



**HAL**  
open science

# Effet des biais cognitifs et de l'environnement sur la qualité des données et des informations

Robin Le Conte Des Floris

► **To cite this version:**

Robin Le Conte Des Floris. Effet des biais cognitifs et de l'environnement sur la qualité des données et des informations. Robotique [cs.RO]. Université Paris sciences et lettres, 2024. Français. NNT : 2024UPSLM004 . tel-04589058

**HAL Id: tel-04589058**

**<https://pastel.hal.science/tel-04589058>**

Submitted on 27 May 2024

**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.



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à Mines Paris

**Effet des biais cognitifs et de l'environnement sur la qualité  
des données et des informations**

Soutenue par

**Robin Le Conte des  
Floris**

Le 2 février 2024

École doctorale n°621

**Ingénierie des systèmes,  
matériaux, mécanique, én-  
ergétique**

Spécialité

**Informatique temps-réel,  
robotique et automatique**

Composition du jury :

Isis Truck Professeur, Université Paris 8	<i>Rapporteur</i>
Alejandro Romero-Torres Professeur, UQAM	<i>Rapporteur</i>
Cécile Chamaret Professeur, École polytechnique	<i>Présidente du jury</i>
Zineb Noumir Docteur, ALTEN	<i>Examineur</i>
Pierre Jouvelot Directeur de recherche, Mines Paris	<i>Directeur de thèse</i>
Cédric Dalmaso Maître de conférences, Mines Paris	<i>Examineur</i>



*Il n'y a pas de faits mais seulement des interprétations.*

Friedrich Nietzsche



# Remerciements

Mes remerciements s'adressent tout d'abord à Pierre Jouvelot, mon directeur de thèse, et à Cédric Dalmasso, mon co-encadrant, pour leur franchise, leur ouverture d'esprit et leurs encouragements. Nos conversations franches et mouvementées ont façonné le sujet de cette recherche et son évolution. Je suis particulièrement reconnaissant envers Pierre Jouvelot pour son soutien, parfois quotidien, dans les moments les plus difficiles de la thèse, en particulier pendant les crises sanitaires liées à la COVID-19.

Ce travail a également été soutenu par ALTEN, et plus particulièrement par la DIN. Leur soutien financier, mais aussi leur accompagnement, ont été essentiels à la réalisation de cette thèse. Je tiens à remercier tout particulièrement Zineb Noumir pour son suivi continu et sa disponibilité à répondre aux problématiques rencontrées.

Je suis également profondément reconnaissant envers ma compagne, Mathilde Bouget, pour son soutien continu, ses relectures attentives, sa patience et sa positivité tout au long de ma thèse. Son aide a été inestimable, en particulier durant la phase de rédaction, où elle a su m'offrir un cadre structuré et rythmer mon travail en me guidant dans l'organisation de mes journées et de mon avancement.

Je souhaite exprimer ma profonde gratitude envers ma famille : mes parents, ma sœur et mon frère, dont le soutien intellectuel et moral a été fondamental tout au long de ce parcours. Leur présence et leur encouragement m'ont été indispensables pour surmonter les défis rencontrés

Une mention spéciale est adressée à Lucas Sguerra, Amdjed Belaref, Vassili Maillet et aux membres du CRI pour les moments partagés à travailler ensemble, nos longues discussions, et nos moments de détente. Ces instants ont été précieux pour moi et ont constitué une source d'inspiration pour ma thèse.

À tous ceux mentionnés et à ceux que je n'ai pas pu nommer, merci du fond du cœur pour votre soutien.



## Résumé

Du point de vue du philosophe Friedrich Nietzsche, il n'y a pas de réalité qui existe en soi, pas de fait brut, pas de réalité absolue : tout ce que nous définissons comme étant la réalité n'est, en fait, que le résultat de processus d'interprétation qui nous sont propres. Or, les données stockées sur les systèmes d'information ne sont souvent rien d'autre que la représentation codée de déclarations émises par des êtres humains, relevant donc intrinsèquement de l'interprétation humaine et étant par conséquent affectées par les mêmes biais et limites qui caractérisent la psyché humaine.

Cette thèse propose un nouveau cadre conceptuel, le modèle "Data Binding and Reification" (DBR), qui décrit le processus d'interprétation des données, puis de réification de l'information, en utilisant une nouvelle approche qui place les mécanismes de perception humaine au cœur de celui-ci. En mobilisant les sciences cognitives et comportementales, cette approche permet d'identifier dans quelle mesure l'intervention de l'homme et la structure de l'environnement auquel il est soumis conditionnent l'apparition de biais cognitifs affectant ces processus. Les résultats expérimentaux valident partiellement ce modèle en identifiant les caractéristiques de l'environnement affectant, dans un contexte organisationnel, le processus de collecte de données et la qualité des informations produites.

Ces travaux ouvrent de nombreuses perspectives, telles que l'élaboration d'une architecture de choix au sens de l'économiste Richard Thaler, pouvant améliorer le processus même de collecte de données en modifiant l'expérience des utilisateurs du système d'information.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	Contexte organisationnel . . . . .	14
1.1.1	Groupe ALTEN . . . . .	14
1.1.2	Qualité des données chez ALTEN . . . . .	15
1.2	Objectif . . . . .	16
1.3	Contributions . . . . .	17
1.3.1	Résultats . . . . .	17
1.3.2	Publications . . . . .	18
1.4	Organisation . . . . .	19
<b>2</b>	<b>Contexte théorique</b>	<b>20</b>
2.1	Concepts de donnée et d'information . . . . .	21
2.1.1	Analyse comparative de la terminologie . . . . .	21
	Approche métaphysique vs. non métaphysique . . . . .	21
	Approche exclusive à l'homme vs. non exclusive . . . . .	22
	Approche fondée sur les sciences cognitives vs. non exclusive . . . . .	22
2.1.2	Théorie de l'information de Shannon . . . . .	23
2.1.3	Théorie linguistique de Chomsky . . . . .	24
2.1.4	Qualité des données du point de vue du consommateur . . . . .	25
2.2	Perception et subjectivité en neurosciences . . . . .	26
2.2.1	<i>Binding Problem</i> . . . . .	26
2.2.2	Subjectivité de la perception . . . . .	27
2.2.3	Association "stimulus-réponse" (S-R) . . . . .	28
2.3	Biais cognitif . . . . .	29
2.3.1	Origines . . . . .	30
2.3.2	Rationalité écologique . . . . .	31
2.3.3	Psychologie évolutionnaire et management de l'erreur . . . . .	31
2.3.4	Biais dans d'autres domaines . . . . .	32
	Statistique, informatique, intelligence artificielle . . . . .	32
	Recherche . . . . .	33
2.4	Acceptation des utilisateurs . . . . .	33
2.4.1	Historique des modèles d'acceptation . . . . .	34
2.4.2	Unified Theory of Acceptance and Use of Technology (UTAUT) . . . . .	35
2.5	Motivation : théorie de l'autodétermination . . . . .	36
2.5.1	Principes . . . . .	37
2.5.2	Deux types de motivations . . . . .	37

	Motivation intrinsèque . . . . .	38
	Motivation extrinsèque . . . . .	38
2.5.3	Régulation et auto-détermination . . . . .	38
2.6	Gamification . . . . .	40
2.6.1	Système d'information motivationnel . . . . .	40
2.6.2	Usages de la gamification . . . . .	41
2.6.3	Incitation au partage d'informations et de connaissance . . . . .	42
<b>3</b>	<b>Modèle <i>Data Binding and Reification</i></b> . . . . .	<b>44</b>
3.1	Donnée et information . . . . .	45
3.1.1	Étymologie . . . . .	46
	Information . . . . .	46
	Donnée . . . . .	46
3.1.2	Choix des définitions . . . . .	47
3.2	Transformation des données et de l'information . . . . .	48
3.2.1	Définition des concepts de <i>binding</i> et de réification . . . . .	48
	<i>Binding</i> . . . . .	48
	Réification . . . . .	49
3.2.2	Facteurs de perception . . . . .	50
3.2.3	Modèle DBR . . . . .	51
3.2.4	Biais cognitifs humains . . . . .	53
3.2.5	Biais dans le modèle DBR . . . . .	55
	Biais généraux . . . . .	56
	Biais cognitifs . . . . .	57
3.3	Qualité des données et de l'information . . . . .	57
3.3.1	Dimensions additionnelles . . . . .	58
	Qualité intrinsèque de l'information . . . . .	60
	Qualité contextuelle de l'information . . . . .	60
	Qualité des données . . . . .	60
	Qualité du système . . . . .	61
3.3.2	Nouvelles caractéristiques . . . . .	61
	Entropie . . . . .	61
	Adaptabilité . . . . .	61
	Délai de réponse . . . . .	62
<b>4</b>	<b>Intégration de l'environnement au modèle DBR : une approche par la théorie ancrée</b> . . . . .	<b>64</b>
4.1	Contexte organisationnel . . . . .	65
4.2	Méthode de recherche . . . . .	66
4.2.1	Sélection de l'échantillon . . . . .	66
4.2.2	Ethique . . . . .	66
4.2.3	Procédure d'entretien . . . . .	67
4.2.4	Collecte des données . . . . .	68
4.3	Résultats . . . . .	69
4.3.1	Codage ouvert . . . . .	69
4.3.2	Codage axial . . . . .	69
	Données et informations . . . . .	70

Collecte de données . . . . .	71
Projet . . . . .	72
Humain . . . . .	73
Parties prenantes . . . . .	74
Culture d'entreprise . . . . .	75
4.3.3 Codage sélectif . . . . .	75
4.4 Alignement avec le modèle DBR . . . . .	76
4.4.1 Réification . . . . .	77
4.4.2 Binding . . . . .	77
4.5 Discussion . . . . .	79
<b>5 Effet de la motivation sur la qualité de l'information : application du modèle DBR</b>	<b>82</b>
5.1 Contexte . . . . .	83
5.2 Outils méthodologiques . . . . .	85
5.2.1 <i>A/B Testing</i> . . . . .	85
5.2.2 Régulation de la motivation en entreprise . . . . .	86
5.3 Protocole expérimental . . . . .	87
5.3.1 Méthodologie . . . . .	87
5.3.2 Définition des critères d'évaluation globaux (OEC) . . . . .	90
5.4 Résultats . . . . .	92
5.4.1 Quantité de données produites . . . . .	92
Temps de réponse . . . . .	93
Taux de participation . . . . .	93
Quantité de données labélisées par personne . . . . .	93
5.4.2 Qualité intrinsèque de l'information . . . . .	93
Succès des projets . . . . .	94
Relation client . . . . .	95
Maturité . . . . .	96
5.5 Discussion . . . . .	96
5.6 Conclusion . . . . .	99
<b>6 Conclusion</b>	<b>101</b>
6.1 Contributions . . . . .	101
6.1.1 Modèle DBR . . . . .	101
6.1.2 Cadre conceptuel de la qualité des données et des informations	102
6.1.3 Effet de la rationalité écologique sur le modèle DBR . . . . .	102
6.1.4 Protocole expérimental pour le modèle DBR . . . . .	103
6.1.5 Effet de la motivation sur la qualité des informations . . . . .	104
6.1.6 Mise en perspective . . . . .	104
6.2 Travaux futurs et limitations . . . . .	105
<b>Références</b>	<b>108</b>
<b>A Données Nvivo</b>	<b>113</b>
<b>B Mails utilisés dans le cadre du protocole d'A/B testing</b>	<b>115</b>

# Table des figures

2.1	Structuration de la thèse par chapitres, objectifs et concepts théoriques pertinents . . . . .	21
2.2	Modèle de recherche (Venkatesh, Morris, Davis, & Davis, 2003) . . . . .	36
2.3	Continuum de la motivation (Denis & Jouvelot, 2005) . . . . .	39
3.1	Processus de binding et réification . . . . .	52
3.2	Echiquier d'Adelson (1995) . . . . .	54
3.3	Modèle conceptuel de la qualité des données de R. Y. Wang et Strong (1996) . . . . .	58
3.4	Modèle unifié de la qualité des données et de l'information . . . . .	59
4.1	Indicateurs de suivi du succès des projets ALTEN . . . . .	65
4.2	Comparaison des KPI de ALTEN et du CHAOS report . . . . .	66
4.3	Informations démographiques . . . . .	66
5.1	Interface SMART de gestion des projets . . . . .	88
5.2	Interface SMART pour un projet particulier . . . . .	89
5.3	Calendrier d'envoi des courriers électroniques . . . . .	90
5.4	Distributions empiriques du succès des projets . . . . .	94
5.5	Distributions empiriques des indicateurs de qualité de la relation avec le client . . . . .	95
5.6	Distributions empiriques des indicateurs de maturité des projets . . . . .	96
5.7	Comparaison des indicateurs de suivi du succès des projets . . . . .	98
B.1	Mail groupe autonome . . . . .	115
B.2	Mail groupe autonome - relance . . . . .	116
B.3	Mail groupe contrôle . . . . .	117
B.4	Mail groupe contrôle - relance . . . . .	118

# Liste des tableaux

2.1	Contextes expérimentaux de la gamification ( <a href="#">Hamari, Koivisto, &amp; Sarsa, 2014</a> ) . . . . .	42
2.2	Mécanismes de gamification ( <a href="#">Friedrich, Becker, Kramer, Wirth, &amp; Schneider, 2020</a> ) . . . . .	43
3.1	Données agrégées du modèle de <a href="#">R. Y. Wang et Strong (1996)</a> . . . . .	62
4.1	Fréquences des thèmes . . . . .	70
4.2	Proportion de verbatims balisés par thème . . . . .	76
4.3	Thèmes et modèle DBR . . . . .	80
A.1	Hierarchie intermédiaire du codage axial . . . . .	113
A.2	Croisement thèmes - valeurs . . . . .	114
A.3	Croisement thèmes - ratio . . . . .	114



# Chapitre 1

## Introduction

Nous vivons à une époque dans laquelle les données numériques ont pris, au fil des années, une importance considérable dans le quotidien de tout un chacun, impactant nos décisions et nos choix ; jamais la frontière entre le monde réel et le monde numérique n'a été aussi ténue.

Pour comprendre ces interactions, le psychologue Stanley Milgram et son équipe de Yale ont, au début des années 90, conceptualisé la notion de "continuum réalité-virtualité", vaste étendue caractérisant les modalités d'existence et dont les deux extrêmes sont un environnement entièrement "présent" et physique, le monde réel, et un environnement entièrement virtuel, le monde numérique. Tout ce qui se trouve entre les deux est considéré comme formant une réalité mixte.

Cette superposition de réalités est fascinante, mais elle soulève également des questions importantes quant à la fiabilité et à la nature des décisions reposant sur ces deux sphères. D'un côté, nous sommes tous familiers avec le monde réel ; c'est l'environnement tangible dans lequel nous vivons, où nous percevons les choses à travers nos sens. D'un autre côté, de nos jours, la réalité virtuelle, apparue avec l'émergence des technologies numériques, prend de plus en plus de place dans notre existence, jusqu'à se superposer à la réalité physique et à influencer nos choix de manière significative.

Du point de vue du philosophe Friedrich Nietzsche, il n'y a pas de réalité qui existe en soi, pas de fait brut, pas de réalité absolue : tout ce que nous définissons comme étant la réalité n'est, en fait, que le résultat de processus d'interprétation qui nous sont propres et, donc, uniques. Or les données stockés sur les systèmes informatiques qui forment l'infrastructure du monde virtuel ne sont souvent rien d'autre que la représentation codée de déclarations émises par des êtres humains, relevant donc intrinsèquement de l'interprétation humaine. Par conséquent, elle sont affectées par les mêmes biais et limites qui caractérisent la psyché humaine, rendant cette superposition de deux mondes, réel et virtuel, aussi imparfaite et subjective que peut l'être l'homme. Malgré cela, nous avons souvent tendance à considérer les données comme des faits incontestables, et à s'y fier aveuglément pour prendre des décisions, alors que celles-ci peuvent, dans certains cas, être même moins fiables que nos propres observations et interprétations.

C'est dire combien il est important de comprendre ce qui différencie fondamentalement donnée et information. L'information est l'interprétation des données, et

c'est à travers cette interprétation que les données acquièrent une signification et une valeur réelles pour celui qui les utilise, mais c'est également là qu'elles perdent leur statut de fait objectif. Appréhender clairement la distinction entre données et informations, point que nous traitons en profondeur dans cette thèse, peut permettre de les utiliser plus efficacement afin de prendre des décisions éclairées en évitant de se laisser aveugler par une confiance excessive dans les données brutes.

Face à ces constats, il devient clair que la qualité des données et des informations produites doit être une préoccupation clé lors de la conception et de l'utilisation d'un système d'informations. Cela passe nécessairement par une prise de conscience des biais humains et des limites de nos propres interprétations et une meilleure compréhension des processus qui sous-tendent la création et l'interprétation des données, éléments cruciaux quand il s'agit de viser une utilisation aussi judicieuse que possible de celles-ci, en tenant compte de leurs imperfections intrinsèques.

C'est dans ce contexte que la compréhension du concept de "biais cognitif", qui repose sur le principe de rationalité limitée des êtres humains, prend toute sa valeur. En explorant les "heuristiques et biais", les psychologues Daniel Kahneman, également économiste et prix Nobel d'économie pour ses travaux sur le sujet, et Amos Tversky ont, en 1979, mis en évidence la manière dont les individus prennent des décisions en tenant compte de leurs ressources disponibles limitées comme le temps, l'information et les capacités cognitives. Cette approche, qui remet en question le principe classique de la rationalité totale des choix des acteurs, a apporté une contribution précieuse à la compréhension du comportement humain, en particulier en soulignant que la prise de décisions est souvent affectée par des biais cognitifs et des limites de traitement de l'information. Ce sont ces notions qui vont être mobilisées ici pour poser les bases du modèle DBR (Data Binding and Reification), l'une des contributions majeures de cette thèse.

## 1.1 Contexte organisationnel

Le modèle DBR introduit ici a émergé dans le cadre d'une recherche fortement ancrée dans une pratique terrain. Il est donc pertinent de préciser dès à présent dans quel contexte organisationnel celle-ci a été produite, et quelles sont les motivations industrielles qui ont poussé la société ALTEN, dont il s'agit, à s'intéresser à cette thématique de recherche.

### 1.1.1 Groupe ALTEN

ALTEN est une société multinationale de services en ingénierie et en technologie basée en France et fondée en 1988. Elle opère principalement dans les secteurs de l'information technologique, de l'ingénierie et de la R&D (recherche et développement). L'entreprise apporte une assistance technique à ses clients pour les aider à développer de nouvelles technologies et à améliorer leurs processus de production et leurs produits. La société propose également des services de conseil pour des projets dans divers secteurs, y compris l'automobile, l'aéronautique, les télécommunications, l'énergie, la santé, la finance et les services publics. Elle compte aujourd'hui plus de 54 100 collaborateurs dans le monde dans près de 30 pays. La société est réputée

pour sa capacité à offrir des solutions innovantes à ses clients grâce à une expertise technique de haut niveau.

ALTEN est une entreprise qui repose, naturellement, en partie sur l'utilisation de données pour guider ses décisions et ses actions. Grâce à son système d'information (SI) sophistiqué, elle suit minutieusement l'évolution des divers projets en cours, générant ainsi une grande quantité de données relatives au suivi de chaque projet. Ces données sont ensuite exploitées pour produire des centaines d'indicateurs clés de performance (*Key Performance Indicator*, ou KPI), qui aident à évaluer et à optimiser la performance des projets. Ces indicateurs portent sur une variété d'aspects, y compris les coûts, les délais, la qualité, la satisfaction client et bien d'autres encore, permettant ainsi une gestion de projet efficace et fondée sur des données concrètes.

Le SI d'ALTEN ne se limite d'ailleurs pas au seul suivi des projets. Il joue également un rôle crucial dans l'évaluation du bien-être des employés. Grâce à la collecte de données sur divers aspects de l'environnement de travail, l'entreprise est en mesure de surveiller le moral des employés, leur satisfaction au travail, leur charge de travail et bien d'autres facteurs qui peuvent avoir un impact sur leur bien-être et leur productivité.

Dans la mesure où les données occupent, chez ALTEN, une place centrale dans les décisions stratégiques, la qualité de ces données est, de fait, une préoccupation majeure. Les données de mauvaise qualité peuvent avoir de multiples impacts négatifs sur l'entreprise, ce qui peut conduire à des erreurs de décision, des risques financiers et une réputation potentiellement endommagée. En effet, si les données utilisées pour guider la prise de décision sont incorrectes ou incomplètes, cela peut conduire à des erreurs de jugement. Par exemple, investir dans un projet considéré, à tort, comme sûr en raison de données inexactes pourrait entraîner des pertes financières significatives pour l'entreprise. De même, si les données concernant la performance d'un projet sont erronées, cela pourrait conduire à des évaluations incorrectes de ce projet, avec des conséquences allant de l'allocation inefficace de ressources à la non-détection de problèmes majeurs.

