotechniques » a été précisée par Le Moigne en regard de la définition suivante : « systèmes composés d'éléments techniques (physiques), humains et organisationnels formellement organisés dans un environnement spécifique sur la base de normes, de règles et de rôles, plus ou moins standardisés, en vue de réaliser des objectifs prédéfinis » (Le Moigne, 1990).

Cet aspect d'interaction a aussi été souligné dans (Hollnagel, *et al.* 2006) ainsi que dans (Ruault, 2008) et (Reiman, 2007), à travers la notion de « complexité des systèmes sociotechniques » notamment caractérisée par l'interaction de facteurs techniques et des facteurs socio-organisationnels. Ainsi, les centrales de production d'électricité telle que celles exploitées par EDF et représentant l'objet de recherche étudié dans cette thèse, sont considérées comme des systèmes sociotechniques complexes. En effet, ce sont des systèmes industriels caractérisés par un couplage complexe entre les composantes techniques, humaines, organisationnelles et l'environnement physique, législatif, écologique et social dans lesquels les interactions entre ces composantes ont lieu (avec toutes les conséquences de ces interactions sur la maîtrise des enjeux auxquels ces systèmes doivent faire face).

De plus, à l'intérieur de ces systèmes complexes, ces interactions évoluent et répondent ainsi aux modifications de ces systèmes consécutivement à la prise en compte de nouveaux besoins/exigences ou à des transformations de leur structure/organisation. Par conséquent, la modélisation de tels systèmes est une tâche délicate, mais nécessaire dans un objectif d'évaluation de leurs performances comme des risques auxquels ils sont confrontés. Ainsi, ce besoin en modélisation requiert de s'attaquer au défi de la complexité, au sens de modéliser les multiples interactions entre les « éléments » du système pour obtenir, par émergence, un modèle du tout, ce qui n'est pas possible à réaliser par une approche de type analytique.

En ce sens, l'analyse des risques a recours à la modélisation de systèmes complexes pour produire une évaluation crédible de leurs risques. Aujourd'hui, cette complexité est liée non seulement au nombre de composants et à leurs interdépendances, mais aussi à l'intégration de l'homme et de l'organisation impliquant le plus souvent une modélisation à dire d'experts. Cette déclinaison supplémentaire de l'interaction conduit à une complexité plus conséquente encore qui se combine à des « propriétés émergentes des systèmes sociotechniques telles que la résilience » (Paries, 2006). Ce concept d'origine physique, mais devenu polysémique après avoir été adopté en sciences sociales, notamment en psychologie et en économie, a été transféré en analyse de risque pour indiquer la « capacité du système (technique et social) à faire face à une perturbation » (Hollnagel, *et al.* 2006).

En effet, selon une approche de type 'sociotechnique' appliquée aux systèmes industriels complexes, l'homme est un complément (et non pas une extension) de la machine. Les approches conventionnelles, développées pour manager les risques et la sûreté d'un système sociotechnique, se focalisent surtout sur l'identification des mécanismes pathogènes. Il s'agit des processus de défaillance ou de dégradation impliquant une ou plusieurs composantes techniques et/ou humaines pouvant compromettre la fonction première pour laquelle le système a été créé. Plus spécifiquement, l'analyse des accidents au service de la sécurité industrielle a permis d'entreprendre une première capitalisation des connaissances traduite dans un modèle général basé sur un ensemble de facteurs organisationnels pathogènes (FOP) (Dien, *et al.* 2004). Néanmoins, l'évaluation des risques d'un système sociotechnique exige une vision objective de la réalité (Back, *et al.*, 2008). Des études récentes, par exemple (Hollnagel et Spezali, 2008), ont confirmé ce constat en soulignant le fait que, bien que les systèmes sociotechniques continuent à se transformer et leurs composantes à devenir de plus en plus couplées, les méthodes d'évaluation des risques et de la sûreté ne se développent pas aussi rapidement. Ainsi, les approches conventionnelles ne sont plus adaptées, ou du moins nécessitent d'être complétées, pour traiter correctement certaines interactions multiples comme celles pouvant impacter la dimension opérationnelle d'un système sociotechnique. En effet, comme le dénonçait (Paté-Cornell et Murphy, 1996), ces méthodes s'appuient sur les mêmes principes que les méthodes d'évaluation de la sûreté technique. En ce sens, et afin de tenir compte des facteurs socio-organisationnels dans une évaluation des risques pour un système socio-technique complexe, les méthodes doivent s'intéresser non seulement aux mécanismes pathogènes, mais aussi à ceux résilients pouvant potentiellement se manifester avant, pendant ou après des scénarios normaux et/ou accidentels/incidentés (Léger, *et al.* 2008a). Cette double considération pathogène - résiliente au sein d'une même analyse de risque est un axe scientifique fort défendu dans cette thèse.

En cohérence avec cette considération d'intégration, de nouvelles approches comme l'ingénierie de la résilience (Hollnagel, *et al.*, 2006) ont proposé de considérer la résilience dans le management des risques en remettant en cause les méthodes traditionnelles. Néanmoins, très peu de travaux ont proposé une formalisation précise de cette résilience et de la façon d'en tenir compte dans un modèle de risques (Woods, 2015). Les problématiques scientifiques à résoudre pour supporter concrètement cette intégration restent donc nombreuses. Par exemple, les travaux de (Diez et Druzdzal, 2006), (Hollnagel, *et al.* 2006) et (Van der Leeuw et Aschan Leygonie, 2000) mettent clairement en évidence des verrous scientifiques à lever relativement à la considération conjointe, dans un modèle unifié, des mécanismes résilients et pathogènes pour une modélisation fidèle de la réalité et une évaluation crédible des risques (majoritairement probabiliste). Pour ce faire, il faut aussi répondre à des questions autour de l'existence de marqueurs permettant de tracer des manifestations de résilience, de la compréhension de l'origine de ces manifestations, etc. Les contributions défendues dans cette thèse contribuent à lever ces verrous dans le contexte particulier de l'analyse intégrée de risques appliquée aux systèmes sociotechniques de type centrale de production d'électricité.

En effet, depuis quelques années, des travaux de recherche ont été initiés dans le cadre d'une collaboration entre le Centre de Recherche en Automatique de Nancy (CRAN), le Département Management des Risques Industriels (MRI) de la R&D d'Electricité de France (EDF-R&D) et l'Institut National pour l'Environnement et les Risques (INERIS) puis pérennisés entre les deux premiers pour aboutir au développement d'une approche de modélisation et d'évaluation probabiliste d'une « barrière humaine » prise dans son contexte organisationnel. Ces travaux ont abouti au développement d'une méthodologie, dite Analyse Intégrée des Risques (AiDR) (Léger, 2009) (Fallet, 2013) mise en œuvre dans le domaine de la chimie (stockage), de la radioprotection et la production d'électricité. C'est une approche d'analyse de risques qui traite des systèmes techniques complexes pris dans leur environnement physique et réglementaire et dont les composants techniques sont soumis à des actions de maintenance et/ou de conduite pour lesquelles l'efficacité est évaluée en fonction de leurs contextes organisationnel et environnemental (Duval, *et al.* 2012). En effet, cette méthodologie repose sur le principe que, dans les systèmes sociotechniques, les conditions de succès du fonctionnement du système dépendent de l'interaction entre les facteurs organisationnels, humains, techniques et environnementaux. Ce principe est défendu dans l'approche dite Système-Action-Management (SAM) proposée par (Paté-Cornell et Murphy, 1996) sur laquelle repose l'AiDR. Selon cette approche, l'analyste considère non seulement les variables techniques, mais aussi humaines et organisationnelles afin d'évaluer, dans un cadre intégré, les différents facteurs de risque provenant de l'exploitation des systèmes sociotechniques.

Plus précisément dans l’AiDR, un modèle de ‘barrière humaine’ (Léger, *et al.* 2008b) a été développée pour évaluer l’efficacité des actions humaines dans leur contexte organisationnel. Le cadre de causalité utilisé repose sur un ensemble de facteurs organisationnels (FO) (Léger, *et al.* 2008a). Des mécanismes pathogènes sont identifiés comme les chemins de causalité liant les facteurs organisationnels aux caractéristiques des actions humaines au niveau du collectif de travail. A chacun de ces mécanismes pathogènes a été associé un ensemble de marqueurs issus des analyses organisationnelles conduites sur plus de soixante-dix incidents/accidents/crises dans tous les secteurs (nucléaire, espace, transport ferroviaire, etc.). L’outil de modélisation retenu pour la représentation des connaissances permettant d’établir les liens de causalité et la quantification finale des risques dans l’AiDR est le formalisme Réseau Bayésien (RBs) (Weber, *et al.* 2012), (Duval, *et al.* 2012), (Weber et Simon 2016).

Les travaux menés dans cette nouvelle thèse CIFRE entre le CRAN et l’EDF, s’inscrivent en continuité des travaux des thèses précédentes (Léger, 2009) et (Fallet, 2013) en proposant une évolution du modèle de barrière humaine prise dans son contexte organisationnel. Cette évolution a pour objet, sans remettre en cause le choix de l’outil RBs pour la modélisation, de répondre au besoin actuel d’EDF qui est de passer d’une *hiérarchisation relative* des actions humaines entre elles (et de leurs caractéristiques) à une évaluation de la *probabilité* de l’efficacité d’une action humaine qui, intégrée au modèle global du système sociotechnique, permettra d’atteindre les probabilités des enjeux d’intérêt retenus pour le système : sûreté, disponibilité, maintien du patrimoine dans la durée. Dans l’AiDR, cette évolution amène à reconsidérer non seulement les mécanismes pathogènes déjà identifiés, mais aussi d’y adjoindre des mécanismes résilients pour aboutir à une évaluation réaliste de la probabilité d’efficacité des actions humaines dans leur contexte organisationnel.

Pour faire évoluer le modèle de l’AiDR, la considération conjointe des influences pathogènes et résilientes requiert de s’attaquer simultanément à des verrous scientifiques plutôt génériques (par exemple, la considération de la résilience dans une analyse de risque), d’autres, directement déduits du contexte applicatif particulier (AiDR) et enfin à des verrous scientifiques associés à l’outil de modélisation retenu. L’objectif est bien évidemment qu’une part de ces contributions puisse être utilisée dans la communauté traitant de l’analyse de risque afin de revendiquer pour cette thèse, un réel travail de recherche finalisé et non pas d’ingénierie.

En regard de ces verrous, les contributions majeures développées dans cette thèse se structurent autour des trois axes suivants :

1. Une amélioration de la structure préexistante du modèle de barrière humaine et organisationnelle. Il s’agit d’aboutir à un modèle par facteurs du

collectif de travail pris dans son contexte organisationnel par « fertilisation » croisée avec les méthodes d’Evaluation Probabiliste de la Fiabilité Humaine (EPFH) qui fournissent des facteurs de forme appelés performance shaping factors (PSF).

2. Une modélisation de l’efficacité humaine encore plus proche de la réalité pour l’Analyse intégrée des Risques en intégrant la résilience dans la modélisation probabiliste.
3. Un traitement global des avis d’expert cohérent avec la structure mathématique du modèle proposé. L’objectif est d’estimer d’une manière objective les paramètres du modèle. Ce traitement se fonde sur la construction d’un questionnaire permettant de ‘guider’ l’expert vers l’évaluation d’effets conjoints issus de l’interaction entre mécanismes pathogènes et résilients.

Ces contributions sont porteuses de plusieurs originalités comme :

L’approfondissement de l’approche par facteurs pour l’évaluation de l’efficacité des barrières humaines prises dans leur contexte organisationnel. Cette investigation répond aux besoins des méthodes d’analyse de risques permettant de traiter à la fois les aspects techniques, humains, organisationnels et environnementaux d’un système sociotechnique complexe pour identifier ses vulnérabilités et y mettre prioritairement les moyens de réduction. Le traitement des facteurs humains et organisationnels au niveau du collectif de travail y est particulièrement traité.

Le développement du concept de mécanisme résilient et sa caractérisation qualitative et quantitative dans une analyse de risques pour contrebalancer les mécanismes pathogènes jusqu’ici considérés et aboutir à une probabilité réaliste de l’efficacité d’une action humaine.

Une structure mathématique des éléments du modèle retenu est proposée selon le principe de mitigation entre influences pathogènes et résilientes.

Un modèle d’éllicitation d’experts fondé sur des probabilités objectives, d’où un questionnaire non pas orienté vers l’estimation des paramètres, mais vers la caractérisation de situations dont la distribution des probabilités objectives fournies par l’expert permet de calculer les paramètres du modèle.

Ainsi dans ce mémoire, les différentes contributions se développent au cours des quatre chapitres organisés comme suit :

Le **Chapitre 1** a pour objectif de mettre en évidence le périmètre de la thèse principalement en lien avec le contexte industriel EDF de maîtrise des risques. Il définit les principaux concepts sur lesquels se base ce travail. En outre, il permet de présenter l’approche AiDR et d’identifier très clairement les besoins exprimés par EDF devant conduire à une évolution de cette approche. Le besoin

majeur est de passer d'une hiérarchisation relative des actions humaines entre elles, de leurs caractéristiques entre elles (formation, aides, gestion collective et dynamique de groupe etc.) à une évaluation de la probabilité de l'efficacité d'une action humaine. Ces besoins sont, finalement, traduits en un ensemble de verrous technologiques précis auquel ce travail de thèse doit apporter des solutions.

Le **Chapitre 2** permet de positionner plus précisément, par rapport à l'existant, chacune des problématiques scientifiques associées aux verrous technologiques mis en évidence dans le chapitre précédent. En ce sens, nous recensons et discutons les différents travaux faits en matière (1) d'évaluation probabiliste de la fiabilité humaine, (2) de considération de la résilience dans une analyse de risque et plus particulièrement dans la modélisation probabiliste, et (3) d'estimation des paramètres basée sur l'élicitation d'avis d'experts. Par rapport à ce recensement, nous nous focalisons ensuite sur les méthodes et/ou approches a priori les plus adaptées pour répondre aux besoins d'EDF, dont la résolution nécessite une évolution de l'AiDR. Ce positionnement construit sur un état de l'art conséquent nous permet d'isoler les verrous scientifiques majeurs qu'il faudra lever pour répondre aux besoins définis dans le chapitre 1.

Le **Chapitre 3** a pour objet de formaliser et justifier chacune de nos contributions construites en réponse aux verrous scientifiques précédents. Dans un premier temps, nous nous focalisons sur la modélisation des mécanismes de résilience et leur interaction avec les mécanismes pathogènes afin d'en évaluer les effets conjoints sur l'efficacité des barrières humaines et organisationnelles. Sur la base de cette formalisation, nous définissons une nouvelle structuration du modèle RB prenant en compte les mécanismes résilients et leur mitigation des mécanismes pathogènes. Cette structuration est basée sur une extension du modèle probabiliste canonique : Leaky Noisy-OR. Ensuite, nous proposons des principes génériques de construction du modèle de barrière humaine. Lors de l'application de ces principes dans le cadre de l'AiDR, nous formalisons l'intégration des PSF au niveau des déterminants de la barrière humaine. Ces facteurs sont issus de l'état de l'art sur les méthodes EPFH analysées au chapitre 2. En outre, pour tenir compte de l'intégration de la résilience et le principe de mitigation, des adaptations et généralisations des variables du modèle sont proposées. Globalement, cela permet d'aboutir à un modèle générique de l'AiDR qui nécessite d'être particularisé sur une action spécifique. Par conséquent, dans la dernière partie, nous nous intéressons à la phase d'estimation des paramètres du modèle et proposons une approche se fondant sur l'élicitation d'avis d'experts pour une estimation objective des paramètres du modèle. Des principes sont présentés en termes de protocoles d'élicitation et de réadaptation des modèles en fonction de la connaissance issue des avis d'experts.

Le **Chapitre 4** concerne la validation de chacune des contributions proposées au chapitre 3 en lien avec les besoins exprimés dans le chapitre 1. Cette validation se construit en appliquant les propositions sur un cas réel de taille réduite permettant de montrer l'applicabilité et la faisabilité de l'ensemble des contributions. Il permet d'illustrer les problématiques solutionnées par nos contributions sur les phases de caractérisation, de modélisation et d'évaluation probabiliste de l'efficacité des barrières humaines et organisationnelles. Cette application nous permet également d'identifier les limites atteintes par nos contributions au regard des hypothèses posées et de dégager de nouvelles perspectives.

Chapitre 1

De la hiérarchisation des actions humaines a une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

1 De la hiérarchisation des actions humaines a une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Introduction

Le présent chapitre a pour objectif de fournir les éléments relatifs au contexte de maîtrise de risques à EDF qui constitue la genèse de cette thèse. Il s'agit donc de définir les points clés et nos objectifs de travail ainsi que de préciser les aspects spécifiques à la problématique de l'évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel. De plus, des définitions essentielles sur lesquelles se construisent nos contributions sont précisées.

En ce sens, la première partie de ce chapitre permet tout d'abord d'introduire le management des risques et les principales compétences et méthodes développées et actuellement utilisées à EDF. Plus particulièrement, nous nous focalisons sur une des problématiques importantes pour EDF qu'est l'intégration de l'homme dans une approche multirisque permettant d'aboutir à une vision globale des risques importants pour la conception et l'exploitation de systèmes industriels comme les centrales de production d'électricité d'EDF (centrales nucléaires, centrales thermiques conventionnelles, etc.).

Ayant défini ce contexte de management des risques à EDF, la seconde partie du chapitre a pour objectif de présenter une de ces approches multirisques : l'Analyse intégrée Des Risques (AiDR). Compte tenu du cadre de modélisation existant, nous identifions certaines limites de l'AiDR dans son exploitation actuelle. Ainsi nous mettons en évidence le besoin pour EDF de parvenir à une évaluation probabiliste réaliste de l'efficacité des barrières humaines éventuellement considérées dans leur contexte organisationnel.

En lien avec ce besoin, la troisième partie de ce chapitre nous amène à identifier un ensemble de verrous techniques qui constituent les éléments essentiels sur lesquels nous apportons nos contributions scientifiques dans la suite du manuscrit.

1.2 Le management des risques à EDF

Le management des risques est défini comme étant un ensemble d'activités coordonnées dans le but de diriger et piloter un organisme vis-à-vis du risque (ISO, 2009). Ce management des risques est la thématique de recherche du département « Management des Risques Industriels » (MRI) créé à EDF R&D en 2001 par l'union de groupes issus des anciens départements « Études de Sécurité et de Fiabilité » (ESF) et SDM « Surveillance, Diagnostic, Maintenance ».

D'un point de vue général, le risque est associé à la notion de « danger éventuel, plus ou moins prévisible » (Roberts 2008). Cette acception négative du risque est partagée dans plusieurs domaines tels que l'industrie, l'économie et la finance, les sciences naturelles, etc. Cependant, l'objet du risque peut changer selon le domaine spécifique. En économie et en finance, par exemple, le risque porte sur la possibilité de perte monétaire due à une incertitude que l'on peut quantifier. Dans l'industrie, le risque est plutôt conçu comme une combinaison de la probabilité et de la (des) conséquence(s) de la survenance d'un événement dangereux spécifié (Magne et Vasseur, 2006).

Par rapport à cette vision générale, les risques qui nous intéressent dans cette thèse sont les *risques industriels* liés à l'occurrence d'incidents et accidents majeurs dans les systèmes exploités par EDF, comme les unités de production d'électricité ainsi que le réseau de transport et donc avec des conséquences sur l'homme, l'environnement et sur les systèmes techniques eux-mêmes.

Ainsi, par rapport à cette vision globale du risque, dans ces travaux nous considérons le **risque** comme :

« L'effet de l'incertain sur l'atteinte des objectifs » (ISO, 2009).

Toujours en référence à la définition contenue dans (ISO, 2009), nous considérons le **risque industriel** comme :

« Un événement accidentel majeur se produisant sur un site industriel et entraînant des conséquences immédiates graves pour le personnel, les populations avoisinantes, les biens ou l'environnement ».

Cette considération du risque à EDF se lie à la notion de **système sociotechnique** compte tenu des systèmes industriels spécifiques à EDF pour la production d'électricité. De façon générale, un système sociotechnique est un *système complexe*, à savoir un système qui possède un grand nombre de composants et est caractérisé par la présence d'interactions fortes entre ces composants (Zio, 2007). La considération 'sociotechnique' a pour origine la rencontre de trois courants de pensée – la psychologie industrielle, la sociologie du travail et les sciences de l'ingénieur – grâce aux recherches réalisées par F. Emery et E. Trist au Tavistock Institute de Londres (Emery et Trist, 1978). En ce sens, Emery et Trist ont souligné la corrélation existante entre le système 'technique' et le système 'social' montrant que l'efficacité globale dépend de l'efficacité conjointe de ces deux systèmes.

Plus récemment, l'expression « systèmes sociotechniques » a été précisée par Le Moigne qui propose la définition suivante : « systèmes composés d'éléments tech-

niques (physiques), humains et organisationnels formellement organisés dans un environnement spécifique sur la base de normes, de règles et de rôles, plus ou moins standardisés, en vue de réaliser des objectifs prédéfinis » (Le Moigne, 1990).

Nous considérons qu'une définition similaire à celle proposée par Le Moigne est appropriée par rapport à la vision du risque adopté à EDF et donc dans cette thèse (en regard des systèmes visés).

Ainsi, dans nos travaux de thèse nous considérons un **système sociotechnique** comme :

Un système où des composantes techniques, humaines et organisationnelles coexistent et sont en interaction dans un environnement spécifique.

Par rapport à ces différentes composantes d'un système sociotechnique et dans un objectif de management des risques, le département MRI a développé des compétences dans les grandes disciplines scientifiques que sont la sûreté de fonctionnement, la fiabilité, les sciences humaines et sociales (ergonomie, sociologie, etc.), les mathématiques appliquées (probabilités, statistiques, etc.) et la simulation numérique.

Aujourd'hui ces compétences sont souvent utilisées dans l'analyse des risques plutôt de façon isolée. Cette façon de procéder n'est plus suffisante pour faire face à la complexité des interactions à maîtriser dans les systèmes sociotechniques. Une synergie de ces compétences est donc nécessaire. Ce constat est conforté par les analyses d'un ensemble d'accidents graves dans l'industrie nucléaire (dont Three Mile Island en 1979, Tchernobyl en 1986 et plus récemment Fukushima en 2011) qui ont mis en évidence, puis confirmé, que les causes des problèmes affectant les composants techniques et celles liées plutôt aux inefficacités aux niveaux humain et organisationnel coexistent et interagissent dans l'occurrence des scénarios accidentels. Ceci souligne l'importance d'une évolution de la vision d'EDF vers une considération conjointe de ces causes dans le management des risques.

Ce besoin d'une synergie entre les différents domaines est d'autant plus important dans la prise en compte de l'environnement physique, réglementaire et du contexte organisationnel. En effet, les changements à ces niveaux influencent fortement les différents enjeux d'un système sociotechnique, principalement la disponibilité, mais aussi la sûreté et imposent par conséquent la considération conjointe de ces influences.

Ces constats d'EDF pragmatiques sont confortés par plusieurs études scientifiques, comme par exemple dans (Galàn, *et al.* 2007) et (Mosleh, *et al.* 2007), qui ont mis en évidence la nécessité de traiter conjointement plusieurs risques dans les systèmes sociotechniques.

Cette synergie, au global, doit aboutir, à EDF, à une maîtrise **multirisque** qui implique de modéliser :

- les risques affectant les *systèmes techniques*, leur probabilité d'occurrence et la gravité de leurs conséquences sur les enjeux du système étudié, considérés ensemble : la disponibilité, la sûreté, le maintien du patrimoine dans la durée et l'environnement ;

- les influences de l'*environnement physique et réglementaire* sur l'occurrence des risques techniques (Duval, et. 2012) ;
- l'impact des *actions de maintenance* et de *conduite*, éventuellement prises dans leur contexte organisationnel, sur les enjeux mentionnés précédemment (Léger, et al. 2008b) (Fallet, 2013).

Cette considération **multirisque** pour les systèmes sociotechniques est partagée dans d'autres secteurs tels que l'aéronautique (Mosleh, et al. 2007), la pétrochimie (Sadorsky, 2001) ou encore dans la construction des grandes infrastructures (Hongwei, 2006). Des travaux en maîtrise des risques comme ceux de (Galàn, et al. 2007) mettent aussi en évidence le fait que, en l'état, les approches **multirisques** existantes permettent de répondre partiellement à ce besoin et des manques ou des limites sont encore à résoudre, notamment dans la prise en compte des composantes humaines et organisationnelles et la quantification de leurs influences sur les risques globaux en regard du contexte d'EDF.

C'est pourquoi, par rapport aux items de cette vision **multirisque**, cette thèse a pour objectif d'attaquer prioritairement les risques liés aux composantes humaines et organisationnelles.

Nous nous focalisons donc dans cette thèse sur le besoin d'étendre la considération de l'humain et de l'organisation dans l'analyse des risques.

1.2.1 Considération « système » des facteurs humains et organisationnels dans les approches multirisques à EDF

La considération des composantes humaines et organisationnelles est un axe de recherche fort à EDF. Plus spécifiquement, cette considération se fait au département MRI au travers de la notion de **fiabilité humaine**.

Plusieurs définitions existent de la fiabilité humaine. Dans le monde industriel, elle est plutôt vue comme la « qualité du comportement qu'attend l'entreprise (et par extension le public et les autorités légales réglementant l'exploitation industrielle) des humains à qui elle confie l'exploitation d'une installation » (Le Bot et Pesme, 2010). Par rapport à cette définition, deux visions de la fiabilité humaine existent. Elles se traduisent dans des approches apparemment opposées (Morin, 2014) : l'approche *mécanique* et l'approche *managériale*.

Selon la vision propre à l'approche *mécanique*, la fiabilité humaine se définit comme la capacité à répondre à des attentes précises et à ne pas dévier des consignes de la même façon que le 'système électrique du démarreur d'un véhicule' (Rasmussen, et al. 1990).

Pour l'approche *managériale*, la fiabilité humaine est plutôt vue comme la capacité de trouver des solutions pour résoudre des problèmes dans des situations imprévisibles, la prise d'initiatives rapides et intuitives avec l'utilisation d'une expérience accumulée à partir de situations vécues (Edmondson, 1996).

De manière plus générale, la fiabilité humaine est associée à l'aptitude d'une entité (individu ou équipe) à effectuer ses tâches de façon sûre, dans les délais et les exigences attendus (AFNOR, 2005).

Une vision proche de cette dernière est proposée par (Hollnagel, 1993) qui définit la fiabilité humaine comme le « niveau de confiance qu'une personne effectue correctement une activité requise du système pendant une période de temps donné (en supposant que le temps est un facteur limitant) sans effectuer aucune activité étrangère qui puisse dégrader le système ».

Cependant, ces définitions ne correspondent pas à la vision d'EDF, du point de vue de la restriction à l'individu et à la tâche élémentaire. L'approche de la fiabilité humaine est systémique et contextuelle, elle évalue l'efficacité de la réalisation des actions requises sur l'installation, en mettant en évidence des mécanismes de défaillance possibles, sans se focaliser uniquement sur l'erreur humaine individuelle.

C'est pourquoi dans cette thèse, nous adhérons plutôt à la définition de proposée par (Hollnagel, 2002). Selon cette définition la **fiabilité humaine** se définit comme :

« Une approche qui tente d'évaluer la *performance humaine* (individuelle et/ou collective) dans un objectif d'efficacité industrielle, et si possible quantifiable ».

Au département MRI cette évaluation de la fiabilité humaine se fait grâce à la compétence en « **évaluation probabiliste de la fiabilité humaine** » (EPFH). Selon la définition retenue dans (Magne et Vasseur, 2006), l'EPFH est une « démarche prévisionnelle appliquée pour évaluer et comprendre la fiabilité des processus des systèmes à risques où intervient essentiellement l'homme ». Au cours du temps, un ensemble de méthodes a été développé permettant de déterminer la *probabilité d'échec* d'une action humaine ou *mission facteur humain* (FH) en exploitation normale, ou en situation incidente/accidentelle dans les industries à risque telles que le nucléaire, la pétrochimie, l'aérospatial, etc. Les probabilités d'échec d'une mission FH permettent, combinées aux approches d'évaluation de la fiabilité des systèmes techniques, de mettre en évidence les « coupes » prépondérantes au sein des études probabilistes de sûreté (EPS) qui peuvent concourir à des conséquences inacceptables du point de vue de la sûreté. À EDF et plus particulièrement à MRI, la méthode MERMOS (Meyer, Le Bot, et Pesme, 2007) permet une approche qualitative et compréhensive de l'échec d'une mission FH et dont la probabilité est à prendre en compte dans les EPS.

Néanmoins, la plupart des méthodes EPFH, dont MERMOS, se focalisent majoritairement sur l'étude des comportements humains en situation de travail sans considérer le contexte organisationnel. Ce positionnement implique des analyses locales détaillées et plutôt centrées sur l'individu et sur le collectif de travail. En effet, ces analyses ne prennent pas en compte les interactions complexes avec le contexte organisationnel. Par conséquent, ces analyses locales s'avèrent insuffisantes face au besoin d'une considération systémique (en référence à la notion de système sociotechnique) de l'homme et plus spécifiquement l'interaction entre les composantes humaine et socio-organisationnelle dans les systèmes sociotechniques (Paté-Cornell et Murphy, 1996).

En ce sens, ces dernières années, la problématique de l'intégration de l'interaction entre ces deux composantes dans les approches multirisques a fait l'objet de

plusieurs recherches à EDF. Une de ces recherches a été menée sur la base d'une méthodologie multirisque proposant des modèles pour la prise en compte des facteurs humains et organisationnels : l'Analyse intégrée des Risques (AiDR).

1.3 L'analyse intégrée des risques (AiDR)

L'Analyse intégrée Des Risques (AiDR) est une méthodologie multirisque formalisée et développée depuis plusieurs années par les deux partenaires de cette thèse (CIFRE), EDF R&D et le CRAN.

D'après (Duval, 2009), cette nouvelle approche dite *intégratrice* entre plusieurs disciplines (la Sécurité de Fonctionnement, l'Analyse de Risque, l'EPFH, l'Analyse Organisationnelle) vise à développer les méthodes, choisir les outils adaptés et les faire évoluer pour réaliser l'analyse de risques des systèmes sociotechniques. Plus spécifiquement, l'AiDR a pour objectif de :

- intégrer des actions humaines de maintenance, de surveillance et de conduite prises dans leur contexte organisationnel ;
- garantir les enjeux clés du système : sa disponibilité prise avec sa sûreté, le maintien de son patrimoine dans la durée, son impact sur l'environnement ;
- hiérarchiser des risques de natures différentes vis-à-vis de ces enjeux dans l'objectif de prioriser les moyens à mettre en œuvre pour les réduire ;
- contribuer à une meilleure culture du risque (AIEA, 1991) et une meilleure communication sur les risques.

La démarche développée est dite « à tiroirs » : en effet elle peut couvrir deux ou trois de ces objectifs ou l'ensemble suivant le besoin et les risques suspectés nécessitant une appréciation.

Le cadre conceptuel de l'AiDR représenté dans la Figure 1 est fondé sur les principes développés dans l'approche System-Action-Mangement (SAM) issue de (Paté-Cornell et Murphy, 1996). Cette approche met l'accent sur les influences du contexte organisationnel vers les actions humaines et, par le biais de ces actions, l'impact sur la disponibilité des systèmes techniques. L'intégration des composantes humaines et organisationnelles dans une même vue globale de l'ensemble des risques a impliqué le choix d'une vision synthétique de ces influences (Léger, *et al.* 2008a), (Léger, 2009).

Dans l'AiDR, ces influences sont représentées par un **modèle de barrière humaine** dans lequel les actions humaines de maintenance et de conduite sont éventuellement prises dans leur contexte organisationnel et environnemental.

Dans ces travaux de thèse, la **barrière humaine** est considérée comme :

L'ensemble des activités (une ou plusieurs opérations) qui s'oppose à l'enchaînement d'évènements susceptibles d'aboutir à un accident, permettant de réduire la probabilité d'occurrence du scénario accidentel ou de limiter ses conséquences (cf. Figure 1.1).

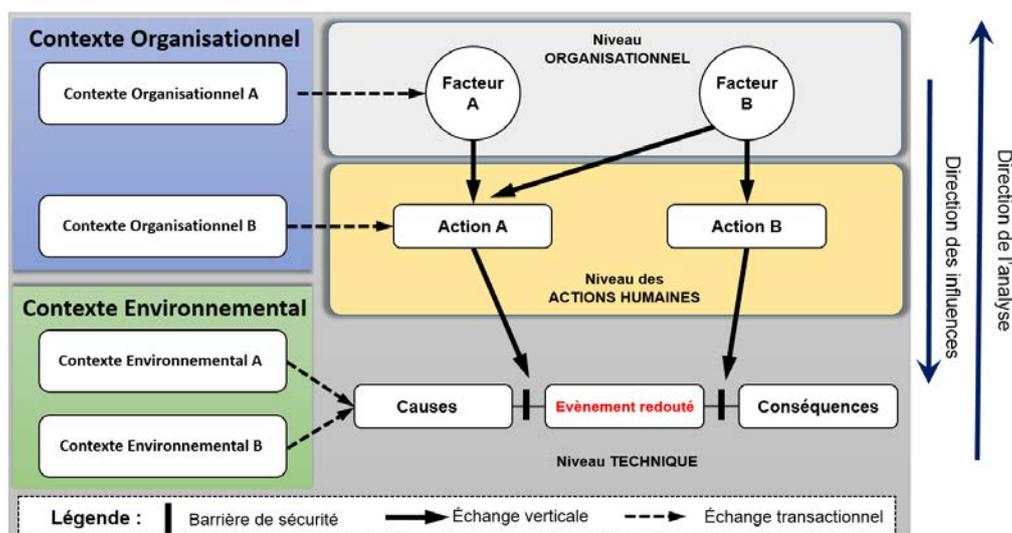


Figure 1.1. Cadre conceptuel de l'AiDR (Léger, 2009).