### 1.1.2 Qualité des données chez ALTEN

Un des objectifs affichés par la société d'ALTEN pour justifier la collecte et le stockage de cette vaste quantité de données projet est l'utilisation potentiellement avantageuse de celles-ci dans l'amélioration de la capacité de l'entreprise à prendre des décisions et à réduire les risques d'erreurs.

Cependant, la poursuite de ces objectifs et les obstacles rencontrés lors de celle-ci ont rapidement soulevé de nouvelles interrogations, préoccupantes. Et si les données collectées par une organisation comme ALTEN n'avaient pas l'objectivité qu'on aurait pu souhaiter pour qu'elles deviennent des alliées impartiales d'une prise de décision efficace et éclairée? En effet, pour que les organisations disposent de systèmes d'information efficaces, il est nécessaire de s'assurer du haut degré de qualité de ces données (R. Wang, Storey, & Firth, 1995). Il est largement accepté que la qualité de l'information déduite directement ou indirectement de ces sources de données affecte significativement la performance et l'efficacité des outils d'aide à la décision utilisés par les organisations pour la conduite éclairée de leurs activités (Raghunathan &

Pham, 1999). Cependant, à moins que la collecte d'informations pertinentes ne soit entièrement automatisée et liée à des processus physiques parfaitement maîtrisés, celle-ci reste étroitement liée aux agents chargés de la mise à jour de ces systèmes d'information.

Or, la plupart des critères d'évaluation et de mesure de la qualité des données ne remettent pas en question la valeur intrinsèque de l'information. En fait, ils ne remettent tout simplement pas en question la qualité de l'information indépendamment des données. Cela est probablement dû au manque de cohérence dans la terminologie et la définition de ces concepts, ce qui obscurcit le rôle relatif joué par chacun d'eux dans le succès des pratiques de *Business Intelligence & Analytics* (BI & A) (Torres & Sidorova, 2019) auprès des entreprises et de leurs managers. Au-delà même de la définition de ces deux concepts, sur laquelle nous allons revenir (voir chapitre 2), il convient de se demander quelle relation existe entre les données et l'information, notamment si, comme le présume l'approche fondée sur les sciences cognitives, ce qui différencie les deux réside dans l'interprétation humaine. De plus, le concept de qualité est abstrait, en particulier lorsqu'il s'agit d'information : qu'est-ce qui fait qu'une information est de "bonne" qualité et, respectivement, qu'est-ce qui fait qu'une donnée est de "bonne" qualité ?

Répondre à ces interrogations devrait pouvoir non seulement permettre de mieux comprendre ce qui fait le succès d'un système d'information, mais également d'identifier des axes permettant d'améliorer la qualité de ces données et informations, et ainsi les prises de décisions qui en découlent. C'est là tout le coeur du travail de recherche présenté ici.

## 1.2 Objectif

L'objectif de cette thèse de *comprendre et conceptualiser les mécanismes qui sous-tendent la création d'informations et de données en utilisant une nouvelle approche, le modèle DBR, qui place l'homme au cœur de son processus*. L'enjeu principal est de parvenir à mobiliser les sciences cognitives et comportementales pour identifier dans quelle mesure l'intervention de l'homme et la structure de l'environnement auquel il est soumis affectent le processus de création de données et d'information, notamment dans un contexte organisationnel.

L'avancée principale de ces travaux est de proposer des outils conceptuels permettant d'améliorer la qualité des données et des informations en limitant les biais cognitifs auquel l'homme est soumis dans le cadre de leur création. Pour cela, nous proposons un protocole expérimental qui permet d'agir sur les caractéristiques de l'environnement, aussi appelé architecture de choix, qui module l'utilisation de biais cognitifs par l'homme, et ainsi affecte la création de données et d'informations.

Pour y parvenir, ces travaux de recherche passeront par plusieurs étapes.

1. La première concerne l'étude fondamentale des concepts de donnée et d'information à travers une analyse approfondie de leur terminologie et de leur étymologie, ainsi que la compréhension de la relation complexe qui existe entre les données et l'information, et de la nature des processus qui régissent cette relation ; Ces processus sont définis comme étant, respectivement, le *binding*

(terme indiquant une liaison, une affectation, et que nous préférons ne pas traduire) et la *reification* (réification) et constituent le socle théorique du modèle *Data Binding and Reification* (DBR) proposé dans ces travaux de recherche.

2. La deuxième étape concerne l'identification et la compréhension des critères permettant de mesurer, respectivement, la qualité des données et celle de l'information. En effet, pour pouvoir correctement évaluer l'effet des biais cognitifs sur la qualité, encore faut-il pouvoir la mesurer. La difficulté réside notamment dans la nécessité de caractériser distinctement la qualité des données et celle des informations, tout en tenant compte de l'enchevêtrement de ces concepts, que seule l'intervention humaine distingue.
3. L'étape suivante sera d'étudier les répercussions du principe de rationalité écologique sur le processus de binding et de réification, et donc sur la création de données et d'informations. Autrement dit, il s'agira d'identifier les caractéristiques de l'environnement dans lequel les individus produisent l'information et saisissent la donnée et de comprendre en quoi celles-ci conditionnent l'apparition de biais cognitifs. Elles seront considérées comme les caractéristiques déterminantes du modèle DBR, qui modulent les processus de binding et de réification.
4. Finalement, la dernière étape consistera à proposer un protocole expérimental permettant de tester les effets sur la qualité des informations de la manipulation des caractéristiques du modèle DBR précédemment identifiées. Cette expérience permet de confirmer partiellement la validité du modèle DBR et d'illustrer son utilité pratique pour améliorer la qualité de l'information et des données dans un contexte organisationnel. Bien sûr, ces travaux préliminaires suggèrent que des expériences similaires devraient être menées sur les autres caractéristiques de l'environnement affectant le binding et la réification pour confirmer intégralement la validité du modèle DBR.

En pratique, il s'agira donc d'étudier les points précédemment soulevés en s'appuyant, à titre d'illustration, sur le système de management de l'information d'ALTEN. En raison de la nature de la problématique et du caractère fondamentalement pluridisciplinaire de cette recherche, ce manuscrit suit principalement une ligne directrice fondée sur les apports des sciences cognitives telles que les neurosciences, la psychologie cognitive et évolutionnaire ou l'économie comportementale pour répondre aux interrogations issues des domaines des sciences de l'information.

## 1.3 Contributions

Les principaux résultats scientifiques apportés par ces travaux de recherche, suivis des publications passées et à venir, sont présentés dans cette section.

### 1.3.1 Résultats

Les principales contributions de ce travail sont les suivantes :

- une définition des concepts de donnée et d’information permettant leur différenciation théorique et fondée sur une étude approfondie de leur étymologie et sur la comparaison des approches similaires existantes ;
- le modèle DBR, un nouveau cadre conceptuel du processus de transformation des données en information et de l’information en données qui permet de caractériser et de conceptualiser ces processus – le manque d’un tel cadre a été noté par d’autres travaux scientifiques en sciences de l’information (Rowley, 2007) (Frické, 2009) ;
- l’identification des caractéristiques de l’environnement qui modulent a priori les processus de binding et de réification, en prenant expérimentalement appui sur le SI de la société ALTEN ;
- la proposition d’un modèle unifié de la qualité des données et de l’information qui s’inspire du modèle de qualité des données de Wang en révisant les différentes caractéristiques et en apportant une différenciation entre la qualité des données et la qualité de l’information ;
- la conception et la réalisation d’un protocole expérimental qui repose, d’une part, sur le modèle DBR pour indiquer les éléments de l’environnement à manipuler pour influencer les processus de binding et de réification et, d’autre part, sur le cadre conceptuel unifié de la qualité des données et de l’information pour mesurer l’effet des changements qui découlent des modifications de l’environnement ;
- enfin, la confirmation que la motivation a un effet significatif sur la qualité intrinsèque des informations produites dans un contexte organisationnel.

### 1.3.2 Publications

Les principales publications présentes et à venir liées à cette recherche sont les suivantes.

1. Article de conférence<sup>1</sup> :
  - Titre : The impact of motivation on the quality of project management data : an email-based communication case study
  - Conférence internationale EURAM 2022<sup>2</sup>
  - Auteurs : Robin Le Conte des Floris, Pierre Jouvelot, Cédric Dalmasso
  - Date et lieu : 16 juin 2022, Zürich, Switzerland
2. Poster de présentation des travaux de recherche (journée ISMME, Mines Paris, 22 juin 2022 ; journée « Mines Paris Research Day : science et industrie ») ;
3. Article de journal de présentation du modèle DBR, en préparation et à soumettre à IJIM (International Journal of Information Management).

---

1. <https://hal-mines-paristech.archives-ouvertes.fr/hal-03662794>

2. <https://conferences.euram.academy/2022conference/>

## 1.4 Organisation

Ce document est structuré en six chapitres principaux, chacun abordant différents aspects de l'étude, à l'exception de l'introduction et de la conclusion, allant de la théorie à l'application.

Le chapitre 2 présente rapidement les concepts théoriques nécessaires pour comprendre la recherche menée ici. Il aborde la différence entre les données et l'information, les théories de l'information de Shannon et de Chomsky, la perception et la subjectivité en neurosciences, les biais cognitifs et leur classification. Il examine ensuite les théories de l'acceptation des utilisateurs, la motivation et l'autodétermination, et l'état actuel de la recherche sur la gamification.

Dans le chapitre 3, le document introduit un nouveau modèle théorique, DBR, qui formalise le processus de transformation des données en informations, et vice versa. Il y est expliqué également en quoi ce nouvel angle pour étudier la relation entre information et donnée implique la prise en compte de l'environnement de saisie et la présence de biais cognitifs. Ce chapitre décrit finalement un nouveau cadre conceptuel de la qualité des données et de l'information et propose de nouvelles dimensions et caractéristiques pour évaluer cette qualité.

Le chapitre 4 utilise une approche par la *Grounded Theory* (théorie ancrée) pour examiner comment la rationalité écologique peut affecter les processus de binding et de réification. Il utilise la théorie ancrée et la méthode d'analyse dite "in vivo" pour collecter, encoder et analyser des données issues d'interviews avec des chefs de projet en charge de la production de données à ALTEN. Les résultats présentent un ensemble de caractéristiques de l'environnement dans lequel les données sont saisies, lequel affecte les processus de binding et de réification.

Le chapitre 5 utilise le modèle DBR et le nouveau cadre conceptuel de la qualité des données et de l'information pour expérimenter en quoi la motivation peut affecter la qualité intrinsèque de l'information. Il utilise des outils méthodologiques spécifiques comme le *A/B testing* pour isoler l'effet de la motivation sur la qualité des informations produites.

# Chapitre 2

## Contexte théorique

Ce chapitre présente un ensemble juxtaposé de théories et notions clés nécessaires à la compréhension de ce document. Ce chapitre n'apporte pas de contribution spécifique et se concentre sur le contexte théorique sur lequel repose ce document. Mais, en fait, une contribution transversale de cette thèse est de montrer qu'appréhender de manière pertinente les notions de donnée, d'information et de leurs qualités respectives nécessite de mobiliser des connaissances de domaines a priori disjoints.

La première section présente les concepts de données et d'informations en commençant par une analyse anthropologique de leur terminologie afin de correctement définir ceux-ci. S'ensuit la présentation des théories de Shannon et de Chomsky, deux approches complémentaires qui traitent différentes facettes de l'information. La question de l'évaluation de la qualité des données est également abordée, en s'appuyant sur le modèle conceptuel de Wang. La deuxième section se focalise sur des concepts issus des neurosciences, notamment le "binding" et la subjectivité de la perception humaine. Le rôle de la mémoire, de la motivation et des émotions dans la formation de notre perception du monde est expliqué, de même que le modèle d'apprentissage fondé sur la "stimulus-réponse". La section suivante aborde le concept de "biais cognitif", qui a marqué un changement de paradigme dans la compréhension traditionnelle de la rationalité. Ce concept met en exergue l'importance des contraintes de temps, d'information et de capacité cognitive, ouvrant la voie à l'émergence de l'économie comportementale.

La quatrième section traite de l'acceptation des utilisateurs et des différents modèles développés dans ce domaine, notamment l'apport de Venkatesh et son modèle unifié. La cinquième section présente la théorie de l'autodétermination de Ryan et Deci, mettant l'accent sur la distinction entre motivation intrinsèque et motivation extrinsèque, ainsi que les mécanismes de régulation de la motivation. Enfin, la sixième et dernière section offre un état de l'art de la *gamification* (ludification), une méthode qui utilise des concepts issus des jeux pour améliorer l'acceptation et l'utilisation des systèmes d'information, avec une attention particulière sur la manière dont elle peut améliorer le processus de partage et de collecte de l'information.

La figure 2.1 indique, de manière concise et structurée, les dépendances entre les différentes théories abordées ici, les objectifs présentés dans l'introduction, et les chapitres à venir de ce document. Si un ou plusieurs sujets sont déjà maîtrisés par le lecteur, la lecture des sections correspondantes (dont le numéro est indiqué sur la

figure 2.1 pour chaque élément), n'est pas indispensable.

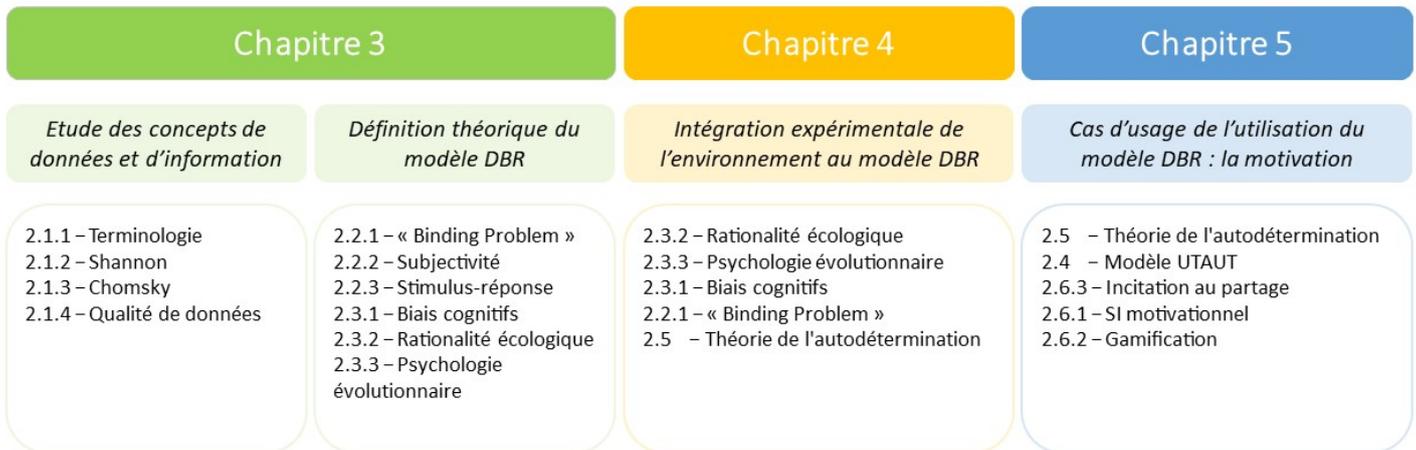


FIGURE 2.1 – Structuration de la thèse par chapitres, objectifs et concepts théoriques pertinents

## 2.1 Concepts de donnée et d'information

Les concepts d'information et de donnée ont été abordés de différentes manières dans la littérature, et il existe une certaine opacité dans la manière de les définir et de mesurer leurs valeurs. Une analyse des différentes terminologies de ces concepts (Zins, 2007) et la présentation des fondements théoriques tels que la mesure de l'entropie de Shanon (Rioul, 2020) ou le modèle conceptuel de Wang (R. Y. Wang & Strong, 1996) vont permettre, dans la suite de ce document, de se positionner vis-à-vis de ces différentes manières de définir et de mesurer la qualité de l'information et de la donnée.

### 2.1.1 Analyse comparative de la terminologie

En 2007, Chaim Zins propose une exploration anthropologique des fondements de la science de l'information grâce à la collaboration de 57 chercheurs de ce domaine provenant de 16 pays différents. Ce travail propose 130 définitions des concepts de donnée, d'information et de connaissance, lesquels sont débattues par les chercheurs sollicités afin d'établir des approches conceptuelles consensuelles permettant de définir ces trois concepts (Zins, 2007).

#### Approche métaphysique vs. non métaphysique

*“Mettre ces trois concepts ("données", "informations" et "connaissances") comme fait ici donne l'impression d'une hiérarchie logique : l'information est construite à partir de données et la connaissance émerge de l'assemblage d'informations. C'est un conte de fées.”* (Raphael Capurro) – Critique de l'approche non métaphysique.

L'approche métaphysique définit, logiquement, ces concepts comme des phénomènes métaphysiques et introduit des postulats tels que « la connaissance est éternelle » ou encore « la connaissance est une entité/objet indépendante ».

Dans le contexte du domaine des sciences de l'information, l'ensemble des membres de l'étude estime qu'une approche non métaphysique est plus appropriée.

### **Approche exclusive à l'homme vs. non exclusive**

*“Les données sont un ou plusieurs types d'ondes ou de particules énergétiques (lumière, chaleur, son, force, électromagnétisme) sélectionnées par un organisme conscient ou un agent intelligent sur la base d'un cadre ou d'un mécanisme inférentiel préexistant dans l'organisme ou l'agent. L'information est le cadre déductif actif ou latent d'un organisme ou d'un agent qui guide la sélection des données pour son propre développement ou sa propre construction. La connaissance est un ou plusieurs ensembles d'informations relativement stables. Un message est un ou plusieurs ensembles de données déduites, glanées à partir de réactions énergétiques externes ou internes.”* (Glynn Harmon) – Approche non exclusive

Parmi les définitions existantes des concepts d'information, de donnée et de connaissance, les approches non exclusives à l'homme considèrent que tous les êtres biologiques, voire plus rarement tous les éléments physiques, sont concernés. À nouveau, dans le contexte des sciences de l'information, presque la totalité des membres de l'étude estime l'approche centrée sur l'homme plus adaptée (Zins, 2007).

### **Approche fondée sur les sciences cognitives vs. non exclusive**

Ces approches sont celles exclusivement réservées à l'homme, jugées les plus pertinentes d'après les 57 membres de l'étude.

*“Dans les systèmes informatiques, les données sont les invariances codées. Dans le discours humain, les données sont celles qui sont déclarées, par exemple, par des informateurs dans le cadre d'une étude empirique. L'information est liée à la signification ou à l'intention humaine. Dans les systèmes informatiques, l'information est le contenu des bases de données, du web, etc. Dans les systèmes de discours humains, l'information est le sens des déclarations telles qu'elles sont voulues par le locuteur/écrivain et comprises/mal comprises par l'auditeur/lecteur. La connaissance est incarnée par l'être humain en tant que capacité à comprendre, expliquer et négocier des concepts, des actions et des intentions.”* (Hanne Albrechtsen) – Approche fondée sur les sciences cognitives.

*“L'information est, en fait, le résultat causal des composants et processus physiques existants. En outre, elle est un résultat émergent de ces entités physiques. C'est ce que révèle la définition systémique de l'information. Il s'agit d'un concept relationnel qui inclut la source, le signal, le mécanisme de libération et la réaction en tant que réactifs.»* (Hjørland) – Approche non exclusive.

Les approches centrées sur l'homme se divisent entre celles qui considèrent les concepts de donnée, d'information et de connaissance comme exclusivement liés aux phénomènes cognitifs et celles qui les considèrent vis-à-vis des phénomènes cognitifs, physiques et biologiques. Ce qui définit, entre autres, l'approche cognitive, c'est le fait que la donnée acquiert le statut d'information seulement au travers de l'interprétation humaine. Autrement dit, l'information n'existe qu'en tant que perception humaine d'un phénomène cognitif, en l'occurrence la lecture d'une donnée.

La majorité des membres de cette étude estime que cette approche est la plus appropriée dans le cadre des sciences de l'information. Cette étude conclut alors en affirmant irrévocablement, sur la base des échanges entre les 57 chercheurs en sciences de l'information, que l'approche non métaphysique, centrée sur l'homme et fondée sur les sciences cognitives est l'approche conceptuelle la plus pertinente (Zins, 2007).