Selon la vision propre à l'AiDR, c'est plutôt l'efficacité qui compte dans un système sociotechnique à risque tel qu'une installation nucléaire, où les actions de conduite et de maintenance (pour un scénario pré-accidentel) doivent être réalisées correctement (Léger, *et al.* 2008a).

Ainsi, l'action humaine est évaluée par rapport à son **efficacité**. Dans cette thèse, **l'efficacité de l'action** est vue comme :

La capacité à réaliser correctement une tâche vis-à-vis de l'atteinte des objectifs.

En effet, l'**efficacité** se différencie de l'efficacité humaine, à savoir la capacité à réaliser une tâche de façon optimale, c'est-à-dire avec le moins de ressources possible ou dans le moins de temps possible.

Dans ce modèle de barrière humaine, l'efficacité de l'action est caractérisée par un ensemble de **déterminants**. Il s'agit de caractéristiques *déterminantes* pour l'efficacité de l'action et spécifiques du collectif de travail (délégation, expérience, gestion collective et dynamique de groupe), (b) relatives aux outils et procédures (aides, capacité à respecter le cahier des charges, contrôle en temps réel, retour d'expérience), (c) associées à l'influence de facteurs externes sur le collectif pendant la réalisation de l'action (facteurs d'environnement). Les déterminants impactent l'efficacité globale de l'action humaine par ses **phases** (préparation, réalisation, clôture). En effet, chaque phase est affectée par un sous-ensemble de déterminants (cf. Table 1.1).

Tableau 1.1. Phases et déterminants de l'action dans le modèle de barrière humaine de l'AiDR

Phases de l'action	Déterminants de l'action
Préparation	Délégation (De) – Formation (Fo) – Aides (Ai)
Réalisation	Expérience (Ex) - Capacité à respecter le cahier de charges (Rcc) - Facteurs contextuels (Fc)- Gestion collective et dynamique de groupe (Gcdg)

Clôture

Contrôle et atteinte des objectifs (Cao) - Retour d'Expérience (Rex)

Pour la considération du contexte organisationnel, l'AiDR se base sur la métaphore médicale proposée par (Reason, 1997). Selon cette métaphore, pour savoir si une organisation est en « bonne santé », il est beaucoup plus simple de rechercher les causes possibles d'une maladie potentielle que de passer en revue les facteurs caractérisant la bonne santé. Ainsi, un ensemble de **facteurs organisationnels pathogènes** (FOP) (Pierlot et Dien, 2007) ont été définis pour caractériser l'efficacité des actions prises dans leur contexte organisationnel (Léger, *et al.* 2008) :

- Difficulté à faire vivre un Retour d'Expérience (REX)
- Défaillance de la Gestion Quotidienne de la Sécurité (GQS)
- Faiblesse des Organismes de Contrôle (OC)
- Mauvais Traitement de la complexité organisationnelle (MT)
- Pressions de Production (PP)
- Faiblesse de la Culture Organisationnelle de Sécurité (COS)
- Absence de Réexamen des hypothèses de conception (AR).

Pour la représentation et quantification des risques au sein de l'approche d'AiDR, (Léger, 2009) a proposé d'utiliser le formalisme réseau Bayésien (RB) à partir d'une comparaison entre différents outils de modélisation (les réseaux de neurones et de Pétri, les RBs, les arbres de défaillances ou encore les chaînes de Markov). De cette comparaison principalement basée sur la possibilité de modéliser dans un graphe des systèmes complexes, de modéliser des connaissances issues de l'élicitation d'experts et surtout de quantifier des distributions à posteriori lors de l'extraction de nouvelles connaissances, c'est l'outil RBs qui ressort comme le plus adapté. Ce choix est une exigence de contrainte dans ces travaux.

Un **réseau Bayésien** (RB) est un *modèle graphique probabiliste* représentant des variables aléatoires sous la forme d'un graphe orienté sans circuit (Jensen, 1996). Pour un domaine donné (par exemple la sûreté de fonctionnement), les relations causales entre variables d'intérêt sont décrites par un graphe. Dans ce graphe, les relations de *cause à effet* entre les variables ne sont pas nécessairement *déterministes*, mais peuvent être *probabilistes*. Ainsi, l'observation d'une ou de plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie seulement la probabilité de les observer (Diez et Druzdzel, 2006).

Ainsi, les facteurs liés aux composantes techniques, humaines, organisationnelles et environnementales constituent les nœuds (variables) et les liens de causes et d'effets entre modes de défaillances et les enjeux du système, les arcs (probabilités conditionnelles entre les variables) de la représentation dans le modèle RB. Un exemple de cette représentation sur l'ensemble des facteurs et leurs liens est présenté à la Figure 1.2. En référence à la structuration sociotechnique proposée dans la Figure 1.1, le système *technique* est modélisé par l'ensemble des composants des sous-systèmes hydrauliques ('SFI', 'SEF' et 'Appoint CVF', à gauche) et électriques ('GEV', 'LGR', 'Tableaux Secourus', etc., à droite) considérés dans l'environnement physique ('Environnement'). Les composants techniques sont soumis à l'occurrence des événements aléatoires environnementales ('Agresseurs') pouvant déclencher des mécanismes

de défaillances : ainsi des barrières humaines considérées dans leur contexte organisationnel (en rouge) sont mises en place pour assurer leurs disponibilité.

L'utilisation du formalisme RB, basé sur la *théorie des probabilités*, permet de modéliser des connaissances de natures qualitatives et quantitatives par liens de causalité entre les variables du modèle de risques et de mener des analyses des causes vers les conséquences (pronostic) ou conséquences vers les causes (diagnostic). En effet, la représentation des connaissances dans un modèle RB est naturelle et assez lisible (par opposition aux réseaux de neurones par exemple, ou mieux des arbres de défaillances) (Weber et Simon, 2016). À chaque variable effet est associée une distribution de probabilité quantifiée à partir d'une table de probabilités conditionnelles (TPC) (Jensen, 1996). La TPC exprime de manière compacte et directe toutes les configurations possibles de causalité exprimées en fonction de l'état des variables cause.

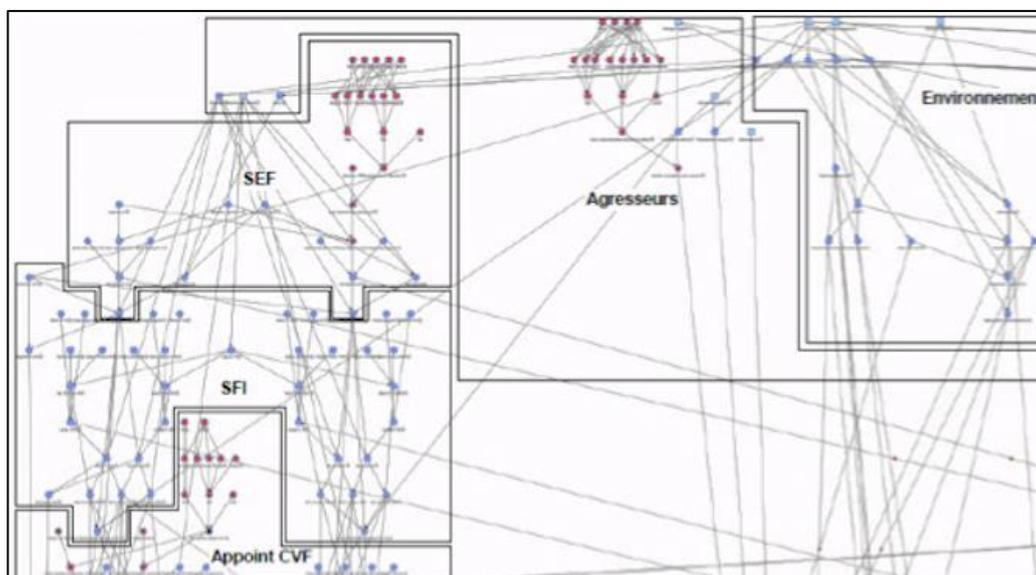


Figure 1.2. Modèle RB dans l'AiDR d'une installation EDF avec des variables techniques, environnementales, humaines et organisationnelles.

En ce qui concerne les barrières humaines, le modèle RB incluant toutes les variables telles que proposées dans (Léger, 2009) est illustré à la Figure 1.3.

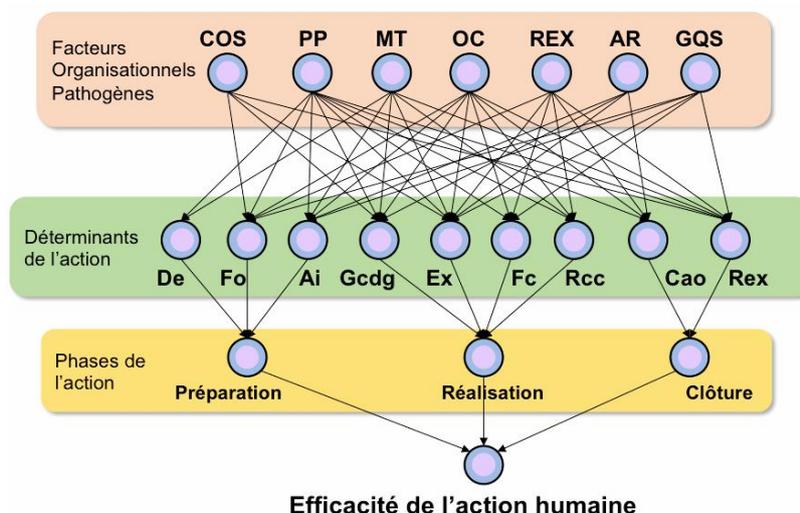


Figure 1.3. Modèle RB de la barrière humaine dans (Léger, 2009).

Dans ce modèle de barrière humaine, les variables se structurent selon les trois niveaux de caractérisation de l'action décrits plus en amont dans cette section (cf. Tableau 1.1). Ainsi, l'action est d'abord factorisée en trois phases 'Préparation', 'Réalisation' et 'Clôture' : elles permettent de collecter les influences des facteurs du collectif de travail (les Déterminants) sur les différents phases de l'action. Si la barrière humaine est prise dans le contexte organisationnel, des liens de causalité basés sur des mécanismes de type pathogène (Léger, 2009) (Pierlot et Dien, 2007) ont été identifiés entre les variables du contexte organisationnel (les FOP) et les Déterminants.

Sur la base des éléments à modéliser, nous confirmons pour cette thèse le choix des RBs comme outil de modélisation de nos contributions afin de les intégrer au modèle initial de l'AiDR.

1.3.1 Intérêt pour les réseaux Bayésiens dans l'AiDR

Dans l'AiDR, les RBs permettent de décrire les scénarios d'accidents/de dégradation en partant de leurs initiateurs jusqu'à leurs conséquences et de quantifier de façon probabiliste les risques. Intuitivement, ils sont à la fois :

- des modèles de représentation des *connaissances* ;
- des « machines à calculer » les *probabilités conditionnelles* ;
- une base pour des systèmes d'*aide à la décision*.

En effet, d'après (Fallet, 2013) pour définir le modèle de risques associé au cadre théorique de l'AiDR présenté ci-dessus, le choix d'un outil de modélisation repose sur la base de trois contraintes principales en lien avec les systèmes sociotechniques complexes que veut traiter l'approche AiDR. Cet outil, sur la base des argumentations défendues dans (Duval, *et al.* 2012), doit principalement :

1. Permettre une *modélisation synthétique* de la complexité des systèmes prenant en compte différents types de variables (quantitatives/qualitatives) multimodales et corrélées entre elles ;
2. Permettre de modéliser des variables *déterministes* et des variables *probabilistes* présentant de l'incertitude aléatoire ;

3. Permettre d'effectuer des analyses des causes vers les conséquences (*pronostic*) et des conséquences vers les causes (*diagnostic*) sur le système modélisé.

Les RB se distinguent principalement des autres outils de modélisation existants par leurs capacités : à modéliser des *systèmes complexes*, à modéliser des variables multimodales de natures différentes, à déterminer précisément la probabilité d'occurrence d'un événement, à effectuer des pronostics et des diagnostics par inférence ou encore, à mettre à jour les probabilités en présence d'observation (distributions de probabilités a posteriori) (Weber, *et al.* 2012). En outre, cet outil est notamment intéressant pour sa polyvalence. En effet, les RB permettent de s'intéresser à des problématiques de pronostic, de diagnostic ou encore d'optimisation, ceci dans différents secteurs d'activité. Ils constituent un outil de modélisation permettant de rendre compréhensible la complexité des systèmes au sein d'une vue unique via une représentation graphique modulaire et compacte. Enfin, l'absence de sémantique spécifique pour son utilisation constitue également un atout pour le choix de ce formalisme dans des problèmes de couplage de disciplines variées telles que la technique, l'humain, l'organisationnel et l'environnemental (Diez et Druzdzel, 2006).

L'utilisation des réseaux Bayésiens comme outil de modélisation de l'AiDR semble donc être – à ce jour – encore le meilleur compromis pour répondre aux contraintes définies précédemment afin de proposer une approche d'analyse multirisque outillée.

Ainsi, pour évaluer les risques associés aux systèmes sociotechniques complexes, l'AiDR procède à l'intégration de ces différentes dimensions dans un même modèle de risque. Cette intégration est possible en utilisant les connaissances particulières à chaque dimension et les relations qui les lient à travers l'outil RB (Weber, *et al.* 2012).

Le modèle RB de barrière humaine de l'AiDR a été déployé dans différents secteurs. Dans le cadre de la thèse (Léger, 2009), par exemple, il a été appliqué à un silo de stockage d'un site industriel chimique. L'exploitation du modèle a aussi été appliquée à des sous-systèmes d'unités de production d'électricité d'EDF dans le cas des travaux de thèse de (Fallet, 2013). Plus récemment, des analyses ont été conduites après l'occurrence d'un événement accidentel (Duval, *et al.* 2012). Ces applications ont fait l'objet de rapports techniques d'EDF qui mettent en évidence des résultats positifs liés à l'utilisation de l'AiDR. Notamment :

- Le constat que *l'intégration des aspects humains et organisationnels* apporte un éclairage quantifié des interactions multiples ayant un impact sur l'efficacité des barrières de sécurité sur les enjeux de disponibilité et de sûreté des systèmes étudiés ;
- La possibilité de *hiérarchiser* les actions d'amélioration apportées aux systèmes.

Cependant, ces applications ont montré que les résultats obtenus ne sont pas complètement satisfaisants vis-à-vis d'EDF.

C'est pourquoi le modèle de barrière humaine utilisé dans l'AiDR doit être amélioré pour répondre aux limites actuelles.

1.3.2 Limites actuelles de l'AiDR et besoins d'EDF

Les expérimentations sur les cas industriels réels discutés dans la section précédente ont conduit les utilisateurs de l'AiDR à exprimer le besoin d'avoir des valeurs *réalistes* des fréquences d'occurrence d'indisponibilité et de sa dégradation en plus de la hiérarchisation des risques qu'elle fournit actuellement.

L'AiDR a donc besoin d'évoluer pour permettre le passage de la hiérarchisation des actions humaines entre elles (et leurs déterminants) à une **évaluation probabiliste de l'efficacité des actions humaines prises dans leur contexte organisationnel** qui constitue l'objectif de cette thèse (cf. Figure 1.4).

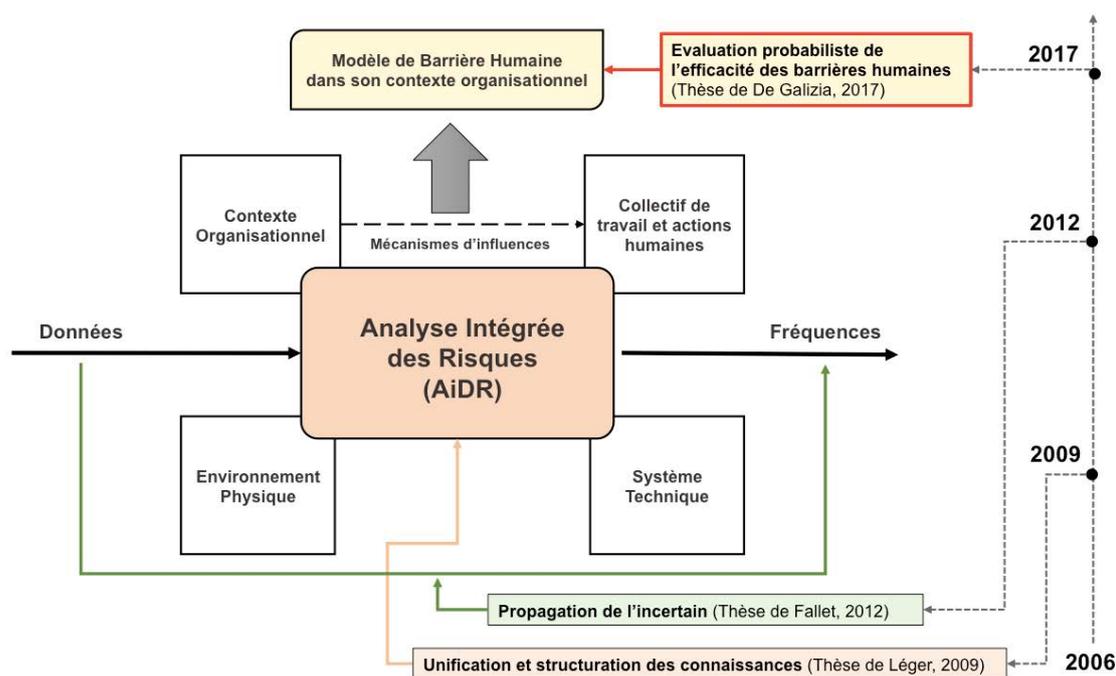


Figure 1.4. Etapes de développement de l'AiDR et identification du nouveau besoin (en rouge).

Ce besoin majoritairement de type industriel/technologique nous a conduit à recenser un ensemble de **verrous techniques** importants pour EDF à attaquer dans ces travaux de thèse pour que les besoins soient satisfaits :

- 1) Tout d'abord, la nécessité de renforcer la caractérisation de l'efficacité de l'action par les déterminants du modèle de barrière humaine. En effet, un potentiel d'amélioration a été identifié concernant les déterminants représentatifs du niveau du collectif de travail. La pertinence et l'orthogonalité ainsi que l'exhaustivité des déterminants sont les éléments clés permettant de mesurer la *robustesse* des résultats produits avec le modèle de barrière humaine et par conséquent des valeurs de fréquence obtenus par l'AiDR ;
- 2) Deuxièmement, le besoin de considérer la *résilience* dans le cadre des relations de causalité affectant l'efficacité des actions humaines dans l'objectif de fournir des probabilités *réalistes* (du point de vue de la sûreté) de l'efficacité d'une action humaine (en couplant pathogène et résilient);

- 3) Une fois intégrée la résilience et ses mécanismes dans la modélisation probabiliste et plus spécifiquement dans les RBs, l'identification *d'un traitement des avis d'experts* leur permettant de s'exprimer de façon relativement simple et cohérente avec la structure mathématique du modèle retenu. L'objectif est d'atteindre une estimation objective des paramètres du modèle RB.

Ces verrous techniques nécessitent d'être précisés pour identifier les éléments clés permettant de satisfaire le besoin de l'AiDR pour une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel.

1.3.3 Expression détaillée des verrous techniques pour l'évolution de l'AiDR

Les verrous techniques identifiés précédemment se structurent en trois items relatifs au besoin de l'AiDR:

1. Renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel ;
2. Intégration de la résilience dans l'AiDR;
3. Traitement des avis d'experts et estimation des paramètres du modèle probabiliste intégrant la résilience.

1.3.3.1 Renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel

De façon générale, la *robustesse* d'un modèle de prédiction peut être définie comme une mesure de sa fiabilité sous différents ensembles de conditions expérimentales (Confaloneri, *et al.* 2010).

Sur la base de cette définition générale, nous pouvons définir la robustesse du modèle de barrière humaine dans l'AiDR comme sa capacité à fournir des résultats fiables et réalistes indépendamment du contexte spécifique considéré.

Bien qu'aucune des expressions pour sa quantification ait été proposée, le manque de robustesse d'un modèle peut être expliquée par certains indicateurs comme l'incohérence dans certaines des relations mathématiques utilisées pour formaliser les connaissances ou par des lacunes dans la connaissance elle-même dans le modèle.

Dans le cas de l'AiDR, les concepteurs de l'AiDR pour produire des résultats fiables et surtout réalistes doivent assurer la robustesse de la connaissance formalisée dans le modèle. Cette robustesse se traduit donc par la pertinence, l'orthogonalité et l'exhaustivité des déterminants de ce modèle (Délégation, Formation, Aides, Expérience, Facteurs environnementaux, Capacité à respecter le cahier des charges, Gestion collective et dynamique de groupe, Contrôle et atteinte des objectifs, Rex). La pertinence ainsi que l'orthogonalité des variables actuellement utilisées et leurs liens avec les FOP (essentiels pour considérer les influences du contexte organisationnel) a été démontrée et validée après plusieurs expérimentations.

Cependant, d'autres facteurs pourraient renforcer l'exhaustivité des facteurs utilisés pour la caractérisation de l'efficacité de l'action humaine. En effet, la considération de l'humain est un axe de recherche important aussi dans d'autres domaines, notamment en EPFH (Mohaghegh et Mosleh, 2009), (Galàn, *et al.* 2007), (Lyons, *et al.* 2004), (Boring, 2010), (Gertman, *et al.* 2005) (Dang, *et al.* 2009) etc. En effet, une vaste communauté de chercheurs existe qui travaille et développe des méthodes d'EPFH basées sur la modélisation des composantes humaine et organisationnelle des systèmes sociotechniques pour l'analyse des risques. Plus spécifiquement une partie des méthodes EPFH reconnues suivent une approche basée sur l'utilisation systématique dans leurs analyses de facteurs d'influence dits **performance shaping factors** (PSF) (Embrey, 1983). Ces PSF sont définis par (Boring, 2010) comme les « caractéristiques qui déterminent la performance humaine et qui constituent un support pour quantifier la probabilité de l'erreur humaine ». Il s'agit de facteurs représentatifs des composantes humaines, du collectif de travail et organisationnelles d'un système sociotechnique.

Par conséquent, ces PSF pourraient être tout à fait pertinents pour la caractérisation de l'action humaine dans le modèle de barrière et de manière spécifique pour la caractérisation de l'efficacité d'une action humaine par les Déterminants.

Ce constat pour renforcer la robustesse du modèle de barrière humaine dans l'AiDR nous conduit à isoler les sous-verrous techniques suivants:

- a. Faire une analogie avec cette caractérisation de l'efficacité d'une action humaine par les déterminants de l'AiDR sur la base des développements faits en EPFH. Notamment, il faudrait se focaliser sur les méthodes EPFH qui possèdent les caractéristiques suivantes : une vision *sociotechnique*, une approche par PSF, une analyse des risques couvrant les niveaux *organisationnel* et du *collectif de travail* (approche multidimensionnelle), une approche de quantification de type bayésien basée principalement sur le recueil d'avis d'experts ;
- b. Isoler un ensemble de PSF pouvant conforter voire compléter cette caractérisation.

1.3.3.2 Intégration de la résilience dans l'AiDR

En l'état, l'approche AiDR se base sur une vision conservatrice plutôt pessimiste de la réalité au sens où certains aspects positifs ne sont pas représentés. En effet, l'efficacité des actions humaines est vue comme l'expression de l'interaction d'influences pathogènes multiples et combinées des facteurs organisationnels sur les déterminants de l'action. Comme il a été souligné dans le § 1.3, ces influences pathogènes témoignent de la dégradation de l'état de santé de l'organisation du système (Pierlot et Dien, 2007). La considération de cette pathogénicité due aux influences du contexte organisationnel est justifiée par les résultats et le REX issus des analyses conduites sur une petite centaine d'incidents/accidents dans des secteurs différents (nucléaire, pétrochimie, aviation, etc.). Elle s'est traduite par l'identification d'un ensemble de contributeurs majeurs, les FOP.

Aujourd'hui, les concepteurs de la méthode AiDR veulent adopter un point de vue plus « fidèle » à la réalité et par conséquent préconise de considérer les mécanismes résilients pouvant affecter positivement l'efficacité des actions humaines au sein d'une

barrière. Cette considération de la résilience permettrait de faire évoluer, ou du moins compléter, les représentations des barrières humaines et organisationnelles d'un système sociotechnique complexe utilisées dans la démarche existante de l'AiDR.

La **résilience** est un terme polysémique, qui, dans les détails, diffère d'un domaine à l'autre (Dauphiné et Provitolo, 2007). En écologie, elle représente la capacité d'un écosystème ou d'une espèce à retrouver un fonctionnement et/ou un développement normal après avoir subi un traumatisme (Mathevet et Bousquet, 2014); en économie, la résilience est plutôt la capacité à revenir sur la trajectoire de croissance après avoir encaissé un choc (Lindbom et Rothstein, 2007) ; en psychologie, la résilience est « la capacité d'une personne ou d'un groupe à se développer bien, à continuer à se projeter dans l'avenir, en dépit d'événements déstabilisants, de conditions de vie difficiles, de traumatismes parfois sévères » (Tisseron, 2014).

Selon (Anderies, *et al.* 2013), la résilience caractérise l'aptitude d'un système à résister à un ensemble de perturbations importantes et surtout imprévisibles. Pour (Pescaroli, 2014), elle se différencie de la notion de robustesse en ce sens que cette dernière considère plutôt le cas de perturbations prévisibles (au sens où on peut le caractériser qualitativement voire quantitativement) d'ampleur a priori modestes. Par ailleurs, les performances attendues dans les deux cas ne sont pas les mêmes : faire preuve de robustesse, « c'est pouvoir maintenir le niveau de performances alors que les conditions exogènes ont un peu changé » ; être résilient, c'est « rebondir pour retrouver son équilibre, c'est la poursuite de la viabilité, au mépris éventuel des performances » (Pescaroli, 2014).

Pourtant, c'est dans (Grotberg, 1996) que nous retrouvons une première définition précise de la résilience, à savoir « une capacité universelle qui permet à une personne, un groupe ou une communauté de prévenir, *réduire* ou *surmonter* les effets néfastes de l'adversité ». Dans l'ouvrage « Resilience engineering », (Hollnagel, *et al.* 2006) proposent différentes définitions de la résilience dont la plus consensuelle est la suivante: « la capacité d'un système ou d'une organisation à réagir et à récupérer après une perturbation, avec un minimum d'effet sur la stabilité dynamique du système ».

La résilience est donc largement associée à la capacité d'un système à « réduire les chances d'un choc, d'absorber un choc s'il se produit et de récupérer rapidement après un choc (rétablir la performance normale) » (Bruneau, *et al.* 2003).

Sur la base des définitions précédentes et plus particulièrement de celle proposée par (Bruneau, *et al.* 2003), dans nos travaux nous considérons la **résilience** comme composée de deux types de mécanisme distincts :

- | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none">- Les mécanismes de mitigation, pour réduire les effets négatifs (pathogènes) causés par les perturbations et/ou les chocs;- Les mécanismes de récupération, pour rétablir une condition nominale (acceptable) après l'occurrence des perturbations. |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Concernant particulièrement la résilience des systèmes sociotechniques, des nombreux travaux comme ceux de (Hollnagel, *et al.* 2006), (Jackson, 2009), (Jackson et Ferris, 2013), (Van der Leeuw et Aschan Leygonie, 2000), (Righi, *et al.* 2015) ont été menés ces dix dernières années (cf. Tableau 1.2).

Tableau 1.2. Cartographie des travaux examinés traitant de la résilience.

Journal	Nombre total de papiers	Théorie de la résilience	Analyse des accidents	Analyse des risques
Cognition, Technology and Work	19	15	4	1
Safety Science	23	11	8	4
Reliability Engineering and Systems Safety	31	17	5	9
Theoretical Issues in Ergonomics Science	5	4	1	-
Applied Ergonomics	4	4	-	-
Human Factors	3	2	2	2

A propos de la résilience dans les systèmes sociotechniques, (Hollnagel, 2006) écrit : « la sûreté et les risques des systèmes complexes ne peuvent pas être prédits ou modélisés sur la base des composants et de leurs interactions, d'autres propriétés doivent nécessairement être considérées aujourd'hui pour une évaluation réaliste ». Ce concept est souligné par (Pariès, 2006), qui voit la résilience comme « une propriété émergente des systèmes complexes ».

Cette propriété émergente peut s'identifier à plusieurs niveaux. En effet, la plupart des travaux retenus à partir des ceux examinés dans le Tableau 1.2 se réfèrent à quatre niveaux de **résilience d'un système sociotechnique**: la résilience au niveau de l'*individu* comme dans (Lo, Pluyter et Meijer, 2015), celle relative au le *collectif de travail* (Duff, *et al.* 2014) et enfin la résilience au niveau de l'*organisation* (Grecco, *et al.* 2013). Certains travaux comme dans (Gilly, *et al.* 2014) ont pris un point de vue encore plus large des systèmes en regardant comment les gouvernements influencent la résilience des grandes organisations.

Par rapport à ces niveaux de résilience dans les systèmes sociotechniques et en lien avec la définition de résilience que nous retenons, nous nous focalisons sur :

Les **mécanismes de mitigation** entre le niveau *organisationnel* et celui du *collectif de travail* dans un système sociotechnique.

La majorité des travaux que nous avons analysés soutiennent l'intérêt pour la résilience par le biais de la complexité (Righi, *et al.* 2015). Une vision de l'analyse des risques y est proposée pour prendre en compte la résilience. Pourtant, très peu de ces travaux montrent les enjeux et les conditions nécessaires pour qu'un système soit résilient face à des situations imprévisibles. Finalement, nous n'avons pas trouvé des travaux abordant la complexité de la **modélisation des mécanismes de mitigation** entre le niveau organisationnel et celui du collectif de travail d'un système sociotechnique dans le cadre probabiliste d'une évaluation des risques.

Par rapport à cette vision de la résilience dans les systèmes sociotechniques, nos investigations se structurent sur deux points :

- Comprendre l'origine des mécanismes résilients et définir des marqueurs permettant de tracer ces mécanismes ;

- Identifier une modélisation permettant de considérer à la fois les mécanismes résilients et pathogènes au sein d'un modèle unifié d'analyse de risques. La représentation de l'interaction de ces mécanismes causals sous la forme de relations probabilistes conditionnelles amène à traiter leur modélisation, leurs paramètres et leur élicitation par avis d'experts afin de les quantifier.

Aussi, nous nous focalisons sur les approches existantes permettant de représenter les mécanismes causals dans le cadre de modélisation probabiliste et de manière spécifique des réseaux bayésiens (RB).

Ces aspects sont à investiguer pour que le modèle de barrière puisse en tenir compte dans les analyses réalisées avec l'AiDR.

En ce sens, cette investigation requiert de s'attaquer aux sous-verrous technologiques suivants pour la considération de la résilience dans l'AiDR :

- a. Investiguer l'origine des **mécanismes résilients** en analysant leurs **marqueurs** ;
- b. Prendre en compte l'**interaction** entre **mécanismes résilients** et **mécanismes pathogènes** pour la traduire dans la modélisation probabiliste et plus spécifiquement dans un modèle RBs pour être appliqué dans le cadre de l'AiDR.

1.3.3.3 Traitement des avis d'experts et estimation des paramètres du modèle probabiliste intégrant la résilience

Les jugements d'experts dans l'AiDR servent à quantifier les distributions de probabilité des TPC du modèle RB. Le recours à l'élicitation d'experts est nécessaire en raison de l'insuffisance du retour d'expérience pour l'analyse des aspects humains et organisationnels de la barrière humaine.

Ces aspects d'élicitation d'experts se combinent à la complexité d'estimation des paramètres du modèle RB. En effet, ces paramètres font partie de la structure mathématique du modèle retenu pour intégrer la résilience. A ce propos, nous rappelons que par rapport à cette intégration de la résilience, nous focalisons notre attention sur les mécanismes de *mitigation* ayant lieu en présence des mécanismes pathogènes pouvant potentiellement produire une action inefficace et par conséquent des scénarios accidentels/incidents. Les experts sont ainsi appelés à s'exprimer sur des situations *connues* dans lesquelles les mécanismes de mitigation peuvent réduire les chances de l'inefficacité des actions humaines causée par les mécanismes pathogènes.

Ainsi, pour consentir une évaluation réaliste de la probabilité d'efficacité des barrières humaines prises dans leur contexte organisationnel, les sous-verrous techniques suivants sont à aborder :

- a. Élaborer un **protocole d'élicitation d'experts** de manière à ce qu'ils puissent se confronter avec un nombre réduit de situations et à des situations peu complexes à évaluer ;
- b. Fonder les jugements des experts sur une **échelle d'élicitation crédible**, c'est-à-dire reconnue dans la communauté d'EPFH;
- c. Estimer les paramètres à partir des jugements d'experts de façon cohérente avec la structure du modèle intégrant la résilience.

1.4 Conclusions

Dans ce premier chapitre, nous avons présenté le contexte EDF en termes de management des risques dans lequel s'inscrivent nos travaux.

La notion de risque est précisée en la centrant sur le risque industriel, car lié à l'occurrence d'accidents et incidents dans les systèmes sociotechniques complexes tels que les centrales de production d'électricité d'EDF. Pour ces systèmes il a été mis en évidence le besoin d'une synergie entre les différentes compétences du management des risques pour en consentir une vision multirisque. Face à ce besoin émergent d'approches multirisques pour les systèmes sociotechniques, nous nous focalisons sur une problématique particulièrement importante à EDF, celle de l'intégration de l'homme et de l'organisation dans une approche globale d'analyse des risques.

Dans ce contexte, l'AiDR développée conjointement par EDF R&D et le CRAN, est une méthode outillée grâce à l'utilisation des Réseaux Bayésiens pour la modélisation et la quantification des influences causales de type pathogène liant la dimension organisationnelle à celle du collectif de travail dans un modèle dit de barrière humaine.

La mise en œuvre de l'AiDR sur des cas industriels réels a montré la crédibilité de l'approche. Toutefois, les résultats de ces applications ne sont pas complètement satisfaisants dans cette considération de l'homme et de l'organisation pour produire des fréquences réalistes en termes de disponibilité et sûreté des systèmes modélisés. Pour améliorer cette considération des aspects liés aux composantes humaine et organisationnelle des systèmes sociotechniques, une évolution du modèle de barrière humaine dans l'AiDR est nécessaire. Plus particulièrement, les utilisateurs de l'AiDR doivent passer de la hiérarchisation relative des actions entre elles à la quantification probabiliste de l'efficacité d'une barrière humaine éventuellement prise dans son contexte organisationnel. Relativement à ce besoin principal, les verrous techniques suivants ont été identifiés pour réaliser cette évaluation probabiliste réaliste :

1. Renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel ;
2. Intégration de la résilience dans l'AiDR ;
3. Traitement des avis d'experts et estimation des paramètres du modèle probabiliste intégrant la résilience.

Dans le chapitre 2 nous examinons ces trois verrous d'un point de vue scientifique sous la forme d'un état de l'art afin d'identifier les verrous scientifiques qui y sont liés.

Cependant, comme le traitement des avis d'expert et l'estimation des paramètres dépendent du modèle mathématique retenu pour intégrer les mécanismes résilients, dans cet état de l'art nous considérons les verrous 2 et 3 simultanément.

Chapitre 2

Des verrous techniques aux verrous scientifiques pour une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

2 Des verrous techniques aux verrous scientifiques pour une évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Introduction

Ce chapitre a pour objectif de faire un « état de l'art » scientifique en regard des trois verrous techniques identifiés dans le chapitre 1 afin d'isoler les verrous scientifiques à lever qui s'y rapportent. Plus particulièrement, cet état de l'art traite du renforcement de la robustesse de la barrière humaine, puis de l'intégration de la résilience dans l'AiDR simultanément au traitement des avis d'expert et l'estimation des paramètres tels que nous l'avons défini au chapitre 1. L'examen conjoint des verrous concernant l'intégration de la résilience dans l'AiDR et le traitement des avis d'expert et l'estimation des paramètres s'impose, car le traitement des avis d'expert et l'estimation des paramètres sont fortement liés au modèle mathématique retenu pour intégrer les mécanismes résilients.

Ainsi, dans la première partie de ce chapitre, nous nous intéressons au renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel. Dans un premier temps, nous considérons la caractérisation de l'efficacité d'une action humaine par ses déterminants sur la base des développements faits en EPFH. Notamment, nous étudions les méthodes EPFH et plus spécifiquement celles qui sont les plus pertinentes à traiter ce verrou. Nous focalisons ensuite sur celles qui possèdent certaines caractéristiques comme un cadre conceptuel de type sociotechnique, une approche de quantification bayésienne. Nous proposons une étude comparative sur cet ensemble de méthodes afin d'isoler les PSF pouvant renforcer la robustesse de la barrière humaine.

Ce chapitre se poursuit en traitant de l'intégration de la résilience dans l'AiDR. Nous investiguons l'origine des mécanismes résilients, *i.e.* les marqueurs qui permettent d'identifier ces mécanismes. Ensuite, nous nous intéressons à la prise en compte de l'interaction entre les mécanismes résilients et pathogènes dans

le cadre probabiliste et plus spécifiquement dans un modèle RB d'une barrière humaine. Nous abordons ces aspects de considération de la résilience en lien avec l'élicitation d'experts et l'estimation des paramètres d'un modèle probabiliste basée sur les jugements d'experts à intégrer de façon cohérente avec la structure du modèle intégrant la résilience.

2.2 Renforcement de la robustesse du modèle de barrière humaine prise dans son contexte organisationnel

Avant de positionner les méthodes EPFH par rapport aux sous-verrous techniques identifiés à la fin du chapitre 1, nous proposons de les présenter par rapport à leurs modèles et cadres théoriques sous-jacents.

2.2.1 Présentation des méthodes d'évaluation probabiliste de la fiabilité humaine

Le développement de la compétence d'EPFH répond à la nécessité de faire face à l'évolution du contexte technique, économique, humain et organisationnel lié aux activités industrielles à risques telles que celles de l'industrie nucléaire, pétrochimique et aéronautique (Mosleh, *et al.*, 2007). Depuis quelques décennies, cette évolution a mis en évidence le rôle important des facteurs humains et socio-organisationnels sur la sûreté et la disponibilité des systèmes complexes où de telles activités sont exploitées (Galàn, *et al.* 2007).

Ce constat a induit – à partir des années 70 – la définition d'un ensemble de méthodes EPFH dans l'objectif de comprendre et analyser les modes de fonctionnement de l'humain et d'intégrer les effets de l'*erreur humaine* sur la disponibilité et la sûreté des systèmes (Reason, 1990). Cette notion d'erreur humaine a évolué dans le temps. Ainsi, différentes approches théoriques ont été proposées se focalisant sur différents aspects. Les modes de cognition, l'intention, l'autonomie dans la prise de décision et le contrôle sont les thématiques principales abordées dans l'approche *cognitiviste* alors que pour l'approche de type *ergonomique* c'est la notion d'interaction homme-machine qui est centrale.

Ces différentes approches théoriques ont conduit à des véritables *modèles* de l'humain (cf. Tableau 2.1) employés dans le cadre des analyses d'EPFH.

Tableau 2.1. Principales approches théoriques et les modèles associés pour supporter l'EPFH

Approche	Contributeurs	Description	Modèles
----------	---------------	-------------	---------

Behavioriste	Swain et Guttman (1983)	Approche « pragmatique » centrée sur les mécanismes comportementaux de l'individu	Modèle de l'homme comme « boîte noire ». Entre le stimulus extérieur et ses réponses ou modes comportementaux
Cognitiviste	Rasmussen (1986)	Approche « heuristique » sur l'erreur individuelle en cours d'action	Modèle Skills Rules Knowledge (S-R-K) . Contrôle cognitif humain et structuré, hiérarchisé, adaptable et capable de traiter les informations en trois « niveaux » différents : le niveau fondé sur les -habiletés (skills), le niveau fondé sur les règles (rules) et le niveau fondé sur les connaissances (knowledge).
	Reason (1990)	Classification symptomatique et pratique de l'erreur humaine limitée aux actions « en sortie » seulement. Les erreurs de « commission » ne sont pas prises en compte	Generic Error Modeling System (GEMS) . Tente de déterminer l'origine des types fondamentaux d'erreurs humaines. Classe les erreurs selon le caractère intentionnel (violations) ou non-intentionnel (ratés, lapsus, fautes).
	Hollnagel (1998)	Approche à la base du développement de la méthode CREAM	Contextual Control Model (COCOM) . Distinction entre modèle de compétence et modèle de contrôle, entre phénotypes corrélés aux erreurs observables et génotypes corrélés aux causes préexistantes
Ergonomique	Leplat (1993)	Approche visant à mettre en relation l'homme et ses moyens, ses méthodes avec le milieu de travail (IHM). Appliquée à la conception des systèmes, elle contribue à la sécurité et l'efficacité des systèmes	Aucun modèle
Sociotechnique	Amalberti (1996)	Approche mixte qui considère le système technique et humain/organisationnel comme un tout	Aucun modèle
Ingénierie de la Résilience	Hollnagel (2006)	Approche qui se focalise sur la capacité de l'homme et l'organisation à contribuer positivement à la sûreté d'un système sociotechnique	Functional Resonance Analysis Method (FRAM) (Hollnagel, <i>et al</i> 2006)
	Bieder (2006)		Modèle de « barrière de sécurité » (Bieder, 2006)

L'approche théorique et le ou les modèles sous-jacents constituent un cadre conceptuel en fonction duquel les méthodes EPFH peuvent se regrouper en deux ou trois générations. Ce regroupement permet de répartir les méthodes sur une classification conventionnellement acceptée à l'heure actuelle et utilisée par exemple dans (Bell et Holroyd, 2009) et (Mosleh, *et al.* 2006).

Focalisée sur la notion d'erreur humaine, les méthodes de *première génération* sont apparues les premières et s'appuient sur l'hypothèse du caractère générique du comportement mécanique (propre à la vision behavioriste) de l'humain en modélisant les interventions des opérateurs par une description décomposée et simplifiée des actions.

Les méthodes de *deuxième génération* (apparues à partir des années 90) s'intéressent davantage aux processus cognitifs permettant d'expliquer et de prédire les défaillances humaines. Selon (NRC, 2005), elles tiennent compte des écarts entre la représentation faite du comportement de l'homme et la réalité (grâce notamment à l'apport d'autres domaines comme l'ergonomie ou la psychologie). Entre autres, elles cherchent à considérer le contexte et les « *errors of commission* ». Dans (Kirwan, *et al.* 2005) il est jugé que les plus notables des méthodes de deuxième génération sont ATHEANA, CREAM et MERMOS, mais que MERMOS est la seule qui est en utilisation régulière (Pesme et Le Bot, 2007).

Les méthodes de *troisième génération*, dont la plus représentative est la méthode NARA (Kirwan, *et al.* 2005), constituent un développement de certaines méthodes de première et deuxième génération. Elles se distinguent par une prise en compte plus importante du contexte organisationnel et environnemental.

Sur la base de ce classement en première, deuxième et troisième génération, nous avons isolé des méthodes EPFH les plus pertinentes à couvrir l'investigation menée par rapport aux sous-verrous identifiés dans le chapitre 1. Ces méthodes ont été choisies principalement en regard de leur couverture d'un ensemble de critères tels que l'applicabilité aux systèmes sociotechniques, la possibilité de quantifier la probabilité d'erreur humaine, la mise en œuvre d'un ensemble de PSF pour la quantification et aussi la popularité de la méthode (comme indicateur de sa crédibilité).

Ces méthodes présélectionnées sont listées dans le Tableau 2.2.

Tableau 2.2. Liste des méthodes EPFH retenues dans cet état de l'art.

Méthode	Génération	Références	Date
THERP	1 ^e	Swain et Guttman – NRC	1983
SLIM-MAUD	1 ^e	Embrey et Kirwan – NRC	1984
HCR	1 ^e	Hannaman et Spurgin – EPRI	1984
TRIPOD-Delta	1 ^e	SHELL	1985
HEART	1 ^e	Williams – NRC	1988
CREAM	2 ^e	Hollnagel – Halden Lab	1994

ATHEANA	2 ^e	Cooper <i>et al.</i> – NRC	1996
MERMOS	2 ^e	Le Bot <i>et al.</i> – EDF R&D	1998
SPAR-H	2 ^e	Gertman <i>et al.</i> – NRC	1999
NARA	3 ^e	Kirwan, <i>et al.</i> - British Energy	2005

Les méthodes du Tableau 2.2, qui ne sont qu'une présélection de méthodes, peuvent ainsi être positionnées par rapport à leur cadre conceptuel (cf. Tableau 2.1) et classées en trois générations. Toutefois, les classifications proposées jusqu'ici ne nous permettent pas de dégager totalement les éléments clés associés à ces sous-verrous.

D'autres types de classification existent notamment celle proposée dans (Dang, *et al.* 2009). Il s'agit d'une taxonomie distinguant les méthodes soit *facto-rielles*, soit *contextuelles* ou bien *basées sur le jugement d'experts*. Sur la base de la classification de (Dang, *et al.* 2009) conjointement à l'utilisation d'un ensemble d'attributs issus d'une revue approfondie de la littérature sur les méthodes EPFH dont (Mosleh, *et al.* 2006) et (Boring 2010), nous avons pu effectuer une investigation plus détaillée des méthodes du Tableau 2.2 (cf. Annexe I).

Nous avons ensuite proposé d'affiner cette classification en nous focalisant sur les méthodes EPFH qui possèdent certaines caractéristiques importantes pour le développement de l'AiDR, notamment :

- Par rapport au Tableau 2.1, un cadre conceptuel dont l'approche théorique soit de type *sociotechnique* ;
- Une approche de *quantification bayésienne* basée principalement sur le recueil et le traitement d'avis d'experts ;
- Une couverture des PSF telle qu'elle permette une caractérisation de l'action sur le plan de l'*organisation* et du *collectif de travail*.

En ce sens, nous proposons d'analyser les méthodes du Tableau 2.2 sur la base de 3 **critères spécifiques** (en lien avec les caractéristiques précédentes) qui sont à décliner selon la taxonomie suivante (cf. Tableau 2.3).

Tableau 2.3. Critères d'analyse des méthodes EPFH (De Galizia, *et al.* 2015b,c).

Objet du critère	Taxonomie
Cadre conceptuel	- Behaviouriste
	- Cognitiviste
	- Ergonomique
	- Sociotechnique
	- Ingénierie de la résilience
Approche quantitative	- Fréquentiste
	- Bayésienne
Couverture des PSF	- Individu
	- Collectif de travail
	- Organisation
	- Environnement

Ces critères se justifient par rapport à des exigences en lien avec les verrous techniques.

Critère n° 1 – Cadre conceptuel

Ce critère permet de déterminer le périmètre des systèmes étudiés et donc les disciplines impliquées dans la démarche. De manière générale, l'ensemble de ces travaux s'accordent sur le fait qu'il est réducteur d'étudier l'efficacité des actions humaines uniquement sous l'angle des aspects individuels cognitifs puisque cela revient à passer sous silence toutes les interactions qui existent entre les différentes sphères humaines, technique, organisationnelle et environnementale et leur influence sur la réussite d'une action humaine.

Nous considérons donc les cadres conceptuels issus du Tableau 1 et positionnons les méthodes du Tableau 2 par rapport à une taxonomie basée sur les items suivants (cf. Tableau 3, Cadre conceptuel) :

- **Behaviouriste** ;
- **Cognitiviste** ;
- **Ergonomique** ;
- **Sociotechnique** ;
- **Ingénierie de la Résilience**.

Critère n° 2 – Approche quantitative

Ce critère permet de déterminer le traitement mis en œuvre pour la quantification des probabilités. En effet, dans le cadre probabiliste deux conceptions sont possibles :

- d'une part des probabilités *fréquentistes* : il s'agit de probabilités dites objectives et a posteriori, visant à dégager les lois stochastiques de processus aléatoires tendanciels dans des statistiques de fréquence à long terme ;
- et d'autre part des probabilités *bayésiennes*, où il s'agit de probabilité subjective, de degré de certitude a priori. Les probabilités subjectives sont ainsi présentées comme étant des « raisons de croire », c'est-à-dire de présenter le calcul des probabilités comme relevant d'une théorie de la connaissance plus que d'une théorie de la nature (Desrosières, 2016).

(Desrosières, 2002) a également souligné que : « La distinction entre les deux faces de la notion de probabilité ne sera formulée explicitement qu'au 19^e siècle, par exemple par Cournot (Exposition de la théorie des chances et des probabilités, 1843) qui opposera « probabilité » et « chance », ou encore : « les probabilités objectives, qui donnent la mesure de la possibilité des choses, et les probabilités subjectives, relatives en partie à nos connaissances, en partie à notre ignorance, variables d'une intelligence à l'autre, selon leurs capacités et les données qui leurs sont fournies ».

Ainsi, d'un côté il y a l'**approche fréquentiste** qui s'intéresse à l'occurrence d'un événement parmi un nombre total et significatif d'observations, et de l'autre l'**approche bayésienne** qui mesure le degré de connaissances subjectives et représentent la traduction chiffrée d'un état de connaissance. En effet, la formule de Bayes (Stigler, 1982) fait apparaître une distinction et une articulation entre probabilité a priori et probabilité a posteriori.

Selon (Diez et Druzdzel, 2006) la notion de probabilité a priori est intéressant puisque cette probabilité est souvent l'avis de l'expert ou du 'sachant' qui donne une première estimation (a priori). La formule bayésienne laisse donc une part à l'intuition, ce que se refusent les probabilités fréquentistes qui ne traitent que des états de nature (Desrosières, 2008).

Ainsi, en lien avec ce critère, les méthodes du Tableau 2 sont classées par rapport à la taxonomie suivante (cf. Tableau 3, Approche Quantitative) :

- **Fréquentiste**
- **Bayésienne.**

Critère n° 3 – Couverture des PSF

Ce critère sert à identifier le périmètre couvert par les PSF.

Tout au long des phases qualitatives et quantitatives, ces facteurs sont utilisés dans un objectif de quantification de la *probabilité d'erreur humaine* (HEP). Par exemple, dans la méthode SPAR-H, ils sont employés en tant que facteurs multiplicateurs d'une HEP nominale. Lorsque les PSF représentent un effet positif, les différents niveaux d'effet pour les PSF correspondent à une valeur inférieure à l'unité. La multiplication d'une HEP nominale par cette valeur fractionnaire associée au PSF sert à réduire la HEP globale. Lorsque les PSF représentent un effet négatif, les différents niveaux d'effet pour les PSF correspondent à une valeur supérieure à l'unité. La multiplication des HEP nominaux par ce nombre entier positif sert à augmenter la HEP globale. Lorsque les PSF sont supposés n'avoir aucun effet, le multiplicateur de PSF est égal à 1, ce qui n'induit aucune augmentation ni diminution de la HEP nominale.

Ces principes de multiplication des PSF et leurs effets sur la HEP nominale sont résumés en Figure 2.1.

$$HEP_{totale} = HEP_{nominale} \times PSF \begin{cases} 0 < PSF < 1 \Rightarrow HEP_{totale} < HEP_{nominale} \\ PSF = 1 \Rightarrow HEP_{totale} = HEP_{nominale} \\ PSF > 1 \Rightarrow HEP_{totale} > HEP_{nominale} \end{cases}$$

Figure 2.1. Schéma générique d'application des PSF (Dang, *et al.*, 2009).

Néanmoins, la plupart des méthodes ne semblent pas couvrir une gamme adéquate de PSF ou tentent de prédire l'efficacité de l'équipe de travail pour toutes les circonstances sans prendre en compte les influences du contexte organisationnel (Galàn, *et al.* 2007). En particulier, certains aspects importants d'un scénario accidentel ne sont pas toujours capturés par ceux qui ne considèrent que l'individu, comme dans les méthodes de première génération (Hollnagel, *et al.* 2006).

Ainsi, nous suggérons de mettre en évidence le niveau de couverture des PSF en classant les méthodes du Tableau 2.2 par rapport à la taxonomie suivante (cf. Tableau 2.3, Couverture des PSF) :

- **Individu ;**
- **Collectif de travail ;**
- **Organisation ;**
- **Environnement.**

Ces trois critères spécifiques sont primordiaux afin d'identifier les méthodes pouvant résoudre le problème du renforcement de la robustesse de la barrière humaine. Aussi, nous les avons appliqués (cf. Tableau 2.4) aux méthodes du Tableau 2.2.

Tableau 2.4. Analyse des méthodes EPFH sur la base des critères du Tableau 2.3.

Méthode	Cadre Théorique	Approche Quantitative	Couverture des PSF
THERP	Behaviouriste	Fréquentiste	Individuel
SLIM-MAUD	Behaviouriste	Fréquentiste	Individuel
HCR/ORE	Behaviouriste	Fréquentiste	Individuel, Collectif travail
HEART	Cognitiviste	Fréquentiste	Individuel, Collectif travail
CREAM	Cognitiviste	Fréquentiste	Individuel, Collectif travail
ATHEANA	Ergonomique Sociotechnique	Bayésienne	Collectif travail, Environnement
MERMOS	Ergonomique Sociotechnique	Fréquentiste	/
SPAR-H	Sociotechnique	Bayésienne	Collectif travail, Organisation, Environnement
NARA	Sociotechnique	Bayésienne	Collectif travail, Organisation, Environnement

Sur la base de l'analyse et par rapport aux trois critères proposés, nous avons mis en évidence (trame de fond en vert) trois méthodes possédant les caractéristiques nécessaires pour résoudre ce premier verrou de l'AiDR, à savoir :

- Une approche de type sociotechnique ;
- Une approche de type bayésien pour la quantification (avis d'experts en entrée);
- Une couverture large des facteurs d'influences, orientée à la prise en compte des dimensions organisationnelles, environnementales et du collectif de travail.

Ces méthodes sont :

- **NARA** (Nuclear Action Reliability Assessment) (Kirwan, *et al.* 2005) ;
- **SPAR-H** (Standardized Plant Analysis Risk - Human Reliability Assessment) (Gertman, *et al.* 2005) ;
- **ATHEANA** (A Technique for Human Event Analysis) (Cooper, *et al.* 1996).

Cette comparaison est une première étape qui doit être poursuivie en se focalisant sur les PSF utilisés par ces méthodes.

2.2.2 Extrapolation des PSF

Pour isoler l'ensemble de PSF pouvant renforcer la caractérisation de l'efficacité de l'action humaine dans l'AiDR, nous proposons de comparer les PSF utilisés dans les méthodes issues du Tableau 2.2 (De Galizia, *et al.* 2015a,b). La NRC dans son rapport sur les bonnes pratiques en matière d'EPFH (NRC, 2005) a indiqué comme terme de comparaison supplémentaire un ensemble de champs d'appartenance des PSF (cf. Tableau 2.5, NRC « Good practices »). Les experts en fiabilité humaine dans les domaines tels que le nucléaire et l'aviation ont en effet regroupé les champs d'appartenance des PSF plus pertinents et permettant une couverture exhaustive lors des analyses d'EPFH.

Tableau 2.5. Comparaison des PSF des méthodes EPFH retenues du Tableau 2.2.

NRC « Good Practices »	NARA	SPAR-H	ATHEANA
<i>Training and Experience</i>	Operator experience	Experience/Training	Formation
<i>Procedures/Administrative Controls</i>	--	Procedures Fitness and Availability	Procedures Quality
<i>Instrumentation</i>	--	--	Material
<i>Time Available</i>	--	Available Time	--
<i>Work Design/Complexity</i>	Work organization	Work Processes Complexity	Tasking and management
<i>Workload/Time Pressure/Stress</i>	Time pressure	Stress/Stressors	Maintenance
<i>Team/Crew dynamics</i>	Team coordination and cohesion	Leadership	Team composition
<i>Available Staffing</i>	--	--	--
<i>Human-System Interface</i>	--	Ergonomics/HMI	HMI Conception
<i>Environment</i>	Poor environment	Work Processes	Error producing circumstances

<i>Accessibility/Operability of Equipment</i>	Adequacy of HMI and Operational Support	--	Equipment properly settled and well arranged
<i>Need for Special Tools</i>	--	--	--
<i>Communications</i>	Quality of Information	Work Processes	Availability of Communication tools
<i>Special (Equipment) Fitness Needs</i>	--	--	--
<i>Safety culture (consideration of 'realistic' accident Sequence Deviations/Deviations)</i>	Safety Culture	--	Quality of safety culture

Parmi ces champs d'appartenance des PSF proposés dans les bonnes pratiques de la NRC, nous avons en mis en évidence un ensemble d'intérêt pour renforcer la robustesse de la barrière humaine (trame de fond en vert). Nous avons aussi souligné (trame de fond en rouge) les champs qui concernent des aspects non considérés dans les analyses de l'AiDR, comme le temps.

Ainsi, par rapport aux champs d'intérêt, nous avons isolé les PSF des trois méthodes considérées qui couvrent le mieux les aspects sous-jacents à ces champs (trame de fond en jaune). Dans deux cas, les méthodes ne couvrent pas les champs NRC – « Need for Special Tools » et « Special (Equipment) Fitness Needs » – et donc nous avons considéré ces champs eux-mêmes comme des PSF à retenir.

Au global, nous avons isolé un ensemble de ces PSF (De Galizia, *et al.* 2015a,b) et fait une traduction de l'anglais au français avec une description synthétique associée pour chaque PSF (cf. Tableau 2.6).

Tableau 2.6. Ensemble de PSF retenus à partir du Tableau 2.5.

Nom du PSF	Description
Cohésion	Capacité à agir comme une équipe, à sacrifier son propre jugement en faveur d'un bon esprit dans le groupe.
Coordination	Niveau auquel les différents rôles et responsabilités individuels sont précisés pour chaque membre de l'équipe.
Composition	Taille, homogénéité/hétérogénéité et compatibilité de l'équipe. Stabilité de l'équipe ou taux de rotation (ex. quand des membres sont remplacés par des membres moins qualifiés, la performance de l'équipe est susceptible de se dégrader).
Disponibilité de communication	La suffisance et la disponibilité des moyens de communication. La proximité physique des opérateurs aux équipements de communication (le cas échéant). Défaillance ou indisponibilité fonctionnelle des systèmes de communication. La probabilité que l'équipement de communication soit sans pilote.

Qualité de communication	Signal déformé ou dégradation en raison de défaillances de l'équipement (par exemple, équipements vieux ou mal entretenus). Anomalies dans la communication ou défauts humains (ex. accent lourd, ambiguïté linguistique, instructions pas claires, etc).
Leadership	Expérience et fiabilité du décideur. La qualité du leadership se mesure par trois éléments: donner les orientations, gagner l'engagement de l'équipe, et surmonter les obstacles pour changer la situation. Le niveau d'engagement se mesure en trois dimensions: identification avec le travail, les collègues et l'organisation.
Conception du processus de travail, affectation des tâches et management	Programmation du travail. Description exhaustive et exacte de la tâche. Rapidité d'affectation du travail aux membres de l'équipe.
Interface Homme-Machine	Qualité de la conception du système pour la facilité et la précision de la perception visuelle, sonore, cognitive de l'information. Pertinence de la répartition du travail entre automates et contrôle manuel de l'opérateur.
Culture et Qualité de Sécurité	Emphase sur la politique qualité / sécurité de l'organisation. Engagement de l'état-major pour assurer une bonne gestion de la sécurité. En réponse, engagement des opérateurs. Violations et erreurs enregistrées dans la durée. Enquêtes sur les accidents ou presque-accidents.
Environnement (de travail et physique)	Espace de travail suffisant et habitable (<i>i.e.</i> , éclairage, température, humidité, vibrations, bruit). Conditions de l'environnement physique.
Disponibilité d'outils spéciaux	Outils nécessaires disponibles, bien organisés et accessibles.
Pertinence et qualité des outils spéciaux	Disponibilité d'outils spéciaux pensés pour certaines tâches spécifiques.
Disponibilité des procédures	Existence et accessibilité des procédures. Accessibilité du contenu (<i>i.e.</i> , indexation des documents).
Pertinence et qualité des procédures	La fidélité des documents (<i>i.e.</i> , adéquation du niveau de détail, exhaustivité des procédures et correspondance à des tâches réelles), lisibilité (<i>i.e.</i> , la mise en page). Ergonomie, facilité à distinguer différentes procédures.

La pertinence des PSF résumés dans le Tableau 6 a été démontrée par le fait qu'ils couvrent les champs d'appartenance indiqués par la NRC et car ils sont partagés, pour la plupart d'entre eux, par toutes les méthodes considérées. Donc, ils peuvent contribuer à renforcer la robustesse du modèle de barrière humaine d'un point de vue de l'exhaustivité de la caractérisation de l'action par ces déterminants.

Cependant, dans l'état actuel, ces PSF ne répondent que partiellement à ce besoin et nécessitent d'être formalisés dans un cadre spécifique tel que celui imposé par la structure du modèle de barrière humaine de l'AiDR.

Ceci nous amène par conséquent à définir un premier verrou scientifique à lever pour répondre au verrou technique identifié dans le chapitre 1.

Verrou scientifique n° 1. Formalisation des PSF pour leur intégration dans la structure du modèle de barrière humaine de l'AiDR.

Ce verrou scientifique étant posé, il faut s'intéresser maintenant au verrou technique concernant l'intégration de la résilience.

2.3 Intégration de la résilience dans l'AiDR

L'intégration de la résilience dans le cadre probabiliste et plus spécifiquement dans les RB est d'autant plus compliquée qu'il y a beaucoup de paramètres à estimer (via élicitation d'experts). Ce double verrou nous amène à traiter les deux aspects en un seul.

Concernant la résilience, un premier sous-verrou à examiner consiste en l'élaboration des marqueurs des mécanismes résilients.

2.3.1 Elaboration des marqueurs des mécanismes résilients

Dans cette thèse, les **marqueurs résilients** se définissent comme toutes informations utiles pour retracer des manifestations de résilience.

Sur ce point, l'ingénierie de la résilience (Hollnagel, *et al.* 2006) suggère d'analyser les événements menés positivement (les succès) pour en tirer des informations sur les mécanismes résilients qu'ont eus lieu et ayant éventuellement conduit à un déroulement positif du scénario.

Malheureusement, cette approche est peu pertinente pour les industries à risque où la plupart des connaissances disponibles provient essentiellement du retour d'expérience « négatif », c'est-à-dire obtenu après avoir examiné les scénarios accidentels/incidents ou les presque-événements (Paries, 2006). En outre, dans des domaines tels l'industrie nucléaire, l'analyse des risques dispose de peu d'informations pour les scénarios imprévisibles (OCDE, 2003). Ce manque d'informations contraint à quitter l'ingénierie de la résilience pour l'investigation des mécanismes résilients.

Cependant, plusieurs auteurs ont confirmé l'importance d'isoler des facteurs de la résilience, comme (Woods et Cook, 2009), (Back, *et al.* 2008) et (Van der Leeuw et Aschan Leygonie, 2000). Dans (Back, *et al.* 2008), un cadre général est proposé où des « stratégies résilientes » sont définies en fonction du niveau de granularité (individuel, petite équipe, niveau du système, etc.). Néanmoins, ces stratégies ne reposent pas sur l'analyse des marqueurs.

De cet état de fait, se construit un deuxième verrou scientifique qu'il faut nécessairement résoudre.

Verrou scientifique n°2. Elaboration et formalisation des marqueurs résilients.

En effet, ce verrou scientifique nous permet de justifier, sur un plan théorique, la présence de mécanismes résilients de la même manière que les marqueurs pathogènes autorisent les concepteurs de l'AiDR à considérer les mécanismes pathogènes éventuellement identifiés.

Nous nous focalisons donc sur la considération conjointe de mécanismes pathogènes et résilients et leur interaction dans le cadre probabiliste.

2.3.2 Interaction entre mécanismes résilients et pathogènes dans le cadre probabiliste

Après avoir mis en évidence le besoin d'élaborer et formaliser les marqueurs résilients, il faut considérer la résilience sous la forme de mécanismes en interaction avec les mécanismes pathogènes dans un contexte de modélisation probabiliste de relations causales.

À ce jour, très peu de travaux existent qui abordent la complexité de la modélisation de la résilience et les interactions entre mécanismes résilients et pathogènes dans un modèle global, notamment dans le cadre des relations non-déterministes d'un modèle probabiliste.

Cette complexité de modélisation d'une interaction entre mécanismes résilients et pathogènes dans un modèle probabiliste nous conduit à traiter plus génériquement les mécanismes de causalité (Anderies, *et al.* 2013) (Mathevet et Bousquet, 2014).

En effet, la notion de causalité est fondamentale dans la modélisation des connaissances, notamment pour l'ingénierie des systèmes (Diez et Druzdzal, 2006). Pendant la construction d'un modèle probabiliste tel que les RB, l'effet de l'influence de certaines variables parent sur une variable fille doit être quantifié en termes probabilistes. Souvent, la direction de l'arc établi entre la variable *parent* et la variable *fille* reflète le sens de l'influence causale. C'est ainsi qu'elle est perçue par le décideur ou l'analyste.

Les experts se réfèrent aux relations de *cause à effet* et donnent souvent une interprétation de causalité dans le processus de quantification des distributions de probabilités conditionnelles (Diez et Druzdzal, 2006). Parfois, la direction de la relation reflète simplement la direction dans laquelle l'éllicitation des probabilités conditionnelles est la plus facile.

Pour résumer, d'une part le cadre de causalité est utilisé pour représenter l'interaction entre les variables d'un modèle probabiliste, par exemple basé sur le formalisme RB ; d'autre part, les modèles probabilistes graphiques, tels que le

RB, permettent de représenter le point de vue subjectif d'un décideur ou d'un expert afin de faciliter l'élicitation (Druzdzet et Simon, 1993).

Ainsi, pour réduire la complexité de l'élicitation lors de la modélisation des mécanismes causals, des modèles canoniques ont été développés simplifiant la construction des modèles probabilistes. En fonction des hypothèses faites en termes de dépendance entre les mécanismes causals, ces modèles permettent d'avoir un nombre plus ou moins important de paramètres pour la définition des relations entre les variables. Le terme canonique est utilisé, car les modèles sont vus comme des unités élémentaires dans la construction de modèles plus complexes.

Il faut donc considérer les mécanismes causals qui peuvent être exprimés par ces modèles probabilistes canoniques afin de mettre en évidence une possibilité de modélisation de l'interaction entre mécanismes pathogènes et résilients.

2.3.3 Les modèles canoniques probabilistes

En général, les *modèles canoniques probabilistes* peuvent se classer en trois catégories :

- les modèles canoniques **déterministes** (MCD) ;
- les modèles canoniques **simples** (MCS) ;
- les modèles à Influences causales **Indépendantes** (ICI).

2.3.3.1 Les Modèles Canoniques Déterministes (MCD)

Les **modèles canoniques déterministes** (MCD) permettent d'établir des relations de causes à effets de type *déterministe*. Considérons un modèle RB constitué par une famille de variables aléatoires binaires dont n variables parent ou cause $\{X_1, X_2 \dots X_n\}$ et une variable fille ou effet Y . La valeur prise par Y est calculée en fonction du vecteur \underline{x} des valeurs des X_i $y = f(x_1, x_2 \dots x_n)$, et la TPC se construit de la façon suivante :

$$P(y|x) = \begin{cases} 1 & \text{si } y = f(\underline{x}) \\ 0 & \text{autrement} \end{cases}$$

Par conséquent, l'avantage principal des MCD est qu'ils n'ont pas besoin de paramètres numériques, car la TPC se base simplement sur la définition de la fonction f .

Le Tableau 2.7 montre quelques exemples de fonctions qui ont été proposées pour les modèles canoniques.