## 2.1.2 Théorie de l'information de Shannon

Le concept d'information n'est pas délimité à un domaine spécifique tel que l'étude des systèmes d'information, la science des données ou la gestion des connaissances. Le concept d'information est utilisé de manière générale dans tout ce qui touche au domaine de la communication. La théorie de l'information de Shannon couvre cet ensemble de champs d'application.

Un des postulats de base de cette théorie est la distinction claire qui est faite entre les rôles de la source, du canal et de la destination, de l'émetteur et du récepteur, ainsi que du signal et du bruit. L'information est donc considérée comme transmise sous forme de message de la source à la destination par l'intermédiaire d'un canal bruité. Afin de minimiser la quantité de données nécessaires à la transmission d'un message, il est suggéré par Shannon de minimiser la redondance présente dans le message, en se fondant sur les propriétés statistiques de la source. Cependant, comme le canal est bruité, la minimisation de la redondance peut conduire à ce qu'un message ne soit pas reconnu, car la redondance aide également à ce que le sens du message puisse être correctement compris malgré le bruit. (Rioul, 2020)

Shannon considère que l'information est essentiellement aléatoire et que son résultat est incertain. Cette incertitude est utilisée pour définir une mesure de l'information transmise, que Shannon appelle "entropie", par analogie avec la notion similaire étudiée en thermodynamique et en physique statistique. De manière intuitive, on peut dire que plus la source émet d'informations différentes, plus l'entropie (ou incertitude sur ce que la source émet) est grande. La quantité d'informations contenues dans un flux de données est généralement caractérisée par son entropie de Shannon (Rioul, 2020).

*"Ma plus grande préoccupation était de savoir comment l'appeler. J'avais pensé à l'appeler 'information', mais le mot était trop utilisé, alors j'ai décidé de l'appeler 'incertitude". Lorsque j'en ai discuté avec John von Neumann, il a eu une meilleure idée. Von Neumann m'a dit : "Tu devrais l'appeler entropie, pour deux raisons. En premier lieu, ta fonction d'incertitude a été utilisée en mécanique statistique sous ce nom, donc elle a déjà un nom. En second lieu, et c'est plus important, personne ne sait vraiment ce qu'est l'entropie, donc dans un débat tu auras toujours l'avantage.".*

(Scientific American, 1971)<sup>1</sup>.

Cependant, cette mesure est insuffisante pour totalement caractériser la notion intuitive d'information, car elle ne tient pas compte de la composante sémantique de l'information, qui permet de la différencier de la donnée. Il sera donc nécessaire, dans notre cas, de compléter la mesure de l'information de Shannon pour couvrir cet aspect sémantique, qui est lié à l'interprétation et à l'intention humaines.

L'une des limites (et forces) de la théorie de Shannon est, donc, qu'elle ignore délibérément les aspects sémantiques de l'information pour se concentrer sur la caractéristique pratique de l'information nécessaire pour transférer celle-ci de A à B. Ainsi, elle reste indifférente à la signification des messages. Cependant, la signification d'un message est fondamentale dans la caractérisation de l'information, comme cela a été précisé précédemment.

### 2.1.3 Théorie linguistique de Chomsky

Une manière de préciser, au-delà de la théorie de Shannon, la distinction entre syntaxe et sémantique de l'information est fournie par la théorie formelle des langages.

La théorie linguistique de Chomsky aide à comprendre cette distinction grâce à son principe de grammaire universelle qui, dans son modèle, se décompose en divers sous-composants : le lexique, la syntaxe, le composant FP (forme phonétique), c'est-à-dire l'interprétation phonétique dans le langage parlé, et le composant FL (forme logique), c'est-à-dire l'interprétation sémantique. Selon Chomsky, tous les langages sont décrits à partir de principes universels, caractérisés par la faculté humaine d'interpréter le langage, et des paramètres qui varient d'une langue à une autre (Blitman, 2020).

Une information donnée, peu importe le langage, si l'on exclut FP dans le cadre d'une communication non verbale, est constituée formellement des sous-composants lexical et syntaxique, ainsi que de l'interprétation sémantique. Cette composante d'interprétation sémantique est ce qui manque à la théorie de Shannon pour obtenir une définition générale de l'information. Il est intéressant de noter que, selon Chomsky, “ *c'est parce que les représentations sémantiques au sens large du locuteur engagent ses croyances sur le monde et varient en fonction de l'intention du locuteur qu'elles échappent à toute modélisation.* ” (Blitman, 2020). Il semble donc que Shannon avait, d'un point de vue formel, vu juste en excluant de sa modélisation de l'information son aspect sémantique.

La théorie des langages de Chomsky peut parfaitement s'appliquer à la donnée, vue comme élément d'un langage, les langages étant considérés comme universels. On pourrait ainsi considérer que les niveaux lexical et syntaxique s'appliquent à la donnée, tandis que le composant FL, ou interprétation sémantique, s'appliquerait uniquement à l'information<sup>2</sup>. En s'appuyant uniquement sur la théorie des langages de Chomsky, une donnée de qualité serait une donnée lexicalement et syntaxiquement correcte, tandis que pour qu'une information soit de qualité, l'interlocuteur

---

1. <https://quotepark.com/fr/auteurs/claude-shannon/>

2. On pourrait, d'une autre manière, voir la sémantique de la donnée comme équivalente à sa seule représentation textuelle, la fonction d'interprétation abstraite étant ici l'identité.

doit partager les croyances sur le monde et comprendre les intentions du locuteur, rendant ainsi l'information vraie et compréhensible à ses yeux.

#### 2.1.4 Qualité des données du point de vue du consommateur

Wang, en 1996, propose un cadre pour définir la qualité des données telle qu'elle est perçue à la fois par le professionnel d'un système d'information et par le consommateur de la donnée (R. Y. Wang & Strong, 1996). La qualité des données est ainsi divisée en quatre dimensions :

1. la qualité des données intrinsèque est la qualité des données en tant que telles et repose sur l'exactitude, l'objectivité, la vraisemblance et la réputation ;
2. la qualité des données contextuelle indique qu'elle doit être prise en compte dans le contexte de la tâche dans laquelle elle doit être utilisée et consiste en la valeur ajoutée, la pertinence, le respect des délais, l'exhaustivité et une quantité de données appropriée ;
3. la qualité des données représentationnelle est liée au format de la donnée et à sa signification et consiste en l'interprétabilité, la facilité de compréhension, la cohérence et la représentation concise ;
4. enfin, la qualité de la donnée liée à l'accessibilité souligne l'importance du rôle du système et consiste en l'accessibilité et en la sécurité de l'accès.

En 2002, ces dimensions sont reprises par le même auteur, en tant que co-auteur d'un second article de recherche, pour définir la qualité de l'information dans le cadre d'un nouveau modèle, PSP/IQ, ce qui signifie *Product and Service Performance model for Information Quality* (Kahn, Strong, & Wang, 2002). Une nouvelle approche est utilisée ici, considérant l'information comme simultanément un produit et un service du point de vue du consommateur. L'information est, dans ce cadre, le produit final fourni au consommateur, et le service est la transformation de la donnée de manière à délivrer une information utile pour le consommateur, en respectant ses attentes et les standards de qualité de l'information. L'objectif est ici notamment d'avoir un modèle davantage centré sur le concept de *fitness for use*, qui met en avant l'importance de prendre en compte le point de vue du consommateur concernant la qualité, car, au bout du compte, c'est lui qui jugera (R. Y. Wang & Strong, 1996)(Kahn et al., 2002).

Cet article reprend, telles quelles, les dimensions décrivant la qualité des données mentionnées dans l'article précédent, mais cette fois-ci pour décrire l'information. Il remplace d'ailleurs, dans la définition de chaque dimension, les utilisations du mot "donnée" par "information" sans expliquer pourquoi ces dimensions seraient applicables de la même manière à l'information et à la donnée. Cela rend discutable la classification qui est faite par la suite de ces dimensions dans le cadre du modèle PSP/IQ. C'est d'autant plus vrai que ces dimensions avaient été définies dans le précédent article grâce à une enquête reposant sur un questionnaire, qui demandait aux répondants de définir spécifiquement la donnée, et non l'information.

## 2.2 Perception et subjectivité en neurosciences

Le *binding* (proche de la notion de lien, mais nous garderons le vocable anglais par la suite) est le nom donné à l'opération cognitive complexe qui lie cognition et monde réel lors de la prise de décision et de la coordination d'actions par un individu ; sa modélisation précise demeure un problème non résolu, d'où le terme de *binding problem*. Une meilleure compréhension de cette notion, généralement abordée en neurosciences, permettra par la suite de mieux clarifier le processus de transformation de la donnée en information, lequel passe par l'interprétation humaine, comme cela a été vu dans la première section. Un des aspects spécifiques du binding, l'unité subjective de perception, pourrait expliquer en partie l'introduction de biais dans les données et la perte d'informations au moment où le système cognitif opère, comme nous le verrons par la suite.

### 2.2.1 *Binding Problem*

Le binding est un processus nécessaire pour que le cerveau puisse prendre des décisions et ordonner des actions, car, même si celui-ci est un système hautement distribué, toutes les informations ne peuvent être traitées et utilisées simultanément. Une certaine forme de tri est donc nécessaire pour séparer les informations relatives à un événement cognitif perçu de celles qui ne le sont pas. De manière similaire, le binding est nécessaire sous la forme d'un rattachement des caractéristiques pertinentes, voire nécessaires, à l'action envisagée (Zimmer, Mecklinger, & Lindenberger, 2006).

En science, on parle de "problème" lorsqu'il n'existe pas de modèle plausible pour son substrat. En ce qui concerne le processus de binding, il n'existe pas de modélisation robuste et universelle du processus de binding, raison pour laquelle les études font référence au *binding problem*. Cependant, les recherches en neurosciences identifient des caractéristiques générales ayant un rôle prépondérant dans le processus de binding (Zimmer et al., 2006).

Le problème du binding peut, de fait, être subdivisé en au moins quatre situations : les considérations générales sur la coordination, l'unité subjective de perception, la liaison des caractéristiques visuelles et la liaison des variables (Feldman, 2013). Nous décrivons ci-dessous ces quatre facettes du problème.

1. Le problème de coordination fait référence à la manière dont le cerveau perçoit de manière unitaire des objets et des activités qui semblent être au même endroit et au même moment. Par exemple, lorsqu'un musicien joue du violon ou de la guitare, il accomplit deux actions distinctes avec chacune de ses mains, sollicitant différents systèmes moteurs. Pourtant, le cerveau les perçoit comme une seule et même action. Il semble que l'attention joue un rôle crucial dans la détermination de ce qui est perçu comme étant lié ou non.
2. L'unité subjective de la perception représente l'un des plus grands mystères des neurosciences et illustre le problème corps-esprit concernant la relation entre notre expérience subjective et la fonction neuronale. Un bon exemple de ce problème est le fait que nous avons une expérience subjective d'une représentation stable, à haute résolution et complète d'une scène visuelle, alors

que les preuves biologiques suggèrent que ce n'est pas le cas. Ce qui est paradoxal, c'est que d'autres exemples suggèrent qu'une perception unifiée n'est pas nécessaire pour prendre une décision, comme c'est le cas pour les personnes qui, dans une entreprise, effectuent de nombreuses petites actions et prennent des décisions dans un but unifié. Les capacités pour toutes ces activités sont distribuées (comme dans le cerveau) sans qu'aucun individu ou petit groupe n'en ait une compréhension complète, et, pourtant, l'organisation prend des mesures unifiées (Zimmer et al., 2006).

3. La liaison des caractéristiques visuelles explique comment le cerveau relie entre elles les caractéristiques visuelles, comme les couleurs, d'un objet.
4. La liaison des variables réfère à la manière dont le cerveau connecte deux variables ou plus, indépendamment de leur valeur. Cette liaison de variables est une caractéristique inhérente au langage et au raisonnement abstrait. À ce jour, il n'existe pas de modèles animaux ou d'expériences disponibles pour étudier ce phénomène. Un concept connexe important est l'unification, soit la liaison de deux variables ou plus, indépendamment de leur valeur. Par exemple, le mot "mouton" peut être utilisé au singulier ou au pluriel. Cependant, dans l'expression "un mouton", le nombre grammatical des deux mots est lié au singulier à cause de la règle d'unification de l'accord.

L'introduction de la notion de binding a donc pour vocation d'expliquer la manière dont des éléments encodés par des circuits cérébraux distincts peuvent être combinés pour la perception, la décision et l'action. Le binding est un processus nécessaire pour que le cerveau puisse prendre des décisions et ordonner des actions. Ce qui est clair, néanmoins, c'est que le processus de binding est conditionné par un certain nombre de facteurs tels que l'action envisagée, l'événement cognitif perçu, le contexte et les connaissances préexistantes.

## 2.2.2 Subjectivité de la perception

Malgré le fait que chaque individu construise sa propre perception du monde, le résultat n'est pas nécessairement spécifique à chaque individu. De manière générale, les différents individus arrivent à avoir une perception du monde plutôt précise et consensuelle, et ce, parce qu'il existe un certain nombre de règles et de contraintes communes à tous les membres d'une même espèce. Cette perception commune est probablement le résultat d'un processus d'évolution de l'homme dans un contexte de sélection naturelle où avoir une perception précise et partagée de l'environnement était critique (Pally & Olds, 2018). Cette perception comporte, cependant, un certain degré d'imprécision. Les illusions d'optique sont souvent utilisées en neurosciences et psychologie pour montrer comment notre cerveau peut être dupé, mais, même dans ce cas, les moyens de le duper sont les mêmes pour tous les individus, et la perception reste consensuelle.

Cependant un certain niveau d'individualité peut être observé dans la perception du monde, notamment quand on s'intéresse à l'influence de la motivation, des émotions et de la mémoire sur la perception (Pally & Olds, 2018).

- La mémoire est prise en compte lors de la perception, car, pour minimiser les efforts associés, le cerveau compare les expériences précédentes à celles en

cours pour essayer d'en trouver une suffisamment similaire et compléter le manque d'informations à disposition pour avoir une perception complète de l'événement cognitif. Ainsi, la perception des individus est fortement biaisée par leurs réminiscences des expériences passées.

- En ce qui concerne la motivation, il est plus probable de percevoir les éléments de l'événement qui sont pertinents pour répondre à nos besoins du moment que ceux qui ne le sont pas. La motivation peut également avoir une influence sur l'attention, qui est un facteur primordial du processus de perception.
- Les émotions influencent comment les stimuli sont interprétés ; par exemple la peur et l'anxiété augmentent la tendance à interpréter un stimulus comme dangereux. Par ailleurs, les individus ayant vécu des traumatismes peuvent perpétuer l'état dans lequel ils se trouvaient au moment du traumatisme, mêlant mémoire et émotions (Pally & Olds, 2018).

On peut dire en un sens qu'il y a donc deux façons de voir la subjectivité : celle qui est générale et qui différencie la réalité physique de la réalité perçue ; et celle, individuelle, qui explique que chaque individu puisse avoir une perception différente de la même réalité. D'un point de vue philosophique, il serait d'ailleurs intéressant de se demander à laquelle de ces deux visions nous faisons appel quand nous reprochons à quelqu'un de ne pas être suffisamment "objectif".

### 2.2.3 Association "stimulus-réponse" (S-R)

Le modèle d'apprentissage S-R, introduit par l'école behavioriste, est présenté comme l'hypothèse selon laquelle l'apprentissage conduit à la formation de connexions stimulus-réponse. Dans le conditionnement classique, ce lien est établi entre le stimulus conditionné<sup>3</sup> et le stimulus non conditionné<sup>4</sup> (par exemple, entre le son et la salivation dans la procédure d'Ivan Pavlov) ; dans le conditionnement instrumental, le lien est établi entre le stimulus discriminant et la réponse (par exemple, entre un son et la pression d'une barre).

Le conditionnement classique est un type d'apprentissage dans lequel un stimulus initialement neutre - le stimulus conditionné - est associé à un stimulus qui provoque une réponse réflexe - le stimulus non conditionné - afin d'entraîner une réponse apprise, ou conditionnée, lorsque le stimulus conditionné est présenté. Par exemple, le son d'une tonalité peut être utilisé comme stimulus conditionné et la salivation liée à la présence de nourriture comme un stimulus non conditionné. Après la présentation de paires répétées, à savoir le signal sonore suivi immédiatement de la nourriture, le signal sonore, qui n'avait initialement aucun effet sur la salivation (c'est-à-dire qu'il

---

3. Le stimulus conditionné est un stimulus neutre qui est associé de manière répétée à un stimulus non conditionné jusqu'à ce qu'il acquière la capacité de susciter une réponse qu'il ne suscitait pas auparavant. Dans de nombreux cas (mais pas tous), la réponse provoquée par le stimulus conditionné est similaire à celle provoquée par le stimulus non conditionné. Une lumière, par exemple, associée de manière répétée à de la nourriture (le stimulus non conditionné), finit par susciter la même réponse que la nourriture (c'est-à-dire la salivation) lorsqu'elle est présentée seule.

4. Le stimulus non conditionné est un stimulus qui provoque une réponse réflexe non conditionnée, comme la contraction de la pupille lors de l'exposition à la lumière ou la salivation lorsque de la nourriture est dans la bouche.

était neutre par rapport à elle), provoquera la salivation même si la nourriture n'est pas présentée.

Le conditionnement instrumental est toute forme de conditionnement dans laquelle la réponse correcte est essentielle pour le renforcement. Le conditionnement instrumental implique généralement des activités complexes pour atteindre un objectif, comme lorsqu'un rat est entraîné à naviguer dans un labyrinthe pour obtenir de la nourriture. Il s'oppose au conditionnement classique, dans lequel le renforcement est accordé indépendamment de la réponse.

Ces principes de conditionnement ont également une influence sur le processus de binding (Colzato, Barone, Sellaro, & Hommel, 2017). Par exemple, en utilisant un conditionnement instrumental similaire à celui utilisé en apprentissage par renforcement, qui récompense la réponse au stimulus souhaité et punit celles qui ne le sont pas, les épisodes S-R récompensés positivement entraînent la création de bindings plus forts.

Cela est notamment confirmé par le fait que le processus de binding est en partie modulé par les fonctions dopaminergiques (Colzato et al., 2017). De manière plus générale, le système dopaminergique joue un rôle important dans la neuromodulation, notamment dans le contrôle moteur, la motivation, la récompense, les fonctions cognitives et les comportements maternels et reproductifs. Ils auront donc aussi un rôle à jouer lors de l'utilisation de systèmes sociaux, comme les systèmes d'information des organisation.

## 2.3 Biais cognitif

L'introduction du terme "biais cognitif" par Kahneman et Tversky dénotait une rupture avec les théories économiques fondées sur une rationalité classique, en reconnaissant les contraintes de temps, d'information et de capacité cognitive nécessaires à la prise de décisions éclairées. Plusieurs concepts ont depuis émergé pour étudier les biais cognitifs : la rationalité écologique, qui voit les biais et les heuristiques comme des outils d'adaptation à l'environnement, et la psychologie évolutionnaire, qui les considère comme des solutions à des problèmes spécifiques à un environnement passé. Les biais cognitifs sont également étudiés dans de nombreux autres domaines, notamment l'apprentissage automatique et l'épidémiologie, où ils sont souvent classés en fonction des étapes du processus de prise de décision.

On donne, ci-dessous, quelques définitions utiles pour mieux comprendre la section qui suit, et le concept de biais cognitif (Wilke & Mata, 2012) :

**rationalité limitée**, *principe selon lequel les organismes ont des ressources limitées, comme le temps, l'information et la capacité cognitive, pour résoudre les problèmes auxquels ils sont confrontés ;*

**biais cognitif**, *erreur systématique de jugement et de prise de décision commune à tous les êtres humains, qui peut être due à des limitations cognitives, à des facteurs motivationnels et/ou à des adaptations à des environnements naturels ;*

**rationalité écologique**, *principe selon lequel il existe une correspondance entre la structure de l'information dans l'environnement et les stratégies de jugement*

*et de prise de décisions des êtres humains et autres organismes ;*

**heuristique**, *mécanisme de jugement ou de prise de décision, ou raccourci cognitif, qui repose sur un petit nombre d'informations et des ressources cognitives modestes.*

### 2.3.1 Origines

Le terme "biais cognitif" a été introduit au début des années 1970 par Kahneman et Tversky dans le cadre de leur programme de recherche sur les "heuristiques et biais". Ce programme se concentrait sur la capacité des individus à prendre des décisions en tenant compte de leurs ressources limitées, comme le temps, l'information ou les capacités cognitives.