Tableau 2.7. Exemples des fonctions utilisées dans les MCD.

Type de fonction	Type de variables	Nom	Définition
Logique	Booléen	NOT	$y \Leftrightarrow \neg x$
		OR	$y \Leftrightarrow x_1 \vee \dots \vee x_n$
		AND	$y \Leftrightarrow x_1 \wedge \dots \wedge x_n$
Algébrique	Ordinal discret Continu	MINUS	$y = -x$
		INV	$y = x_{max} - x$
		MAX	$y = \max(x_1, \dots, x_n)$
		MIN	$y = \min(x_1, \dots, x_n)$

Cependant, les MCD ne répondent pas au besoin de modéliser les interactions entre mécanismes pathogènes et résilients car ces interactions sont de type non-déterministe.

Ainsi, nous nous intéressons à d'autres modèles canoniques où les relations non-déterministes sont exprimées par des *paramètres*.

2.3.3.2 Les Modèles Canoniques Simples (MCS)

Les **modèles canoniques simples** (MCS) constituent le modèle le plus générique de construction de la TPC dans un modèle RB par la définition d'un ensemble de paramètres représentant les relations probabilistes. Dans ce cas, il est envisageable d'utiliser un MCS, dont la structure interne est illustrée par la Figure 2.2.

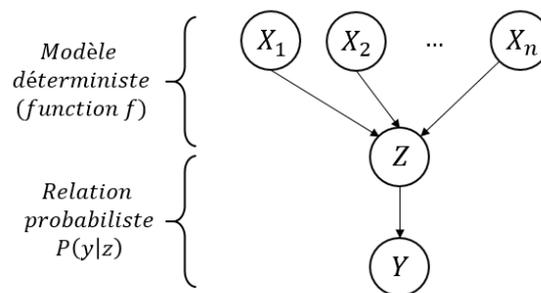


Figure 2.2. Structure interne d'un MCS.

Toutefois, en l'absence d'une connaissance suffisante du système permettant de faire certaines hypothèses sur les mécanismes causaux sous-jacents, le *nombre* de paramètres nécessaire à la construction de la TPC est *exponentiel* au nombre de variables cause.

Par conséquent, il est très difficile de définir tous les paramètres associés à une TPC avec un grand nombre de variables cause. Cette complexité de construction de la TPC se lie à la difficulté des experts à définir trop de paramètres.

Ce fait nous conduit à ne pas considérer les MCS comme adaptés pour modéliser l'interaction entre mécanismes résilients et pathogènes à cause de cette complexité liée à l'élicitation d'experts.

Nous examinons donc les modèles canoniques ICI.

2.3.3.3 Les Modèles Canoniques « ICI »

Les **modèles canoniques « ICI »** intègrent le concept d'*Influences Causales Indépendantes* (ICI). Ce concept repose sur l'hypothèse qu'aucune dépendance n'existe entre les mécanismes causals qui relient les variables causes X_i à Y .

Les modèles ICI peuvent être classés en « Noisy » et « Leaky », le deuxième étant un cas particulier du premier. D'un côté, les modèles ICI « Noisy » se construisent à partir des MCD en introduisant n variables auxiliaires $\{Z_1, Z_2 \dots Z_n\}$, comme illustré dans la Figure 2.3.

De cette manière, Y devient une fonction déterministe des Z_i et la valeur de chaque Z_i dépend d'un paramètre (du modèle probabiliste) des X_i .

L'hypothèse ICI impose alors des contraintes sur les valeurs $P(z_i|x_i)$ qui peuvent être résumées par :

- $P(z_i|\neg x_i) = 0$;
- $0 < P(z_i|x_i) \leq 1$

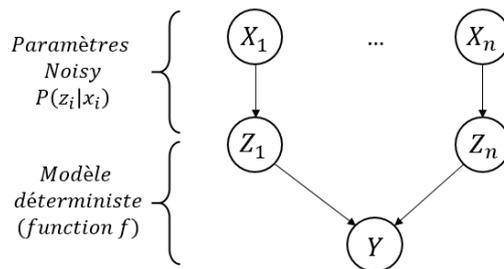


Figure 2.3. Structure auxiliaire d'un modèle canonique ICI Noisy.

La distribution de probabilités conditionnelles $P(y|x)$ s'obtient par la marginalisation des Z_i :

$$P(Y = \neg y|x) = \sum_z P(Y = \neg y|z) \cdot P(z|x)$$

Où $P(z|x) = \prod_i P(z_i|x_i)$.

Donc, la distribution de probabilité de Y est définie par :

$$P(Y = \neg y | \underline{x}) = \sum_{z | f(z) = y} \prod_i (z_i | x_i)$$

La propriété ICI se traduit par l'absence dans la Figure 2.3 des liens causals $X_i \rightarrow Z_j$ et $Z_i \rightarrow Z_j$, avec $i \neq j$. Autrement dit, les mécanismes causals produisant $Y = \neg y$ sont *indépendants* entre eux. Ce fait permet de définir seulement n paramètres pour un modèle avec n variables cause. Cet aspect est d'autant plus important qu'on considère le traitement des avis d'experts pour la paramétrisation d'un modèle RB contenant plusieurs structures ICI.

En effet, il a été démontré par (Diez et Druzdzel, 2006) que l'utilisation des modèles ICI Noisy rend plus facile la construction des TPCs dans un modèle RB, car les experts ont besoin de fournir un nombre limité de probabilités, qui est dans ce cas linéaire avec le nombre de variables cause (cf. Figure 2.4).

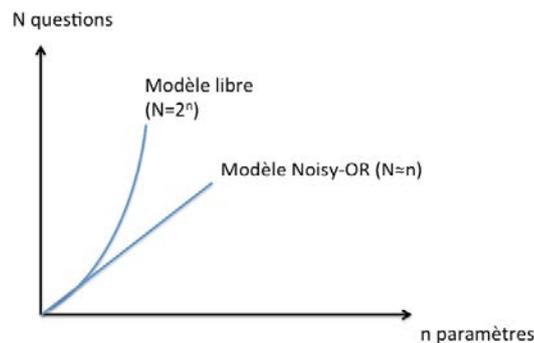


Figure 2.4. Nombre de questions pour l'élucation en fonction du nombre de paramètres nécessaires.

Pour leur part, les modèles ICI « Leaky » (cf. Figure 2.5) représentent une extension des ICI Noisy. En effet, ils prennent en compte d'autres variables cause qui ne sont pas énumérées explicitement ou dont le mécanisme causal n'est pas suffisamment connu pour être intégré dans le modèle. Ainsi, une variable auxiliaire supplémentaire est ajoutée dans la structure du modèle dont le paramètre f représente l'effet de mécanismes causals de « fuite » (de l'anglais « leak »). Le résultat de cette extension est que l'effet $Y = \neg y$ peut se produire même quand aucune variable cause n'est présente.

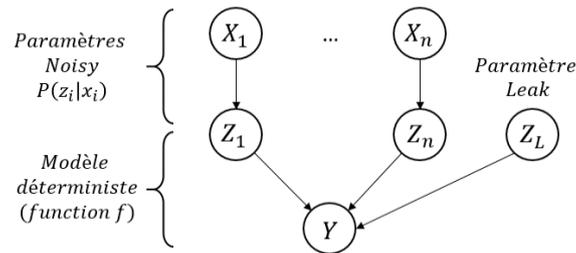


Figure 2.5. Structure auxiliaire de la version Leaky d'un modèle Noisy-OR.

Dans le cas des modèles ICI Leaky, seulement $n + 1$ paramètres sont nécessaires pour générer la TPC pour un modèle avec n variables causes.

En se basant sur ces considérations, les modèles ICI Leaky ont été choisis par (Léger, 2009) pour la représentation des relations causales dans la barrière humaine de l'AiDR, notamment pour la représentation des mécanismes pathogènes entre FOP et Déterminants. En effet, ces modèles permettent de :

- définir tous les mécanismes causals *non-déterministes* avec un nombre réduit de paramètres ;
- de prendre en compte une probabilité résiduelle (la fuite) quand la connaissance des mécanismes causals n'est pas complète.

Bien que d'autres approches existent pour la définition des TPCs dans un modèle basé sur les RBs (Mkrtychyan, *et al.* 2015), ces deux avantages ont permis d'identifier le modèle ICI Leaky comme étant le modèle le plus adapté pour répondre au besoin de l'AiDR de caractériser les influences causales de type pathogène des FOP sur les Déterminants.

Nous présentons ci-dessous le modèle ICI (Leaky) Noisy-OR utilisé actuellement dans la barrière humaine.

2.3.4 Le modèle « Leaky » Noisy-OR

L'interprétation causale du modèle Noisy-OR (ou, de façon complémentaire, du modèle Noisy-AND) est que chaque X_i , dont les états sont $+x_i = \text{présent}$, $\neg x_j (\forall j, j \neq i)$ représente une cause de Y et chaque Z_i indique si X_i a produit l'effet $Y = \neg y$, *i.e.* l'état indésirable.

Le terme « Noisy » fait référence alors à la possibilité que certaines des causes ne produisent pas l'état indésirable même quand elles sont présentes. De fait, $\neg z_i$ signifie que X_i n'a pas produit Y , soit parce que X_i était absent ou parce qu'un certain inhibiteur I_i a empêché que X_i produise Y . Si l'on désigne par q_i la probabilité que l'inhibiteur I_i soit actif, alors la probabilité c_i que X_i cause l'effet $Y = \neg y$, lorsqu'il est présent, est :

$$c_i = P(+z_i | +x_i, \neg x_j (\forall j, j \neq i)) = 1 - q_i$$

Bien entendu, lorsque X_i est absent, il ne peut pas produire l'effet $Y = \neg y$, *i.e.* :

$$P(Y = \neg y | \underline{x}) = 1 - \prod_{i \in I_+(\underline{x})} q_i = 1 - \prod_{i \in I_+(\underline{x})} (1 - c_i)$$

Considérons le cas d'un modèle RB composé par deux variables cause X_1 et X_2 et une variable effet Y . Alors l'équation précédente conduit à la TPC illustrée dans le Tableau 2.8.

Table 2.8. TPC pour un modèle Noisy-OR avec deux variables cause.

$P(+y x_1, x_2)$	x_1	$\neg x_1$
x_2	$c_1 + (1 - c_1) \cdot c_2$	c_2
$\neg x_2$	c_1	0

Comme il a été montré dans la section 2.3.3, il est important de disposer d'une version Leaky (*i.e.* avec fuite) du modèle ICI car il est généralement impossible d'énumérer explicitement toutes les variables cause possibles conduisant à l'effet $Y = \neg y$.

Ainsi, dans la TPC nous considérons une probabilité de fuite c_L dans l'équation ci-dessous :

$$P(Y = \neg y | \underline{x}) = 1 - P(\neg z_L) \prod_{i=1}^n P(\neg z_i | x_i) = 1 - (1 - c_L) \prod_{i \in I_+(\underline{x})} (1 - c_i)$$

Sur cette base et pour le même exemple considéré précédemment (cf. Tableau 2.9), la TPC pour un modèle Leaky Noisy-OR devient :

Table 2.9. TPC pour un modèle Leaky Noisy-OR avec deux variables cause.

$P(+y x_1, x_2)$	x_1	$\neg x_1$
x_2	$1 - (1 - c_1) \cdot (1 - c_2) \cdot (1 - c_L)$	$c_2 + (1 - c_2) \cdot c_L$
$\neg x_2$	$c_1 + (1 - c_1) \cdot c_L$	c_L

Malgré tous les avantages illustrés jusqu'ici justifiant le choix de (Léger, 2009) d'utiliser le modèle Leaky Noisy-OR dans la barrière humaine de l'AiDR, sa structure est limitée aux mécanismes de type pathogène. En effet, chaque variable cause présente ajoute un effet pouvant seulement augmenter la probabilité $P(Y = \neg y)$.

Pourtant, pour répondre aux besoins actuels d'EDF (cf. chapitre 1), il est nécessaire de prendre en compte aussi des mécanismes de causalité résilients qui pourraient mitiger une ou plusieurs combinaisons de ces mécanismes pathogènes. Le résultat d'une telle interaction causale doit conduire à une diminution de la probabilité jointe de l'état indésirable $Y = \neg y$.

L'intégration des mécanismes résilients et leur interaction avec ceux pathogènes dans la construction de la TPC du modèle RB soulève une difficulté majeure liée d'une part à l'élicitation d'experts pour la caractérisation des situations complexes produites par cette interaction et d'autre part à la définition des paramètres de la nouvelle structure proposée pour intégrer les mécanismes résilients.

Ainsi, ces réflexions nous conduisent à définir un troisième, un quatrième et un cinquième verrou scientifique à résoudre pour répondre aux deux verrous techniques identifiés dans le chapitre 1 et traités simultanément dans ce chapitre.

Verrou scientifique n°3. Formalisation de l'interaction entre mécanismes résilients et pathogènes dans le cadre des relations non-déterministes.

Verrou scientifique n°4. Définition d'un protocole de traitement des avis d'experts.

Verrou scientifique n°5. Définition d'une approche d'estimation des paramètres pour le modèle intégrant cette interaction entre mécanismes résilients et pathogènes.

2.4 Conclusions

Dans ce chapitre, nous avons analysé les travaux scientifiques en lien avec les verrous techniques identifiés dans le chapitre 1.

En ce sens, par rapport au verrou concernant la robustesse du modèle de barrière humaine, nous avons examiné la caractérisation de l'action par les développements faits en matière d'évaluation probabiliste de la fiabilité humaine (EPFH). Nous nous sommes focalisés sur un ensemble de méthodes EPFH considérées comme pertinentes pour répondre à ce verrou et les avons comparées sur la base des trois critères spécifiques (le cadre conceptuel, l'approche quantitative, la couverture des PSF). L'application de ces critères a permis de mettre en évidence des méthodes plus en phase avec l'AiDR : NARA, ATHEANA, SPAR-H. À partir de ce sous-ensemble de méthodes, nous avons conduit une étude spécifique portant sur les PSF qui a permis d'en isoler une dizaine. La pertinence des PSF isolés ayant déjà été démontrée, nous avons conclu qu'ils peuvent contribuer

à renforcer la robustesse du modèle de barrière humaine, notamment vis-à-vis de l'exhaustivité de la caractérisation de l'action au niveau du collectif de travail. Cependant, ces PSF nécessitent d'être formalisés pour leur intégration dans la structure de barrière humaine de l'AiDR (verrou scientifique n° 1).

La seconde partie a été consacrée aux deux verrous techniques de l'intégration de la résilience dans l'AiDR, le traitement des avis d'experts et l'estimation des paramètres. Ainsi, concernant l'intégration de la résilience, nous avons constaté le manque d'une formalisation claire des mécanismes résilients. Tout d'abord, d'un point de vue de ses marqueurs (verrou scientifique n°2) puis comme mécanisme causal en interaction avec d'autres mécanismes pathogènes dans le cadre probabiliste (verrou scientifique n°3). En effet, une telle considération de la résilience dans le cadre probabiliste et plus spécifiquement dans les RBs, mène à traiter de la modélisation des connaissances en termes de mécanismes causals. Notamment, des modèles probabilistes canoniques existent et permettent de représenter les connaissances en entrée issues d'avis d'experts. Toutefois, ces modèles ne permettent pas de prendre en compte l'interaction entre mécanismes pathogènes (possible jusqu'ici au travers du modèle canonique Leaky Noisy-OR) et résilients. Lorsque la formalisation d'une telle interaction est possible dans un cadre probabiliste, il subsiste des difficultés majeures pour les experts dans la caractérisation des scénarios produits par cette interaction. Ainsi, il est nécessaire de formaliser un protocole permettant d'élucider les experts par rapport à ces interactions (verrou scientifique n°4). De plus, le traitement de ces avis d'expert doit se poursuivre par une approche d'estimation des paramètres dans la nouvelle structure mathématique intégrant la résilience (verrou scientifique n°5).

Par conséquent, de cet état de l'art, cinq verrous scientifiques ont été identifiés :

- Formalisation des PSF pour leur intégration dans la structure du modèle de barrière humaine de l'AiDR ;
- Elaboration et formalisation des marqueurs résilients ;
- Formalisation de l'interaction entre mécanismes résilients et pathogènes dans le cadre des relations non-déterministes ;
- Définition d'un protocole de traitement des avis d'experts ;
- Définition d'une approche d'estimation des paramètres pour le modèle intégrant cette interaction entre mécanismes résilients et pathogènes.

Ces verrous scientifiques sont repris au chapitre 3 pour fonder nos contributions.

Chapitre 3

Contributions à l'évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

3 Contributions à l'évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Introduction

Ce chapitre présente les contributions relatives aux verrous scientifiques formulés dans le Chapitre 2. Par conséquent ces contributions adressent d'abord l'intégration de la résilience et ensuite la proposition d'un nouveau modèle de barrière humaine permettant, d'une part, de renforcer la robustesse de la caractérisation de l'action humaine dans l'AiDR et, d'autre part, de prendre en compte les mécanismes résilients. Enfin l'évaluation probabiliste à partir d'une élicitation pertinente d'avis d'experts est abordée pour conclure sur une résolution du problème d'estimation des paramètres du modèle RB.

En ce sens, la première partie du chapitre se focalise sur l'intégration de la résilience dans l'AiDR. Nous commençons par la formalisation des mécanismes résilients sur un plan qualitatif. Cette formalisation se fait dans un premier temps en présentant le concept de marqueur résilient. Ensuite, des éléments sont présentés sur l'interaction entre mécanismes résilients et pathogènes dans la modélisation probabiliste. Les principes de cette interaction sont formalisés de manière générique, puis appliqués dans le cadre probabiliste et plus spécifiquement dans le cas des RB. Il s'agit de proposer un modèle incluant le principe de mitigation propre aux mécanismes résilients.

Le chapitre se poursuit par la proposition d'un nouveau modèle de barrière humaine. Tout d'abord nous présentons les principes génériques de construction d'un modèle de barrière humaine. Dans l'application de ces principes génériques au cas de l'AiDR nous proposons la formalisation de l'ensemble des PSF retenu à partir de l'état de l'art pour leur intégration dans la structure du modèle de barrière humaine de l'AiDR. Cela répond en effet au besoin de renforcer la caractérisation de l'action par ses déterminants. Ensuite, pour tenir compte des mécanismes résilients dans l'AiDR, nous proposons de généraliser les facteurs organisationnels, puis d'adapter les liens de causalités entre facteurs organisation-

nels et déterminants de l'action. Au global, l'ensemble de ces propositions de formalisation dans le cadre de l'AiDR aboutissent à créer un "méta-modèle" qui sera exploité par instanciation, pour développer des modèles de barrières liés à d'actions humaines spécifiques.

La mise en œuvre du modèle RB est traitée à la fin du chapitre. Nous proposons d'abord un protocole d'élicitation d'avis d'experts basé sur la confrontation des experts avec des situations peu complexes. Pour crédibiliser les jugements d'experts, nous intégrons dans ce protocole une échelle d'élicitation reconnue dans les méthodes d'EPFH et notamment à EDF. Nous concluons le chapitre avec une approche proposée permettant d'estimer les paramètres à partir des jugements d'expert en cohérence avec la structure du modèle RB proposé.

3.1 Intégration de la résilience dans l'AiDR

Cette intégration de la résilience dans l'AiDR requiert avant tout de formaliser les marqueurs résilients, puis de prendre en compte l'interaction entre mécanismes résilients et pathogènes dans le cadre probabiliste et plus spécifiquement dans un modèle RB pour être appliquée dans l'AiDR.

3.1.1 Élaboration des marqueurs résilients

Comme nous l'avons précisé au chapitre 1, une analyse multirisque des événements accidentels/incidentés dans les systèmes sociotechniques a permis d'identifier, en plus des mécanismes de défaillance technique, des *mécanismes pathogènes* concernant les composantes humaines et organisationnelles impliquées dans les scénarios accidentels. Le retour d'expérience analysé à partir de ces scénarios a conduit à l'élaboration des marqueurs pathogènes (Dien, *et al.* 2004) spécifiant les conditions qui 'activent' les mécanismes pathogènes entre 'Facteurs Organisationnels' et 'Déterminants' de l'action.

Cependant, certaines de ces analyses comme celle de l'accident ferroviaire de Paddington (Dien, *et al.* 2004) avaient conduit à établir des mesures permettant de s'affranchir de lacunes de la réglementation à l'époque en pleine évolution en Grande-Bretagne.

Nous nous appuyons sur cette idée que d'autres informations peuvent être identifiées à partir de l'analyse accidentelle sur des aspects de résilience. En effet, un ensemble spécifique de compétences, de pratiques et d'attitudes avait été identifié dans le cas de Paddington relativement à des comportements potentiellement

résilients des opérateurs et capables d'atténuer les effets produits par les mécanismes pathogènes.

Ainsi, comme pour les marqueurs pathogènes, pour que ce potentiel se traduise par des **mécanismes résilients**, des marqueurs résilients – ressources appropriées, caractéristiques du système, structures organisationnelles – devraient préciser les conditions qui doivent se présenter pour qu'un système ou une organisation puisse réagir de façon résiliente.

Concernant uniquement les phénomènes de mitigation sans considérer les mécanismes de récupération (pour les raisons déjà citées au chapitre 1, à savoir le choix dans l'AiDR de ne pas prendre en compte le temps), l'approche retenue pour élaborer les marqueurs résilients est de nous référer aux marqueurs pathogènes associés aux mécanismes pathogènes.

Ainsi, l'hypothèse suivante est posée :

Hypothèse n° 1.

Pour les situations *attendues*, il existe une relation entre les **marqueurs résilients** et les **marqueurs pathogènes**.

Dans la Figure 3.1, toutes les situations possibles sont identifiées en fonction de la combinaison du niveau de résilience du système et de la vraisemblance des événements.

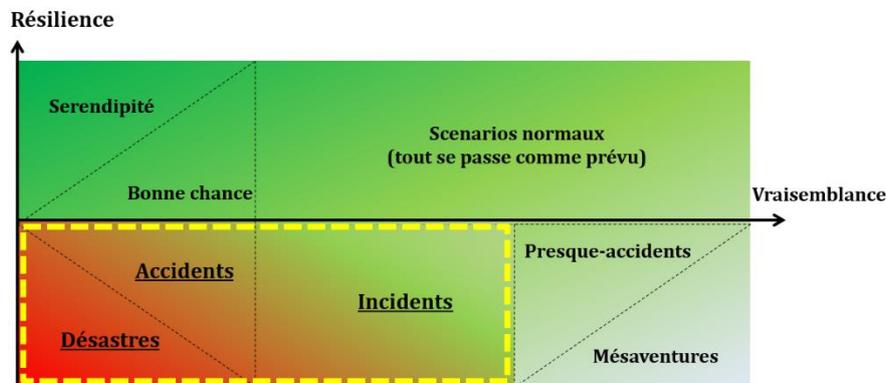


Figure 3.1. Diagramme résilience – vraisemblance et scénarios possibles (Paries, 2006).

Les régions désignées par « Serendipité », « Bonne chance » et « Scénarios normaux » (c'est-à-dire des situations définies par une résilience élevée du système et des vraisemblances élevées ou faibles des événements dangereux) et « presque-accidents » et « accidents » (c'est-à-dire, des situations définies par une faible résilience, mais une vraisemblance élevée des événements dangereux) sont celles que l'ingénierie de la résilience vise à étudier, car elles représentent « ce que nous cherchons à instruire dans cette caractérisation de la résilience ».

Néanmoins, la seule source d'information disponible permettant d'élaborer les marqueurs de la résilience est celle mise en évidence dans la Figure 3.1 (encadré en pointillés jaunes) et les situations notées par « Désastres », « Accidents » et « Incidents » (c'est-à-dire des situations caractérisées par une faible résilience et une faible vraisemblance). Comme les deux types de marqueurs, pathogènes et résilients, se réfèrent à des perturbations que le système a déjà expérimentées, une relation à minima 'logique' peut être établie entre eux.

Pour formaliser les marqueurs de manière à intégrer le modèle RB support de l'AiDR, considérons deux ensembles génériques de marqueurs associés à un facteur organisationnel générique X :

- M^r , un ensemble composé par des marqueurs résilients désignés par $m_i^{r \rightarrow X}$ et
- M^p , un ensemble composé par des marqueurs pathogènes désignés par $m_j^{p \rightarrow X}$.

Le symbole \rightarrow indique que les marqueurs sont liés à un facteur X .

Cette proposition de formalisation part du concept de complémentarité, c'est-à-dire d'une relation existant entre un marqueur résilient $m_i^{r \rightarrow X}$ et un marqueur pathogène 'joint' $m_j^{p \rightarrow X}$.

Cependant, il est réducteur, ou du moins insuffisant, de considérer le complémentaire des marqueurs pathogènes pour en déduire les marqueurs résilients. Nous proposons donc d'étendre cette relation :

- pour un sous-ensemble de marqueurs pathogènes $\underline{m}_j^{p \rightarrow X}$ (tous référant au même facteur X) ;
- il existe une fonction g de mitigation permettant de prendre en compte une relation générique entre un sous-ensemble $\underline{m}_j^{p \rightarrow X}$ et un marqueur résilient.

Par conséquent, un marqueur résilient peut être élaboré non pas comme la partie complémentaire du/des marqueur/s pathogène/s correspondant/s, mais plus génériquement comme une fonction de mitigation d'une partie des marqueurs pathogènes du facteur X .

Cette fonction de mitigation conduit à un ensemble de marqueurs résilients qui permettent de remonter à un (ou plusieurs) facteur(s) X . En effet, un facteur X n'est pas nécessairement résilient ou pathogène pour toutes les situations : les effets de sa présence (ou absence) dépendent du contexte (par exemple le type de système) ou de l'action spécifique.

Dans cette fonction de mitigation, nous devons considérer éventuellement toute information inconnue (ou non considérée) qui peut aider à mieux caractériser les marqueurs résilients (De Galizia, *et al.*, 2016b,c).

Nous avons supposé précédemment que des marqueurs résilients sont identifiés en référence à des situations attendues. Il est envisageable d'imaginer la validité des marqueurs élaborés pour des situations *imprévisibles*. Cela signifie étendre l'Hypothèse 1 à des situations inconnues. Afin de mieux comprendre les raisons de cette extension, l'analogie suivante avec le domaine médical est proposée (Dien, *et al.* 2004).

« Considérons un patient qui souffre d'une maladie connue et correctement identifiée. Sur la base d'un ensemble de symptômes (conduisant à des marqueurs pathogènes spécifiques), le médecin prescrit des mesures curatives. Le même médecin peut éventuellement fournir au patient des recommandations sur la façon d'agir dans l'avenir face à des situations similaires. Ces recommandations permettent au patient d'être, dans ce sens, 'résilient' s'il se trouvait à nouveau dans les mêmes conditions donc «attendues». Néanmoins, rien n'empêche que les mêmes recommandations puissent aider le patient à effectuer de manière résiliente dans des circonstances «inattendues» que le médecin lui-même ne connaissait pas. »

Suite à l'approche décrite et puis illustrée dans l'analogie ci-dessus, les marqueurs résilients sont identifiés à partir des marqueurs pathogènes pour un ensemble de facteurs organisationnels (De Galizia, *et al.*, 2016b).

Par exemple, les marqueurs résilients sont identifiés à partir des marqueurs pathogènes pour le FOP 'MT', comme indiqué dans le Tableau 3.1.

Tableau 3.1. Exemples de marqueurs résilients obtenus selon l'approche dans la (3.1).

Mauvais traitement de la complexité organisationnelle	
Marqueurs Pathogènes	Marqueurs Résilients
<ul style="list-style-type: none"> - Éloignement géographique des centres décisionnels. - Difficultés à joindre les instances décisionnelles 	<ul style="list-style-type: none"> - Bonne communication entre centres décisionnels et centres opérationnels
<ul style="list-style-type: none"> - Non-prise en compte de la complexité et de l'opacité par le management. - Confiance excessive du management - Attitudes en positions idéologiques 	<ul style="list-style-type: none"> - Capacité du management à s'interroger
<ul style="list-style-type: none"> - Circuits « bottom-up » inexistantes ou biaisés (communication interne insuffisante ou biaisée) 	<ul style="list-style-type: none"> - Capacité du management à écouter les remontées du terrain
<ul style="list-style-type: none"> - Conflits intra et extra-organisationnels 	<ul style="list-style-type: none"> - Capacité des responsables à converger vers une vision partagée
<ul style="list-style-type: none"> - Interférences entre services et tâches nuisant à la sécurité ('co-activité') 	<ul style="list-style-type: none"> - Capacité à prendre en compte la co-activité

La formalisation jusqu'ici proposée pour l'élaboration d'un marqueur résilient sur la base des marqueurs pathogènes nous permet d'identifier, dans le modèle de barrière humaine, la présence de mécanismes résilients. Cependant, nous ne connaissons pas à ce stade la nature de l'interaction entre mécanismes résilients éventuellement identifiés et les mécanismes pathogènes.

Ainsi, nous proposons de formaliser cette interaction dans le cadre probabiliste pour l'intégrer dans l'AiDR. Avant de passer directement au cadre probabiliste, nous fondons les principes génériques de cette interaction.

3.1.2 Formalisation de l'interaction entre mécanismes résilients et pathogènes

Considérons un ensemble générique de variables aléatoires binaires composé de n variables $\{X_1, \dots, X_n\}$ dont les états possibles sont $(x, \neg x)$ et m variables Y_l ($l \in \{1, \dots, m\}$) dont les états sont $(y, \neg y)$. Supposons que k parmi les n variables – dénotées X_i^p ($i \in \{1, \dots, k\}$) – ont une influence pathogène sur les variables Y_l , et $n - k$ – dénotées X_j^r ($j \in \{k + 1, \dots, n\}$) – ont une influence résiliente sur Y_l .

Ainsi, les hypothèses suivantes ont été posées pour combiner l'état des variables X_i^p et X_j^r à des mécanismes d'influence résiliente ou pathogène sur Y_l .

Hypothèse n° 2.

Un mécanisme d'influence *résiliente*, dit **mécanisme résilient**, se produit si un sous-ensemble de *marqueurs résilients* $\underline{m}_j^r \in M^r$ est identifié attestant que X_j^r a une *influence résiliente* sur Y_l . Alors la présence de ce mécanisme est représentée par $X_i = X_j^r = \text{présent}$.

Hypothèse n° 3.

Un mécanisme d'influence *pathogène*, dit **mécanisme pathogène**, se produit si un sous-ensemble de *marqueurs pathogènes* $\underline{m}_i^p \in M^p$ est identifié attestant que X_i^p a une *influence pathogène* sur Y_l . Alors la présence de ce mécanisme est représentée par $X_i = X_i^p = \text{absent}$.

Par conséquent, les variables X_j^r – par défaut dans l'état 'absent' – ont une influence résiliente quand elles passent à l'état 'présent' ; tandis que les variables X_i^p – par défaut dans l'état 'présent' – ont une influence pathogène quand elles passent à l'état 'absent'. Sur la base des hypothèses 1 et 2, et en fonction de la présence ou l'absence des marqueurs associés, des mécanismes pathogènes et résilients peuvent se produire, interagir et leur influence jointe peut avoir un effet sur la variable Y_l .

Ainsi, nous proposons de modéliser le mécanisme d'interaction entre les influences résilientes et pathogènes affectant les Y_l par une fonction générique I suivante :

$$Y_l = I(X_i^p, X_j^r, \varepsilon) \quad (3.1)$$

où le terme externe ε doit être inclus dans la formalisation pour rendre le modèle plus générique et représenter des influences pathogènes inconnues qui produisent également $Y_l = \neg y$.

Puisque la résilience est considérée comme ayant un effet de mitigation, la fonction générique I dans l'équation (3.1) doit traduire ce *principe de mitigation*.

Considérons dans la Figure 3.2 l'impact de k variables X_i^p sur Y_l . Quand les X_i^p se trouvent dans l'état 'absent', elles ont une influence pathogène (flèches rouges) produisant des effets indésirables désignés par E_i .

Ces effets E_i ont comme résultat que Y_l se trouve dans l'état indésirable $\neg y$. Alors, l'interaction des X_i^p avec $n - k$ variables X_j^r qui se trouvent dans l'état 'présent' et donc ayant une influence résiliente (flèches vertes) a comme résultat la mitigation des effets E_i et donc la probabilité que Y_l se trouve dans l'état indésirable $\neg y$ diminue.

Sur cette base, l'hypothèse suivante est posée résumant le mécanisme qui vient d'être formulé :

Hypothèse n° 4.

L'interaction entre les *mécanismes pathogènes* des X_i^p et *résilients* X_j^r se fonde sur le **principe de mitigation** consistant en l'atténuation des *effets pathogènes* E_i potentiellement produits par les X_i^p .

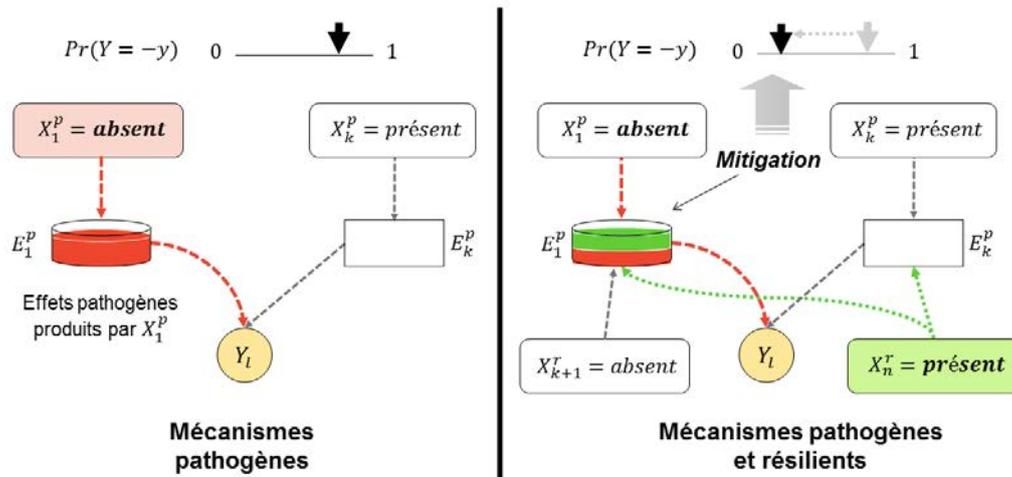


Figure 3.2. Mitigation des effets pathogènes produits sur des X_i^p par les X_j^r .

Supposons qu'en l'absence d'influences pathogènes, d'autres effets pathogènes résiduels puissent produire $Y_l = -y$. Alors, pour une interaction générique entre une variable X_i^p et X_j^r , le Tableau 3.2 résume l'ensemble des situations qui peuvent se produire et l'effet joint de l'interaction sur Y_l .

Tableau 3.2. Effets du mécanisme de mitigation sur les influences pathogènes (implicites ou explicites).

Effet conjoint sur Y_l	X_i^p	
	Présent (pas d'influence)	Absent (influence pathogène)
X_j^r		
Présent (influence résiliente)	<i>Mitigation des effets résiduels</i>	<i>Mitigation des effets pathogènes</i>
Absent (pas d'influence)	<i>Effets résiduels</i>	<i>Effets pathogènes</i>

Le principe de mitigation que l'on vient de formuler a besoin d'être traduit dans un cadre mathématique pour quantifier les effets de la mitigation entre les variables du modèle de barrière humaine.

Ainsi, pour réaliser une intégration de ces formulations dans l'AiDR, nous proposons d'appliquer ces principes dans le cadre probabiliste et plus spécifiquement par un RB.

3.1.3 Application de l'interaction entre mécanismes résilients et pathogènes dans les RBs

L'objectif de cette application est d'évaluer l'effet conjoint de mécanismes causaux multiples – dont certains sont pathogènes et d'autres résilients – sur la probabilité $P(Y_l = \neg y)$.

Considérons donc un ensemble de variables aléatoires binaires composées par n variables causes $\{X_1, \dots, X_n\}$ dont les états possibles sont $(x, \neg x)$ et m variables effets Y_l ($l \in \{1, \dots, m\}$) dont les états sont $(y, \neg y)$. Nous supposons également que k parmi les n variables causes sont notées X_i^p ($i \in \{1, \dots, k\}$) en raison de leur influence pathogène et $n - k$ notées X_j^r ($j \in \{k + 1, \dots, n\}$) en raison de leur influence résiliente.

Dans ce contexte, les deux hypothèses suivantes sont posées (De Galizia, *et al.*, 2016a) :

Hypothèse n° 5.

Quand des *mécanismes pathogènes* se produisent (hypothèse 2), la probabilité $P(Y_l = \neg y)$ que Y_l se trouve dans l'état indésirable $\neg y$ **augmente**.

Hypothèse n° 6.

Quand des *mécanismes résilients* se produisent en présence de mécanismes pathogènes (hypothèses 2 et 3), la probabilité $P(Y_l = \neg y)$ **diminue** sur la base du **principe de mitigation**.

En effet, lorsque des mécanismes résilients produits par les variables X_j^r se produisent, l'effet de la mitigation – diminution de $P(Y_l = \neg y)$ – n'affecte pas directement la variable effet Y_l . De ce fait, le principe de mitigation n'a de sens que quand des mécanismes pathogènes se sont produits. Par conséquent, les mécanismes résilients sont plutôt perçus comme des inhibiteurs des mécanismes pathogènes.

Toutes ces hypothèses nous permettent de modéliser ce principe de mitigation dans le modèle RB dont la représentation est proposée dans la Figure 3.3.

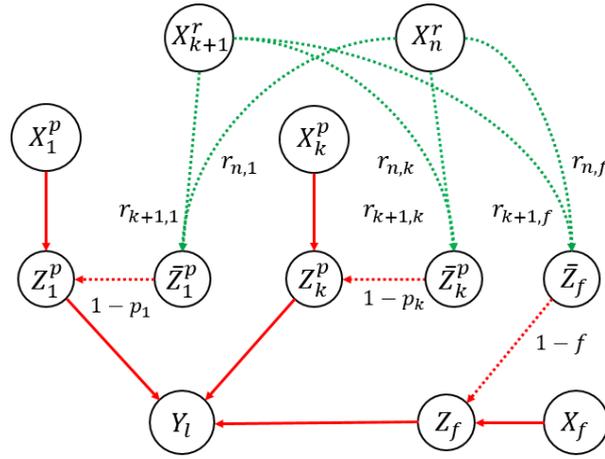


Figure 3.3. Principe de mitigation dans un modèle RB (De Galizia, et al. 2015a,b).

Considérons donc un modèle RB composé par n variables causes $\{X_1, \dots, X_n\}$ dont :

- un ensemble X^p de k variables X_i^p ($i \in \{1, \dots, k\}$) qui produisent l'effet $Y_l = \neg y$ avec une probabilité p_i ;
- un ensemble X^r de $n - k$ variables X_j^r ($j \in \{k + 1, \dots, n\}$) qui mitigent l'effet du mécanisme pathogène produit par X_i^p avec une probabilité $r_{j,i}$.

En fonction de l'état présent ou absent des variables X_i^p et X_j^r , les paramètres associés apparaissent dans une équation générique qui exprime les probabilités conditionnelles à chaque ligne de la TPC.

Ainsi, nous proposons d'utiliser des facteurs multiplicateurs pour traduire la présence et l'absence de chaque type des variables:

$$R_j = \begin{cases} 1 & \text{si present} \\ 0 & \text{si absent} \end{cases}$$

$$P_i = \begin{cases} 1 & \text{si absent} \\ 0 & \text{si present} \end{cases}$$

Alors, l'équation générale (De Galizia, et al. 2016c) suivante permet de générer la TPC du modèle RB intégrant le principe de mitigation :

$$P(Y = \neg y | \underline{X^p}, \underline{X^r}) =$$

$$1 - \left\{ 1 - \left[\prod_{j=k+1}^n (1 - r_{j,f} \cdot R_j) \right] \cdot f \right\} \cdot \prod_{i=1}^k \prod_{j=k+1}^n [1 - (1 - r_{j,i} \cdot R_j) \cdot p_i \cdot P_i] \quad (3.3)$$

Pour traduire le principe de mitigation dans le modèle RB proposé, nous proposons des règles pour assurer une cohérence entre la structure mathématique et le principe de mitigation jusqu'ici formulé.

Ainsi, pour intégrer ce principe de mitigation dans un modèle RB nous proposons de :

- 1) Se conformer à un ensemble d'inégalités (règles de cohérence) dans la construction du modèle RBs. Ces contraintes confinent chaque ligne de la TPC dans des intervalles de probabilité déduits du principe de mitigation. Elles assurent que chaque situation résultant de ce principe soit cohérente avec la structure mathématique du modèle ;
- 2) Utiliser des variables discrètes ordinales comme variables effet dans le RBs. Afin de pouvoir établir l'effet de la mitigation d'une influence résiliente sur une influence pathogène, il doit y avoir une relation d'ordre entre les états de la variable « effet » qui est affectée.

Afin de montrer comment ces règles s'appliquent à la construction de la TPC, considérons un modèle RB composé par une variable d'influence pathogène X_1^p , une variable d'influence résiliente X_2^r et une variable effet Y . Des mécanismes pathogènes résiduels de fuite sont considérés par un terme f .

La TPC (cf. Tableau 3.3) de l'exemple considéré s'obtient par l'application de l'équation (3.3).

Tableau 3.3. TPC pour un cas avec deux variables causes X_1^p et X_2^r .

X_1^p	X_2^r	$P(Y = \neg y X_1^p, X_2^r)$	
$\neg x$	x	$1 - (1 - (1 - r_{2f}) \cdot f) \cdot (1 - (1 - r_{21}) \cdot p_1)$	$\varphi_{1,2}$
	$\neg x$	$1 - (1 - f) \cdot (1 - p_1)$	φ_1
x	x	$1 - (1 - (1 - r_{2f}) \cdot f)$	φ_2
	$\neg x$	$1 - (1 - f) = f$	φ_0

Nous proposons de désigner par $\varphi_{i,j}$ la probabilité conditionnelle associée à la combinaison d'états des variables $X_1^p = \neg x$ et $X_2^r = x$ à chaque ligne de la TPC.

Ainsi, nous proposons un ensemble d'inégalités (règles de cohérence) exprimant une relation d'ordre entre les $\varphi_{i,j}$. Les inégalités ont les significations suivantes :

Situation 0 – Mécanismes pathogènes de fuite: $0 \leq \varphi_0 \leq 1$

Où $\varphi_0 = P(Y = \neg y | X_1^p = x, X_2^r = \neg x)$

i.e., quand aucun mécanisme pathogène considéré dans le modèle n'est présent, la probabilité de produire $\neg y$ est au mieux égale à 0. Cependant des mécanismes pathogènes non énumérés et représentés par f (modèle avec fuite) peuvent produire $\neg y$, alors φ_0 est comprise entre 0 et 1;

Situation 1 – Mitigation des mécanismes pathogènes de fuite : ($\varphi_2 \leq \varphi_0$)

Où $\varphi_2 = P(Y = \neg y | X_1^p = x, X_2^r = x)$.

i.e., l'effet des mécanismes pathogènes de fuite a moins de probabilité de produire $\neg y$ à cause du principe de mitigation entre X_2^r et f . Cela signifie que si X_2^r mitige l'effet des mécanismes pathogènes, la probabilité que $Y = \neg y$ doit être au pire aussi élevée que pour la situation 0. Dans ce cas, la situation 1 représente la condition la plus favorable ($\neg y$ étant l'effet indésirable et donc la condition défavorable pour Y). Par contre, si aucune mitigation n'est portée par X_2^r , alors $\varphi_2 = \varphi_0$;

Situation 2 – Mécanismes pathogènes explicites et de fuite : ($\varphi_0 < \varphi_1 \leq 1$)

Où $\varphi_1 = P(Y = \neg y | X_1^p = \neg x, X_2^r = \neg x)$.

i.e., dans cette situation la probabilité de produire $\neg y$ est maximale. L'effet conjoint des mécanismes pathogènes explicites (X_1^p) et de fuite (f) augmente la probabilité d'avoir $Y = \neg y$ et aucun mécanisme résilient n'est présent. Cette situation représente la condition la plus défavorable ;

Situation 3 – Mitigation des influences pathogènes explicites et de fuite :

($\varphi_2 \leq \varphi_{12} \leq \varphi_1$)

Où $\varphi_{12} = P(Y = \neg y | X_1^p = \neg x, X_2^r = x)$.

i.e., la présence du mécanisme résilient diminue la probabilité de produire $\neg y$. L'interaction avec les mécanismes pathogènes explicites (X_1^p) et de fuite (f) conduit, dans le meilleur des cas, à avoir une probabilité $\varphi_{12} = \varphi_2$; le pire des cas fournit une probabilité $\varphi_{12} = \varphi_1$.

Donc, nous avons formalisé le principe de mitigation comme étant le principe qui représente l'interaction entre mécanismes pathogènes et résilients. Ce principe a été traduit dans le cadre probabiliste et plus spécifiquement dans les RBs afin d'être appliqué au modèle de barrière humaine de l'AiDR.

Pour consentir l'application de ces principes au cas de l'AiDR, le modèle actuel de barrière nécessite d'être adapté. Ainsi, nous proposons de fonder les principes de construction d'un nouveau modèle de barrière permettant, dans le cas de l'AiDR, d'intégrer les mécanismes résilients.

3.2 Proposition d'un nouveau modèle de barrière humaine

L'objectif à ce stade est de proposer un nouveau modèle de barrière tout d'abord en fondant les principes génériques de construction de la barrière puis par l'application de ces principes dans le cas de l'AiDR. Cette application permettra globalement de générer un méta-modèle à instancier sur une action spécifique.

Ainsi, lors de cette application dans l'AiDR, nous formalisons les PSF issus de l'état de l'art, tout d'abord par leur intégration dans la structure de la barrière humaine au niveau des déterminants de l'action, puis par l'identification des liens de causalité entre les PSF et les variables actuellement existantes dans le cadre de causalité du modèle. Ensuite, pour tenir compte des mécanismes résilients dans l'AiDR, nous proposons de généraliser les facteurs organisationnels, puis d'adapter les liens de causalités entre facteurs organisationnels et déterminants de l'action.

Avant de présenter les éléments de cette formalisation, nous présentons donc les principes génériques de construction d'un modèle de barrière humaine prise dans son contexte organisationnel.

3.2.1 Principes génériques de construction du modèle de barrière humaine

Donnons pour commencer une définition du modèle générique de **modèle de barrière humaine** tel que nous la considérons dans la thèse :

Modèle servant à représenter et structurer des connaissances issues des composantes *humaines, du collectif de travail, organisationnelle et environnementale* des systèmes sociotechniques nécessaires à une évaluation qualitative et quantitative de l'efficacité d'une action humaine éventuellement prise dans son contexte organisationnel. Ce modèle doit permettre, dans le cadre d'une analyse globale des risques, d'évaluer l'impact des actions humaines sur la disponibilité des composants ou des sous-systèmes techniques étudiés.

Dans le modèle de barrière humaine, un cadre de causalité y est intégré permettant de mettre en relation les variables à des niveaux différents du modèle, ceux de l'action humaine, d'une part, et de l'organisation, d'autre part. Ces relations sont génériques, mais doivent être traduites en paramètres numériques pour servir à l'évaluation du risque. Cette dimension numérique probabiliste doit faire sens avec les données du terrain fournies par les experts. Les liens de causalité représentent les relations logiques entre efficacité de l'action, phases et déterminants ainsi que les mécanismes pathogènes ou résilients de type non-déterministe

entre collectif de travail et organisation. L'ensemble des effets, se propageant grâce aux relations de causalité, se mesure sur l'efficacité de l'action.

De façon générique, toutes les variables de la barrière humaine sont des variables aléatoires discrètes. Les variables de la barrière sont regroupées en classes. Chaque classe identifie un niveau différent de caractérisation de l'action dans le modèle de barrière humaine. Comme dans (Léger, 2009), les variables composant chaque niveau de la barrière sont non-corrélées entre elles.

Nous présentons ci-dessous ces classes.

3.2.1.1 Classes de variables

Dans la barrière humaine prise dans son contexte organisationnel les variables appartiennent à quatre classes :

- **Efficacité de l'action**
- **Phases de l'action**
- **Déterminants**
- **Facteurs Organisationnels**

Comme nous l'avons précisé dans la section 3.1, il s'agit de variables discrètes. Pour chaque classe de variables, nous précisons l'état de référence (ER), c'est-à-dire l'état naturellement envisagé. Analysons en détail les spécificités de chacune de ces classes.

Classe 1 – « Efficacité de l'action »

La classe « Efficacité de l'action humaine » représente une action menée par un collectif dans les systèmes sociotechniques. Son efficacité est évaluée par rapport à la probabilité d'atteindre les objectifs (cf. définition d'efficacité donnée dans le chapitre 1). Dans le modèle de barrière humaine, l'action est vue comme un processus mais elle est réalisée à un niveau local. Elle se déroule dans un intervalle de temps donné et pour en faciliter l'analyse, nous pouvons segmenter le processus en phases opérationnelles.

La granularité de l'efficacité a été restreinte à deux états : efficace (ER) ou inefficace. Ce qui spécifie l'espace des possibles dans le cadre probabiliste.

Classe 2 – « Phases » (de l'action)

Les phases permettent de représenter les actions humaines à un niveau plus opérationnel. D'après (Léger, 2009), cette représentation permet de structurer la connaissance et la collecte des informations lors des entretiens de terrain.

Les phases peuvent avoir deux états : bien faite (ER) ou mal faite. Ce qui spécifie, là encore, l'espace des possibles dans le cadre probabiliste.

Classe 3 – « Déterminants » (de l'action)

Les déterminants représentent les caractéristiques et les comportements des collectifs de travail (Léger, 2009). Ils servent à focaliser l'étude à un niveau plus macroscopique et non plus localisé sur une action plus spécifique. Des relations logiques sont établies entre les déterminants et les phases.

Les déterminants peuvent avoir deux états : adéquat (ER) ou inadéquat.

Classe 4 – « Facteurs Organisationnels »

Les facteurs organisationnels permettent d'effectuer une étude de l'organisation à un niveau macroscopique. Il s'agit de facteurs qui peuvent être à l'origine de mécanismes pathogènes ou résilients avec des effets qui se propagent sur les facteurs du collectif de travail, les phases de l'action et, finalement sur l'action analysée. Ces facteurs représentent l'état de santé de l'organisation en fonction de la présence de certains marqueurs résilients (bonne santé de l'organisation) ou pathogènes (mauvaise santé de l'organisation).

Les facteurs organisationnels peuvent avoir deux états : présent ou absent. Néanmoins, à ce stade nous ne précisons pas encore l'ER de ces variables. Cet aspect sera clarifié dans la suite de cette section.

3.2.1.1 Espace des possibles

Les variables de ces classes sont associées à des *propriétés*. Par exemple, une variable de type déterminant comme 'Formation' peut être associée à des propriétés tels la durée, l'efficacité, le résultat, etc. Une propriété présente à la fois deux ou plusieurs *modalités*. Ces modalités correspondent à différentes situations ou différents états dans lesquels les variables, par rapport à la propriété, se retrouvent ou sont qualifiées. Par exemple, la durée de la 'Formation' peut être 'longue', 'moyenne', 'courte'.

Cependant, le nombre des modalités d'une propriété varie selon le degré de granularité et les informations disponibles. Le cas le plus simple correspond à deux modalités.

Une propriété est considérée comme étant :

- *Qualitative*, lorsque ses modalités ne sont pas mesurables. A priori, si aucun ordre ne peut être établi entre ses modalités, elles sont dites nominales (ex. les modalités 'vert', 'rouge' et 'bleu' pour la propriété couleur). Toutefois, dans certains cas, il est possible d'établir une relation d'ordre ou de préférence entre les modalités, alors ces variables sont dites ordinales (ex. les modalités 'mauvaise', 'moyenne', 'bonne' pour la propriété qualité référée à la variable 'Formation') ;

- *Quantitative*, lorsque ces modalités sont mesurables. Autrement dit, à chaque modalité correspond une valeur et entre ces valeurs peuvent être appliquées des opérations algébriques pour faire des calculs. Par conséquent, elles sont qualifiées de cardinales.

Les modalités d'une propriété sont considérées comme étant exhaustives et mutuellement exclusives, c'est-à-dire qu'une propriété ne peut prendre qu'une et une seule des modalités à un instant donné.

Tous ces principes doivent être appliqués dans un cadre particulier pour consentir la construction d'un méta-modèle à déployer sur des études spécifiques.

Nous considérons comme cas d'application celui de l'AiDR.

3.2.2 Application des principes génériques pour la construction d'un méta-modèle de barrière humaine dans le cadre de l'AiDR

Cette partie vise à définir, à partir des principes génériques de construction du modèle de barrière humaine décrits dans la section précédente, leur application dans le cas de l'AiDR.

Cette application se fait par rapport aux trois niveaux du modèle de barrière humaine :

- le niveau entre les variables des classes 'Efficacité de l'action' et 'Phases' ;
- le niveau entre les variables des classes 'Phases' et 'Déterminants' ;
- le niveau entre les variables des classes 'Déterminants' et 'Facteurs Organisationnels'.

Pour chacun de ces niveaux, nous rappelons l'ensemble des variables de chaque classe de l'AiDR et leurs spécificités telles que cela est défini dans les travaux de thèse (Léger, 2009), puis nous nous focalisons sur les relations de causalité entre variables.

Lors de la description de ces niveaux, nous proposons :

- de formaliser et d'intégrer les PSF issus du chapitre 2 dans la classe 'Déterminants' ;
- puis, pour tenir compte des mécanismes résilients, de généraliser les facteurs organisationnels et adapter leurs liens de causalité avec les Déterminants.

D'un point de vue général, l'application des principes génériques permet de définir un méta-modèle de barrière humaine dans l'AiDR à instancier sur un cas particulier d'action humaine.

3.2.1.1 Niveau I : Segmentation de l'action en trois phases

Dans le cas du modèle de barrière humaine utilisé dans le cadre de l'AiDR, la variable de la classe « Efficacité de l'action » représente une action de conduite ou de maintenance.

Pour ce nouveau modèle de barrière humaine, nous reprenons la segmentation de l'action humaine en trois phases (cf. Tableau 3.4) telle qu'elle a été développée dans (Léger, 2009). Ces trois phases représentent une instance de la classe 'Phase' décrite dans la sous-section 3.3.1.1.

Tableau 3.4. Définition des phases de préparation, de réalisation et de clôture d'après (Léger, 2009).

Phase de l'action	Définition
Préparation	Étape qui permet l'organisation, la spécification et la caractérisation de toutes les conditions nécessaires à la bonne exécution de l'intervention (par le collectif de travail). Elle permet de choisir les moyens les mieux adaptés aux différentes exigences de coûts et de délais. La préparation des moyens matériels, par exemple, a pour but de permettre au personnel d'intervention de disposer, au moment du besoin, des matériels et outillages nécessaires.
Réalisation	Étape qui permet d'implanter l'intervention. Elle peut se décomposer en : 1. La mise en œuvre des moyens (faire exécuter les travaux prescrits avec les moyens en personnels et en matériels requis). 2. La coordination des moyens (assurer la réalisation des travaux dans les délais prescrits, dans le respect des règles et consignes spécifiques, de sécurité et d'interférences entre plusieurs interventions en particulier).
Clôture	Étape qui permet d'assurer la bonne intégration de l'intervention dans le système et de confirmer sa continuité. Elle peut se décomposer en différentes sous-étapes : 1. Le contrôle des travaux (conformité aux spécifications, au cahier des charges, aux règles formelles ou informelles de la qualité et des règles de l'art). 2. La réception des travaux (un travail est terminé lorsque les équipements ont satisfait aux tests prescrits, que le compte rendu d'intervention a été validé, que les données documentaires ont été collectées et que la zone de travail a été restituée). 3. L'analyse des informations collectées en cours et en fin de travaux (donner les informations en cours et en fin de travaux qui sont nécessaires à leur coordination, et au contrôle, et qui servent à alimenter les rapports et historiques de données capi-

talisables actualisées, qualitatives et quantitatives) et le retour d'expérience (rassembler toutes les informations exploitables pour capitaliser et valoriser l'expérience à des fins de progrès).

Pour les liens de causalité entre l'efficacité de l'action et les phases dans le cas de l'AiDR, nous proposons la règle logique suivante :

L'action humaine est efficace si TOUTES les phases sont dans l'état bien faite (ET-logique).

La formalisation des variables et les liens de causalité à ce premier niveau du modèle de la barrière humaine sont traduits dans le modèle RB illustré dans la Figure 3.4.

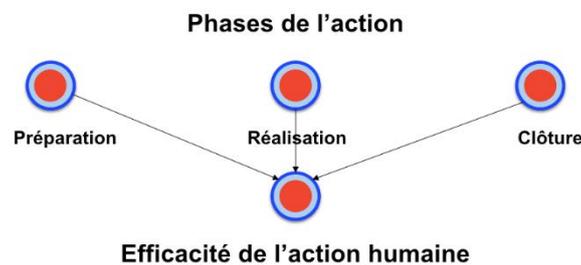


Figure 3.4. Modèle RB du niveau I de la barrière humaine dans l'AiDR.

3.2.1.1 Niveau II : Formalisation des PSF pour leur intégration dans la structure du modèle de barrière humaine de l'AiDR

Au niveau II de la barrière, les phases sont en relation avec les variables de la classe « Déterminants ». Nous proposons donc de formaliser les variables de cette classe et leurs relations de causalité avec les phases. Ce niveau de la structure du modèle de barrière humaine est renforcé par l'intégration de l'ensemble de PSF isolé à partir de l'état de l'art. Nous proposons donc de formaliser ces PSF pour les intégrer dans ce nouveau modèle de barrière humaine de l'AiDR.

Nous précisons les variables de la classe « Déterminants » qui sont en relation avec les phases à ce niveau de la barrière. Plus spécifiquement, nous procédons à leur formalisation sur deux niveaux différents :

- D'une part, nous reprenons la définition des variables de type « Déterminants » retenue de (Léger, 2009) et sur cette base nous la précisons par l'intégration des PSF issus du chapitre 2 dans la structure de

ces déterminants (cf. Tableau 3.5). L'allocation des PSFs a ainsi permis de proposer une nouvelle définition des Déterminants 'Gcdg' et 'Aides' ;

- D'autre part, nous proposons de compléter ce niveau de caractérisation de l'action humaine par l'ajoute du PSF 'Conception du processus de travail, attribution des tâches et management'. En effet, ce PSF a été isolé des autres, car il permet de prendre en compte des aspects concernant la conception et l'organisation du processus de travail par le management, ces aspects étant absents dans le modèle utilisé avant cette thèse (cf. Tableau 3.6).

Tableau 3.5. Structuration des 'Déterminants' sur la base de l'intégration des PSF.

Déterminant	Libellé du PSF	Description
Délégation (De)	-	Fait de remettre la responsabilité d'une tâche à une autre personne, généralement un subordonné. L'hypothèse sous-jacente consiste à considérer que la personne déléguée est compétente pour réaliser la tâche. Le jugement porte sur la décision prise.
Formation (Fo)	-	Ensemble des activités (mise en place par l'entreprise) visant à assurer l'acquisition des capacités pratiques, des connaissances et des attitudes requises pour occuper l'emploi (et donc de s'assurer de l'aptitude des opérateurs à réaliser les actions).
Expérience (Ex)	-	La connaissance acquise par la pratique accompagnée d'une réflexion sur cette pratique.
<i>Gestion Collective et Dynamique de groupe (Gcdg)</i>	<ul style="list-style-type: none"> ▪ <i>Composition</i> ▪ <i>Coordination</i> ▪ <i>Cohésion</i> ▪ <i>Environnement de travail</i> ▪ <i>Leadership</i> 	<p>Capacité du collectif à s'adapter à une situation spécifique et à composer une dérive potentielle pour atteindre les objectifs.</p> <p>Cette capacité dépend d'un ensemble de sous-éléments tels que la composition du collectif de travail, la cohésion entre les membres et leur coordination. En outre, la gestion collective repose sur la capacité du leader d'assurer ces éléments.</p>

<i>Aides (Ai)</i>	<ul style="list-style-type: none"> ▪ <i>Disponibilité des procédures</i> ▪ <i>Qualité et pertinence des procédures</i> ▪ <i>IHM</i> ▪ <i>Communication</i> ▪ <i>Disponibilité d'outils spéciaux</i> 	Ensemble des procédures, outils et moyens (finalisés à la prise de décision) utilisés par le collectif en support à leur activité et dont la disponibilité et la qualité sont fondamentales pour atteindre les objectifs. La procédure est définie comme l'ensemble des documents opératoires prescriptifs (obligatoires comme les consignes, les gammes d'essais, etc.) et non prescriptifs (non obligatoires) utilisés par les opérateurs en support à leur activité. Les outils spéciaux se définissent comme l'ensemble des autres documents (modes d'emploi, etc.) et appareillages venant également en support à l'activité des opérateurs.
Contrôle et Atteinte des Objectifs (Cao)	-	Contrôle : évaluation de la conformité du résultat réel par rapport au résultat attendu, en considérant les moyens mis en œuvre pour atteindre ce résultat ; Atteinte des objectifs : les moyens mis en œuvre pour juger de la visibilité des résultats d'une action, sans considérer ceux utilisés pour atteindre ce résultat.
Retour d'Expérience (Rex)	-	Démarche structurée de capitalisation et d'exploitation des informations issues de l'analyse d'événements positifs et/ou négatifs. Elle met en œuvre un ensemble de ressources humaines et technologiques qui doivent être managées pour contribuer à réduire les répétitions d'erreurs et à favoriser certaines pratiques performantes
Facteurs Contextuels (Fc)	-	Éléments externes influençant le collectif de travail durant la réalisation d'une action et sur lesquels on a peu de prise. Exemples : les conditions de travail, les autres actions en cours, un local enfumé, l'éclairage, etc.

Tableau 3.6. Conception du processus de travail, attribution des tâches et management.

Libellé du PSF	Description
<i>Conception du processus de travail, attribution des tâches et management (Ctm)</i>	<p>Ensemble des activités servant à déterminer le flux de travail, les besoins d'équipement et, les exigences de mise en œuvre d'un processus particulier. Il consiste aussi en la répartition du travail dans les ordres de travail (capacité à respecter le cahier des charges) et en passant ces ordres aux unités concernées. Chaque ordre comprend habituellement les instructions détaillées suffisantes pour permettre l'exécution des tâches et accomplir la mission avec succès.</p> <p>La conception du 'processus de travail' nécessite généralement un certain nombre d'outils, notamment des organigrammes, logiciels de simulation d'un modèle processus, etc.</p>

Il s'agit donc de passer de la structure proposée avant cette thèse (cf. tableau 3.7) à une structure où les influences des Déterminants sur les phases sont celles illustrées dans le Tableau 3.8.

Tableau 3.7. Liens d'influence entre déterminants et phases.

Phases de l'action	Déterminants de l'action
Préparation	Délégation (De) – Formation (Fo) – Aides (Ai)
Réalisation	Expérience (Ex) - Capacité à respecter le cahier de charges (Rcc) - Facteurs contextuels (Fc)- Gestion collective et dynamique de groupe (Gcdg)
Clôture	Contrôle et atteinte des objectifs (Cao) - Retour d'Expérience (Rex)

Tableau 3.8. Conception du processus de travail, attribution des tâches et management.

Phases de l'action	Déterminants de l'action
Préparation	Délégation (De) Formation (Fo) Aides (Ai) <i>Conception du travail, attribution des tâches et management (Ctm)</i>
Réalisation	Expérience (Ex) Facteurs contextuels (Fc) Gestion collective et dynamique de groupe (Gcdg)
Clôture	Contrôle et atteinte des objectifs (Cao) Retour d'Expérience (Rex)

Pour les liens de causalité entre les phases et les déterminants dans le cas de l'AiDR, nous proposons la règle logique suivante :

Chacune des **phases** est *bien faite* si **AU MOINS UN DES déterminants** est dans l'état *adéquat (OU-logique)*.

Ce deuxième niveau de la barrière se traduit dans le modèle RB illustré en Figure 3.5.

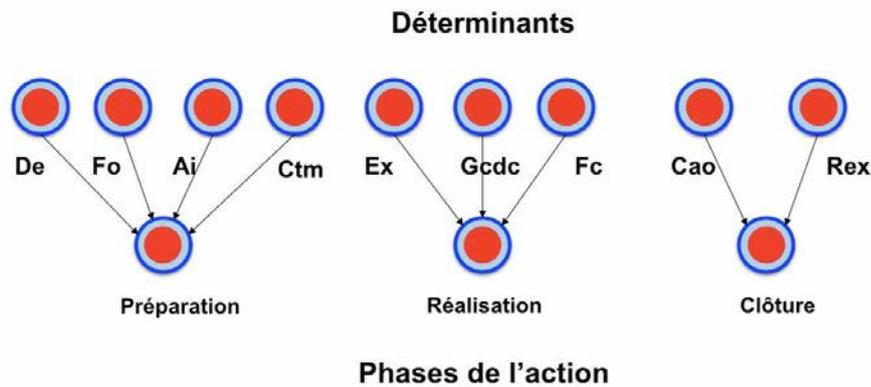


Figure 3.5. Modèle RB du niveau II de la barrière humaine dans l'AiDR.

Compte tenu de la formalisation des PSF et leur intégration dans la barrière humaine, nous proposons maintenant de définir les variables qui composent la classe 'Facteurs Organisationnels' dans le cas de l'AiDR.

3.2.1.3 Niveau III : Généralisation des Facteurs Organisationnels pour la considération des mécanismes résilients

Dans l'AiDR les variables de la classe « Facteurs Organisationnels » sont reprises de (Léger, 2009).

Initialement définis dans une logique d'influence purement pathogène (et par conséquent nommés FOP comme nous l'avons déjà dit plus en amont dans ce chapitre), nous en proposons une généralisation pour aller vers une formulation permettant de prendre en compte les mécanismes résilients que nous avons formalisés dans la section 3.2.