Cette approche était en rupture avec le principe de rationalité classique utilisé dans la plupart des théories économiques de l'époque. Au lieu de cela, Kahneman et Tversky ont proposé le principe de rationalité limitée, qui prend en compte les contraintes auxquelles un individu peut être confronté lorsqu'il prend une décision, tel que le fait que les ressources comme le temps, l'information et la capacité cognitive soient limitées. Ce changement de paradigme a donné naissance à un nouveau champ d'études en économie, l'économie comportementale, qui formule de nouvelles théories économiques en s'appuyant sur la psychologie et les sciences cognitives ([Wilke & Mata, 2012](#)).

Kahneman et Tversky ont été récompensés en 2002 par le prix de la Banque de Suède en sciences économiques en mémoire d'Alfred Nobel (communément appelé "prix Nobel d'économie") pour leur théorie des perspectives. Cette théorie est fondatrice de l'économie comportementale, une branche de l'économie qui s'appuie sur la psychologie et les sciences cognitives, et constitue l'une des premières théories économiques construites à partir de travaux expérimentaux. D'ailleurs, les neurosciences sont souvent perçues comme définissant les sciences cognitives, mais c'est, en fait, une discipline très spécifique qui s'intéresse au fonctionnement de notre cerveau avec un prisme un peu limité et qui n'est souvent pertinente qu'en lien avec d'autres disciplines pour répondre à des problématiques plus complexes.

Les sciences cognitives représentent, en fait, l'ensemble des disciplines qui s'intéressent au cerveau, au comportement et à la psychologie humaine. En effet, dans les sciences cognitives, il y a plusieurs niveaux d'analyse, tels que le niveau neuronal, le niveau cérébral et le niveau comportemental, qui fournissent différents types de réponses à différentes questions. Les niveaux neuronal et cérébral sont très intéressants, mais ne sont pas toujours pertinents pour toutes les problématiques. Par exemple, observer les réactions du cerveau par imagerie par résonance magnétique (IRM) peut être pertinent pour étudier un phénomène de bas niveau comme la vision, qui implique des zones cérébrales spécifiques. Cependant, pour des phénomènes de haut niveau tels que la prise de décision face à des situations sociales ou politiques spécifiques, cette approche peut ne pas fournir de réponses suffisantes, car ces situations impliquent un grand nombre de zones cérébrales.

### 2.3.2 Rationalité écologique

Le concept de rationalité écologique décrit la correspondance entre une structure et une représentation de l'information dans l'environnement, d'une part, et l'algorithme simple de prise de décisions de l'individu telles que les heuristiques, d'autre part. Le principe de la rationalité limitée est fortement lié à l'idée selon laquelle le système cognitif est fondamentalement adapté à son environnement, que ce soit grâce à un apprentissage ou à l'évolution de l'espèce. À partir des années 1990, le programme *Fast and frugal heuristics* présenté par Gigerenzer, Todd et l'ABV Research Group met en avant ce principe de rationalité écologique, soit en quoi le succès de stratégies de raisonnement dépend de la structure de l'environnement (Wilke & Mata, 2012). Ce programme recherche et analyse les caractéristiques de l'environnement auxquelles les individus font face pour identifier le ou les modèles cognitifs de raisonnement qui fonctionnent bien dans cet environnement, pourquoi ils sont choisis et pourquoi ils fonctionnent bien.

Cette approche est un peu à l'opposé de l'approche de Kahneman et Tversky, puisqu'on considère ici les biais cognitifs et les heuristiques non pas comme le résultat des limites d'un individu, mais comme des outils permettant de s'adapter à un environnement et à trouver une solution à un problème. Ces deux approches opposées sont sources de clivage dans la littérature, avec ceux qui affirment que l'esprit humain n'est pas infaillible et commet des erreurs et ceux qui ont tendance à affirmer qu'il est le plus souvent précis et exact. Quoi qu'il en soit, cette approche a le mérite d'avoir mis l'accent sur l'importance de l'environnement pour qualifier un biais cognitif ou une heuristique. Le modèle DBR y fait référence, dans le chapitre 3.

### 2.3.3 Psychologie évolutionnaire et management de l'erreur

Alors que la rationalité écologique décrit la correspondance entre l'esprit humain et l'environnement, la rationalité évolutionnaire affirme que, ce qui est important, c'est la correspondance avec les environnements passés dans lesquels l'homme a évolué en tant qu'espèce. La psychologie évolutionnaire tente ainsi de comprendre les origines des biais cognitifs (Haselton, Nettle, & Andrews, 2015). De la même manière que dans l'approche suivant le principe de rationalité écologique, en psychologie évolutionnaire, les biais cognitifs ne sont pas vus comme des défauts de conception, sur les bases de standards de logique et de précision, mais comme des outils permettant de résoudre des problèmes. En psychologie évolutionnaire, les biais cognitifs rappellent que les traits cognitifs peuvent être évalués de nombreuses manières différentes (suffisance logique, précision, rapidité du processus, etc.). Ainsi, en psychologie évolutionnaire, les traits cognitifs ne sont pas évalués sur la base de la précision ou de la logique, mais sur la base de la capacité à résoudre un problème de manière spécifique à l'environnement dans lequel l'individu est placé (Haselton et al., 2015). Avec cette approche, la présence d'un biais cognitif peut être le signe que le trait cognitif sollicité n'a pas été conçu pour la tâche demandée, mais reste efficace pour ce pour quoi il a été conçu, c'est-à-dire répondre à des problèmes liés à un environnement, potentiellement passé, dans lequel l'espèce a évolué. Par ailleurs, certains scientifiques du domaine de la psychologie évolutionnaire ont suivi l'approche du programme des heuristiques et biais de Kahneman et Tversky, qui

consiste à utiliser les erreurs pour définir les biais cognitifs, mais avec une différence dans la manière de les appréhender. En effet, ils ont introduit le rôle de la théorie des coûts dans les études des biais cognitifs en avançant l'argument selon lequel éliminer toutes les erreurs est pratiquement impossible, mais que les coûts associés à certaines erreurs peuvent conduire les organismes à commettre systématiquement un type d'erreur plutôt qu'un autre (Wilke & Mata, 2012). C'est la base de la théorie du management de l'erreur, qui introduit un nouveau type de biais, les biais de management de l'erreur.

Ainsi, les biais cognitifs sont, en psychologie évolutionnaire et plus précisément en théorie du management de l'erreur, répartis en trois types (Haselton et al., 2015) :

- les heuristiques, qui sont des biais qui proviennent des contraintes cognitives et de traitement de l'information humaine et qui, dans la plupart des cas, fonctionnent correctement, mais qui dysfonctionnent de manière systématique dans certains autres ;
- les biais de management de l'erreur, qui augmentent le nombre d'erreurs mais entraînent des erreurs peu, ou moins, coûteuses.
- les artefacts, qui sont des biais apparents qui résultent de l'application de normes et standards inappropriés, ou du placement d'un humain dans une structure d'environnement non naturelle.

### 2.3.4 Biais dans d'autres domaines

Comme nous l'avons expliqué précédemment, les sciences cognitives et comportementales sont souvent sollicitées pour des problématiques pluridisciplinaires, afin de répondre à des questions complexes. Il n'est donc pas surprenant que l'on retrouve le concept de biais cognitif dans de nombreux autres domaines pour expliquer les erreurs dans la résolution de problèmes. Cela dit, on observe une différence dans la manière d'identifier et classifier les biais entre les études appartenant strictement à un domaine des sciences cognitives et comportementales et celles les sollicitant pour des problématiques issues d'autres domaines.

Ces études spécifiques à une problématique identifient les biais, cognitifs ou non, qui affectent ainsi que les caractéristiques de l'environnement qui entraînent le plus d'erreurs ou les erreurs les plus coûteuses.

**Statistique, informatique, intelligence artificielle** Dans le domaine de l'apprentissage automatique, lorsque l'objectif est de réaliser un outil informatique d'aide à la décision, la principale préoccupation concerne les biais affectant les décisions fondées sur l'intelligence artificielle (IA). Les biais retenus seront alors ceux liés à la qualité des données, à la performance de l'algorithme et aux ingénieurs travaillant sur l'outil de décision. Ils seront ensuite classés en fonction des différentes étapes du processus (Srinivasan & Chander, 2021), à savoir :

- la création des données, avec des biais liés à l'échantillonnage, aux appareils de mesure ou à l'étiquetage des données ;

- la formulation du problème, avec des biais tels que ceux liés à la spécification du problème, à sa présentation avec, notamment, l'effet de cadrage (*framing effect bias*);
- l'analyse des données, avec des biais relatifs à la sélection des données, des variables et de l'algorithme d'apprentissage automatique;
- la phase de validation et de test, avec des biais relatifs à l'évaluation humaine de l'outil et au choix des échantillons de données utilisés pour tester son efficacité.

**Recherche** Dans le domaine de la recherche en épidémiologie ou de la recherche expérimentale, en général, on retrouve également une classification des biais relative au processus de production de la recherche scientifique et des analyses statistiques (Balakrishnan et al., 2014), avec :

- les biais de revue de littérature, ainsi que les biais liés à la publication elle-même et aux attentes des éditeurs;
- les biais de conception et d'exécution de l'étude tels que les biais de sélection et d'échantillonnage;
- Les biais de collecte et d'analyse des données, notamment liés aux appareils de mesure, à la définition du processus de collecte de données ou à la stratégie d'analyse des données, qui sont également liés aux biais de l'observateur ou du sujet;
- les biais d'interprétations qui proviennent des spéculations et inférences faites par les chercheurs, telles que l'influence des opinions, suppositions et présomptions préalables des chercheurs.

Comme le montrent les deux exemples précédents, une approche fréquemment utilisée pour identifier et classer les biais consiste à les évaluer en fonction du processus qui mène à l'objectif recherché, comme la réalisation d'une étude statistique, la conception d'un programme informatique ou la mise en place d'un outil d'aide à la décision. Les biais sont ensuite classés selon les étapes du processus, en privilégiant ceux qui ont le plus d'impact sur la qualité du produit final.

Bien que cette méthode puisse sembler diverger de l'approche traditionnelle en sciences cognitives, elle s'y inscrit finalement. Effectivement, l'évolution des perspectives dans ce domaine met de plus en plus l'accent sur les attributs de l'environnement, comme le démontre le concept de rationalité écologique. Les chercheurs s'intéressent donc de plus en plus à la structuration et à la présentation de l'information dans l'environnement, ainsi qu'au processus de décision et d'action. Le modèle DBR décrit dans le chapitre 3 s'inscrit également dans ce mouvement.

## 2.4 Acceptation des utilisateurs

L'utilisation efficace des technologies, et plus précisément du système d'information, a toujours été une préoccupation importante pour les organisations (R. Wang et al., 1995). Il y a généralement deux manières d'aborder la question de son utilisation. D'abord, on peut se demander si le système d'information est effectivement

utilisé par les utilisateurs, ou, autrement dit, s'il est accepté. Cette approche repose sur la mesure de l'intention de l'utilisation et de l'utilisation effective du système d'information. Cette approche est désignée par le terme d'“acceptation des technologies par les utilisateurs”, et c'est celle que nous abordons dans cette section. Ensuite, la question est de savoir si son utilisation est efficace ou non. Cette efficacité dépendrait essentiellement de la motivation de l'utilisateur à utiliser le système d'information, qui est en partie corrélée à la manière dont celui-ci est conçu (Koivisto & Hamari, 2019). Nous aborderons cette seconde approche dans les deux sections suivantes consacrées à la motivation et à la gamification.

Le manque d'acceptation par les utilisateurs a longtemps été un obstacle au succès des nouveaux systèmes d'information. Pour expliquer et conceptualiser l'acceptation d'un nouveau système d'information, ou d'une nouvelle technologie, par les utilisateurs, de nombreux modèles ont été développés. Venkatesh a consacré une importante partie de sa carrière à les étudier, les comparer et à en proposer un nouveau les unifiant. Dans un premier temps, nous ferons un rapide résumé historique des modèles existants à partir de deux articles détaillés ci-dessous (Alomary & Woollard, 2015) (Venkatesh et al., 2003), puis nous présenterons en détail l'apport du modèle unifié de Venkatesh et al.

### 2.4.1 Historique des modèles d'acceptation

L'historique des modèles d'acceptation des utilisateurs de systèmes d'information inclut plusieurs approches théoriques développées pour comprendre et prédire l'acceptation des technologies par les utilisateurs. Parmi ces modèles, on retrouve : la *Theory of Reasoned Action* (TRA), qui analyse les motivations sous-jacentes à l'action en se fondant sur l'attitude et les normes subjectives ; le Modèle de l'acceptation des technologies (TAM-TAM2), qui prédit l'utilisation et l'acceptation des systèmes d'information en se focalisant sur l'utilité perçue, la facilité d'utilisation perçue et les normes subjectives ; le Modèle motivationnel (MM), qui repose sur les théories de la motivation intrinsèque et extrinsèque de Ryan et Deci ; la Théorie du comportement planifié (TPB), qui ajoute le concept de *Perceived Behavioral Control* pour évaluer la perception de contrôle sur l'utilisation ; le Modèle TAM et TPB combiné, qui hybride les concepts de TPB avec l'utilité perçue de TAM ; le Modèle d'utilisation du *Personal Computing* (MPCU), issu de la théorie du comportement humain de Triandis ; la Théorie de la diffusion d'innovations (IDT), qui analyse les innovations en se fondant sur l'avantage relatif, la facilité d'utilisation, l'image et la visibilité, la compatibilité, la démontrabilité des résultats, et le volontariat de l'utilisation. Pour finir cet historique, la Théorie sociale cognitive (SCT) se fonde sur la notion de l'observation des autres dans les interactions sociales pour comprendre l'acquisition de connaissances d'un individu en lien avec l'utilisation d'un système d'information (Venkatesh et al., 2003)(Alomary & Woollard, 2015).

On le voit, la liste est longue et une approche unifiée de ces modèles était clairement nécessaire. Venkatesh et al (Venkatesh et al., 2003) présentent un comparatif des modèles d'acceptation des utilisateurs de systèmes d'information pour créer un modèle unifié, l'*Unified Theory of Acceptance and Use of Technology* (UTAUT). L'étude a été menée dans quatre organisations différentes, mesurant les concepts

des modèles à trois moments distincts. Les résultats ont révélé que chaque modèle avait au moins un concept central significatif influençant l'acceptation à toutes les périodes de mesure. Le modèle unifié intègre les concepts centraux les plus explicatifs de chaque modèle, donnant lieu à quatre concepts directement liés à l'acceptation : l'espérance de performance, l'espérance d'effort, l'influence sociale et les conditions facilitantes. Trois concepts indirects sont également pris en compte : l'attitude vis-à-vis de l'utilisation, l'auto-efficacité et l'anxiété. Les modérateurs tels que l'âge, le volontariat, le sexe et l'expérience ont une influence sur un ou plusieurs de ces concepts centraux. Ce modèle unifié offre une approche plus complète pour comprendre l'acceptation des utilisateurs de systèmes d'information.

## 2.4.2 Unified Theory of Acceptance and Use of Technology (UTAUT)

Une comparaison des modèles d'acceptation des utilisateurs de systèmes d'information, présentés dans la sous-section précédente, a donc été étudiée pour créer un modèle unifié, l'Unified Theory of Acceptance and Use of Technology (UTAUT). L'étude a été menée dans quatre organisations variées en termes de technologie, de nature d'utilisation, d'organisation et de cœur de métier. Les mesures ont été recueillies à trois moments différents à l'aide d'un questionnaire reflétant les concepts principaux des huit modèles étudiés. Le temps d'utilisation des nouvelles technologies a également été mesuré, avec un suivi sur une période de 6 mois.

L'analyse a été faite en utilisant la régression des moindres carrés partiels (PLS) pour évaluer la robustesse du modèle. Les résultats ont révélé que chaque modèle avait au moins un concept central significatif. Ainsi, les concepts les plus explicatifs ont été intégrés dans le modèle unifié, qui comprend finalement 4 concepts ayant un impact direct sur l'acceptation des utilisateurs, 3 ayant un impact indirect et 4 modérateurs qui ont une influence sur un ou plusieurs des concepts centraux.

- **Concepts principaux** : Espérance de performance, espérance de l'effort, influence sociale et conditions facilitantes.
- **Concepts indirects** : Attitude vis-à-vis de l'utilisation, auto-efficacité et anxiété.
- **Modérateurs** : Âge, volontariat, sexe, expérience.

L'espérance de performance est définie comme le degré auquel un individu pense que l'utilisation du système va l'aider à obtenir des gains dans le cadre de son travail. On retrouve des concepts parents dans les modèles antérieurs : utilité perçue (TAM/TAM2, C-TAM-TPB), motivation extrinsèque (MM), adaptabilité métier (MPCU), avantage relatif (IDT), résultat attendu (SCT).

L'espérance d'effort est définie comme le degré auquel un utilisateur perçoit la facilité d'utilisation du système. Ses concepts parents sont, respectivement, la facilité d'utilisation perçue (TAM/TAM2), la complexité (MPCU) et la facilité d'utilisation (IDT).

L'influence sociale est vue comme le degré auquel un individu perçoit l'importance de l'opinion des autres concernant le fait qu'il devrait ou non utiliser le système. Là encore, on peut y voir l'influence des concepts parents antérieurs : normes

subjectives (TRA, TAM2, TPB/DTPB, C-TAM-TPB), facteurs sociaux (MPCU) et image (IDT).

Enfin, les conditions facilitantes correspondent au degré auquel un individu pense qu'il peut être aidé par une organisation et une infrastructure technique pour utiliser le système, avec comme concepts parents la perception de contrôle sur l'utilisation (TPB/DTPB, C-TAM-TPB).

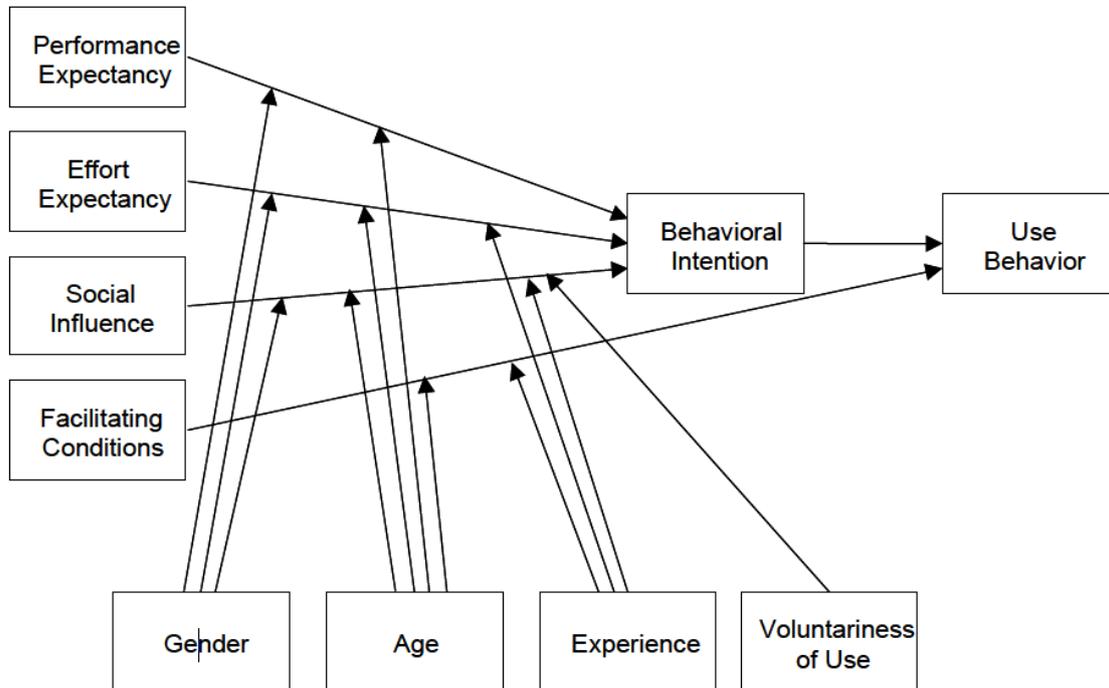


FIGURE 2.2 – Modèle de recherche (Venkatesh et al., 2003)

Sur la figure 2.2 se trouve un schéma du fonctionnement du modèle et les influences des modérateurs sur chacun des concepts centraux. Les différents facteurs mis en évidence dans ces réflexions sur la notion de modèle d'acceptation vont se trouver naturellement mobilisés dans notre réflexion sur la meilleure manière d'obtenir des données fiables au sein des systèmes d'information.