Par conséquent, nous proposons ci-dessous la généralisation de l'ensemble de variables de type « Facteurs Organisationnels » dans le cas de l'AiDR (les modifications apportées par rapport à (Léger, 2009) sont en italique) :

- **Culture de sûreté (CS)** : *ensemble de connaissances, de pratiques et de valeurs partagées et soutenues de façon volontaire par les membres de l'organisation. Elle vise à assurer que la sûreté des installations bénéficie à tout moment (et en toutes circonstances) de toute l'attention requise dans l'atteinte des objectifs ;*
- **Culture de production (CP)** : *ensemble de connaissances, de pratiques et de valeurs partagées et soutenues de façon volontaire par les membres de l'organisation. Elle vise à assurer que la production des installations bénéficie à tout moment (et en toutes circonstances) de toute l'attention requise dans l'atteinte des objectifs. Des injonctions visant à passer outre ou à volontairement ignorer certaines dimensions de la sûreté de manière à favoriser les critères*

de rentabilité de court terme caractérisent son influence pathogène. Il s'agit alors de toutes les pressions de production rencontrées par les membres de l'organisation au détriment de la sûreté (les pressions de production apparaissent lorsque les activités de production ne sont plus contrebalancées par la culture de sûreté) ;

- **Retour d'expérience (REX)** : *consiste à recueillir et analyser les dysfonctionnements qui ont pu survenir dans l'activité. Les connaissances accumulées doivent permettre d'alimenter la réflexion collective pour tirer parti des erreurs passées et aboutir ainsi à des propositions sur des mesures correctives dans l'organisation et le contrôle de l'activité. Ces améliorations visent à prévenir l'occurrence de nouveaux incidents et accidents. Ce processus est itératif et dynamique, c'est ce que l'on retrouve dans le vocable 'faire vivre' un retour d'expérience ;*
- **Traitement de la complexité organisationnelle (TCO)** : *dispositions pour faciliter les relations de travail et les décisions ainsi que les communications portant sur les risques et la sûreté. Cette complexité organisationnelle, que l'ensemble de la littérature des sciences de l'organisation considère comme endogène à l'existence même de tels systèmes sociotechniques, peut avoir plusieurs origines. Elle est liée (1) à la complexité technologique, (2) à l'émergence irrésistible de cloisonnements et (3) à la multiplication des parties prenantes rendant la coordination inter-organisationnelle très délicate. Ces causes originelles peuvent bien sûr se combiner ;*
- **Gestion quotidienne de la sûreté (GQS)** : *mise en œuvre pratique des exigences de sûreté au sein de l'organisation. Elle vise en particulier à assurer une bonne adéquation entre tâches à accomplir et compétences des personnels, à fournir une formation adaptée aux personnels, à s'assurer de la transmission des savoir-faire de sûreté tout au long des parcours professionnels, y compris pour la sous-traitance ;*
- **Réexamen des hypothèses de conception (RHC)** : *définition et prise en compte d'hypothèses de dimensionnement (technique et social). Ces hypothèses sont basées sur une vision du fonctionnement futur du système et sont nécessaires à la conception de tout système technologique ;*
- **Organismes de contrôle (OC)** : *on distingue les autorités de contrôle (internes ou externes) aux différents niveaux de l'organisation suivants : niveau installation (ex. EDF : MSQ), le niveau 'corporate' (ex. EDF : inspection nucléaire), les autorités de sûreté (DRIRE, IRSN), les corps d'inspection des ministères de tutelle (ex : ministère du travail).*

Par rapport à ces définitions *neutres* des facteurs organisationnels, deux modes de propagation ou mécanismes d'influence sont possibles :

- **Mécanismes pathogènes**, s'ils ont une influence pathogène sur les déterminants. Dans ce cas, l'état de référence du facteur organisationnel est par défaut présent et son absence déclenche la pathogénicité ;
- **Mécanismes résilients**, s'ils ont une influence résiliente sur les déterminants. Dans ce cas, l'état de référence du facteur organisationnel est par défaut absent et sa présence déclenche la résilience.

Ainsi, sur la base :

- d'une part, de la formalisation de la présence (ou absence) des liens de causalité entre Déterminants et FOP (résumés dans le Tableau 3.9) proposée avant cette thèse suite à des multiples interrogations conduites à partir des travaux de (Pierlot et Dien, 2007) et puis poursuivies dans les thèses de (Léger, 2009) et (Fallet, 2013),
- et d'autre part, de l'intégration du nouveau PSF 'Ctm' et la généralisation de facteurs organisationnels comme conséquence à l'intégration des mécanismes résilient,

nous proposons de redéfinir ces liens (cf. Tableau 3.10).

Tableau 3.9. Construction des liens de causalité d'après (Léger, 2009).

Déterminants	Facteurs Organisationnels Pathogènes						
	COS	PP	MT	OC	REX	AR	GQS
De		X	X				
Ai	X	X			X	X	
Fo		X	X	X	X		X
Rcc	X	X	X	X			X
Ex		X					X
Gcdg	X	X	X		X		X
Fc		X	X		X	X	
Cao		X		X	X	X	
Rex	X	X	X		X	X	X

Dans le Tableau 3.10, nous avons mis en évidence – en rouge et en *italique* – les facteurs organisationnels issus de la généralisation et le PSF 'Ctm' intégré au niveau des déterminants.

Tableau 3.10. Construction des liens de causalité pour le nouveau modèle de barrière humaine dans l'AiDR.

Déterminants	Facteurs Organisationnels						
	<i>CS</i>	<i>CP</i>	<i>TCO</i>	<i>OC</i>	<i>REX</i>	<i>RHC</i>	<i>GQS</i>
De	X	X	X				X
Ai	X	X		X	X	X	
Fo		X	X	X	X		X
Cmt	X		X				X
Ex		X					X
Gcdg	X	X	X		X		X
Fc		X	X		X	X	
Cao		X		X	X	X	
Rex	X	X	X		X		X

Après avoir formalisé les liens de causalité entre les phases et les déterminants selon le principe de mitigation, nous proposons la structure du modèle RB du troisième niveau de la barrière humaine illustré par la Figure 3.6.

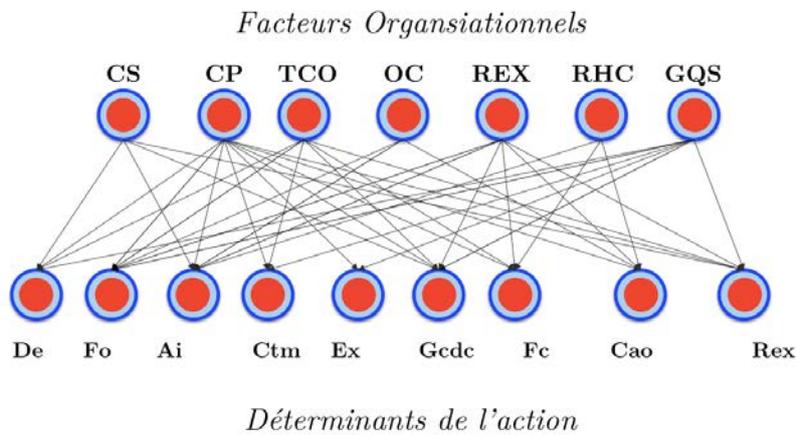


Figure 3.6. Modèle RB du niveau III de la barrière humaine dans l'AiDR.

À la lumière de cette formalisation de nouveaux liens de causalité, nous proposons dans la section suivante une description détaillée et réadaptée des influences entre les facteurs organisationnels et les Déterminants.

En effet, suite à la généralisation des facteurs organisationnels permettant d'intégrer les mécanismes résilients, il est maintenant nécessaire de proposer

une adaptation des liens de causalité identifiés dans (Léger 2009) et une formulation opportune des nouveaux liens.

Nous proposons donc, dans la partie suivante, une formulation adaptée de ces liens en regard de l'intégration des mécanismes résilients dans l'AiDR.

3.2.1.3 Niveau III : Adaptation des liens de causalité entre Facteurs Organisationnels et déterminants pour intégrer les mécanismes résilients

L'objectif de cette partie est de proposer un cadre de causalité générique, *i.e.* applicable à plusieurs types de structures organisationnelles et à différents types de mécanismes d'influence (pathogènes et résilients). Ce cadre peut être particularisé dans différents secteurs, selon la situation analysée et une action spécifique. Cette particularisation se traduit, dans le modèle de barrière humaine prise dans son contexte organisationnel, par la suppression éventuelle de certaines variables du type 'Facteurs Organisationnels' et par conséquent des liens de ces variables sur celles du type 'Déterminant'.

Ainsi, nous formalisons les liens de causalité génériques proposés entre les 'Facteurs organisationnels' et les 'Déterminants' sur lesquels des mécanismes pathogènes et résilients peuvent être identifiés en fonction du scénario particulier considéré.

Délégation (De)

La justification de liens de causalité identifiés avec le déterminant De est la suivante :

- **TCO**: la capacité à identifier les personnes compétentes, dans chacun des domaines, peut dépendre de la configuration organisationnelle. Cette configuration peut à la fois favoriser ou pénaliser la capacité à déléguer ;
- **CP** : face à des objectifs constants, le temps, les budgets, les moyens et les ressources peuvent varier. Selon qu'ils augmentent ou diminuent, l'effet peut être favorable ou pénalisant pour certaines décisions et donc sur la capacité de délégation ;
- **CS** : la définition de ce facteur est axée sur la présence ou l'absence de la culture de sûreté au sein de l'entreprise et donc sur la validité des choix du décideur. Ce dernier peut être plus ou moins sensibilisé à certaines notions de sûreté, et par conséquent être plus ou moins capable de remettre la responsabilité d'une tâche à une autre personne ;
- **GQS** : ce facteur se réfère à l'adéquation des tâches à accomplir et des compétences des personnels (résultats de décisions prises par l'encadrement). Une

gestion plus ou moins favorable de l'adéquation tâches/compétences peut avoir une influence sur la capacité de délégation du décideur.

Aides (Ai)

La justification de liens de causalité identifiés avec le déterminant Ai est la suivante :

- **CS** : si les risques sont correctement identifiés, éventuellement à l'aide d'une liste exhaustive des évènements pour lesquels le système a une réponse prête, alors la définition des plans d'action (programmes de maintenance, etc.) sera adaptée aux risques réellement présents dans l'installation. Dans le cas contraire, les mesures prises pour l'exploitation de l'installation peuvent se révéler inadaptées ;
- **OC** : le contrôle fréquent et l'adaptation des procédures ont un impact très positif sur la qualité et la disponibilité des aides. Dans le cas contraire, des dérives potentielles (banalisation des déviations) peuvent survenir et causer des problèmes à ce niveau. On considérera a minima les organismes de contrôle interne ;
- **REX** : si les méthodes d'analyse du REX sont adaptées et fonctionnelles et si des problèmes au niveau des aides (outils, documents, procédures ...) ont été identifiés durant une action et qu'ils sont résolus au moment de refaire cette action, alors les problèmes ne risquent pas d'apparaître à nouveau du fait du traitement de ce dysfonctionnement ; si ces méthodes étaient insuffisantes, incomplètes, les aides ne seraient pas adaptées et les risques pourraient réapparaître ;
- **CP** : face à des objectifs constants, le temps, les budgets, les moyens et les ressources peuvent varier. Selon qu'ils augmentent ou diminuent, l'effet peut être favorable ou pénalisant sur l'adéquation des outils et des procédures à réaliser ;
- **RHC** : si la définition des plans d'actions (programmes de maintenance, etc.) n'évolue pas au même rythme que les modifications du système alors les documents utilisés pour préparer l'action seront adaptés ; l'absence ou insuffisance d'analyse de ces plans après la détection de défauts de conception ou la non mise à jour des activités après des changements au niveau technique peuvent entraîner des effets pénalisants en cascade non négligeables.

Formation (Fo)

La justification de liens de causalité identifiés avec le déterminant Fo est la suivante :

- **GQS** : si les plans de formations sont réactualisés (en fonction des évolutions du système sociotechnique) et réalisés par des personnes compétentes alors la formation du personnel d'intervention sera toujours adaptée ; dans le cas contraire, les formations pourraient s'avérer partielles ou insuffisantes, etc. ;
- **OC** : si la fonction de 'validation des programmes de formations' est correctement réalisée par les organismes de contrôle (concernés), alors ceci assure un bon déroulement des activités ; dans le cas contraire (la fonction de validation n'est pas réalisée ou que partiellement), les formations pourraient être inadaptées et/ou incomplètes pouvant ainsi aboutir à des actions erronées.
- **TCO**: La complexité organisationnelle doit être telle que la communication et la diffusion d'information impactent positivement la planification, la participation et le contenu des sessions de formation. Toutefois, une certaine configuration organisationnelle complexe peut impacter négativement la diffusion des informations et des communications, ce qui peut selon les cas favoriser ou pénaliser la planification, la participation et le contenu des sessions de formation ;
- **REX** : faire vivre le retour d'expérience de façon systématique au niveau organisationnel peut avoir pour conséquence l'identification de certaines situations à risques biaisées et donc une sensibilisation du personnel d'intervention en ce qui concerne ces situations. Si cela n'est pas vrai, la formation du personnel peut être inadéquate ;
- **CP** : face à des objectifs constants, le temps, les budgets, les moyens et les ressources peuvent varier et selon qu'ils augmentent ou diminuent, l'effet peut être favorable ou pénalisant respectivement pour le nombre et la durée des formations.

Expérience (Ex)

La justification de liens de causalité identifiés avec le déterminant Ex est la suivante :

- **CP** : face à des objectifs constants, le temps, les budgets, les moyens et les ressources sont tels qu'ils permettent l'acquisition d'une expérience pour réaliser correctement les tâches. Une inadéquation des conditions peut conduire à ne plus gagner l'expérience nécessaire à la réalisation des tâches. L'expérience résultante peut être plus ou moins adéquate à réaliser les tâches ;
- **GQS** : si la gestion quotidienne de la sûreté est correctement réalisée, alors le savoir-faire (objectif du 'faire-faire', sous-traitance, etc.) devient une valeur positive qui est accrue et partagée. Ainsi, le niveau d'expérience augmente. Si les anomalies de sûreté répétées ne sont pas corrigées alors ce niveau peut diminuer, avec des conséquences sur la réalisation des activités.

Facteurs de contexte (Fc)

La justification de liens de causalité identifiés avec le déterminant Fc est la suivante :

- **TCO** : la complexité organisationnelle est axée sur les notions de communication et de coordination. Un certain type de configuration organisationnelle peut favoriser ou pénaliser des difficultés de type logistique ou relationnel entre les différentes entités (présence de cloisonnements), ce qui peut avoir une influence positive ou négative sur l'action. Exemple : le fait de transporter une pièce d'un local à l'autre nécessite de coordonner différents services, de mettre à disposition le local et l'outillage, etc. ;
- **CP** : face à des objectifs constants, le temps, les budgets, les moyens et les ressources peuvent varier et selon qu'ils augmentent ou diminuent, on peut avoir une influence favorable ou pénalisante sur les facteurs de contexte. Par exemple, des réductions de budget peuvent conduire à un environnement dégradé pour la réalisation des actions (propreté) ;
- **REX** : si les informations relatives aux incidents, presque accidents et accidents dus à des événements contextuels, sont recensées pour des actions similaires, alors le personnel d'intervention sera plus sensibilisé à ces notions. Dans le cas contraire, si ces informations ne sont pas prises en compte (ou si elles ne le sont que partiellement), les facteurs de contexte pourraient être dégradés ;
- **RHC** : actualiser les informations concernant le système (aspects techniques, humains ou organisationnels) quand les conditions d'exploitation sont modifiées ou partiellement différentes permet de faire évoluer les modes d'intervention du personnel et donc leur contexte. Le fait de ne pas mettre à jour ces informations peut nuire au contexte d'une bonne réalisation de l'action.

Gestion collective et dynamique de groupe (Gcdg)

La justification de liens de causalité identifiés avec le déterminant Gcdg est la suivante :

- **TCO** : selon que les interfaces entre services du site (relations d'influence entre services, responsables, etc.) sont bien définies ou non, la circulation des informations peut avoir un impact favorable ou pénalisant sur la réactivité du collectif de travail ;
- **CS** : si la culture de sûreté a sa place dans l'organisation alors le groupe de travail sera préparé face à des conditions dangereuses qui peuvent conduire à une dérive potentielle ;

- **GQS** : la composante 'métier' contenue dans la définition du facteur de groupe est fortement liée à la 'transmission du savoir-faire' dont la gestion quotidienne de la sûreté est responsable ;
- **REX** : la « Gestion collective et dynamique de groupe » se positionne au niveau du métier (retour d'expérience du groupe) alors que le facteur organisationnel est vu comme un processus global à l'entreprise. De manière générale : comme on s'appuie sur les opérateurs (qui se font leur propre retour d'expérience), l'état du processus de retour d'expérience ne devrait pas influencer Gcdg. Mais si le processus de retour d'expérience (global au site) est bien fait, cela peut faciliter la gestion collective (du fait du climat dans lequel évoluent les opérateurs) ;
- **CP** : face à des objectifs constants, le temps, les budgets, les moyens et les ressources peuvent varier et selon qu'ils augmentent ou diminuent le climat relationnel dans le groupe peut favoriser ou nuire respectivement à la réalisation de l'action par le collectif.

Contrôle et atteinte des objectifs (Cao)

La justification de liens de causalité identifiés avec le déterminant Cao est la suivante :

- **OC** : l'amélioration continue (qui inclut la notion de contrôle en temps réel) se définit en partie par la notion de contrôle. De plus, le point clé du processus de qualité est de participer à l'amélioration continue. Aujourd'hui, en interne, les organismes de contrôle sont submergés ; on peut également ajouter les problèmes de collusion entre les services qui peuvent influencer le déroulement des contrôles ;
- **REX** : le travail effectué durant le processus de retour d'expérience permet d'identifier les problèmes et de poser des solutions pour y pallier (donc de participer à l'amélioration du fonctionnement du système). Si le processus du retour d'expérience est bien intégré dans la politique organisationnelle alors les solutions proposées peuvent servir à favoriser l'atteinte des objectifs ;
- **CP** : face à des objectifs constants, le temps, les moyens et les ressources peuvent varier. Selon qu'ils augmentent ou diminuent, la redéfinition des actions prioritaires peut être lourde de conséquences et plus ou moins productrice de valeurs ;
- **RHC** : si les hypothèses de conception sont questionnées, cela signifie que l'organisation est dans un contexte favorable au contrôle et à l'interrogation sur l'atteinte des objectifs.

Retour d'expérience (Rex)

La justification de liens de causalité identifiés avec le déterminant Rex est la suivante :

- **CS** : si la conception de la sûreté/sécurité des cadres et des experts est flexible et valorisée, les analyses de risques sont réactualisées ou réexaminées compte tenu de changements techniques et/ou organisationnels et du REX traité lors d'une action ; toutefois, certains marqueurs abordent des notions qui peuvent aller à l'encontre du retour d'expérience (blâme, dilution des activités de sûreté, non prise en compte des alertes). En particulier pour la culture du blâme : les gens auront beaucoup de mal à faire état des difficultés donc à réaliser concrètement le REX ;
- **GQS** : l'analyse d'événements liés au fonctionnement du système, activité centrale pour faire vivre le retour d'expérience, est fonction de l'évolution du personnel et des connaissances qu'il a du système. De bonnes connaissances sur le système facilitent l'analyse et la création d'une liste des événements exhaustifs ; dans le cas contraire, l'appel à du personnel de substitution, de remplacement, non qualifié, rend moins efficace la pratique du REX ;
- **TCO** : le retour d'expérience est une démarche qui nécessite un ensemble de ressources humaines. Une certaine configuration organisationnelle peut impacter de manière favorable ou pénalisante au niveau de la communication et de la coordination, ce qui peut avoir une influence positive ou négative sur la constitution de ce retour d'expérience ;
- **REX** : le travail effectué durant le processus de retour d'expérience permet d'identifier les problèmes et de poser des solutions pour y pallier (donc de participer à l'amélioration du fonctionnement du système). Si le processus du retour d'expérience est bien intégré dans la politique organisationnelle alors les solutions proposées peuvent servir à favoriser la pérennité du retour d'expérience liée à une action ;
- **CP** : le retour d'expérience est un processus qui nécessite du temps. Face à des objectifs constants, le temps, les budgets, les moyens et les ressources peuvent varier et influencer positivement ou négativement la démarche.

Conception du processus de travail, attribution des tâches et management (Ctm)

La justification de liens de causalité identifiés avec le déterminant Ctm est la suivante :

- **TCO** : si les dispositions au niveau de l'organisation sont définies pour faciliter les relations de travail et les décisions, ainsi que les communications portant sur les risques alors la conception du travail sera conséquente et efficace. Dans le cas contraire, la non-prise en compte de la complexité et l'opacité des

différents acteurs peuvent nuire à une structuration adaptée et claire du travail et des tâches ;

- **CP** : face à des objectifs constants, le temps, les moyens et les ressources peuvent varier. Selon qu'ils augmentent ou diminuent, la définition des actions prioritaires peut avoir des conséquences positives ou négatives ;
- **GQS** : si la gestion quotidienne de la sûreté est correctement réalisée alors les exigences de mise en œuvre d'un processus particulier sont satisfaites et toutes les tâches nécessitant des outils et logiciels spéciaux sont prévues et organisées.

Au global, l'application des principes génériques de construction du modèle de barrière humaine présentés dans la section 3.2 de ce chapitre et simultanément la formalisation des PSF, la généralisation des facteurs organisationnels et l'adaptation des liens de causalité, ont conduit, dans l'AiDR, à un méta-modèle de barrière humaine prise dans son contexte organisationnel (cf. Figure 3.7).

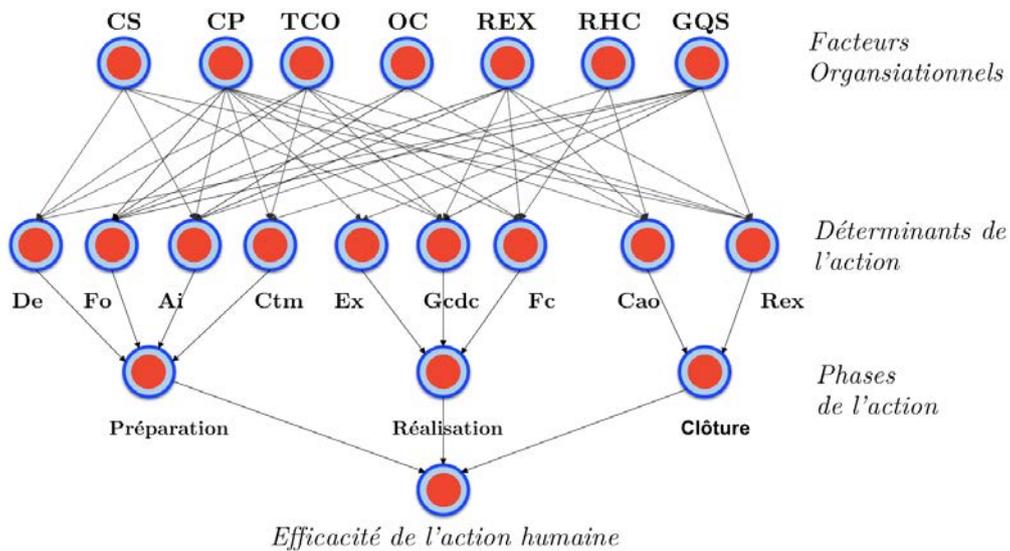


Figure 3.7. Méta-modèle de barrière humaine prise dans son contexte organisationnel de l'AiDR

Ce **méta-modèle** de barrière humaine de l'AiDR nécessite d'être instancié sur un cas particulier pour conduire à une évaluation probabiliste de l'efficacité d'une action humaine spécifique prise dans son contexte organisationnel.

Ainsi, nous proposons des principes permettant de guider l'utilisateur de l'AiDR pour créer un modèle de barrière humaine sur un cas particulier.

3.3 Principes d'instanciation du méta-modèle de barrière humaine de l'AiDR pour créer un modèle particulier

Nous proposons ici les principes d'instanciation du méta-modèle Réseaux Bayésiens de barrière humaine de l'AiDR permettant l'évaluation probabiliste réaliste de l'efficacité de la barrière éventuellement prise dans son contexte organisationnel dans le cas d'une action spécifique.

Ces principes ont l'objectif de définir un protocole de traitement des avis d'experts nécessaires pour évaluer les liens de causalité entre les variables du méta-modèle, puis d'estimer des paramètres permettant de construire les TPC et par conséquent de quantifier les distributions de probabilité dans le cas spécifique considéré. Ces principes tiennent compte des hypothèses faites concernant l'intégration de la résilience (principe de mitigation des mécanismes pathogènes) et donc elles sont pertinentes par rapport à la structure mathématique imposée pour réaliser cette intégration et au besoin d'estimation des paramètres lors de l'extraction des connaissances via élicitation d'experts.

3.3.1 Définition d'un protocole de traitement des avis d'expert

Les deux aspects que nous formalisons à ce stade : traitement des avis d'experts et paramétrisation du modèle, doivent être résolus en réduisant la difficulté des experts pour se prononcer sur des situations complexes produites par l'interaction entre mécanismes pathogènes et résilients. Si les experts étaient en mesure de caractériser quantitativement chacun des liens de causalité identifiés dans la section précédente et donc définir de façon précise et objective chacun des paramètres de la TPC, alors il n'y aurait pas de verrou scientifique. Malheureusement, l'estimation directe des paramètres est complexe, car le principe de mitigation conduit à des effets combinés sur la variable finale, ce qui rend difficilement quantifiables les paramètres par l'expert.

Face à cette complexité, nous proposons à l'expert d'évaluer des probabilités directement sur la variable finale. Ces probabilités font référence à des situations bien identifiées.

Nous proposons donc les choix suivants pour le traitement des avis d'expert :

- 1) les experts doivent être élicités sur des situations où un nombre réduit de mécanismes pathogènes et résilients sont présents (situations de complexité réduite) ;

- 2) les situations évaluées doivent être assez probables conduisant à des situations déjà rencontrées par l'expert ;
- 3) les experts doivent s'exprimer de façon libre sur une échelle de probabilité de référence fondée sur un retour d'expérience reconnu par la communauté.

Concernant le troisième point, nous proposons d'exploiter une grille d'élicitation de référence pour le traitement des avis d'experts. Cette grille est obtenue par l'extension de la grille utilisée dans le cadre des analyses EPFH (cf. Tableau 3.11) avec la méthode MERMOS (Le Bot, *et al.* 2002).

Tableau 3.11. Grille d'élicitation utilisée dans MERMOS.

Grille d'élicitation de MERMOS	
Echelle nominale	Probabilité
Très probable	0.9
Probable	0.5
Assez probable	0.3
Peu probable	0.1
Assez peu probable	0.03
Très peu probable	0.01

Cette extension de la grille utilisée dans MERMOS est nécessaire, car elle manque de précision notamment entre 0.5 (Probable) et 0.9 (Très Probable). En fait, une variété majeure de situations très complexes peut se produire en raison de l'interaction entre mécanismes résilients et pathogènes et le principe de mitigation modélisant cette interaction. Par rapport à cette variété des situations produites, les experts ont besoin d'affiner leurs jugements et donc de disposer d'une grille plus fine.

Par conséquent, nous proposons un niveau intermédiaire – appelé « Fortement Probable » – entre les niveaux 0.5 et 0.9. Cette extension nous permet d'aboutir à une grille adaptée (cf. Tableau 3.12) pour le traitement des avis d'experts lors de l'instanciation du méta-modèle de barrière humaine dans le cas de l'AiDR.

Tableau 2.12. Grille d'élicitation de référence dans le cadre de l'AiDR.

Grille d'élicitation dans l'AiDR	
Echelle nominale	Probabilité
Très probable	[0.9 – 1.0]
<i>Fortement Probable</i>	0.7
Probable	0.5
Assez probable	0.3

Peu probable	0.1
Assez peu probable	0.03
Très peu probable	0.01

Ainsi, au travers de cette grille de référence et dans l'objectif de quantifier des situations peu complexes, un certain nombre de questions sont à poser aux experts pour quantifier les paramètres du modèle. En ce sens, il faut préciser que les questions posées lors de l'extraction des connaissances ne sont pas générées de façon automatique (aléatoire), mais elles sont construites sur la base de la capacité de l'expert à se faire une représentation la plus objective possible des situations à évaluer. Ceci a abouti à des questions relativement simples tout en étant basées sur une sémantique pertinente de l'expert.

Globalement, il s'agit en effet de résoudre un *problème sous contraintes*. En fonction du rapport entre le nombre de questions et le nombre de paramètres à estimer, trois problèmes doivent être résolus : déterminé, sous-déterminé et sur-déterminé.

Dans cette thèse, nous avons traité le cas d'un problème déterminé en raison de l'applicabilité de la démarche proposée et pour assoir la crédibilité du processus pour l'estimation des paramètres par l'élicitation d'experts. En effet :

- le cas sur-déterminé est moins intéressant, car il oblige à poser plus de questions ;
- le cas sous-déterminé impliquant des solutions paramétriques qui ne sont pas non plus d'intérêt dans l'application du modèle.

Ainsi, relativement à un cas déterminé, nous proposons les principes d'estimation des paramètres pour instancier le méta-modèle de barrière humaine de l'AiDR sur un cas spécifique.

3.3.2 Définition d'une approche d'estimation des paramètres pour le modèle intégrant l'interaction entre mécanismes résilients et pathogènes

Pour l'estimation paramétrique, nous avons procédé avec deux approches différentes : dans un premier temps, une technique de criblage a été envisagée afin d'observer la relation entre les contraintes imposées dans la construction du problème et les solutions admissibles. Ensuite, nous sommes passés à une technique d'optimisation mathématique pour l'estimation des paramètres à partir des avis d'experts.

Nous présentons d'abord la première de ces deux techniques.

3.3.2.2 Approche par criblage et création d'arbres décisionnels

L'approche par criblage consiste à calculer des arbres décisionnels où les paramètres moyens du modèle sont calculés à partir de l'exploration d'un espace de valeurs admissible par la structure mathématique du modèle. Cet espace est calculé sur la base de :

- n , le nombre de paramètres à estimer ;
- n , le nombre de questions à poser (pour un cas déterminé) et donc de réponses quantifiées b_i ;
- v , le nombre de niveaux discrets dans l'espace de probabilité $[0,1]$;

Sur cette base, l'exploration a pour objectif de générer toute la combinatoire des paramètres en fonction des réponses admises, permettant ainsi de cibler les valeurs des paramètres de la TPC. Ainsi, l'algorithme de criblage suivant a été appliqué pour :

- calculer la matrice initiale M ayant v^n lignes et $2 * n$ colonnes. Chaque ligne de M est calculée par le produit cartésien de valeurs de paramètres. Elle constitue l'espace global des solutions ;
- procéder de façon itérative pour :
 - sur chaque colonne associée à la réponse b_i : effectuer une réorganisation de cette colonne et sur la base de la grille d'élicitation utilisée (cf. Tableau 3.11), regrouper les réponses b_i dans des intervalles I de valeurs;
 - extraction de la sous-matrice M_i contenant l'ensemble des valeurs des paramètres associées aux réponses b_i dans l'intervalle i_i et estimation de la valeur moyenne pour chaque paramètre p_i .

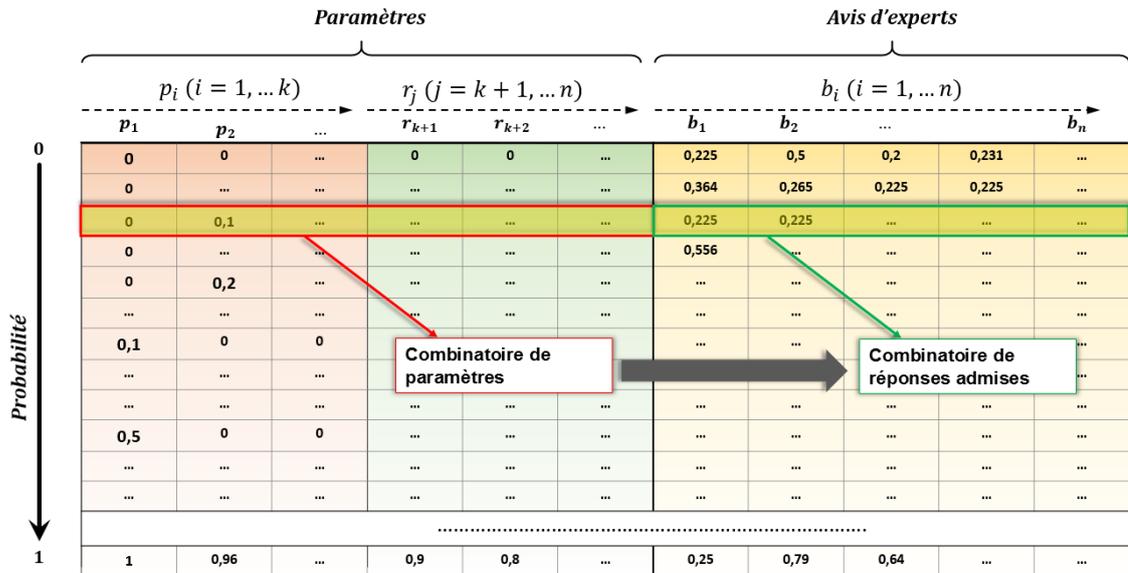


Figure 3.8. Matrice initiale M correspondant à l'espace des solutions du problème.

Les intervalles ont été ainsi définis dans le Tableau 3.11 ci-dessous à partir de la grille d'élicitation de l'AiDR (cf. Tableau 3.13).

Tableau 3.13. Intervalles associés aux niveaux de la grille d'élicitation.

A]1.00 - 0.90]
B]0.90 - 0.70]
C]0.70 - 0.50]
D]0.50 - 0.30]
E]0.30 - 0.10]
F]0.10 - 0.03]
G]0.03 - 0.01]

À titre illustratif, considérons un cas avec trois facteurs organisationnels, dont deux produisent des mécanismes pathogènes avec une probabilité représentée par les paramètres p_1 , p_2 et l'autre produit un mécanisme résilient avec paramètres $r_{3,1}$, $r_{3,2}$ et $r_{3,f}$ mitigeant les mécanismes pathogènes. Nous montrons ainsi comment cette procédure est déployée pour estimer les 5 paramètres associés aux mécanismes pathogènes et résilients produits par les facteurs organisationnels, et en utilisant 7 niveaux de discrétisation de l'espace de probabilité $[0,1]$ de la grille d'élicitation de référence. Au global, nous avons généré une matrice initiale M de $7^5=16807$ lignes et $2 \cdot 5=10$ colonnes.

Dans la Figure 3.9, le criblage est appliqué à la colonne des réponses b_1 de la matrice initiale M calculée pour ce cas.