## 2.5 Motivation : théorie de l'autodétermination

La théorie de l'autodétermination de Ryan et Deci (Ryan & Deci, 2000) permet de distinguer différents types de motivation chez les individus et analyse les facteurs psychologiques innés permettant de maximiser la motivation chez ceux-ci. Par ailleurs elle propose des méthodes dites de "régulation" permettant de motiver "intrinsèquement" les individus, ce qui est considéré comme la situation la plus productive. La motivation est un facteur qui impacte la subjectivité et l'attention des individus dans l'utilisation des technologies de l'information et qui joue également un rôle clé dans les processus de partage de connaissances, ce qui explique que ce concept soit sollicité dans de nombreux travaux en sciences de l'information.

### 2.5.1 Principes

La théorie de l'autodétermination, *Self Determination Theory* (SDT), adopte une approche qui s'inscrit dans le cadre d'une psychologie cognitive qui étudie la pensée humaine sous le jour du traitement de l'information. C'est aussi une théorie humaniste qui étudie les tendances inhérentes au développement des hommes en fonction de leurs besoins psychologiques innés ; elle utilise ces derniers comme base explicative de leur motivation volontaire et éclairée, ou, autrement dit, auto-déterminée. Elle s'appuie, par ailleurs, sur une démarche expérimentale similaire à celle utilisée en théorie de la décision, théorie des jeux ou encore en économie comportementale. La SDT a pris forme à la suite d'une expérience qui a mis en avant la diminution de l'engagement volontaire des sujets pour une activité pourtant jugée intéressante s'ils ont été, dans un premier temps, incités à s'y livrer contre une rétribution, ici un dollar par énigme résolue.

La SDT opte pour une approche différenciée de la motivation pour permettre la mise en évidence de facteurs déterminants relatifs aux caractéristiques de l'agent, du contexte et de l'action. Dans cette optique de différenciation, deux types de motivation sont distingués (Ryan & Deci, 2000) :

- la motivation intrinsèque caractérise un engagement libre de l'individu pour ce qu'une activité lui apporte en elle-même : plaisir, satisfaction, détente, etc. ;
- la motivation extrinsèque caractérise un engagement incité par des facteurs externes à l'activité elle-même, comme les récompenses, la pression sociale, la compétition, etc.

La raison pour laquelle la théorie de la motivation suscite autant d'intérêt est avant tout sa principale conséquence pratique : la motivation est productive. Que ce soit la production de biens physiques ou intellectuels, la motivation est le précurseur de l'action, de sa direction et de son intensité. La possibilité de cibler et comprendre les facteurs qui permettent de la faire varier devient alors vitale. La démarche empirique empruntée par la SDT conduit d'abord à la mise en évidence de facteurs psychologiques innés (Ryan & Deci, 2000) tels que :

- le besoin de compétence en adéquation avec celui attendu ;
- les interactions sociales ;
- l'autonomie.

La présence de ces facteurs est nécessaire pour atteindre un état propice à la maximisation de la motivation, i.e., "la libre disposition de soi" (traduction alternative de *self-determination*).

### 2.5.2 Deux types de motivations

Comme vu précédemment, la SDT opte pour une approche différenciée de la motivation pour permettre la mise en évidence de facteurs déterminants relatifs aux caractéristiques de l'agent, du contexte et de l'action. Dans cette optique de différenciation, deux types de motivation sont distingués.

## Motivation intrinsèque

En 1985, la théorie de l'évaluation cognitive (CET ou *Cognitive Evaluation Theory*) est présentée par Ryan et Deci. Celle-ci s'intéresse exclusivement aux facteurs expliquant la variabilité de la motivation intrinsèque et aux besoins fondamentaux d'autonomie et de compétence.

- Le premier facteur relève de la présence d'événements contextuels sociaux qui, à travers leurs différentes formes – feedback, communication, récompense –, peuvent conduire au sentiment d'être compétent, ce qui favorise la motivation intrinsèque. Inversement, un feedback négatif sur la performance conduit à un état de démotivation.
- Par ailleurs, les deadlines, les menaces, les directives, les objectifs imposés et la pression managériale diminuent la motivation intrinsèque. Ceci s'explique par le fait que ce sont des incitations externes – comme peuvent l'être des récompenses matérielles – et qui, donc, relèvent davantage de la motivation extrinsèque.

Cependant, le sentiment de performance adéquat à la difficulté (compétence) ne permet pas nécessairement d'améliorer la motivation intrinsèque en l'absence d'autonomie ou, autrement dit, s'il n'est pas relié au locus de causalité. L'autonomie est révélée par la connaissance de ses centres d'intérêt, valeurs, désirs, et l'opportunité de choisir soi-même la direction de son action (choix du domaine d'activité, par exemple).

Enfin, les interactions sociales semblent également avoir un rôle important pour permettre à la motivation intrinsèque de s'exprimer. Notamment, le sentiment d'être soutenu par son entourage plutôt que d'être ignoré ou dénigré a un lien direct avec la réussite ou l'échec d'une action. Bien sûr, un certain nombre d'actions sont effectuées plus efficacement de manière isolée; cependant, même dans ce cas, des rapports sociaux sécurisants semblent avoir un effet positif sur la motivation intrinsèque.

## Motivation extrinsèque

Bien que la motivation intrinsèque soit considérée comme la plus performante, une grande partie des actions que nous devons effectuer ne sont pas motivées intrinsèquement. Typiquement, les étudiants n'apprécient pas nécessairement tous les cours et, de toute façon, n'ont pas nécessairement une connaissance précise de leurs centres d'intérêt et de ce qui est bon pour leur développement personnel. Par ailleurs, un certain nombre de personnes ne travaillent pas par passion ou par choix, mais par obligation ou pour des raisons strictement pécuniaires. Il devient alors intéressant, notamment pour les entreprises, d'étudier les différentes formes de motivation extrinsèque des agents et donc, en agissant sur celles-ci, de chercher à augmenter la productivité des collaborateurs.

### 2.5.3 Régulation et auto-détermination

Dans le cadre de la SDT, la motivation intrinsèque est considérée comme étant plus productive que la motivation extrinsèque. Il devient alors intéressant de tenter

de parvenir à comprendre ce qui peut permettre de se rapprocher de cette motivation intrinsèque. D'après la SDT, un individu qui n'est pas motivé intrinsèquement peut cependant devenir auto-déterminé, un état proche de la motivation intrinsèque ; la SDT étudie notamment les influences de l'environnement social sur le processus le permettant. Deux facteurs essentiels sont identifiés :

- l'internalisation, ou processus permettant à l'individu de se sentir concerné par une valeur ou un objectif qui n'est pas le sien ;
- l'intégration, ou processus par lequel l'individu s'approprie une valeur ou un objectif au point qu'il devienne sien.

La théorie de l'intégration organismique (OIT) est une sous-théorie de la SDT qui s'intéresse aux différentes formes de motivation découlant du processus d'intégration et d'internalisation décrit précédemment.

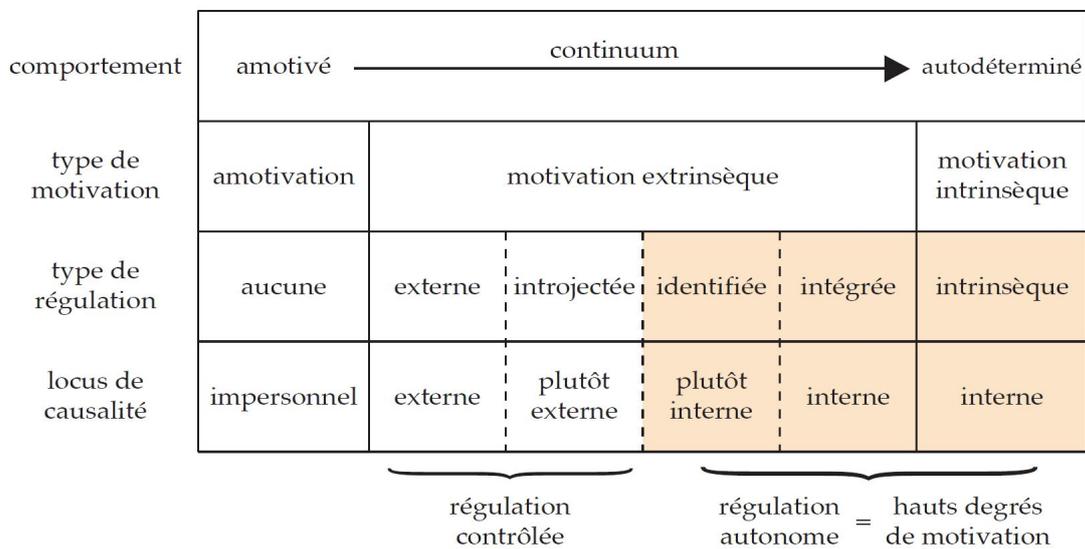


FIGURE 2.3 – Continuum de la motivation (Denis & Jouvelot, 2005)

Un des raffinements du modèle de Ryan et Deci propose de considérer plusieurs niveaux de motivation extrinsèque, caractérisés par des régulations différentes du comportement (voir Figure 2.3).

- La régulation externe désigne une action dirigée vers l'obtention d'une récompense ou l'évitement d'une punition.
- La régulation introjectée apparaît en réaction à une pression sociale pour éviter la culpabilité ou l'anxiété, ou pour satisfaire son ego, recevoir des compliments, etc.
- La régulation identifiée est atteinte quand l'individu internalise l'activité comme étant importante pour son développement ;
- Enfin, la régulation intégrée est révélée quand l'activité est complètement intégrée par l'individu, ce qui se traduit par une adéquation entre l'activité effectuée et les valeurs, les besoins et la philosophie de vie de l'individu.

En parallèle, certaines études ont suggéré qu’il pouvait être pertinent de combiner les régulations externe et introjectée pour former une composante contrôlée de motivation.

On pourrait imaginer combiner la régulation identifiée, intégrée et intrinsèque pour créer une forme mixte de motivation plus adaptée à la diversité que peut prendre l’activité professionnelle d’une personne au sein d’une organisation, qui peut comporter, en pratique, des activités intrinsèquement régulées et d’autres, extrinsèquement.

Pour faciliter l’internalisation et l’intégration, les interactions sociales et le rapport aux autres priment, contrairement à la motivation intrinsèque où ce sont les compétences et l’autonomie qui jouent le premier rôle. La première explication pour laquelle un individu internalise une activité à laquelle il participe pour des raisons initialement externes est le sentiment que celle-ci est perçue comme efficace et utile par les personnes de son entourage. En conséquence, un niveau de compétence inadéquat dans le cadre de son environnement social entraîne une perte de sens de son activité, et donc une dé-motivation. Donner accès à l’autonomie s’avère également nécessaire à l’intégration, permettant à l’individu de comprendre le sens de son activité et ainsi de la synthétiser avec ses propres objectifs et valeurs.

Pour finir, la SDT préconise, pour maximiser la motivation d’un individu, de se tourner d’abord vers le contexte social dans lequel il est placé et, ensuite, de prendre en compte son environnement de développement afin de déterminer l’encadrement adéquat de l’autonomie, des compétences et de ses interactions.

## 2.6 Gamification

La gamification est une méthode qui consiste à incorporer des mécanismes et concepts issus des jeux vidéo dans d’autres domaines comme le travail. La gamification fait partie des méthodes permettant de passer d’un système d’information dit “utilitariste” à un système d’information “motivationnel”, dans lequel les facteurs psychologiques liés à la motivation des individus ont été pris en compte dans les phases de design et d’implantation du SI afin de susciter l’adhésion des utilisateurs.

De nombreux articles analysent les effets de la gamification, ses limites, et les moyens de la mettre en place (voir, par exemple, ([Hamari et al., 2014](#))). Ce qui est intéressant avec la gamification dans le cadre du sujet de cette thèse, c’est la manière dont elle peut être utilisée pour améliorer le processus de partage de connaissance et de collecte de l’information par des acteurs de l’entreprise ([Friedrich et al., 2020](#)).

### 2.6.1 Système d’information motivationnel

Historiquement, on a pu considérer que deux types de système d’information existaient ; les système d’information “utilitaristes” et les système d’information “hédoniques”. La fonction du premier, le système d’information utilitariste, est liée à la productivité et a généralement pour objectif d’améliorer l’efficacité d’une tâche donnée. L’utilisation du système a donc par nature une utilité externe à l’utilisation même du logiciel. Par exemple, un Progiciel de gestion intégré a pour unique objectif d’améliorer la gestion d’une organisation. À l’inverse, un système d’information

hédonique a pour fonction principale de créer un loisir ou un divertissement pour l'utilisateur. Son utilité est donc liée directement au plaisir que procure l'utilisation même du système.

Le terme de système d'information motivationnel, ou système d'information mixte, a été introduit pour la première fois par les travaux de Koivisto et Hamari (Koivisto et Hamari (2019)). Comme l'indique son deuxième nom, ce type de système est apparu « plus tard » par la convergence entre les systèmes d'information utilitaristes et hédoniques. L'objectif est généralement d'améliorer l'efficacité d'une tâche en améliorant le plaisir issu de l'exécution de la partie de son processus impliquant l'utilisation du système d'information motivationnel. On retrouve comme exemples ici des systèmes tels que LinkedIn, Stack Overflow ou Volkswagen Group Loyalty Program. Dans ce cas, on peut parler de systèmes à l'origine utilitariste empruntant des concepts au design de SI hédonique. À l'inverse, on peut également observer l'apparition de systèmes hédoniques convergeant en intégrant un objectif externe au système. On parlera ici essentiellement de « serious games » comme par exemple Borderland Sciences, une des plus grandes réussites de ce genre qui a pour objectif le séquençage d'ADN de bactéries issues d'excréments humains.

Dans les deux cas, la gamification est présentée comme une méthode permettant de motiver intrinsèquement un individu à utiliser un système d'information. En effet, Koivisto et Hamari tentent à travers leurs recherches de faire un rapprochement entre les différents types de systèmes d'information et la théorie de la motivation de Ryan et Deci (Ryan & Deci, 2000). Ils font notamment l'hypothèse que l'utilisation des systèmes d'information utilitaristes implique l'utilisation de la motivation extrinsèque chez l'utilisateur ; à l'inverse, les systèmes d'information hédoniques découleraient de la motivation intrinsèque de l'utilisateur, même si ce n'est pas exclusivement. Les systèmes d'information motivationnels étant la convergence de ceux-ci, ils impliquent simultanément les motivations intrinsèque et extrinsèque de l'utilisateur. Maximiser cette motivation avec des méthodes de régulation de la motivation pourrait permettre de maximiser l'efficacité d'utilisation du système d'information (Hamari et al., 2014)(Ryan & Deci, 2000).

Ainsi, bien que la méthode de conception de systèmes d'information motivationnels initialement proposé par Koivisto et Hamari soit la gamification, des méthodes appartenant à n'importe quel domaine peuvent également être utilisées, pourvu qu'elles permettent d'améliorer la motivation de l'utilisateur du système.

## 2.6.2 Usages de la gamification

La gamification a très largement été étudiée et testée pour la conception de système d'information motivationnels, et son efficacité a été évaluée par de nombreuses recherches. Les travaux étudiant les différentes méthodes de gamification utilisées, ainsi que leur efficacité dans divers contextes expérimentaux, ont été référencés en grande partie dans l'article publié en 2014 par Hamari et al (Hamari et al., 2014).

Ils mettent notamment en avant les effets psychologiques, les comportementaux attendus et les mécanismes motivationnels utilisés, tout en soulignant l'importance des facteurs contextuels tels que l'environnement social et le caractère volontaire de l'accomplissement de la tâche. Finalement, ils présentent une variété de mécanismes

de gamification précis, testés dans les études empiriques, tels que des mécaniques de cumul de points, de classements, de badges, ainsi que toutes les variantes les plus courantes (figure 2.1).

Contexte	Mécanisme
Commerce	Points
Education et apprentissage	Classement
Santé et sport	Badges et succès
Système intra-organisationnel	Niveaux
Partage	Histoire et thème
Développement durable	Objectif
Travail	Retour d'expérience
Innovation	Récompenses
Collecte de données	Progrès
	Défi

TABLE 2.1 – Contextes expérimentaux de la gamification (Hamari et al., 2014)

Le nombre de contextes dans lesquels la gamification a été expérimentée met en avant le large spectre d'application que peuvent avoir ces méthodes. En l'occurrence la figure 2.1 référence la plupart des contextes expérimentaux de la gamification, et nous pouvons observer parmi ceux-ci la collecte de données et le partage, qui seront sollicités dans le cadre de cette thèse.

### 2.6.3 Incitation au partage d'informations et de connaissance

Un des objectifs clés des systèmes d'information est d'objectiver la somme des connaissances formelles et informelles associées à toute organisation. Le domaine du management de la connaissance (*Knowledge Management*, ou KM), au coeur de toutes les organisations, comprend le processus de création, de stockage, de transfert et d'application de la connaissance dans une entreprise. D'un point de vue technologique, il s'appuie sur un système de management de la connaissance (KMS), qui est un système d'information qui fournit un support technique aux employés dans le processus de partage de connaissance (*Knowledge Sharing*, ou KS).

Néanmoins, un KMS ne suffit pas à lui seul à permettre un partage de connaissance efficace. De nombreuses barrières peuvent freiner ou même annuler son efficacité. Inversement, des incitations ou mécanismes motivationnels peuvent être mis en place pour améliorer son efficacité. Le sujet de cette thèse est en partie lié au partage de connaissance, et au système de management de la connaissance, ou des informations, des entreprises. Et connaître les barrières qui s'opposent à l'implémentation efficace de tels systèmes, mais également les moyens existants permettant d'encourager le partage de l'information et de la connaissance, s'avère utile.

La première barrière peut être une culture de l'entreprise inappropriée vis-à-vis du partage de connaissance. Un exemple est la culture du "knowledge is power" qui peut non seulement pousser les employés à ne pas partager leurs connaissances, mais aussi pousser des individus à empêcher les autres employés d'accéder à la connais-

sance. D'autres cultures d'entreprises peuvent être néfastes, comme une culture de compétition entre les employés plutôt que d'entraide.

La deuxième barrière est le manque de compréhension de la pertinence du partage de connaissances, ou encore la sous-estimation de la valeur de sa propre connaissance vis-à-vis des autres. Enfin, on peut citer la privatisation intra-organisationnelle de certaines connaissances, réservées uniquement à une population restreinte. Un accès facile et gratuit à la connaissance pour toute l'entreprise est évidemment favorable au partage de connaissance. Inversement, les sous-groupes se réservant l'accès à une partie de la connaissance ont un impact négatif sur celui-ci.

La motivation des acteurs à partager la connaissance, elle-même dépendante de nombreux facteurs, a une forte corrélation avec le partage de connaissances. Elle est notamment décisive en ce qui concerne la qualité des connaissances et informations partagées.

Les études identifient un large spectre de mécanismes motivationnels affectant le KS allant de l'altruisme à des aspects sociaux tels que la réputation, la reconnaissance ou des facteurs externes tels que des récompenses. La réciprocité, par exemple, serait particulièrement efficace en déclenchant des sentiments tels que la gratitude, l'obligation ou la confiance. L'ensemble des motivations peuvent être identifiés et analysés à travers le spectre de la motivation de Ryan et Deci (Ryan & Deci, 2000). Ce spectre de la motivation est traduit en un framework de mise en place d'incitations permettant d'identifier le type d'incitations mis en place et son effet sur les individus. Tout système d'incitation sera alors considéré comme étant la somme de toutes les incitations externes et internes (intrinsèques et extrinsèques) qui renforcent les comportements désirés (objectifs de l'entreprise) et diminuent les comportements non désirés.

TABLE 2.2 – Mécanismes de gamification (Friedrich et al., 2020)

Motivation pour le partage	Mécanisme de gamification
Altruisme	Feedback
Contribuer au succès de l'entreprise	Feedback, graphiques de performance
Fun/amusement	Challenge, feedback, compétition
Visibilité des accomplissements	Feedback, graphiques de performance
Réciprocité	Feedback, récompenses
Communauté/participation	Feedback, statut
Réputation	Feedback, statut, récompense
Démonstration de compétence	Graphiques de performance, statut
Reconnaissance	Feedback, récompenses, statut

Ainsi, pour chaque type de motivation au partage d'information et de connaissances, des mécanismes d'incitation, ou de gamification, peuvent être identifiés, tel que cela a été fait de manière théorique par Friedrich et al. (2020), et est représenté ici dans le tableau 2.2

## Chapitre 3

# Modèle *Data Binding and Reification*

Au fil du temps, les chercheurs ont élaboré divers cadres pour évaluer la qualité des données et de l'information, en particulier dans le contexte des systèmes d'information (SI) (R. Y. Wang & Strong, 1996), (R. Wang et al., 1995). Néanmoins, ces tentatives sont souvent entravées par un amalgame fait entre la qualité des données et la qualité de l'information, deux concepts qui sont généralement utilisés de manière interchangeable, conduisant à des ambiguïtés et des malentendus.