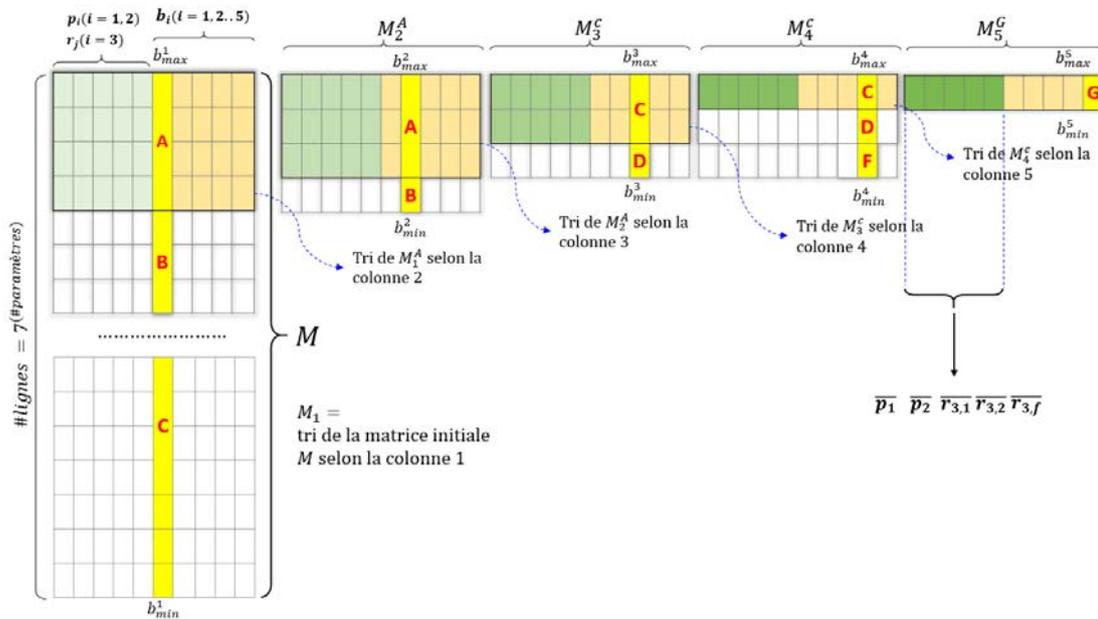


Figure 3.9. Criblage appliqué à une colonne de réponses pour un cas à 5 paramètres.

Ainsi, un ensemble de paramètres moyens est identifié en correspondance de chaque combinatoire de réponses issues du criblage. Considérons par exemple les intervalles associés à chaque colonne b_i dans la Figure 3.9, une première combinatoire de réponses serait $b_1 = A, b_2 = A, b_3 = C, b_4 = C, b_5 = G$.

L'application de la procédure permet de générer un arbre décisionnel tel qu'il est montré en Figure 3.10. Cet arbre nous permet ainsi de disposer d'un ensemble de paramètres moyens pour chaque combinaison possible de réponses.

Avis d'experts					Paramètres moyens				
b_1	b_2	b_3	b_4	b_5	\bar{p}_1	\bar{p}_2	$\bar{r}_{3,1}$	$\bar{r}_{3,2}$	$\bar{r}_{3,f}$
A	A	A	A	A	0,38	0,68	0,88	0,38	0,38
				B	0,21	0,34	0,38	0,99	0,14
			B	A	0,87	0,52	0,12	0,38	0,38
				B	0,01	0,02	0,56	0,78	0,24
		B	A	A	...				
				B	...				

Figure 3.10. Exemple d'un arbre décisionnel.

Au-delà de tout aspect relatif aux coûts computationnels, cette procédure s'avère plutôt intéressante et représente une solution simple à appliquer pour des situations caractérisés par un nombre de variables limité. Néanmoins, l'exploita-

tion de cette technique pour des cas avec plus de variables (et donc plus de paramètres à estimer) comporte quelques limites, notamment la discrétisation de l'espace de probabilité des paramètres. En effet, cette dernière étant arbitraire, elle ne peut pas être aussi fine que l'on souhaite pour couvrir l'espace réel admissible par le modèle. Ce constat a été vérifié pour un cas relativement simple, qui, après une discrétisation plus fine, a conduit à des arbres décisionnels plus complets. En même temps, il est très difficile de retrouver les b_i admissibles dans la grille de référence.

Compte tenu des limites rencontrées avec la technique de criblage et de l'imprécision due à une discrétisation nécessaire à contrôler le coût computationnel, il a été envisagé d'analyser une autre approche basée sur l'optimisation mathématique permettant d'utiliser directement les b_i de la grille de référence en entrée pour estimer les paramètres du modèle. Cette technique nécessite cependant une formulation du problème et un choix de l'algorithme d'optimisation plus complexes.

3.3.2.1 Optimisation mathématique

Cette deuxième approche consiste à traiter l'estimation des paramètres du point de vue de l'**optimisation mathématique**.

Dans notre cas, le problème à résoudre est un problème de *minimisation sous contraintes*. Ce problème consiste à minimiser l'erreur commise par un algorithme d'optimisation dans l'estimation des paramètres d'un système d'équations non-linéaires dont les termes connus sont fournis en entrée via l'élicitation d'experts. En effet, le système d'équations non-linéaires traduit en forme mathématique l'ensemble des questions posées à l'expert pour la quantification probabiliste des situations à évaluer.

Ainsi, nous proposons une approche d'optimisation qui se définit sur la base des éléments suivants :

- 1) Le **critère** ψ , est la *fonction-objectif* à minimiser pour l'estimation des paramètres. Considérons un vecteur $\underline{\theta} = \{\theta_1, \dots, \theta_n\}$ représentatif des paramètres à estimer :

$$\psi(\underline{\theta}) = \|\underline{g}(\underline{\theta}) - \underline{b}\|^2 \quad (3.4)$$

où \underline{g} représente le système d'équations non linéaires. Ainsi, dans un espace cartésien, le critère devient :

$$\psi(\underline{\theta}) = \sum_{i=1}^n (g(\theta_i) - b_i)^2 \quad (3.5)$$

Ce critère consiste en la minimisation du carré des écarts entre les valeurs \underline{b}_i fournies par l'expert et ceux calculés à partir des valeurs $\underline{\theta}_i$. La surface du critère dépend à la fois des questions posées (discriminations des effets individuels des mécanismes pathogènes et résilients) et des réponses fournies par l'expert (qui est contraint par la grille d'élicitation proposée) (cf. Figure 3.11).

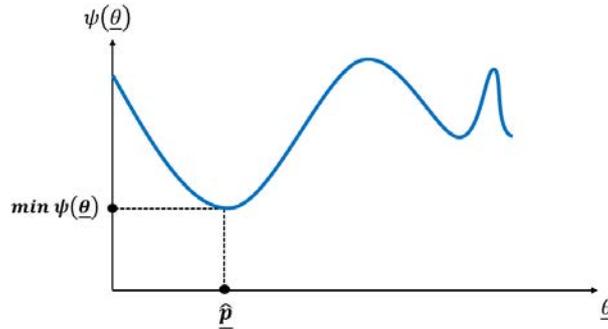


Figure 3.11. Représentation bidimensionnelle du critère dans l'espace décrit par les θ_i . Les paramètres à estimer correspondent au point où la fonction a son minimum (Jorge et Stephen, 2006).

Selon le conditionnement du critère (questions posées et précision des réponses), la surface du critère change et nous avons un minimum global (en cas de bon conditionnement) ou des minima locaux (en cas de mauvais conditionnement).

- 2) Les **contraintes sur les paramètres** θ_i : pour que les θ_i représentent des probabilités, les paramètres à estimer doivent rester dans l'intervalle $[0,1]$: $0 \leq \theta_i \leq 1$ avec $i = 1, \dots, n$.
- 3) Des **inégalités de contrainte** entre les φ_i calculées à partir des paramètres estimés θ_i . Elles permettent d'assurer la cohérence du modèle avec les principes de mitigation des facteurs résilients.

Le vecteur $\underline{\hat{b}}$ des réponses calculées à partir des paramètres estimés θ_i est obtenu :

$$\underline{\hat{b}} = g(\underline{\theta}_i) \quad (3.6)$$

Ces réponses $\underline{\hat{b}}$ sont utilisées dans un *critère de qualité de l'estimation* (CQE), comme suit :

$$CQE = \psi(\underline{\theta}) = \|\underline{\hat{b}} - \underline{b}\|^2 = \sum_{i=1}^n (\hat{b}_i - b_i)^2 \leq n \quad (3.7)$$

Ce critère est appliqué afin de valider les résultats par rapport aux réponses fournies au départ.

Cette interaction avec l'expert constitue une étape d'un protocole d'élicitation (cf. Figure 3.12). Ce protocole permet à l'expert de valider ou invalider les réponses b_i (fournies en entrée) par rapport aux nouvelles réponses \hat{b}_i (calculées à partir des paramètres estimés).

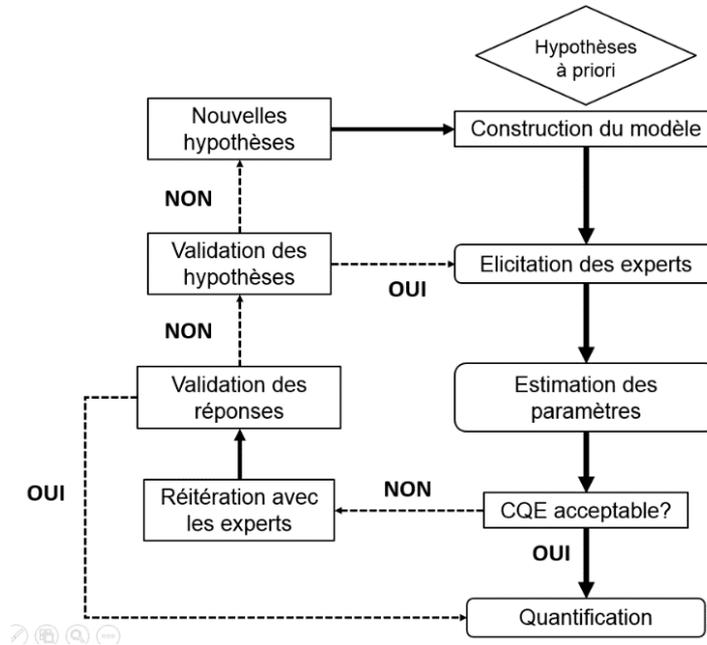


Figure 3.12. Protocole d'élicitation d'experts dans l'AiDR.

Si l'expert ne valide pas les b_i , une deuxième élicitation est faite pour converger vers un ensemble de réponses dont il est plus sûr. Si par contre, l'expert non seulement décide de revenir sur ses réponses, mais en plus remet en cause les hypothèses (pathogènes et résilientes) alors une analyse des influences identifiées au départ est nécessaire. Cette remise en question des hypothèses initiales et par conséquent la reconfiguration de la barrière humaine au niveau des facteurs organisationnels permet d'aboutir à plus proche de la réalité vécue par les experts.

Grâce à ces itérations avec l'expert et à son expérience, ce protocole permet au modélisateur de tendre vers un modèle exact de la probabilité d'efficacité de l'action.

3.4 Conclusions

Dans ce chapitre nous avons présenté les contributions scientifiques de cette thèse. Ces contributions répondent aux verrous scientifiques identifiés à la fin du chapitre 2, que nous rappelons ici :

1. Formalisation des PSF pour leur intégration dans la structure du modèle de barrière humaine de l'AiDR ;
2. Élaboration et formalisation des marqueurs résilients ;
3. Formalisation de l'interaction entre mécanismes résilients et pathogènes dans le cadre des relations non-déterministes ;
4. Définition d'un protocole de traitement des avis d'experts ;
5. Définition d'une approche d'estimation des paramètres pour le modèle intégrant l'interaction entre mécanismes résilients et pathogènes.

Par rapport à ces verrous scientifiques, nous avons tout d'abord traité l'intégration de la résilience dans l'AiDR. Nous avons donc élaboré des marqueurs résilients à partir des marqueurs pathogènes issus d'un REX disponible après l'analyse de plusieurs accidents/incidents. Ensuite, nous avons formalisé la résilience et ses mécanismes dans le cadre de la modélisation probabiliste. La modélisation de ces mécanismes et l'interaction avec ceux pathogènes, déjà inclus dans la barrière humaine, ont été abordées en adoptant un point de vue partiel, celui de la *mitigation*. Le principe de mitigation a été ainsi formalisé comme étant le mécanisme d'interaction entre variables pathogènes et résilientes produits par les facteurs organisationnels et dont l'impact est évalué sur les déterminants de l'action. Ces principes et la modélisation de cette interaction ont été formalisés plus spécifiquement dans le cas d'une modélisation probabiliste basée sur le formalisme des RBs. Ainsi, un nouveau modèle générique de barrière a été proposé pour définir les mécanismes causaux dans le cadre de relations non-déterministes qui caractérisent les interactions entre ces deux niveaux de la barrière humaine.

Compte tenu de cette formalisation des mécanismes résilients, dans la seconde partie de ce chapitre nous avons proposé des principes génériques de construction du modèle d'une barrière humaine. Nous avons donc appliqué ces principes au cas de l'AiDR. Lors de cette application, l'ensemble de PSF isolé dans le chapitre 2 à partir des méthodes EPFH a été intégré dans la barrière. Cette intégration a permis de répondre au besoin de renforcer la caractérisation de l'action par ses déterminants dans le modèle de barrière de l'AiDR. Notamment, nous avons proposé et intégré le PSF 'Conception du processus de travail, attribution des tâches et management' au niveau du collectif de travail et de management (déterminants). Il permet dans le nouveau modèle de barrière humaine de l'AiDR de prendre en compte des aspects managériaux sur la conception du processus de travail et l'attribution des tâches.

De plus, à la lumière de mécanismes résilients formalisés dans la première partie du chapitre, nous avons proposé de passer d'un ensemble de facteurs organisationnels pathogènes (FOP) à des facteurs organisationnels génériques. Cette généralisation permet de prendre en compte une multitude de situations qui ne pouvaient pas être prises en compte avant. Le cadre de causalité liant facteurs

organisationnels et déterminants a été redéfini à la lumière de ces généralisations. Globalement, ces principes génériques et l'adaptation des variables particulières utilisées dans le cadre de l'AiDR, aboutissent à un méta-modèle de barrière humaine à déployer sur une action spécifique.

Enfin, en raison des contraintes que cette prise en compte de la résilience comporte au niveau de la construction de la TPC du modèle RB de la barrière, nous avons proposé des principes pour la définition d'un traitement des avis d'expert et l'estimation des paramètres pour une évaluation réaliste des probabilités de l'efficacité des actions humaines. Cette estimation a été proposée selon deux approches différentes : d'une part, avec une technique de criblage dédiée à la création d'arbres décisionnels, d'autre part, par une technique d'optimisation fondée sur la minimisation d'une fonction-objectif définie sur la base du questionnaire conçu par l'utilisateur pour guider l'expert à la caractérisation de situations issues du terrain. Malgré la nécessité d'utiliser un algorithme d'optimisation, cette dernière approche semble préférable, car la connaissance apportée par l'expert permet de quantifier les paramètres de façon objective et permet aussi d'adapter le modèle RB sur la base d'un protocole itératif avec l'expert. Cependant, l'informatisation actuelle de cette technique d'optimisation n'a pas permis de solutionner l'estimation des paramètres pour un modèle complet de barrière humaine dans l'AiDR. En effet, seulement 4 facteurs organisationnels parmi les 7 possibles ont pu être considérés pour caractériser le contexte organisationnel de l'action à évaluer. A terme, une autre voie d'informatisation de cette technique est donc à étudier.

Chapitre 4

Validation des propositions à un cas réel d'EDF

Les résultats issus de cette phase de validation étant confidentiels, l'ensemble du chapitre applicatif de cette thèse fait l'objet d'une clause de confidentialité vis-à-vis d'EDF.

Le cas d'application sera présenté dans les revues qui suivront la thèse.

Conclusions et perspectives

Conclusions et perspectives

Cette thèse de doctorat CIFRE s'inscrit dans le cadre d'une collaboration entre le Centre de Recherche en Automatique de Nancy (CRAN) et le Département Management des Risques Industriels (MRI) de la R&D d'Electricité de France (EDF R&D).

Elle s'est basée sur les travaux que depuis plusieurs années EDF et le CRAN conduisent et qui ont permis d'aboutir à une méthodologie multidisciplinaire globale dite Analyse Intégrée des Risques (AiDR) (Léger 2009), (Fallet 2013) opérationnelle à ce jour. Il s'agit d'une approche d'analyse de risques qui traite des systèmes sociotechniques complexes pris dans leur environnement physique et réglementaire et dont les composants techniques sont soumis à des actions de maintenance et/ou de conduite pour lesquelles l'efficacité est évaluée en fonction de leur contexte organisationnel et environnemental (Duval, *et al.* 2012).

Plusieurs expérimentations sur des cas industriels réels dans différents secteurs, tels que l'électrique et le chimique, ont mis en évidence certaines limites dans le traitement des facteurs humains et socio-organisationnels du modèle de barrière humaine de l'AiDR. Notamment, il a été souligné le besoin de dépasser une certaine vision « dysfonctionnelle » adoptée au début du développement de l'AiDR par rapport à la considération des dimensions humaine et organisationnelle et de s'attaquer à la considération, malgré toute sa complexité, d'aspects de causalité plutôt résilients pouvant conduire à une évaluation réaliste de l'efficacité d'une action humaine dans les systèmes sociotechniques.

Ces travaux de thèse ont eu pour objectif d'aboutir à cette évaluation réaliste de la probabilité de l'efficacité des barrières humaines prises dans leur contexte organisationnel. Pour satisfaire à cet objectif, des contributions ont été apportées par rapport à un ensemble de verrous technologiques principaux:

1. Un renforcement de la structure du modèle de barrière humaine pour aboutir à un modèle par facteurs – organisationnels et du collectif de travail (non individuels) – enrichis par l'identification et l'intégration d'un ensemble de facteurs de forme, appelés performance shaping factors (PSF), issus des méthodes d'Evaluation Probabiliste de la Fiabilité Humaine (EPFH) ;
2. Prendre en compte la résilience dans la modélisation probabiliste pour atteindre une évaluation de l'efficacité humaine le plus fidèle possible à la réalité;

3. Un traitement global des avis de l'expert permettant d'estimer d'une manière objective les paramètres du modèle proposé.

En regard de ces verrous, les contributions majeures développées dans cette thèse se structurent autour des trois axes suivants :

- Une confortation du modèle de barrière humaine dans la description des déterminants de l'action, *i.e.* les facteurs représentatifs du collectif de travail et management, et leurs liens de causalité avec les facteurs de l'organisation. Cette confortation est le résultat d'une fertilisation-croisée entre l'AiDR et les méthodes EPFH utilisant les PSFs ;
- La modélisation de la résilience et ces mécanismes dans le cadre des relations causales non-déterministes, puis la formalisation de ces mécanismes dans un modèle RB permettant de prendre en compte l'interaction entre mécanismes pathogènes et résilients et leurs effets conjoints sur l'efficacité d'une action humaine ;
- Un traitement des avis d'expert d'abord et l'estimation paramétrique du modèle ensuite permettant de rendre exploitable le modèle de barrière humaine, à savoir de construire les tables de probabilités conditionnelles (TPCs) et produire des résultats réalistes en vue d'une prise en compte des probabilités d'efficacité des actions humaines lors d'une évaluation des risques avec l'AiDR.

Ces contributions sont porteuses de plusieurs originalités comme :

- L'approfondissement de l'approche par facteurs pour l'évaluation de l'efficacité des barrières humaines prises dans leur contexte organisationnel. Cette investigation répond aux besoins des méthodes d'analyse de risques permettant de traiter à la fois les aspects techniques, humains, organisationnels et environnementaux d'un système sociotechnique complexe pour identifier ses vulnérabilités et y mettre prioritairement les moyens de réduction. Le traitement des facteurs humains et socio-organisationnels y est particulièrement traité ;
- Le développement du concept de mécanismes résilients et leur caractérisation qualitative et quantitative dans une analyse de risques pour mitiger les effets perturbateurs produits par les mécanismes pathogènes jusqu'ici considérés et aboutir à une probabilité réaliste de l'efficacité d'une action humaine ;
- Une structure mathématique (des éléments du modèle retenu) proposée selon le principe de mitigation entre mécanismes pathogènes et résilients ;

- Un modèle d'élicitation d'experts fondé sur des probabilités objectives, d'ou un questionnaire non pas orienté vers l'affectation directe d'une valeur de probabilité aux paramètres, mais vers la caractérisation de situations bien définies et dont la représentation par un expert est relativement simple. L'expression objective de l'expert sur la probabilité de ces situations permet, grâce à ce modèle, de procéder à l'estimation des paramètres.

Pour résumer, la première plus-value scientifique de nos travaux réside dans la proposition d'une approche à la fois de traçage et de modélisation des mécanismes résilients au niveau humain et organisationnel au sein d'un système sociotechnique. À cette considération de la résilience a suivi toute la difficulté d'une double considération des mécanismes pathogènes et résilients dans le cadre de causalité des relations non-déterministes de la barrière humaine. Il en résulte un deuxième résultat majeur de la thèse : un modèle d'interaction entre ces deux types de mécanismes et la possibilité d'intégrer cette interaction dans le cadre probabiliste.

Afin de les valider, ces contributions ont été mises en œuvre sur un cas réel de barrière humaine importante dans un scénario d'inondation externe d'une unité de production d'électricité d'EDF. Cette application de taille réduite, mais permettant tout de même d'illustrer les problématiques solutionnées par ces contributions, a été l'occasion de démontrer la faisabilité de nos propositions et la cohérence des résultats par rapport à ceux obtenus avant leur mise en place.

Sur la base de ces résultats et des objectifs initiaux, des perspectives à court terme se dégagent pour réduire l'incertitude sur l'évaluation probabiliste de l'efficacité de l'action humaine :

1. L'application du nouveau modèle global de la barrière de l'AiDR à des cas de barrières de plus en plus complexes (7 facteurs organisationnels). Ceci nécessite l'extension de l'approche d'optimisation mathématique développée dans la thèse afin de permettre l'estimation d'un nombre plus important de paramètres. Cette extension peut se traduire par le choix d'une méthode d'optimisation différente, comme les méthodes pour la résolution de problèmes non convexes de programmation non linéaire de nombres entiers mixtes (MINLP) décrites dans (Bao, *et al.* 2015). Pour ce genre de problèmes, des solveurs sont proposés tel que le solveur « BARON » (Sahinidis, 1996) ;
2. Le passage à l'échelle industrielle par l'intégration de ces barrières dans la modélisation globale de risques d'un système déjà traité par l'AiDR dans (Simon, *et al.* 2014). L'objectif est d'étudier l'impact de la probabilité réelle d'efficacité des barrières humaines considérées dans le système sur les différents risques et prioriser les moyens par rapport à ces risques.

Au-delà des perspectives liées aux objectifs initiaux de la thèse, d'autres à long terme sont à investiguer :

3. La considération des résultats de l'AiDR en termes d'efficacité de la barrière dans l'établissement des scénarios fait par l'EPFH à EDF. En effet, il est possible de prendre en compte les déterminants de la barrière humaine issue de cette thèse ayant contribuant le moins à l'efficacité de l'action et déduire des scénarios pour l'EPFH sur la base de ces déterminants.
4. L'intégration de la résilience peut être complétée en considérant les mécanismes de récupération en plus des mécanismes de mitigation étudiés dans la thèse. Cette considération implique d'intégrer la modélisation des aspects temporels dans le méta-modèle de barrière humaine de l'AiDR. Cette intégration pourrait s'appuyer sur les travaux de (Codetta-Raiteri, *et al.* 2014).

Bibliographie personnelle

- De Galizia A., *et al.* 2015a. Investigation avancée des méthodes d'étude probabiliste d'efficacité des actions humaines dans les systèmes sociotechniques. *11ème Congrès International Pluridisciplinaire en Qualité, Sécurité de Fonctionnement et Développement Durable*. 17 – 19 Mars 2015, Nancy, France.
- De Galizia A., *et al.* 2015b. Advanced Investigation of HRA Methods for Probabilistic Assessment of Human Barriers Effectiveness in Complex Systems for a given Organizational and Environmental Context. *12th International Topical Congress on Probabilistic Safety Assessment and Analysis PSA 2015, American Nuclear Society*, 25 – 30 Avril 2015, Sun Valley, Etats Unis.
- De Galizia A., *et al.* 2015c. Contribution à une méthodologie de quantification de la probabilité de l'efficacité des actions humaines prises dans leur contexte organisationnel et environnemental. *Note H-T51-2015-04332-FR*, EDF Lab Paris-Saclay.
- De Galizia A., *et al.* 2016a. Modeling Non-deterministic Causal Mechanisms involving Resilience in Risk Analysis. *8th IFAC/IEEE congress on Manufacturing Modeling Management and Control*. 27 - 30 Juin 2016, Troyes, France.
- De Galizia A., *et al.* 2016b. Markers and Patterns of Organizational Resilience for Risk Analysis. *13th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems*. 29 August – 2 Septembre 2016, Kyoto, Japon.
- De Galizia A., *et al.* 2016c. Evaluation probabiliste d'une barrière humaine prise dans son contexte organisationnel. *20ème Congrès Lambda-Mu sur Maîtriser les risques dans un monde en mouvement*. 11 – 13 Octobre 2016, Saint Malo, France.
-

Bibliographie

- AFNOR, 2005. Sécurité fonctionnelle - Systèmes instrumentés de sécurité pour le secteur de l'industrie de procédés.
- AIEA, 1991. Culture de sûreté, Rapport du Groupe consultatif international pour la sûreté nucléaire, Vienne.
- Allaire, G., 2005. Analyse numérique et optimisation: Une introduction à la modélisation mathématique et à la simulation numérique. Editions Ecole Polytechnique.
- Anderies, J., *et al.*, 2013. Aligning key concepts for global change policy: robustness, resilience, and sustainability. *Ecology and society*, Volume 18(2), p. 8.
- Bao, X., *et al.*, 2015. Global optimization of nonconvex problems with multilinear intermediates. *Mathematical Programming Computation*, 7(1), 1-37.
- Bell, J., *et* Holroyd, J. 2009. Review of human reliability assessment methods. *Health and Safety Laboratory*, United Kingdom.
- Boring, R. L., 2010. How Many Performance Shaping Factors are Necessary for Human Reliability Analysis? Dans: *Proceedings of the 10th International Probabilistic Safety Assessment et Management Conference (PSAM10)*, Seattle, WA.
- Codetta-Raiteri, *et al.*, 2014. Approaching dynamic reliability with predictive and diagnostic purposes by exploiting dynamic Bayesian networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*.
- Confalonieri, R., *et al.*, 2010. A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions. Dans: *Ecological Modelling*, Volume 221(F), pp. 960-964.
- Conrady S. et Jouffe L., 2015. Bayesian Networks & BayesiaLab – A Practical Introduction for Researchers.
- Dang, E., *et al.* 2009. International HRA Empirical Study-Phase 1 Report. NU-REG/IA-0216. Washington, DC: US NRC.
- Dauphiné, A. et Provitolo, D., 2007. La résilience: un concept pour la gestion des risques. *Annales de géographie*, Volume 2, pp. 115-125.

- Diez, F. et Druzdzel, M., 2006. Canonical Probabilistic Models for Knowledge Engineering. *Technical Report CISIAD-06-01*, UNED, Madrid, Spain.
- Druzdzel, M., et Simon, A. (1993, July). Causality in Bayesian belief networks. Dans : *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, pp. 3-11. Morgan Kaufmann Publishers Inc.
- Duff, S., *et al.*, 2014. The diagnosis and measurement of team resilience in sociotechnical systems. *2014 7th IEEE International Symposium on Resilient Control Systems (ISRCS)*, pp. 1-5.
- Duval, C., 2014. Analyse de risques de perte de la Source Froide de Cruas incluant les aspects humains et organisationnels, *Note Technique EDF R&D*, Clamart.
- Duval, C., *et al.*, 2012. A Bayesian network based integrated risk analysis (IRA) approach for industrial systems: application to the heat sink system and prospects development. *Proceedings to the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability*, Volume 226(5), pp. 488-507.
- Edmondson, A. C., 1996. Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *The Journal of Applied Behavioral Science*, Volume 32(1), pp. 5-28.
- Embrey, D. E., 1983. Use of performance shaping factors and quantified expert judgment in the evaluation of human reliability: An initial appraisal, NUREG/CR-2986. Brookhaven National Lab., Upton, NY.
- Emery, F. et Trist, E., 1978. Analytical model for sociotechnical systems. Dans: *Sociotechnical systems : A Sourcebook*, pp. 120-133.
- Eriksson, D. M., 1997. A Principal Exposition of Jean-Louis Le Moigne's Systemic Theory. *Cybernetics and Human Knowing*, Volume 4 , pp. 2-3.
- Fallet, J., 2013. Contribution à la modélisation et au traitement de l'incertain, Thèse de doctorat, Université de Lorraine.
- Galàn, S., *et al.*, 2007. Incorporating organizational factors into probabilistic safety assessment of nuclear power plants through canonical probabilistic models. *Reliability Engineering and System Safety*, Volume 92, pp. 1131-1138.
- Gertman, D. I., *et al.*, 2005. Human error quantification using performance shaping factors in the SPAR-H method. *Proceedings of the Human Factors*

- and Ergonomics Society Annual Meeting*, Volume 52 (21), pp. 1733-1737. SAGE Publications.
- Gilly, J. P., *et al.*, 2014. Resilience of organisations and territories: The role of pivot firms. *European Management Journal*, Volume 32(4), pp. 596-602.
- Grecco, C. H. *et al.*, 2013. A Fuzzy Logic Based Method to Monitor Organizational Resilience: Application In a Brazilian Radioactive Facility. *2013 International Nuclear Atlantic Conference-INAC*, pp. 24-29.
- Hollnagel, E., 2006. Resilience : the challenge of the unstable.
- Hollnagel, E., *et al.*, 2006. Resilience engineering: concepts and precepts. Aldershot, UK: Ashgate.
- Hongwei, H., 2006. State of the art of the research on risk management in construction of tunnel and underground works. *Chinese Journal of Underground Space and Engineering*, Volume 2(1), pp. 13-20.
- ISO, G., 2009. ISO 31000, AFNOR.
- Jensen, F. V., 1996. An introduction to Bayesian networks. London UCL press.
- Kirwan, B., 1997a. Validation of human reliability assessment techniques: part 1 – validation results. *Safety Science*, Volume 27(1), pp. 25-41.
- Kirwan, B., 1997b. Validation of human reliability assessment techniques: part 2 – validation results. *Safety Science*, Volume 27(1), pp. 43-75.
- Kirwan, B., *et al.*, 2005. Nuclear action reliability assessment (NARA): a data-based HRA tool. In *Safety and Reliability*. Volume 25 (2), pp. 38-45. Taylor & Francis.
- Laprie, J. C. *et al.*, 1995. Guide de la sûreté de fonctionnement, Toulouse, Cépaduès.
- Le Bot, P. et Pesme, H., 2010. The Model of Resilience in Situation (MRS) as an idealistic organization of at-risks systems to be ultra safe. PSAM 2010, Seattle.
- Le Moigne, J. L., 1990. La modélisation des systèmes complexes. Paris: Bordas, Dunot.
- Léger, A., 2009. Contribution à la formalisation unifiée des connaissances fonctionnelles et organisationnelles d'un système industriel en vue d'une évaluation quantitative des risques et de l'impact des barrières envisagées. Thèse de doctorat, Université Henri Poincaré-Nancy I.
- Léger, A. *et al.*, 2008a. Modeling of human and organizational impacts for system risk analyses. *9th International Probabilistic Safety Assessment and Management Conference*. Hong-Kong, Chine.

- Léger, A. *et al.*, 2008b. A safety barriers-based approach for the risk analysis of socio-technical systems. *IFAC Proceedings*, Volume 41(2), pp. 6938-6943.
- Léger, A. *et al.*, 2009. Methodological developments for probabilistic risk analyses of socio-technical systems. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, Volume 223(4), pp. 313-332.
- Lindbom, A. et Rothstein, B., 2007. La résilience du modèle suédois de Welfare dans l'économie mondialisée. *Revue internationale de politique comparée*, Volume 13(3), pp. 429-445.
- Lo, J. C., *et al.*, 2015. Individual Markers of Resilience in Train Traffic Control The Role of Operators' Goals and Strategic Mental Models and Implications for Variation, Expertise, and Performance. *The Journal of the Human Factors and Ergonomics Society*, p. 0018720815622357.
- Lyons, M., *et al.*, 2004. Human reliability analysis in healthcare: A review of technique. *International Journal of Risk & Safety on Medicine*.
- Magne, L. et Vasseur, D., 2006. Risques industriels, complexité, incertitude et décision: une approche interdisciplinaire. Paris: Lavoisier.
- Magne, L., 2010. Histoire sémantique du risque et de ses corrélats. Dans: *Journées d'histoire de la comptabilité et du management*.
- Mathevet, R. et Bousquet, F., 2014. Résilience et environnement: penser les changements socio-écologiques. Paris: Buchet Chastel.
- Meyer, P., Le Bot, P. et Pesme, H., 2007. MERMOS : an extended second generation HRA method. *2007 IEEE 8th Human Factors and Power Plants and Hprct 13th Annual Meeting*, pp. 276-283. Monterey CA.
- Mkrtchyan, L., Podofillini, L., & Dang, V. N., 2015. Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability engineering & system safety*, Volume 139, pp. 1-16.
- Mohaghegh, Z. et Mosleh, A., 2009. Incorporating organizational factors into probabilistic risk assessment of complex socio-technical systems: Principles and theoretical foundations. *Safety Science*, October, Volume 47(8), p. 1139–1158.
- Morin, E., 2014. Le paradigme perdu. Seuil.
- Mosleh, A., *et al.*, 2007. Human reliability analysis methods. Selection guidance for NASA. Office of Safety and Mission Assurance, Washington, DC, 123.