En effet, l'absence de consensus quant à la définition précise de "donnée" et "information" et leur relation complexe et enchevêtré accentue ces confusions. Cette situation a des répercussions directes sur l'appréciation de leur rôle respectif dans l'efficacité d'un système de management de l'information, ralentissant ainsi le progrès dans ce domaine et soulignant la nécessité de recherches complémentaires pour clarifier ces concepts.

C'est pourquoi, dans ce chapitre, nous avons pris l'initiative de fournir une définition que nous espérons claire des concepts de donnée et d'information et d'examiner leur relation, notamment dans le cadre de réflexions autour des SI. Nous nous reposerons ensuite sur le domaine des sciences cognitives, et notamment celui des neurosciences, pour approfondir notre compréhension de cette relation complexe entre donnée et information, régie en grande partie par l'intervention humaine. Cet apport nous permettra notamment de définir le processus de transformation des données en informations, et inversement, que nous nommerons, respectivement, binding et réification et qui seront les piliers du modèle *Data binding and reification* (DBR). Ce qui rend ce modèle innovant mais surtout adapté à la problématique organisationnelle de ce mémoire, c'est qu'il donne à l'homme une place centrale dans la conceptualisation de ces processus. En conséquence, DBR nous permet d'intégrer aux mécanismes qui régissent la création d'informations et de données la possibilité que les limitations cognitives de l'homme, et l'utilisation de biais cognitifs et heuristiques par celui-ci, puissent avoir un effet sur celle-ci.

Cette compréhension plus poussée des concepts de donnée et d'information, et de leurs processus de transformation, a permis la formalisation d'un cadre conceptuel "unifié" de la qualité des données et de l'information inspiré des travaux de Wang (R. Y. Wang & Strong, 1996), mais marquant, cette fois, une démarcation

que nous voulons claire entre ces deux concepts.

### 3.1 Donnée et information

Dans l'ensemble, la littérature suggère que la qualité des données et la qualité de l'information sont, en fait, deux concepts distincts et que tous deux sont des antécédents importants de la réussite de la Business Intelligence and Analytics (BI & A) (Ariyaratna & Peter, 2019), (Ewen, Armstrong, Parker, & Seddon, 2012), (Zins, 2007). Plusieurs études ont tenté de proposer des cadres permettant de caractériser la qualité de la donnée ou de l'information dans un contexte de SI (R. Y. Wang & Strong, 1996), (R. Wang et al., 1995), (Kahn et al., 2002), (Delone & Mclean, 2003), mais, dans la plupart de ces articles, les auteurs semblent largement confondre qualité des données et qualité de l'information en utilisant l'un de ces deux termes pour décrire la même chose ou en les utilisant de manière interchangeable.

Cette confusion est due, selon nous, notamment au manque de consensus pour définir les concepts de donnée et d'information. En effet, le manque de cohérence dans la terminologie et la définition des concepts obscurcit le rôle relatif joué par chacun d'entre eux dans le succès de la BI& A, créant ainsi le besoin d'une recherche supplémentaire pour spécifier le rôle de ces concepts plus clairement (Torres & Sidorova, 2019).

Une des contributions présentées dans ce chapitre est de tenter de distinguer les notions de donnée et d'information, et de caractériser et comprendre la relation qui unit ces deux concepts. Pour simplifier la compréhension de ce qui suit, nous introduisons dès maintenant les définitions utilisées dans cette étude de ces deux concepts, ce qui devrait permettre d'avoir une intuition sur la différence entre ces deux concepts. Le cheminement suivi et les sources utilisées pour en venir à ces définitions sont détaillés par la suite.

- La donnée est ce qui est connu, déterminé, ou déclaré ; en informatique, il s'agit là de la représentation codée de ce qui est connu, déterminé, ou déclaré. Par exemple, de tels données peuvent être des mesures de température provenant de capteurs physiques, des résultats d'enquête effectuée sur un échantillon de population, des noms de personnes, des adresses ou encore des numéros d'identification INSEE.
- L'information est l'interprétation, ou encore la signification, donnée par l'homme aux données. L'information est le résultat de l'analyse, de la compréhension et de la contextualisation d'une ou de plusieurs données. Par exemple, les données brutes d'une enquête sur la satisfaction des clients peuvent être transformées en informations significatives sur les tendances de la satisfaction des clients et les domaines à améliorer.

Ces définitions illustrent l'approche clairement cognitive prise ici, car la donnée acquiert le statut d'information seulement en se rapportant à la "signification et à l'intention humaine" (Zins, 2007).

### 3.1.1 Étymologie

L'étude étymologique de ces deux mots peut permettre de mieux comprendre ce qui, historiquement, a pu causer le manque de robustesse dans la terminologie et les définitions de ces concepts.

#### Information

Le mot “information” vient du vieux français “enformacion/informacion/information”, qui signifie “enquête criminelle, instruction et investigation”. À partir du 14e siècle, il signifie “élément informatif, data, connaissance” et, au pluriel, “action de former quelque chose ou de donner une forme à quelque chose”. À partir du 15e siècle, son pluriel signifie “ensemble de connaissances liées à un sujet précis”. Son étymon, le latin “informatiō(n)” signifie “aperçu, concept, idée”, mais, également, “enseignement” (Harper, 2023)(Oxford English Dictionary, 2023).

Étymologiquement, le mot information est ainsi très lié à l'homme, ce qui rejoint les définitions de l'information centrées sur l'homme, telles que celles fondées sur les neurosciences. Que ce soit une idée, un concept, une investigation ou l'action de donner une forme à quelque chose, l'information est ici le fruit d'une transformation par l'homme de quelque chose, comme une donnée telle que définie précédemment. (Oxford English Dictionary, 2023) (Zins, 2007)

Finalement, on remarque que, paradoxalement, l'information n'est que peu utilisée dans le cadre de la communication, alors qu'aujourd'hui, et notamment à partir de l'introduction de la théorie de l'information en 1950, elle lui est très liée. La raison se trouve dans le fait que l'évolution du mot “information” a été fortement influencée par son association au mot “informer”, qui a une signification beaucoup plus axée sur la communication (Zins, 2007).

#### Donnée

Le mot latin “datum” signifie “chose donnée” et sera repris en 1640 par la langue anglaise pour signifier, alors, “un fait donné ou déterminé”. Le mot “data” n'est, lui, autre que le pluriel du mot “datum”, qui signifie alors “de multiples faits” (Harper, 2023).

Dans le langage commun, cette distinction de nombre a pratiquement disparu. L'utilisation du mot “data” (en anglais) comme nom commun singulier est largement répandue et s'est étendue à tous les domaines, sauf celui de la médecine (McAlister, 2016).

Il est également intéressant de noter qu'étymologiquement, le mot “data” n'était pas lié à l'informatique, et que c'est à partir de l'invention de l'ordinateur, en 1946, que le mot “data” est utilisé pour la première fois pour signifier “une information transmissible et stockable dans un ordinateur” (Harper, 2023). Le mot “data” sera, par la suite, à nouveau emprunté pour définir de nombreux concepts en informatique, tels que *data processing*, en 1954, ou *database*, en 1962.

La disparition du pluriel dans “data” dans la langue commune, ainsi son utilisation dans le domaine de l'informatique a posteriori pour définir un bon nombre de

concepts informatiques, a probablement participé au manque de cohérence observé aujourd’hui dans la terminologie et la définition de ce concept.

### 3.1.2 Choix des définitions

Le choix ou la proposition d’une définition dépend essentiellement de son adéquation avec le sujet et le domaine d’études, mais également, de manière presque inévitable, des croyances personnelles de chacun. En l’occurrence, le sujet traité ici porte sur les conséquences de l’interaction homme-machine sur la qualité des données et des informations. Parmi les nombreuses approches existantes, lorsqu’il s’agit de définir l’information, celle appelée approche centrée sur l’homme, fondée sur les sciences cognitives, est celle qui semble la plus appropriée pour étudier ce type d’interaction.

En effet, comme cela a été vu en détail dans le chapitre consacré au contexte théorique, parmi les approches centrées sur l’homme, il existe une opposition entre l’approche cognitive qui est exclusive à l’homme et celle qui ne l’est pas (Zins, 2007). Ici, c’est l’approche cognitive qui est retenue, car elle non seulement semble la plus appropriée pour le sujet, mais est également la plus alignée avec le reste du contexte théorique sollicité dans le reste du document. Par ailleurs dans le domaine des sciences de l’information, c’est l’approche cognitive la plus fréquemment adoptée, raison pour laquelle, c’est celle qui a été choisie par la quasi-totalité des 57 chercheurs de ce domaine lors de l’étude réalisée par Zins (Zins, 2007).

De manière simplifiée, dans l’approche cognitive, la donnée est ce qui est connu (“donné”) ou déterminé, et l’information est le fruit de l’interprétation humaine, liée donc à la signification et/ou à l’intention humaine. Cette définition est fortement inspirée de celle proposée par Albrechtsen (Zins, 2007), qui distingue clairement celle utilisée dans le discours humain de celle utilisée dans les systèmes informatiques, ce qui fait écho à l’analyse précédente de l’étymologie du mot “data”. Cependant, bien qu’il soit nécessaire, étant donné l’évolution étymologique de ce mot, de distinguer la définition générale de la définition informatique, ce n’est pas le cas du mot information. C’est pourquoi la définition proposée dans ce document est légèrement différente de celle proposée par Hanne Albrechtsen.

Dans cette définition de la donnée et de l’information, l’accent est mis sur le principe, présent dans l’approche fondée sur les neurosciences, selon lequel la donnée acquiert le statut d’information seulement à travers l’interprétation humaine qui permet de donner un sens à la donnée. Cette manière de définir l’information présuppose que ces deux concepts sont liés et existent l’un par rapport à l’autre, ce qui est une approche partagée par de nombreux experts du domaine des sciences de l’information (Rowley, 2007). Approfondir cette relation et étudier le processus qui permet de passer de l’un à l’autre dans un cadre organisationnel est une des contributions majeures de ce manuscrit.

Ainsi, l’information issue de l’interprétation de données va dépendre, ici, en partie de l’individu. Par exemple, une donnée indiquant la température en degrés dans une pièce n’aura de sens que pour une personne qui connaît cette unité d’échelle de température, et elle n’aura de la valeur que si cette personne ne connaissait pas déjà la température de cette pièce et si cette valeur lui est pertinente, en fonction de la

sensibilité thermique de cet individu.

## 3.2 Transformation des données et de l'information

Quand il s'agit d'étudier la relation qui existe entre la donnée, l'information et la connaissance, de nombreux théoriciens, que ce soit en informatique, en sciences de l'information ou en bibliothéconomie, se reposent sur la hiérarchie ou pyramide "data-information-knowledge-wisdom" (DIKW). Dans sa version originale, ce modèle conceptuel a été critiqué et reformulé à plusieurs reprises (Rowley, 2007) (Frické, 2009). Cela étant dit, la très grande majorité des travaux qui ont repris ou utilisé le DIKW ont presque systématiquement exclu le "Wisdom", car, pour beaucoup, celui-ci n'est pas nécessaire pour résoudre les problèmes qu'ils adressent ou n'a tout simplement pas sa place dans un tel modèle (Frické, 2009), devenant donc la hiérarchie simplifiée DIK.

Dans la hiérarchie DIK, auquel nous nous limitons ici, la donnée prévaut à l'information, et l'information à la connaissance. Les définitions de la donnée et de l'information sont très proches de celles proposées dans ce document, faisant une différence fonctionnelle entre la donnée et l'information, la donnée devenant une information à partir du moment où elle est utilisée pour répondre à une question. La connaissance, elle, est considérée comme une compétence, le "knowhow", et permet notamment de transformer les informations en instructions. Les définitions proposées précédemment des concepts de donnée et d'information, choisies par les experts en sciences de l'information (Zins, 2007), partagent également cette différence fonctionnelle. Nous allons, ci-dessous, proposer une version raffinée de ces notions, plus à même de modéliser les processus informationnels présents dans les organisations.

### 3.2.1 Définition des concepts de *binding* et de réification

Il existe un consensus sur le fait que les notions de donnée, d'information et de connaissance doivent être définies l'une par rapport à l'autre. Ce consensus réaffirme la validité d'un modèle conceptuel qui relie les concepts de donnée, d'information et de connaissance (Rowley, 2007)

Il y a cependant moins d'accord sur la nature des processus qui convertissent les données en informations, et les informations en connaissances, au point qu'il n'est pas clair s'il y a en fait trois concepts distincts (Rowley, 2007). Cet enchevêtrement des concepts peut être exploré à deux niveaux - la relation entre les données et l'information, et la relation entre l'information et la connaissance.

Une des contributions apportées ici, via les notions de *binding* et de réification, est de proposer un nouveau modèle conceptuel décrivant le processus de transformation de la donnée en information et de l'information en connaissance, modèle dont notre étude montre la pertinence dans un cadre organisationnel.

#### *Binding*

Que ce soit dans l'approche fondée sur les sciences cognitives (Zins, 2007) ou celle de la théorie linguistique de Chomsky (Blitman, 2020), le passage de la donnée à

l'information est caractérisé par le moment où intervient l'interprétation de l'homme. Tant que celle-ci n'est pas présente, les données ne sont qu'un ensemble d'invariants<sup>1</sup>, ou leurs représentations codées, c'est à dire un ensemble de symboles respectant une certaine forme lexicale et syntaxique.

Au moment de l'interprétation d'une donnée, celle-ci se lie à une multitude d'autres éléments tels que les connaissances préexistantes, le contexte, les émotions ou encore la motivation en ce que l'on appelle une expérience subjective unique. Ce processus, qui consiste à prendre comme base de raisonnement et de prise de décision une expérience subjective unique plutôt que l'ensemble de toutes les informations disponibles de manière parallèle et indifférenciée, est appelé en neurosciences le 'binding'<sup>2</sup> et semble approprié pour désigner le processus qui permet de passer de données à information (Feldman, 2013).

Ainsi, quand un individu accède à une donnée, celle-ci devient une information dont la perception, et donc l'expérience subjective, va dépendre de facteurs généraux, parce qu'il existe un certain nombre de règles et de contraintes communes à tous les membres d'une même espèce, et d'autres, individuels, différents d'un individu à un autre. Ainsi, bien que la plupart des êtres humains sont capables de voir et reconnaître un oiseau, un individu qui a peur des oiseaux pourrait le percevoir de manière différente, le voir plus grand, ou plus dangereux.

## Réification

Si une donnée devient une information quand elle est interprétée par un humain, qu'en est-il du processus inverse? Une donnée au sens informatique provient soit d'un appareil de mesure, soit d'un être humain qui la saisit sur une interface homme-machine telle qu'un ordinateur. Comme nous l'avons vu, l'homme raisonne et agit sur la base de ses expériences subjectives uniques, issues du processus de binding, lequel sera donc l'élément abstrait sur la base duquel un individu saisit une donnée. Une donnée au sens informatique étant la représentation codée d'invariants, il s'agit donc de donner des caractéristiques statiques à cette expérience subjective afin de la saisir en tant que donnée. Ce processus qui consiste à transformer un élément abstrait en élément concret s'appelle la réification, qui, en philosophie, désigne "le processus par lequel on transforme quelque chose de mouvant, de dynamique en être fixe, statique" (Larousse, s. d.-b) Ainsi, dans le cadre d'une organisation du type de l'entreprise ALTEN, un chef de projet va, sur la base d'un ensemble de KPI, de retours clients et d'échanges avec les équipes et différents responsables, développer une expérience subjective qui se traduira, par exemple, par une opinion vaguement positive d'un projet. Il aura ensuite la possibilité, voire l'obligation, de stocker cette expérience subjective dans une base de données, par exemple en lui donnant une note de 1 à 10 ou en lui attribuant une couleur, qui pourrait être ici de 7 ou vert, par exemple.

Étant donné que, pour agir et prendre des décisions, l'humain passe nécessairement par un processus de binding (Zimmer et al., 2006) et que saisir une donnée est

---

1. Le terme "invariant" désigne ce qui demeure constant, identique à soi-même dans une structure, un système (Larousse, s. d.-a).

2. Nous avons choisi de garder le terme anglais pour cette notion et omettons par la suite de signifier qu'il s'agit d'un terme non français en utilisant la graphie usuelle des mots.

une action, qui d’ailleurs implique souvent une décision, la réification ne va jamais sans le binding. La saisie d’une donnée, quelle qu’elle soit, par un individu passe donc, fondamentalement, par un processus double de binding/réification.

### 3.2.2 Facteurs de perception

Le fonctionnement du binding étant logiquement commun à tous les membres de l’espèce, il peut donc être considéré comme un facteur de perception général. Bien qu’il n’existe pas de modélisation robuste et universelle du processus de binding, certains facteurs ayant un rôle prépondérant dans un processus de binding peuvent être identifiés (Zimmer et al., 2006), (Feldman, 2013), (Pally & Olds, 2018).

- L’évènement cognitif va conditionner la séparation des informations relatives à l’évènement cognitif perçu de celles qui ne le sont pas. De manière similaire, l’action envisagée va orienter le rattachement des caractéristiques pertinentes voire nécessaires à l’action envisagée.
- Les connaissances préexistantes sont liées à l’évènement cognitif et à l’action envisagée, puisque le cerveau va sélectionner les informations, parmi les connaissances préexistantes, les plus pertinentes pour effectuer l’action, mais également celles qui sont le plus fortement liées à l’évènement cognitif vécu. Cette forme de binding est appelée “binding relationnel” (Zimmer et al., 2006).
- Parce que l’attention est considérée comme la “glue” permettant la liaison entre les différentes caractéristiques impliquées dans l’évènement, elle est souvent considérée comme une précondition nécessaire au binding.
- Le contexte de l’évènement cognitif joue aussi un rôle important, notamment dans la mesure où le cerveau doit être capable, dans une certaine mesure, de distinguer l’objet de l’évènement cognitif de son contexte. En fonction de la tâche à effectuer, le contexte peut faire partie partiellement, intégralement, ou pas du tout de l’objet ; et plus celui-ci doit être séparé de l’objet, plus le cerveau doit être capable de le distinguer.

L’évènement cognitif, l’action envisagée, le binding relationnel, l’attention et le contexte peuvent être considérés comme des facteurs de perception généraux, puisqu’ils ne dépendent pas directement de l’individu<sup>3</sup> et sont inhérents au processus de binding. Il faut toutefois préciser que le binding relationnel étant un fonctionnement commun à tous les individus, il s’agit là d’un facteur général, mais que la connaissance pré-existante en elle-même étant liée à l’individu, elle peut donc être considérée comme un facteur de perception individuel.

D’autres facteurs de perception tels que la mémoire, la motivation et les émotions peuvent expliquer en partie que des individus différents puissent avoir une perception différente d’un même évènement cognitif, et sont considérés comme des facteurs individuels<sup>4</sup>.

---

3. On suppose ici l’absence de pathologie cognitive chez l’individu, ainsi qu’une uniformité culturelle (ou alors on se doit d’envisager la culture même comme faisant partie du contexte de l’évènement cognitif).

4. Cette liste n’est, bien sûr, pas exhaustive puisqu’à nouveau, il n’existe pas de modélisation robuste du processus de binding (Pally & Olds, 2018).

- La mémoire doit être prise en compte, car, pour minimiser les efforts de perception, le cerveau compare les expériences précédentes à celles en cours pour essayer d'en trouver une suffisamment similaire et simplement compléter le manque d'informations à disposition pour avoir une perception complète de l'évènement cognitif sans avoir à traiter l'ensemble de l'évènement à nouveaux frais. Ainsi, la perception des individus est fortement biaisée par leurs mémoires des expériences passées.
- En ce qui concerne la motivation, il est plus probable de percevoir les éléments de l'évènement qui sont pertinents pour répondre à des besoins propres que ceux qui ne le sont pas. La motivation peut également avoir une influence sur l'attention, qui est un facteur primordial du processus de perception.
- Les émotions, enfin, influencent comment les stimuli sont interprétés ; par exemple, la présence de peur et d'anxiété chez l'individu augmente la tendance à interpréter un stimulus quelconque comme dangereux. Par ailleurs, les individus ayant vécu des traumatismes peuvent avoir tendance à réactiver l'état mental dans lequel ils étaient au moment du traumatisme, mêlant mémoire et émotion (Pally & Olds, 2018).