- OCDE, 2003. Les Risques Emergeants au XXIe Siècle.
- Paries, J., 2006. Complexity, emergence, resilience. Dans: *Resilience Engineering. Concepts and precepts*. Hampshire: Ashgate.
- Paté-Cornell, M. et Murphy, D., 1996. Human and Management factors in probabilistic risk analysis: the SAM approach and observations from recent applications. *Reliability Engineering and System Safety*, Volume 53, pp. 115-126.
- Peretti-Watel, P., 2010. La société du risque. La découverte.
- Pescaroli, G., 2014. Pathogenic vulnerabilities and resilient factors in systems and populations experiencing a cascading disaster. Deliverable 2.1.
- Picoche, J., 1992. Étymologie du français, Paris: Le Robert, coll. « Les usuels ».
- Pierlot, S. et Dien, Y., 2007. From organizational factors to an organizational diagnosis of the safety. *2007 European safety and reliability conference (ESREL)*, Norway.
- Pradier, P. C., 2004. Histoire du risque. Dans: *Historia de la Probabilidad y la Estadística*. Delta Publicaciones, pp. 171-186.
- Rasmussen, J., *et al.*, 1990. Human error and the problem of causality in analysis of accidents. *Philosophical Transactions of the Royal Society B: Biological Sciences*, Volume 327(1241), pp. 449-462.
- Reason, J., 1997. Managing the risks of organizational accidents. Aldershot, United Kingdom: Ashgate.
- Reiman, T., 2007. Assessing Organizational Culture in Complex Sociotechnical Systems. Dans: *Methodological Evidence from Studies in Nuclear Power Plant Maintenance Organizations*. Finlande, Espoo: VTT.
- Ruault, J., 2008. La place de l'humain dans le contexte des systèmes de systèmes. Dans: *Systèmes de systèmes : concepts et illustrations pratiques*. Hermès Science Lavoisier, Paris.
- Sadorsky, P., 2001. Risk factors in stock returns of Canadian oil and gas companies. *Energy economics*, Volume 23, pp. 17-28.
- Sahinidis, N., 1996. BARON : a general purpose global optimization software package. *Journal of Global Optimization*, 8 :201–205.
- Tisseron, S., 2014. La résilience: « Que sais-je? ». Presses universitaires de France.

- Van der Leeuw, S. E., et Aschan-Leygonie, C., 2005. A long-term perspective on resilience in socio-natural systems. Micro, meso, macro: addressing complex systems couplings. World Scientific, Hackensack, New Jersey, pp. 227-264.
- Villemeur, A., 1988. Sûreté de fonctionnement des systèmes industriels. Dans: *Collection de la direction des études et recherches d'Électricité de France*. Paris: Eyrolles, p. 744.
- Villemeur, A., 1992. Reliability, Availability, Maintainability and Safety Assessment, Assessment, Hardware, Software and Human Factors. Volume 2, Wiley.
- Weber, P., *et al.*, 2012. Overview on Bayesian networks. Engineering Applications of Artificial Intelligence, Volume 25, pp. 671-682.
- Weber, P. et Simon, C., 2016. Benefits of Bayesian Network Models, ISBN: 978-1-84821-992-2, pp. 74-89, August 2016, éd. Wiley-ISTE.
- Woods, D. et Cook, R., 2009. Incidents – markers of resilience or brittleness? Ashgate, Aldershot, UK.
- Zio, E., 2007. From complexity science to reliability efficiency: a new way of looking at complex network systems and critical infrastructures. *International Journal of Critical Infrastructure*, Volume 3(3), pp. 488-508.

Annexes

Annexes

I Les méthodes EPFH analysées

Critères de sélection des méthodes

Comme précisé au début de ce document, nous traitons dans cette étude des systèmes sociotechniques complexes où la prise en compte des actions humaines et de l'environnement est essentielle dans le processus de gestion des risques. Dans l'industrie nucléaire, comme dans d'autres domaines, les méthodes et outils d'évaluation de la fiabilité humaine ont été mis en œuvre et intégrés à l'analyse des risques. La classification proposée dans la section 2.1.4 permet une taxonomie de ces méthodes basées sur leur évolution temporelle. Toutefois il est encore difficile de comprendre exactement les raisons de l'application de certaines plutôt que d'autres. Dans l'objectif de répondre à ce besoin, d'autres critères ont été identifiés pour sélectionner les méthodes HRA en raison de leur mise en œuvre réelle dans les EPFH.

Ces critères ont été identifiés à partir d'une investigation de nature bibliographique sur la littérature – anglais en particulier – portant sur l'analyse de la fiabilité humaine dans des domaines comme l'aérospatial, le nucléaire, le pétrochimique etc. Notamment, dans (Mosleh, *et al.*, 2006) la NASA envisage une sélection de méthodes EPFH afin d'étudier les risques liés à l'erreur humaine dans les missions spatiales. Certains besoins identifiés dans (Léger, 2009) ont aussi contribué à préciser ces critères. D'autres travaux publiés dans le département MRI portant sur la validation de méthodes EDF comme dans (Le Bot, *et al.*, 2002,) ou sur un « état de l'art de l'EPFH », ont été retenus afin de déterminer les critères proposés. Ils sont brièvement présentés dans le Tableau I.1.

Tableau I.1. Critères de sélection des méthodes EPFH d'intérêt.

Critère	Description
La méthode est-elle applicable aux systèmes nucléaires?	Nous nous adressons principalement à des méthodes EPFH doivent être applicables aux installations nucléaires existantes.
La méthode quantifie-t-elle :	La méthode EPFH doit inclure des procédures pour la quantification de la probabilité d'erreur humaine (HEP). Pour être le plus adaptable aux besoins de

<ul style="list-style-type: none"> la probabilité d'erreur humaine (HEP) ? la probabilité d'échec d'une action FH éventuellement prise de son contexte organisationnel ? 	<p>l'AiDR, la méthode (ou un ensemble de méthodes choisies) doit inclure des procédures d'identification d'erreur, la modélisation d'erreur, et détermination des HEP, plutôt que seulement une évaluation qualitative des scénarios probables de conduire à des erreurs.</p>
La méthode fournit-elle une liste des PSF?	La plupart des méthodes applique les PSF dans le processus de quantification, mais pour certaines parmi elles ces facteurs sont inconnus. D'autres n'utilisent pas une liste fixe de PSF. Ainsi, nous nous intéressons aux méthodes qui présentent au moins l'ensemble des PSF utilisé pour les calculs.
Quelle est la variabilité des sources de données?	La méthode AiDR doit s'adapter à des données provenant d'une grande variété de sources, y compris les simulateurs, les études de performance humaine, et potentiellement, les avis d'experts. Les méthodes HRA devraient avoir la même flexibilité pour les infos d'entrée afin de pouvoir les comparer.
La méthode identifie-t-elle des « errors of omission » et « errors of commission » ?	La méthode HRA devrait traiter à la fois les deux types d'erreurs. L'AiDR vise à traiter les opérations en conduite nominale et d'accident (Léger, 2009). Les opérations nominales se réfèrent aux activités humaines exercées dans une situation où il n'y a pas les défaillances du système (maintenance). Les opérations d'urgence se réfèrent aux activités humaines réalisées avec au moins la perception de l'échec du système complet ou partiel. Cela peut nécessiter l'utilisation de procédures d'exploitation d'urgence et pourrait changer le caractère de la mission.
La méthode est-elle citée dans la littérature?	Une revue de la littérature est considérée comme critère pour juger la « popularité » et par conséquent une certaine crédibilité de la méthode analysée.

Dans le tableau A.2, nous avons listé les méthodes EPFH qui répondent au mieux à ces critères vis-à-vis des exigences de développement concernant l'AiDR. Elles sont listées par ordre chronologique.

Tableau I.2. Liste (chronologique) des méthodes EPFH issues des critères du Tableau A.1.

Méthode	Références	Date
THERP*	Swain et Guttman – Rapport WASH1400/ NUREG 75/014	1983
SLIM-MAUD*	Embrey et Kirwan – NRC NUREG/CR-3518	1984
HCR*	Hannaman et Spurgin – EPRI RP 2170-3	1984
TRIPOD-Delta*	SHELL	1985
HEART*	Williams – NRC	1988
CREAM**	E. Hollnagel – Halden Lab. (Norvège)	1994
ATHEANA**	Cooper <i>et al.</i> – NRC NUREG/CR-6350	1996

MERMOS**	Le Bot <i>et al.</i> – EDF	1998
SPAR-H**	Gertman <i>et al.</i> – NRC NUREG/CR-6883	1999
NARA***	British Energy (UK)	2005

* Méthodes de première génération

** Méthodes de deuxième génération

*** Méthodes de troisième génération

De plus, les méthodes EPFH envisagées ont les caractéristiques suivantes : (1) relativement faciles à utiliser ; (2) capables de fournir une procédure explicite pour l'estimation des HEP ; et (3) ne nécessitent pas une vaste quantité de données et informations pour l'analyse des tâches (quand elle est utilisée dans les modes de dépistage).

Dans les paragraphes suivants, nous présentons plus en détail les différentes méthodes sélectionnées. En effet, il est nécessaire d'affiner la classification proposée dans la section 2.1.5 sur la base d'un ensemble d'attributs ou caractéristiques descriptives. Ces attributs permettent d'identifier les éléments clés pour une évaluation plus détaillée de chaque méthode. Dans la section 2.1.6, ces attributs sont discutés et appliqués aux méthodes sélectionnées.

Sélection et présentation des méthodes

Attributs de description des méthodes

Après une revue de la littérature sur les méthodes EPFH (Mosleh *et al.*, 2006) (Boring, 2010) une liste de caractéristiques ou « attributs » a été identifiée de sorte que chacune de ces méthodes peut être décrite et comparée à une variété de méthodes différentes. La liste de ces critères a été affinée en tenant compte des informations dont nous avons besoin pour traiter les problèmes de l'AiDR.

Ainsi, parmi les attributs qu'ont été identifiés pour comparer les méthodes, nous en avons sélectionnés 9 répondants à ces besoins. Ils sont listés ci-dessous :

- (1) Contexte de développement ;
- (2) Dépistage ;
- (3) Paramètres, PSFs et modèles causals ;
- (4) Décomposition des tâches et procédure de calcul des HEP ;
- (5) Applications ;
- (6) Disponibilité ;
- (7) Validation ;
- (8) Avantages ;
- (9) Limites.

Les attributs de description sont décrits plus en détail dans le Tableau A.3.

Tableau I.3. Attributs descriptifs des méthodes EPFH sélectionnées pour l'étude.

Attribut	Description
Contexte de développement	L'histoire de développement de la méthode et le domaine dans lequel elle a été mise au point. Un tel contexte a souvent un impact significatif sur la structure de la méthode, les hypothèses, les résultats, et la mesure d'application à d'autres domaines et contextes.
Paramètres, PSFs et modèles causals	La plupart des méthodes HRA utilise un ensemble fixe de PSF pour l'analyse prédictive (c'est à dire, quantifier les HEPs) ou rétrospective (identifier les causes principales). Certaines méthodes permettent à l'analyste de spécifier l'ensemble des PSF en fonction de l'analyse des tâches étudiées. Pour une analyse rétrospective, un modèle de causalité est nécessaire à l'analyste. Nous considérons les méthodes qui fournissent un ensemble de PSF, en précisant explicitement leurs dépendances. Celles qui présentent seulement un ensemble de PSF sans expliquer les mécanismes liant ces facteurs ne sont pas retenus pour avoir un modèle de causalité. Seules les méthodes qui fournissent plusieurs couches et dépendances PSFs explicites sont créditées d'avoir un modèle de causalité.
Dépistage	Certaines méthodes exigent des détails, des efforts et des temps considérables pour effectuer une analyse. Dans certains cas, l'application de telles analyses approfondies peut ne pas être possible (par exemple, pour les nouveaux modèles) ou nécessaire (événements peu importants dans les EPS). Pour cette raison, certaines méthodes fournissent une procédure pour une analyse de dépistage qui nécessite moins d'information et d'effort.
Décomposition des tâches	La décomposition des tâches est un processus clé avec lequel une méthode EPFH décompose les activités humaines d'intérêt dans une liste de sous-tâches qui correspondent à des « unités d'analyse » de la méthode ou tâches de base. Certaines méthodes d'étude exigent et fournissent des lignes directrices pour une telle décomposition, d'autres pas.
Procédure de calcul des HEP	Désigne les détails de la procédure et des données utilisées par la méthode dans le calcul des HEP.
Application	Les secteurs dans lesquels la méthode EPFH a été/est appliquée (recherche, industrie etc.).
Disponibilité	La plupart des méthodes sont accessibles au public. Cependant, certaines nécessitent des outils d'analyse particuliers. Dans certains cas, des données supplémentaires peuvent être nécessaires pour effectuer une analyse. L'acquisition de ces outils et des données pourrait entraîner des coûts supplémentaires.
Validation	Il s'agit pour la plupart des méthodes de validation empirique.
Avantages	/
Limites	/

En outre, nous proposons de classer les méthodes EPFH sur la base d'une taxonomie souvent utilisée dans les études portant sur des méthodes et outils de la fiabilité humaine. Nous reprendrons ainsi celle utilisée dans (Dang, *et al.*, 2009). Il s'agit d'une classification qui nous permettra d'identifier les points

communs et les divergences entre un groupe de méthodes ayant au moins une base commune. A cette base correspondront les méthodes qui apportent des connaissances vis-à-vis des trois axes de développement qui nous intéressent : le traitement des PSF et d'autres facteurs utilisés pour la quantification, la formalisation de jugements d'experts et toutes connaissances de nature subjective, la prise en compte du contexte (organisationnel et environnemental) pour l'analyse.

La classification repose ainsi sur la distinction en trois grandes familles :

- méthodes factorielles ;
- méthodes contextuelles ;
- méthodes basées sur jugement d'experts.

Dans les sections qui suivent, nous affectons les méthodes du Tableau A.3 aux différentes familles et chaque méthode est détaillée suivant les attributs descriptifs précédents.

Méthodes FACTORIELLES

Les méthodes dites factorielles utilisent principalement des facteurs dits PSF (Performance Shaping Factors) ou facteurs de performance. Les PSF sont des facteurs présents tout le temps, liés aux dimensions humaines, organisationnelles et de l'environnement de travail, qui peut donner certains types d'erreurs. Ces méthodes considèrent que ces facteurs ont des répercussions directes sur l'exécution de la tâche.

Il s'agit par exemple du temps disponible pour réaliser l'action, de l'interface homme-machine, de la formation, des procédures, de l'organisation et de la complexité de la tâche. L'effet de ces facteurs est pris en compte dans la quantification des erreurs.

Nous avons identifié 3 méthodes factorielles parmi celles qui sont présentées dans le Tableau A.2 : THERP, HCR et SPAR-H.

THERP

Libellé	Technique for Human Error Rate Prediction
Contexte de développement	Méthode basée sur les théories de (Swain et Guttman, 1983).
Paramètres, PSF et	TRC: Time-Reliability Correlation

modèles causals	Trois PSF sont prédéfinis (training level, stress, and expérience). La méthode permet à l'utilisateur d'ajouter des PSF supplémentaires.
Dépistage	Méthode de dépistage des activités de deux types : actions basées sur les règles et actions basées sur le diagnostic.
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Décomposition des tâches en éléments (identification des PSFs) • Affectation des HEP à chaque élément • Estimation des effets des PSF sur chaque élément (par jugements d'experts, modèles mathématiques, TRC) • Evaluation des effets de dépendance entre les tâches • Modélisation avec un arbre d'événements • Quantification de la probabilité d'erreur de la tâche
Application	Pré/Post-Accidentels
Disponibilité	THERP-Handbook (manuel) : en libre accès
Validation	NRC, (Swain et Guttmann, 1983)
Avantages	<ul style="list-style-type: none"> • Bien utilisée dans la pratique • Possibilité de vérification • Basée sur des bases des données présentes dans le manuel THERP-Handbook
Limites	<ul style="list-style-type: none"> • Mise en œuvre difficile

Tableau A.5. Synthèse de la méthode THERP.

HCR

Libellé	Human Cognitive Reliability
Contexte de développement	Méthode basée sur l'hypothèse que la probabilité de réussite ou d'échec de l'opérateur effectuant une tâche critique au sens du temps dépende fortement du processus cognitif mis en place pour la prise des décisions déterminantes pour la tâche même.
Paramètres, PSF et modèles causals	<ul style="list-style-type: none"> • PNR : Probabilité de Non-Réponse d'une équipe de pilotage à un stimulus • $T_{1/2}$: Temps moyen (médiane) nécessaire pour répondre à ce même stimulus (détecter, traiter l'information, élaborer une réponse adéquate) • PSFs : Performing Shaping Factors. <ul style="list-style-type: none"> ➤ K1 = Expérience de l'opérateur ➤ K2 = Niveau de stress ➤ K3 = Qualité de l'interface opérateur-installation
Dépistage	Non
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> ▪ Analyse de la situation d'intérêt et relation par rapport aux trois niveaux de prise de décision du modèle SKR ▪ Combinaison des PSFs et calibration des courbes « temps de réponse » présentes en littérature ▪ Détermination du temps moyen de réponse nominale $T_{1/2}$ nominal ▪ Adaptation du $T_{1/2}$ (nominal) au contexte spécifique par la formule suivante : $T_{1/2} = T_{1/2} \text{ nominal} \times (1 + K1) (1 + K2) (1 + K3)$

	<ul style="list-style-type: none"> ▪ Calcul de la fenêtre temporelle T nécessaire à l'opérateur pour accomplir la tâche considérée ▪ Estimation du temps de Non-Réponse sur la base de la formule : $\Delta T / T^{1/2}$ ▪ Détermination de la PNR courbe sélectionnée <p>Estimation des HEP par utilisation d'autres méthodes HRA</p>
Application	Industrie nucléaire. Recherche, Post-accidentelle
Disponibilité	Non
Validation	Non
Avantages	<ul style="list-style-type: none"> ▪ Prise en compte du temps ▪ Le modèle SKR est bien intégré dans la méthode et il explique bien les trois modes de prise de décision ▪ Technique simple et rapide à utiliser
Limites	<ul style="list-style-type: none"> ▪ La relation entre les tâches et les trois modes S-K-R est difficile à établir. Ceci peut être cause d'erreurs ▪ Estimation de la HEP pas complète ▪ Les mêmes courbes identifiées pour l'analyse sont utilisées pour des types d'échec très différents ▪ Types de PSFs limités à trois. Autres facteurs existent qui peuvent affecter la performance humaine ▪ Estimation de HEP très sensible au calcul du $T^{1/2}$

Tableau A.6. Synthèse de la méthode HCR.

SPAR-H

Libellé	Standardized Plant Analysis Risk - Human Reliability Assessment
Contexte de développement	Méthode fondée sur l'ergonomie, plutôt qualitative et basée sur un processus d'affectation de la tâche humaine sous analyse à l'une de deux catégories génériques : action et diagnostic.
Paramètres, PSF et modèles causals	<ul style="list-style-type: none"> • PSFs: Performing Shaping Factors (Available time, Stress and stressors, Experience and training, Complexity, Ergonomics and human-machine interface, Procedures, Fitness for duty, Work Processes) (Gertman, <i>et al.</i>, 2004)
Dépistage	Elle ne fournit pas une procédure pour des analyses de dépistage
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Description générique de l'événement de base en considération • Identification de la catégorie correspondante à l'événement de base • Estimation des HEPs sans dépendance (nominale) grâce à la procédure de multiplication des PSFs implémentée dans la feuille de calcul de SPAR-H • Détermination de la HEP totale en utilisant les HEPs d'action, de diagnostic et jointe • Identification du niveau approprié de dépendance • Estimation sur la feuille de calcul de SPAR-H de la probabilité d'échec de la tâche sur la base du niveau de dépendance
Application	Industrie nucléaire
Disponibilité	Manuel complet sur le site de l'U.S.NRC

Validation	En cours par la NRC
Avantages	<ul style="list-style-type: none"> ▪ Les PSF sont très exhaustifs ▪ Prise en compte dans la quantification des dépendances entre les erreurs
Limites	<ul style="list-style-type: none"> ▪ Inadaptée aux situations réelles ▪ Degré de spécification des PSFs non satisfaisant pour des analyses très détaillées

Tableau A.7. Synthèse de la méthode SPAR-H.

Méthodes CONTEXTUELLES

Les méthodes contextuelles modélisent l'activité en utilisant principalement le concept de « EPC » (Error Producing Conditions) ou conditions favorables aux erreurs. Il s'agit des propriétés du contexte lié à l'historique de l'installation, à l'organisation du système, aux caractéristiques de l'interface, et qui influencent la nature et le contenu de la réalisation de la tâche confiée à l'opérateur.

Elles sont utilisées pour identifier les erreurs et proposer des mesures permettant de les réduire. Ces méthodes considèrent que l'erreur est essentiellement due au contexte de l'activité. Nous avons identifié 5 méthodes contextuelles parmi celles du tableau 2.

HEART

Libellé	Human Error Assessment And Reduction Technique
Contexte de développement	<p>Méthodologie basée sur les hypothèses suivantes :</p> <ul style="list-style-type: none"> ▪ La fiabilité humaine est, en principe, dépendante de la nature générique de la tâche à effectuer ; ▪ Dans les conditions idéales, le niveau de fiabilité humaine est atteint de manière compatible avec une vraisemblance nominale donnée dans certaines limites probabilistes ; ▪ Dans les conditions réelles, la fiabilité humaine estimée peut être dégradée dans la mesure dont les EPCs identifiées peuvent s'appliquer
Paramètres, PSF et modèles causals	EPC: Error Producing Conditions. Les EPC sont déterminées sur la base des PSF et des conditions du système
Dépistage	La méthode ne fournit pas explicitement une procédure de dépistage. Toutefois, elle est relativement simple à utiliser et ne requiert pas beaucoup des ressources
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Classification de la tâche à analyser selon 9 Generic Task Types (GTT) qui sont décrites dans HEART (Williams, 1986) • Affectation de la HEP nominale à la tâche considérée

	<ul style="list-style-type: none"> • Identification des EPC impactant la fiabilité de la tâche et la proportion évaluée de l'impact (APOA, Assessed Portion of Affect) pour chaque EPC • Calcul de la HEP
Application	Nucléaire civil et autres industries (chimique, médical, aéronautique etc.) du Royaume-Uni
Disponibilité	En libre accès dans les papiers publiés. Une copie du manuel est disponible sur demande (British Energy)
Validation	Validation empirique effectuée par (Kirwan, 1997a)
Avantages	<ul style="list-style-type: none"> ▪ Ressources nécessaires limitées ▪ Versatile et facile à utiliser par ingénieurs ou ergonomistes
Limites	<ul style="list-style-type: none"> ▪ Modélisation de la dépendance de l'erreur pas incluse ▪ Possibilité de calculer des HEPS très différentes à partir d'une description de la tâche qui n'est pas suffisamment claire ▪ Erreurs potentielles dans l'affectation des EPC aux tâches, l'estimation de l'APOA et la décomposition des tâches

Tableau A.8. Synthèse de la méthode HEART.

ATHEANA

Libellé	A Technique for Human Event Analysis
Contexte de développement	<p>Méthodologie basée sur l'hypothèse suivante : l'erreur humaine est le résultat de « Error Forcing Context » (EFCs), c'est-à-dire des combinaisons de conditions liées à l'installation et autres facteurs d'influence (PSFs) qui peuvent créer le contexte favorable à l'erreur. Ces facteurs sont pris parmi les suivants:</p> <ul style="list-style-type: none"> ▪ Conditions de l'installation ▪ Actions de sûreté ▪ Scénario ▪ Mécanismes d'erreurs ▪ Interaction homme-machine <p>Première méthode traitant les erreurs de commission (EOC)</p>
Paramètres, PSF et modèles causals	<p>EFC : Error Forcing Context. Ils se distinguent en deux types: EFC qui caractérise les initiateurs ou séquences accidentelles; les EFC qui caractérisent le système.</p> <p>Les EFC sont identifiés par les experts par la combinaison de PSF et les conditions du système dans le scénario analysé</p>
Dépistage	Non
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Intégration des enjeux d'intérêt pour l'analyse • Identification des événements d'échecs humains et/ou actions peu sûres • Pour chaque événement ou action, identification des causes qui le/la produisent. Ceux-ci sont les éléments d'un EFC – conditions de l'installation et PSFs • Quantification des EFC et de la probabilité de chaque action analysée

	<ul style="list-style-type: none"> • Evaluation des résultats par rapport aux enjeux
Application	Recherche
Disponibilité	En libre accès sur http://www.nrc.gov/
Validation	Non
Avantages	<ul style="list-style-type: none"> ▪ Traitement du contexte par les EOC ▪ Etude systématique des causes des échecs humains ▪ Résultats qualitatifs très détaillés sur les conditions menant à l'erreur. Outil pour l'aide à la conception ▪ Prise en compte des processus cognitifs
Limites	<ul style="list-style-type: none"> ▪ Trop dépendante de l'expertise ▪ Manuel assez complexe ▪ Quantification faible et dépendante des jugements d'experts

Tableau A.9. Synthèse de la méthode ATHEANA.

CREAM – Basic Version

Libellé	Cognitive Reliability and Error Analysis Method
Contexte de développement	<p>Méthodologie « bidirectionnelle » dans le sens qui peut être utilisée à la fois pour faire des analyses rétrospectives ou des prévisions de performance humaine. Elle est basée sur :</p> <ul style="list-style-type: none"> ▪ le modèle de Hollnagel COCOM (Contextual Control) qui fait la distinction entre compétence et contrôle. ▪ Le concept de CPC (Common Performance Conditions)
Paramètres, PSF et modèles causals	<p>CPC : Common Performance Conditions COCOM : Contextual Control Modes. Les modes de contrôle sont de 4 types (stratégique, tactique, opportuniste, interférente). A chaque mode de contrôle est associé un intervalle de fiabilité de la performance</p>
Dépistage	<p>Elle est conçue pour faire des analyses de dépistage. Sur la base de règles simples (combinaison d'états des PSF), CREAM est capable de déterminer des intervalles de valeurs pour les HEP</p>
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Analyse des tâches • Affectation d'un niveau de CPC à chaque tâche • Les CPC sont associées à trois descripteurs et à chaque descripteur correspond un effet attendu sur la fiabilité de la tâche à performer • Calcul du score total correspondant aux CPC identifiées • Identification du mode de contrôle en relation avec le score obtenu • L'intervalle de fiabilité associé au mode de contrôle identifié est l'estimation de la HEP
Application	Nucléaire civil. Un exemple pour un accident de train (Marseguerra et al, 2007)
Disponibilité	Principes d'usage (Hollnagel, 1998). CREAM Navigator, en ligne sur le site de 'University of Illinois'
Validation	En cours
Avantages	<ul style="list-style-type: none"> ▪ Evaluation de l'importance d'une tâche dans un contexte spécifique ▪ Intégration d'un modèle cognitif des décisions et actions humaines
Limites	/

Tableau A.10. Synthèse de la méthode CREAM.

MERMOS

Libellé	Méthode d'Evaluation de la Réalisation des Missions Opérateur pour la sûreté
Contexte de développement	<p>Méthodologie systémique basée sur le concept de « mission FH (Facteur Humain) ». Les concepts sous-jacents incluent :</p> <ul style="list-style-type: none"> ▪ CICAs : Caractéristiques Importantes de la Conduite Accidentelle. Elles modélisent les modes d'organisations cohérentes du système au niveau collectif. ▪ EOS : Emergency Operations System (équipes d'exploitation, procédures, interface homme-machine, facteurs organisationnels, milieu de travail). ▪ SAD : (Stratégie, Action, Diagnostic), les trois fonctions impliquées dans l'exploitation de la mission FH par les EOS
Paramètres, PSF et modèles causals	/
Dépistage	/
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Identification et définition des missions FH par l'analyse fonctionnelle du système après l'événement initiateur • Décomposition des missions FH selon SAD et identification des CICAs • Création des scénarios d'échec • Analyse qualitative et quantitative des missions FH
Application	Nucléaire civil. Méthode utilisée couramment par EDF pour les EPS de niveau 1 des unités N4 (les réacteurs français plus récents)
Disponibilité	Propriété d'EDF
Validation	Validée (Le Bot, <i>et al.</i> , 2002)
Avantages	<ul style="list-style-type: none"> ▪ Méthode de deuxième génération apportant des nouveaux concepts à l'analyse de la fiabilité humaine ▪ EDF a obtenu des résultats satisfaisants en appliquant MERMOS
Limites	<ul style="list-style-type: none"> ▪ Pas en libre accès ▪ Pas pour la conduite normale (à développer)

Tableau A.11. Synthèse de la méthode MERMOS.

NARA

Libellé	Nuclear Action Reliability Assessment
Contexte de développement	Méthodologie basée sur HEART et demandée par British Energy afin d'effectuer les EPS et l'analyse de la fiabilité humaine des installations du Royaume-Uni. Elle consiste en une version améliorée de HEART et est adaptée spécifiquement au domaine nucléaire.

Paramètres, PSF et modèles causals	GTT: Generic Task Types (Kirwan, <i>et al.</i> , 2005) EPC: Error Producing Conditions (équivalents aux PSF)
Dépistage	La méthode ne fournit pas explicitement une procédure de dépistage. Toutefois, elle est relativement simple à utiliser et ne requiert pas beaucoup de ressources.
Décomposition des tâches et procédure de calcul des HEP	(Cf. HEART)
Application	Spécifique pour l'industrie nucléaire. Autres applications : aérospatiales (NASA)
Disponibilité	Non. Propriété de British Energy
Validation	Basée sur HEART (validation empirique)
Avantages	<ul style="list-style-type: none"> ▪ A priori, elle est spécifique pour le nucléaire ▪ Elle se base sur une méthode validée (HEART)
Limites	<ul style="list-style-type: none"> ▪ Manque d'un guide d'utilisation de la méthode

Tableau 1. Synthèse de la méthode NARA.

Méthodes basées sur les JUGEMENTS D'EXPERTS

Les méthodes de ce type mettent l'accent sur les probabilités d'erreur, déterminées à partir des estimations par jugements d'experts. La fiabilité des estimations dépend fortement de la complexité des situations analysées, de la sélection des experts mêmes et de la manière dont les jugements sont agrégés. Nous avons identifié trois méthodes parmi celles qui sont présentées dans le Tableau 3.

SLIM – MAUD

Libellé	Success Likelihood Index Method using Multi-Attribute Utility Decomposition
Contexte de développement	Méthode basée sur les jugements d'experts et composée par : <ul style="list-style-type: none"> ▪ SLIM : ensemble de procédures qui produisent des jugements d'experts afin d'estimer les HEPs ▪ MAUD : logiciel à l'aide de SLIM intervenant sur l'identification des PSFs et empêchant les biais
Paramètres, PSF et modèles causals	PSF: Performing Shaping Factors SLI: Success Likelihood index
Dépistage	Non
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Sélection d'un groupe d'experts • Définition des situations et des tâches • Identification des PSFs • Classification des erreurs sur la base des PSFs • Vérification de l'indépendance • Procédures de pondération

	<ul style="list-style-type: none"> • Calculs des SLIs • Conversion des SLIs en probabilités d'erreur par une relation logarithmique de la forme : $\text{Log } p(\text{succès}) = a(\text{SLI}) + b$ • Estimation de la HEP = $1 - p(\text{succès})$ • Analyse d'incertitude • Analyse de sensibilité afin de réduire les erreurs
Application	Analyse accidentelle, conduite normale et de maintenance
Disponibilité	<ul style="list-style-type: none"> ▪ SLIM : en libre accès ▪ MAUD : propriété de London School of Economics, possibilité d'acheter le logiciel entier (SLIM-MAUD)
Validation	NRC (Kirwan, 1997b)
Avantages	<ul style="list-style-type: none"> • Flexibilité générale de l'approche • Efficacité des évaluations en termes coûts-avantages
Limites	<ul style="list-style-type: none"> • Manque de données de calibration des SLIs • Nature arbitraire du choix des PSFs • Un groupe d'experts est nécessaire • Tendances aux biais

Tableau A.13. Synthèse de la méthode SLIM-MAUD.

TRIPOD – Delta

Libellé	Tripod - Delta
Contexte de développement	Méthode basée sur un modèle de sécurité organisationnelle conçu pour rechercher a posteriori les causes organisationnelles d'un accident (TRIPOD-Beta) et servant à identifier les défaillances organisationnelles d'un système industriel afin d'anticiper la survenue d'accidents (TRIPOD-Delta) (Hudson, Reason, Bentley, et Primrose, 1994).
Liste des PSF et modèles causals	BRF (Basic Risk Factors) (Cambon et Guarnieri, 2008)
Dépistage	Non
Décomposition des tâches et procédure de calcul des HEP	<ul style="list-style-type: none"> • Préparation de l'étude d'enquête en sélectionnant 275 questions à partir d'une base centrale (Base Delta – 1500 questions) • Collecte des informations • Validation statistique des réponses et analyse des résultats • Interview et propositions d'actions correctives
Application	Sécurité et santé du travail
Disponibilité	Oui
Validation	Oui
Avantages	/
Limites	/

Tableau A.13. Synthèse de la TRIPOD-Delta.



UNIVERSITÉ
DE LORRAINE



Ecole Doctorale IAEM Lorraine - Département de Formation Doctorale en
Automatique

Thèse

Présentée pour l'obtention du grade de

Docteur de l'Université de Lorraine

Spécialité Automatique, Traitement du Signal et des Images, Génie Informatique

par

Antonello De Galizia

Évaluation probabiliste de l'efficacité des barrières humaines prises dans leur contexte organisationnel

Soutenue publiquement le 28 Février 2017 à l'Université de Lorraine devant le jury composé de :

Rapporteurs	Frédéric VANDERHAEGEN	Professeur à l'UVHC de Valenciennes
	Eric ZAMAI	Enseignant-Chercheur à l'INP de Grenoble
Examineurs	Benoit IUNG	Professeur à l'Université de Lorraine, directeur de thèse
	Philippe WEBER	Professeur à l'Université de Lorraine, co-directeur de thèse
	Christophe SIMON	Maître de Conférences à l'Université de Lorraine, co-encadrant de la thèse
	Carole DUVAL	Ingénieur de recherche expert chez EDF R&D, co-encadrant de la thèse
	Jean-Yves DANTAN	Professeur à l'Ecole Nationale Supérieure d'Arts et Métiers Paristech
	Enrico ZIO	Professeur à l'Ecole Polytechnique de Milan et à École Centrale-Supélec, Paris
Invitée	Dounia TAZI	Responsable du pôle expertise et accompagnement de l'ICSI
	Emmanuel SERDET	Ingénieur de recherche chez EDF R&D, co-encadrant de la thèse