### 3.2.3 Modèle DBR

Comme vu ci-dessus, la perception, et plus précisément le processus de binding, est avant tout conditionné par l'évènement cognitif, l'action envisagée, le binding relationnel, l'attention et le contexte, qui sont des facteurs généraux qui ne varient pas en fonction de l'individu. En parallèle, un certain nombre de facteurs individuels tels que les connaissances pré-existantes, la mémoire, les émotions et la motivation influencent la perception et expliquent que celle-ci puisse être différente d'un individu à l'autre (Pally & Olds, 2018).

L'évènement cognitif et l'action envisagée peuvent être prédéterminés en fonction de ce que l'on souhaite observer. Dans l'optique de caractériser le processus de conversion de la donnée en information et de l'information en donnée, l'évènement cognitif correspondant est formé de la consultation d'une donnée, l'action envisagée étant l'assimilation de l'information qui en découle, suivie, par exemple, de la réification de celle-ci en une nouvelle donnée. L'attention est, quant à elle, probablement portée sur la donnée consultée, puisque c'est une condition au binding. Cet entrelacement entre binding et réification constitue le moteur du modèle *Data Binding and Reification* (DBR), présenté dans la figure 3.1.

Il convient de noter que le processus de binding d'une donnée en information pourrait également avoir été effectué avec l'objectif de prendre une décision stratégique ou simplement de répondre à une interrogation, et non pour la création d'une nouvelle donnée comme évoqué précédemment ; le binding visant comme action envisagée la réification de l'information en données est donc ici un cas spécifique.

La conceptualisation du processus de binding et de réification représente, dans notre contexte, l'interaction d'un individu avec un système d'information (IHM) pour d'abord consulter une donnée, ce qui constitue l'évènement cognitif, pour ensuite la réifier en une nouvelle donnée, qui est ici l'action envisagée (voir la figure 3.1). Si le processus est suivi chronologiquement, l'individu consulte la don-

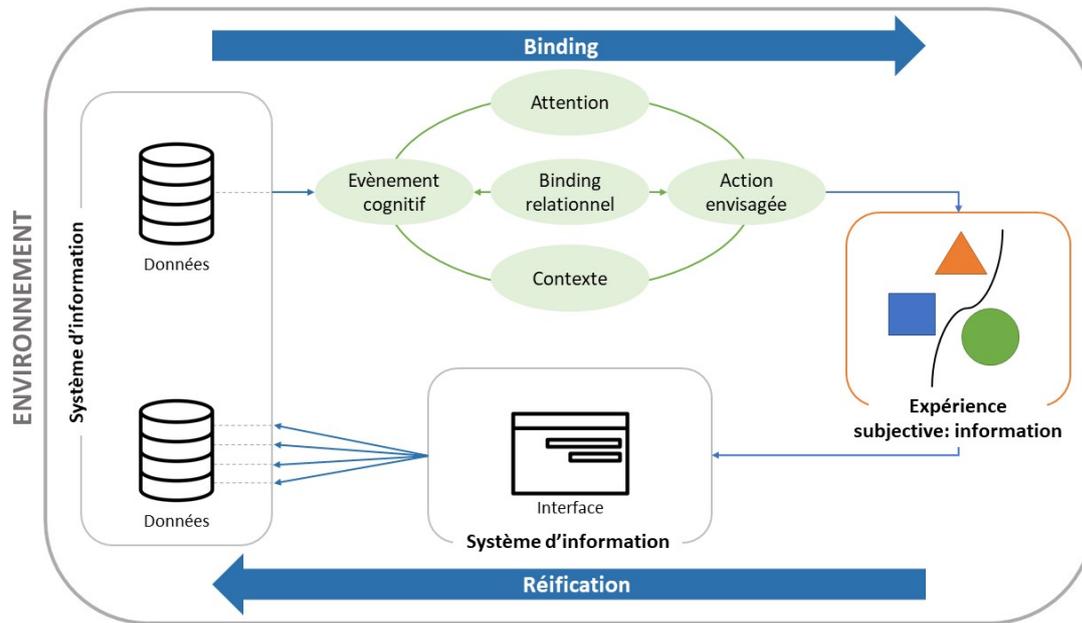


FIGURE 3.1 – Processus de binding et réification

née, son cerveau l’interprétant alors en opérant un binding, ce qui résulte en une “expérience subjective unique”. Si on se réfère à l’approche neuroscientifique, qui soutient que la donnée acquiert le statut d’information à travers l’interprétation humaine, cette expérience subjective est “l’information” (Zins, 2007). Celle-ci permet ensuite d’effectuer l’action envisagée, ici la création d’une nouvelle donnée, en opérant cette fois-ci une réification de cette information, ce qui revient à lui attribuer des caractéristiques fixes, dont la syntaxe est généralement imposée par le système d’information, notamment pour la saisie de données structurées (figure 3.1).

Dans cette figure, l’expérience subjective unique est représentée par un objet unique constitué d’un ensemble de formes abstraites n’ayant pas de signification claire (figure 3.1). Cette représentation abstraite représente notamment le fait que l’information est abstraite, et que le processus de réification permet, comme son nom l’indique, de lui attribuer des caractéristiques fixes. L’explication, en ce qui concerne le caractère unique de l’expérience subjective, réside dans le fait que, bien que le cerveau soit un système hautement distribué traitant un grand nombre d’informations simultanément, l’homme ne peut percevoir le monde que de manière unifiée. Cette particularité est appelée “unité subjective de la perception” et constitue notamment un exemple du problème corps-esprit concernant la relation entre notre expérience subjective et la fonction neuronale (Feldman, 2013).

Finalement, l’environnement, représenté dans la figure 3.1 par la bordure grise, fait référence à la structure de l’environnement dans lequel les processus de binding et de réification ont lieu. Ce que signifie précisément “environnement” et les implications de sa structure sur les processus de binding et de réification, sont décrits en détail dans les deux sous-sections suivantes, “Biais cognitifs” et “Biais dans le contexte DBR”. Celui-ci peut avoir un effet sur les facteurs individuels de perception évoqués précédemment, dont, rappelons le, la liste n’est pas exhaustive. D’ailleurs, le chapitre suivant est exclusivement consacré à l’identification de la structure de

cet environnement dans un contexte organisationnel de saisie de données et à son impact sur les processus de binding et de réification.

### 3.2.4 Biais cognitifs humains

Des définitions mêmes du binding et de la réification découlent le fait que des biais affectent très probablement, respectivement, l'information et la donnée pendant l'exécution de ces processus. Il serait même envisageable que le processus de binding soit l'une des explications, du point de vue des sciences cognitives, de l'origine des biais cognitifs liés à la perception, mais ce n'est, en l'occurrence, pas le sujet de notre attention ici. Faire le lien entre ces processus et l'apparition des biais nécessite, tout d'abord, de bien comprendre ce qu'on entend par "biais", comment on les identifie et comment on évalue leurs importances.

Le mot "biais" est utilisé dans de nombreux domaines scientifiques et contextes organisationnels et ne désigne pas toujours exactement la même chose pour chacun d'entre eux. De manière générale, un biais peut être toutefois être défini comme tout écart d'une quantité mesurée ou calculée par rapport à sa valeur réelle (vraie), de sorte que la mesure ou le calcul ne soit pas représentatif de l'élément considéré (*APA Dictionary of Psychology - Bias, s. d.*). Plus précisément, quand les biais sont liés à l'intervention de l'homme, on parle alors de biais cognitif.

Un exemple célèbre est l'illusion de l'échiquier d'Adelson, proposée en 1995 par le neuroscientifique américain Ted Adelson au MIT (voir la figure 3.2). Dans cette expérience, on vous demande d'identifier les couleurs des cases A et B et de déterminer si elles sont différentes. Surprenamment, bien que les cases A et B soient de la même couleur, la perception courante est que la case B semble plus claire que la case A. Cette différence perçue soulève la question : pourquoi voyons-nous ces cases comme étant de couleurs différentes alors que nous savons qu'elles sont identiques ?

La réponse réside dans la manière dont nous interprétons les images. Nous ne les percevons pas de manière isolée, mais en utilisant nos connaissances préalables sur le monde. Nous savons, par exemple, qu'un échiquier est composé d'une alternance de cases claires et foncées en diagonale. Ainsi, le cerveau ajuste notre perception pour correspondre à cette connaissance, uniformisant ce que nous voyons et compensant les variations de luminosité pour créer une image plus stable. Cette expérience illustre bien la différence et la relation entre les données et l'information, qui est en l'occurrence caractérisé par l'interprétation humaine. Ce que nous voyons n'est pas simplement des faits, mais plutôt des interprétations basées sur nos connaissances du monde.

Comme cela a été présenté dans le chapitre présentant le contexte théorique, le concept de biais cognitif a été proposé par Kahneman et Tversky dans le cadre d'un programme pluridisciplinaire traitant des biais et heuristiques (Wilke & Mata, 2012). Leurs travaux ont été à l'origine de la remise en question de la rationalité classique, ce qui a donné naissance à une nouvelle branche de l'économie, l'économie comportementale. Cette approche s'appuie sur le concept de rationalité limitée, qui postule que l'homme prend des décisions rationnelles compte tenu de ses ressources limitées, que ce soit l'information incomplète à sa disposition ou ses limitations cognitives pour la traiter et prendre une décision. Les biais cognitifs sont alors définis

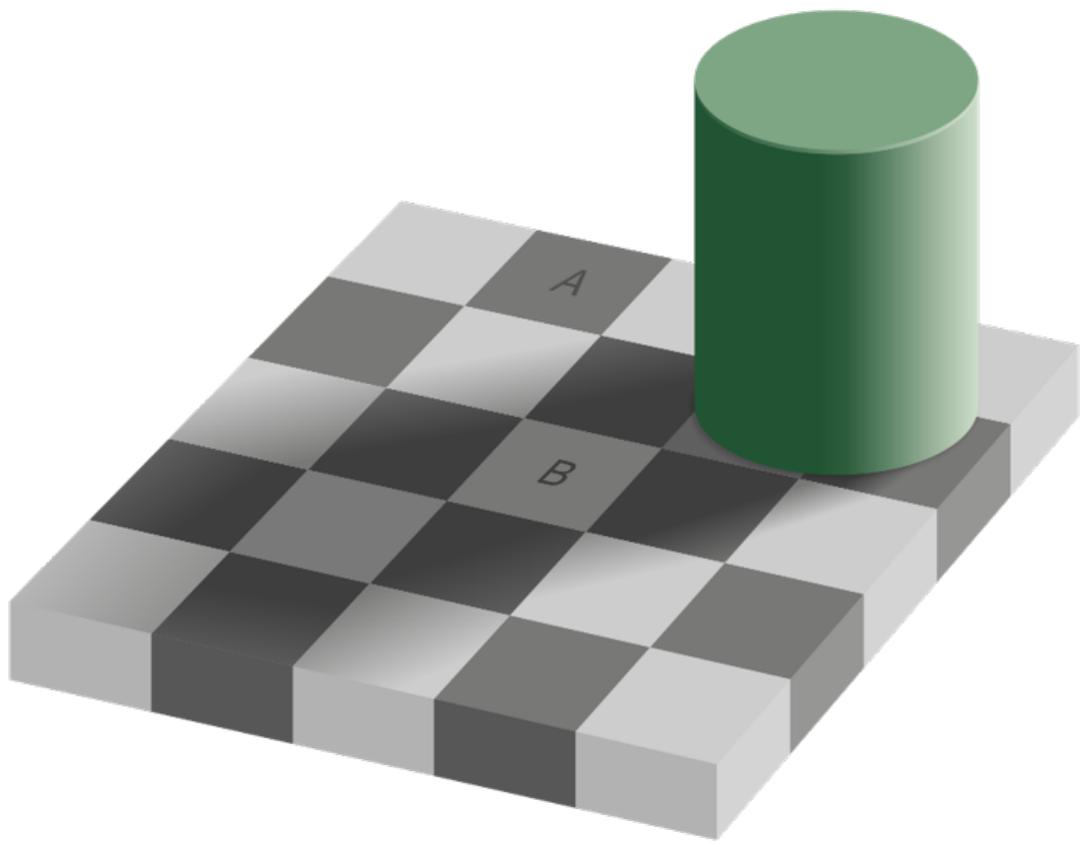


FIGURE 3.2 – Echiquier d'Adelson (1995)

comme des erreurs systématiques de jugement et de prise de décision communes à tous les êtres humains, lesquelles peuvent être dues à des limitations de ressources, entre autres cognitives.

Ce programme sur les heuristique et biais s’inspire de la recherche sur les biais perceptifs et propose que le système cognitif humain, comme le système perceptif, soit appréhendé comme conçu pour faire des déductions sur le monde extérieur en se fondant sur des indices imparfaits qui peuvent conduire à des erreurs dans certaines situations. C’est précisément ce que décrit le processus de binding que nous avons proposé dans le cadre du modèle DBR à un niveau plus fin, et c’est d’ailleurs la raison pour laquelle nous nous intéressons ici aux biais et, plus précisément, aux biais cognitifs. En effet, ce qui nous intéresse, c’est en quoi les processus de binding et de réification peuvent être à l’origine d’erreurs de jugement et de prise de décision, donc de biais, et de les identifier.

La méthode proposée par Kahneman et Tversky pour identifier et mesurer les biais a servi de base pour de nombreuses recherches futures, même pour ceux critiquant leur approche, tel que le programme “fast and frugal” ([Wilke & Mata, 2012](#)). Cette méthode peut être décrite comme suit. Tout d’abord, les participants se voient présenter un problème de raisonnement auquel correspond une réponse normative estimée comme étant la bonne réponse du point de vue statistique, en partant du principe que l’individu est rationnel. Ensuite, les réponses des participants sont comparées à la solution impliquée par ces normes, et les écarts systématiques (biais) constatés entre les réponses et les solutions normatives sont répertoriés. Enfin, ces biais sont expliqués comme étant la conséquence de l’utilisation d’heuristiques ou de biais cognitifs pour résoudre ces problèmes.

À défaut de révolutionner radicalement la méthode, de nombreuses recherches postérieures proposeront cependant d’aborder les biais et heuristiques de manière différente, en ne les considérant plus comme des défauts de conception mais davantage comme des méthodes de raisonnement permettant aux individus de résoudre efficacement des problèmes en fonction des caractéristiques de leur environnement. Cette approche s’appuie notamment sur le principe de rationalité écologique, qui a le mérite de mettre en avant l’importance des caractéristiques déterminantes de l’environnement dans le processus de définition et d’identification des biais ([Wilke & Mata, 2012](#)) ([Haselton et al., 2015](#)).

Une grande partie des travaux récents sur ce sujet suit l’approche de la psychologie évolutionnaire, qui s’appuie également sur le principe de rationalité écologique, mais en rappelant que les biais cognitifs et autres méthodes de raisonnement humain se sont développés pour fonctionner dans des environnements dans lesquels l’homme a évolué, du passé donc, et peuvent donc s’avérer inefficaces dans des environnements qui ne sont pas “naturels”.

### 3.2.5 Biais dans le modèle DBR

Dans la plupart des travaux tentant d’identifier des biais dans divers contextes organisationnels ou scientifiques, ceux présentés comme n’étant pas directement des biais cognitifs sont souvent en fait des éléments de l’environnement ayant une influence sur la prise de décisions par les individus, entraînant un écart systématique

de jugement. Comme cela est expliqué par le principe de rationalité écologique et discuté dans l'approche évolutionnaire, ces éléments de l'environnement peuvent être considérés comme les alter-ego des biais cognitifs. En effet, les biais cognitifs et heuristiques n'étant qu'une réponse particulière à une structure d'environnement, identifier les caractéristiques de l'environnement pouvant expliquer une certaine variabilité dans les réponses des individus revient à identifier, d'une manière différente mais duale, des biais cognitifs (Haselton et al., 2015) (Wilke & Mata, 2012).

## Biais généraux

Il convient toutefois de préciser que tous les biais ne sont pas nécessairement liés à l'homme, même si ils le sont pratiquement tous au moins indirectement. En effet, la présence d'un appareil de mesure défectueux dans un système est probablement dû à une erreur de conception ou à son nonremplacement avant la fin de sa période d'amortissement. De la même manière, des écarts statistiques dans des données peuvent être dus à une erreur dans la conception du pipeline de pré-traitement de ces données ou à un problème d'échantillonnage. Ces biais sont dus à des erreurs humaines, mais sont des conséquences indirectes de celles-ci, et ne seront donc pas considérées comme des biais cognitifs ici.

Nos recherches bibliographiques n'ont pas permis, à ce jour, de trouver de mention de représentation robuste et universelle de tous les biais, laquelle permettrait ainsi de différencier les biais cognitifs et les autres types de biais. Il semblerait même que, dans la plupart des cas, les biais cognitifs, les heuristiques, les caractéristiques de l'environnement et les autres biais ne soient pas spécifiquement distingués les uns des autres. De nombreux travaux identifient puis présentent tous les biais en fonction d'autres critères, par exemple en fonction du processus qui mène à l'objectif souhaité ou, dans certains cas, la production de données de qualité (Delgado-Rodriguez, 2004), (Srinivasan & Chander, 2021), (Balakrishnan et al., 2014).

Notre recherche s'intéresse, entre autres, à l'importance de l'interaction entre l'homme, le processus de production de données et d'informations et la qualité de celles-ci. Les biais cognitifs sont donc une, voire la, préoccupation principale de cette étude, et il semble nécessaire de pouvoir les distinguer de ceux qui n'y sont pas directement liés l'homme.

Dans de nombreuses études, un type particulier de biais revient fréquemment : les biais statistiques (Delgado-Rodriguez, 2004), (Srinivasan & Chander, 2021), (Balakrishnan et al., 2014). Pour des raisons de simplicité, tous les biais qui ne sont pas liés directement à l'homme, ou à l'environnement, seront considérés, ici, comme des biais statistiques.

- Biais cognitif. Il s'agit là d'une erreur systématique de jugement et de prise de décision commune à tous les êtres humains, qui peut être due à des limitations cognitives, à des facteurs de motivation et/ou à des adaptations aux environnements naturels. Les biais cognitifs sont étudiés dans les domaines des sciences cognitives et comportementales, mais également dans d'autres domaines scientifiques, où ils sont mis en avant en fonction de leur pertinence pour répondre aux problématiques choisissent.
- Biais statistique. Les biais statistiques peuvent être, par exemple, dus à un

appareil de mesure défectueux, l'application d'un coefficient statistique inapproprié ou encore des biais dans l'échantillonnage et dans la population censée être représentative.

## Biais cognitifs

L'évolution des courants de pensée relatifs aux biais cognitifs accorde de plus en plus d'importance aux caractéristiques de l'environnement, comme le souligne le concept de rationalité écologique. Les chercheurs se penchent ainsi davantage sur la structure et la représentation de l'information dans l'environnement, ainsi que sur le processus de prise de décisions et d'actions. Les travaux les plus récents ayant marqué l'économie comportementale et la psychologie sont sans doute ceux de Thaler ([Thaler & Sunstein, 2009](#)). D'ailleurs, Thaler adopte également cette approche en se concentrant presque exclusivement sur l'influence de la structure de l'environnement sur la prise de décision. Dans le cadre de sa théorie du *nudge* (coup de pouce), il développe le concept d'"architecture de choix". Il suggère également la possibilité d'agir sur la structure de l'information dans l'environnement pour influencer subtilement les décisions des individus. Il qualifie cette méthode de "paternalisme doux", car elle permet aux individus de prendre de meilleures décisions en limitant l'influence des biais et des limitations du système cognitif sur la prise de décisions, tout en leur laissant le choix final.

Cette thèse adopte une approche écologique similaire à celle que propose Richard Thaler, et s'intéresse donc davantage à la structure de l'environnement, ou architecture de choix, sous-jacente aux biais, qu'aux biais en eux-mêmes. Cette approche a notamment comme avantage de permettre, dans un contexte donné et une fois les caractéristiques de l'environnement identifiées, d'agir sur l'architecture de choix pour améliorer la prise de décisions des individus. Ainsi, dans le chapitre 4, nous nous proposons d'identifier les caractéristiques de l'environnement qui conditionnent la production de données, et donc le processus de binding et réification. Ensuite, dans le chapitre 5, nous testons l'effet exercé par la motivation sur la production de données. La motivation est, en effet, identifiée dans le chapitre 4 comme étant une caractéristique de l'environnement affectant le binding et la réification. De plus, de nombreuses recherches la présentent comme un facteur clé dans la production de données de qualité ([Friedrich et al., 2020](#)), ([Ryan & Deci, 2000](#)), et l'utilisation efficace du système d'information est étroitement liée à ce paramètre ([Koivisto & Hamari, 2019](#)), ([Venkatesh et al., 2003](#)).

## 3.3 Qualité des données et de l'information

Les sections précédentes ont proposé des définitions croisées de ce que sont l'information, la donnée et le processus qui permet de les produire. Nous avons également souligné que, quand la production de celles-ci implique l'intervention d'humaine, elles peuvent être affectées par un certain nombre de biais cognitifs ou statistiques. Ceux-ci ont, naturellement, une influence certaine sur leurs qualités.

De la même manière que le manque de cohérence dans la terminologie et la définition des concepts de donnée et d'information a obscurci le rôle relatif joué

par chacun d’entre eux, il a également rendu délicate la proposition d’un modèle conceptuel décrivant la qualité des données et des informations en identifiant le rôle de chacun distinctement.

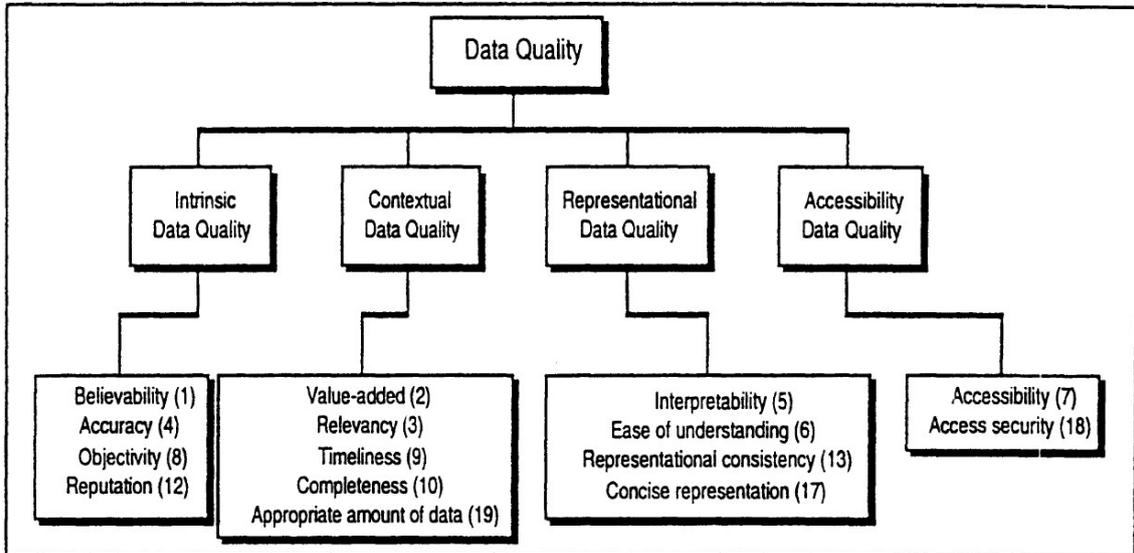


FIGURE 3.3 – Modèle conceptuel de la qualité des données de [R. Y. Wang et Strong \(1996\)](#)

Le cadre conceptuel de la qualité des données proposé par Wang ([R. Y. Wang & Strong, 1996](#)) (voir figure 3.3) repose sur un certain nombre de caractéristiques réparties en quatre dimensions régissant la qualité des données sur la base de l’analyse des résultats d’une large enquête. Cependant, à nos yeux, l’auteur ne prend pas clairement en compte la différence entre donnée et information soulignée ci-dessus, probablement à cause du manque de cohérence dans la terminologie, comme cela a été vu précédemment. Il en résulte une utilisation indifférenciée des mêmes caractéristiques et dimensions pour définir, d’abord, la qualité des données ([R. Y. Wang & Strong, 1996](#)), puis, plus tard, la qualité de l’information ([Kahn et al., 2002](#)), le tout fondé sur la base du même recueil de données effectué en 1996, ce qui semble contradictoire.

Nous proposons ici une version unifiée du cadre conceptuel de Wang (voir figure 3.4), englobant un périmètre plus large, cadre dans lequel la distinction entre la donnée, l’information et le système est faite. Ce nouveau modèle repose sur le même recueil de données ([R. Y. Wang & Strong, 1996](#)), dont les résultats agrégés sont rappelés en table 3.1. Néanmoins, et de manière cruciale, chaque aspect du modèle sera au préalable analysée de manière à déterminer si elle concerne plutôt l’information ou la donnée, en se fondant sur le principe, présent dans l’approche cognitive, selon lequel la donnée n’acquiert le statut d’information qu’à travers l’interprétation humaine ([Zins, 2007](#)).

### 3.3.1 Dimensions additionnelles

Le choix des dimensions d’analyse, i.e., des grands axes structurants qui correspondent à la troisième ligne du schéma de la figure 3.4, est fait, dans ce nouveau

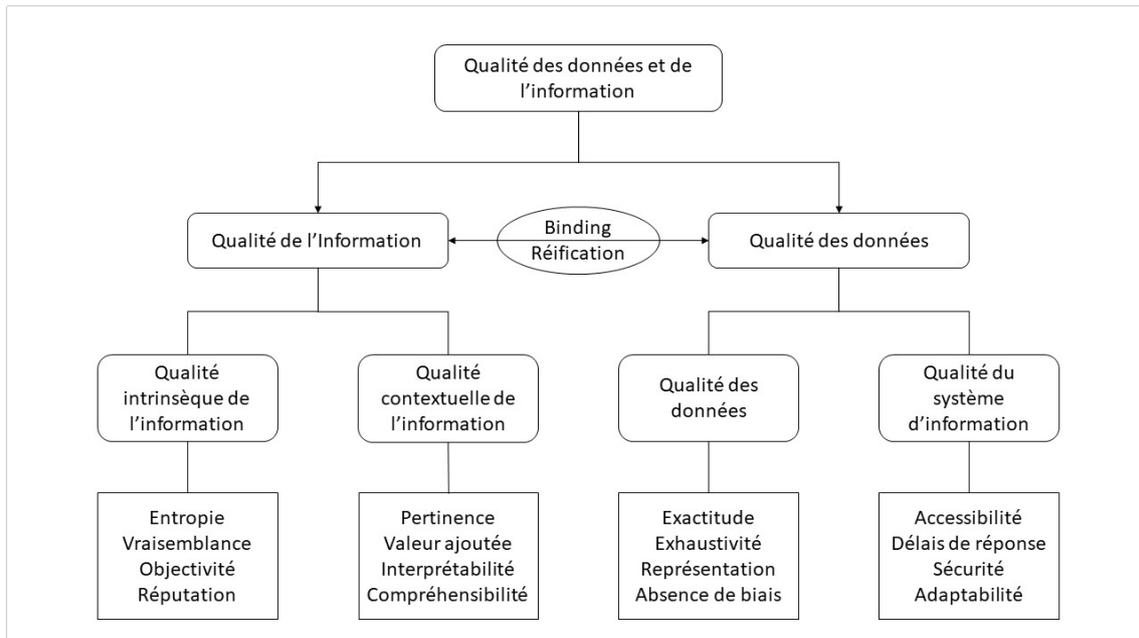


FIGURE 3.4 – Modèle unifié de la qualité des données et de l'information

modèle conceptuel de la qualité des données et de l'information, de manière à préserver une grande partie du modèle original de Wang, tout en mettant l'accent sur la distinction entre information et données.

Tout d'abord, les qualités intrinsèque et contextuelle de la donnée deviennent les qualités intrinsèque et contextuelle de l'information. Ce changement provient du constat simple que la majorité des caractéristiques (dernière ligne de la figure 3.3) définissant originalement ces deux dimensions concernent l'information au sens cognitif du terme, et non la donnée. Les caractéristiques ne se référant pas à l'information mais plutôt à la donnée ou au système, telles que l'exhaustivité, la représentation ou le délai de réponse, ont été déplacées dans les dimensions concernant respectivement la donnée et le système d'information.

Toutes les caractéristiques concernant la qualité de la donnée à proprement parler ont été regroupées dans une unique méta-dimension appelée "qualité des données", qui concerne l'aspect statistique, informatique ou technique de la donnée, en général mesurable, et qui est relativement indépendant de l'interaction avec l'homme. On y retrouve donc des caractéristiques connues telles que la précision ou l'exhaustivité des données.

Enfin la dimension initialement appelée "accessibilité de la donnée", faisant référence au système d'information, devient "qualité du système d'information" et comprend les mêmes caractéristiques que dans le modèle original avec l'ajout des caractéristiques "délai de réponse" et "adaptabilité". Nous proposons donc ici d'élargir cette dimension pour englober les caractéristiques relatives à l'efficacité du système d'information de manière générale.

Étant donné que ce travail conserve en partie les caractéristiques du modèle original de Wang, seules les caractéristiques qui ont été modifiées ou déplacées sont discutées ci-dessous. Les nouvelles caractéristiques sont décrites dans la sous-section suivante.

## Qualité intrinsèque de l'information

La vraisemblance et la réputation semblent toutes les deux être des caractéristiques directement liées à l'interprétation humaine et, en particulier, au processus de binding. Dans le cas de la vraisemblance, la donnée est liée aux croyances de l'individu sur le monde et ses connaissances afin de déterminer si l'information est crédible et vraie à ses yeux (Blitman, 2020); de manière similaire, la réputation perçue va dépendre des connaissances de l'individu concernant la source et sa réputation.

L'objectivité est sans aucun doute la caractéristique la plus difficile à classer, et la raison en est qu'elle peut concerner à la fois la donnée et l'information telles qu'elles sont définies par Wang (R. Y. Wang & Strong, 1996). En effet, elle décrit le fait que la donnée ou l'information soit objective et non biaisée; or, bien que l'objectivité soit un concept strictement lié à l'homme, les biais peuvent être humains, mais pas seulement. En effet, on peut observer un biais provenant d'un appareil de mesure ou d'une erreur de calcul dans le traitement de la donnée, par exemple.

On va donc conserver ici la caractéristique "objectivité", dont la fonction sera de caractériser la minimisation des biais humains affectant l'information. Les biais non humains tels que les erreurs de mesure ou de traitement de la donnée seront caractérisés par une nouvelle caractéristique nommée "absence de biais", qui sera déplacée dans la dimension "qualité technique des données".

## Qualité contextuelle de l'information

La valeur ajoutée ainsi que la pertinence sont des caractéristiques qui n'ont de sens qu'à travers un individu. La valeur ajoutée tout comme la pertinence dépendent des besoins de l'individu, de ses objectifs et de ses connaissances. Si la personne reçoit une information concernant la situation météorologique en Afrique alors que son travail est de gérer une équipe d'ingénieurs, cette information ne sera probablement pas pertinente et n'aura probablement pas de valeur ajoutée. En revanche, avoir une information sur les tendances de la satisfaction des membres de son équipe pourrait lui permettre d'anticiper un turnover trop élevé, ce qui serait pertinent et aurait probablement de la valeur ajoutée. Ces deux caractéristiques sont donc plus adaptées pour qualifier la qualité de l'information.

L'interprétabilité représente le fait d'utiliser un langage, des symboles ou des unités de mesure connues et comprises par l'individu et est, donc, de fait, liée à l'individu et à son interprétation sémantique telle que définie par Chomsky (Blitman, 2020). La compréhensibilité est une combinaison abstraite de nombreux facteurs. Néanmoins, la compréhension est inhérente à l'homme, ou du moins au vivant; elle concerne bien l'interprétation de la donnée et donc relève de l'information.

## Qualité des données

Cette nouvelle dimension regroupe les caractéristiques concernant strictement la qualité des données, c'est à dire l'exactitude, l'exhaustivité, la représentation et l'absence de biais liés à la machine. La caractéristique "représentation" est l'agrégation des caractéristiques précédentes "représentation concise" et "représentation consistante" (voir figure 3.3 et table 3.1).

## Qualité du système

Les notions d’“accessibilité” et de “délai de réponse” étaient déjà classifiées dans la précédente dimension “accessibilité des données”, qui a été élargie et renommée “qualité du système d’information”. Le délai de réponse est une nouvelle caractéristique que nous avons extraite des libellés correspondant à l’ancienne caractéristique “accessibilité”. Parmi celles-ci, comme on peut le voir dans la table 3.1, on trouve “rapidité d’accès”, notion qui aujourd’hui est devenue une métrique importante dans la conception des systèmes d’information en général. Nous proposons donc d’en faire une caractéristique à part entière, nommée “délai de réponse”. L’adaptabilité, une nouvelle caractéristique détaillée dans la sous-section suivante, est issue de l’agrégation des caractéristiques “flexibilité” et “facilité des opérations” (voir table 3.1).

### 3.3.2 Nouvelles caractéristiques

Notre nouveau cadre conceptuel repose en partie, on l’a vu, sur une extension des caractéristiques du modèle de Wang. Ces additions sont dues à l’élargissement du périmètre du modèle et requièrent de couvrir les écarts qui en résultent. Elles prennent cependant bien racine dans les données initialement collectées par Wang (voir table 3.1) et sont le résultat d’une interprétation différente de celle-ci, fondée sur le modèle DBR.

#### Entropie

Une des caractéristiques du modèle initial est “quantité appropriée de données”, mais, en l’occurrence, celle-ci ne devrait pas uniquement concerner la donnée. Comme le propose la théorie de l’information de Shannon, une importante quantité de données ne signifie pas nécessairement une importante quantité d’informations, et peut même correspondre à une quantité nulle d’information. Cela est expliqué par le fait que les données peuvent être caractérisées par une très forte redondance, auquel cas chaque donnée ne contiendra potentiellement pas d’information, si celle-ci est prévisible (Rioul, 2020). C’est la valeur de l’entropie qui permet de caractériser la quantité d’information que peut potentiellement délivrer une source de données. Cependant, il convient de rappeler que ce concept fait fi de l’interprétation sémantique, qui, dans notre cas en particulier, est primordiale pour déterminer la quantité d’informations perçues par l’interlocuteur ; il convient donc ici de définir l’entropie comme étant la quantité d’informations perçues par l’interlocuteur, dépendant donc de ses connaissances préexistantes (Blitman, 2020).

#### Adaptabilité

On peut observer dans les données initiales du modèle de Wang, en table 3.1, que les caractéristiques “flexibilité” et “facilité des opérations” ne sont pas présentes dans le modèle final de Wang. Ces caractéristiques ont probablement été supprimées parce que le domaine d’application initialement envisagé était moins large que dans le cas présent. Dans notre cas, pour décrire notre modèle unifié, qui s’intéresse également à la qualité du système d’information pour produire des données et délivrer

Caractéristique	Étiquettes
Vraisemblance	-
Valeur ajoutée	Avantage compétitif; Ajout de valeur dans les activités
Pertinence	Applicable, pertinent, intéressant, utilisable
Exactitude	Précis, correct, irréprochable, fiable, intègre
Facilité de compréhension	Facile à comprendre, claire, lisible
Interprétabilité	Interprétable
Objectivité	Non biaisé, objectif
Actualisé	Age de la donnée; mise à jour
Exhaustivité	Étendue, profondeur et champ d'application
Réputation	Réputation de la source de la données, réputation de la donnée
Représentation consistante	Donnée toujours présentée au même format, représentée et formatée de manière consistante, compatible avec les données plus anciennes
Représentation concise	Bien présentée, concise, représentée de manière compacte, bien organisée, esthétique, format de la donnée, bien formatée, forme
Quantité de données appropriée	-
Accessibilité	Accessible, requêtable, rapidité d'accès, disponibilité, mise à jour
Traçabilité	Bien documenté, facilement traçable, vérifiable
Coût	Coût de la précision et de l'exactitude, coût de la collecte de données, rapport qualité-prix
Flexibilité	Adaptable, flexible, extensible, expansible
Facilité des opérations	Facilement requêtable, changé, mis à jour, téléchargé, manipulé, agrégé, reproduit, intégré, customisé
Variété des données et des sources	-
Sécurité d'accès	Sécurisé, la concurrence n'y a pas accès, de nature propriétaire, accès peut être restreint

TABLE 3.1 – Données agrégées du modèle de [R. Y. Wang et Strong \(1996\)](#)

des informations, ces caractéristiques semblent pertinentes. Cela dit, elles semblent partager beaucoup de points communs, à en juger par les étiquettes collectées et visibles dans la deuxième colonne de la table 3.1. Nous proposons donc d'introduire une caractéristique unique les regroupant, nommée "adaptabilité", qui désigne de manière générale la facilité et la flexibilité avec laquelle le système d'information peut s'adapter aux besoins à la fois de saisie de données et de livraison d'informations.

### Délai de réponse

De nos jours, dans le développement des systèmes d'information en général, avec l'évolution des technologies de la communication, les utilisateurs s'attendent à ce que le temps de réponse de l'interface soit très court voire imperceptible. De manière générale, il est communément accepté que l'utilisateur s'attend à avoir un délai de réponse inférieur à 100 ms, même si ce seuil varie légèrement en fonction du type

d'interaction avec le système (visuel, audio, tactile, etc.) ([Attig, Rauh, Franke, & Krems, 2017](#)).

Le délai de réponse semble donc une nouvelle caractéristique qui semble s'imposer et qui se trouvait dans les étiquettes correspondant à l'ancienne caractéristique "accessibilité". Parmi celles-ci, comme on peut le voir dans la table 3.1, on trouve "rapidité d'accès", qui sera ici renommé "délai de réponse".

## Chapitre 4

# Intégration de l'environnement au modèle DBR : une approche par la théorie ancrée

Dans ce chapitre, nous poursuivons l'approche fondée sur le principe de rationalité écologique défini dans le chapitre 3. En suivant cette approche, pour identifier d'où proviennent les biais qui affectent la qualité des données et des informations, il convient de connaître la structure de l'environnement dans lequel les collecteurs de données sont amenés à les créer.

A priori, les biais cognitifs et heuristiques peuvent être (inconsciemment) utilisés par l'homme dans deux circonstances : lors d'interprétation de données en informations, caractérisée par le processus de binding ; lors de la saisie de données, caractérisée par le processus de réification. Comme nous l'avons vu dans le chapitre 3, un certain nombre d'autres variables, en partie liées à l'individu, peuvent également affecter le processus de binding. Il est cependant possible de limiter leurs effets en déterminant l'action envisagée, l'événement cognitif, le contexte et une partie des connaissances préexistantes au préalable.

Le présent chapitre présente les résultats d'une étude de cas visant à mettre en oeuvre, et ainsi valider expérimentalement, le cadre conceptuel de la qualité des données et de l'information fondé sur le modèle DBR présenté dans le chapitre 3. En particulier, on s'intéresse à l'impact de l'environnement, en sélectionnant ici un profil d'individu spécifique, dans un contexte déterminé. Dans ces conditions, il ne reste en effet qu'à identifier les caractéristiques de l'environnement ayant un effet sur les processus de binding et de réification, et donc sur la création d'informations, et la saisie de données. À l'issue de cette étude expérimentale, nous espérons mieux comprendre dans quelle mesure l'intervention de l'homme et de la structure de l'environnement auquel il est soumis affecte le processus de création de données et d'information, notamment dans un contexte organisationnel.

En pratique, un ensemble d'entretiens semi-structurés a été réalisé auprès de chefs de projet d'ALTEN responsables, entre autres, de la production des données relatives aux projets d'ingénierie, en se concentrant exclusivement sur l'aspect subjectif de l'évaluation globale des projets. Les autres métriques d'évaluation comme les éléments financiers et ressources humaines sont, a priori, moins sensibles aux biais

étudiés ici et n’ont donc pas été abordés. Après avoir été informés de l’existence de différences apparentes dans l’évaluation de projets similaires (voir ci-dessous), les participants ont été invités à expliquer ces variations. Ces entretiens ont été réalisés sans que les participants aient été préalablement initiés aux concepts de qualité des données et de l’information tels que présentés dans les chapitres précédents.

## 4.1 Contexte organisationnel

Lors de l’analyse préliminaire des données projet stockées au sein du SI d’AL-TEN, en l’occurrence divers indicateurs qualitatifs de suivi de gestion de projets d’ingénierie (temps, qualité et satisfaction concernant la prestation), nous avons identifié ce qui paraissait être un problème de qualité de l’information disponible ; les valeurs des indicateurs renseignés mensuellement par les chefs de projet à propos des risques et des états d’avancement pour chaque projet se caractérisaient par des taux de réussite très élevés (voir figure 4.1), ainsi que par une très faible dispersion, et donc une faible entropie au sens de Shannon (voir chapitre 2).

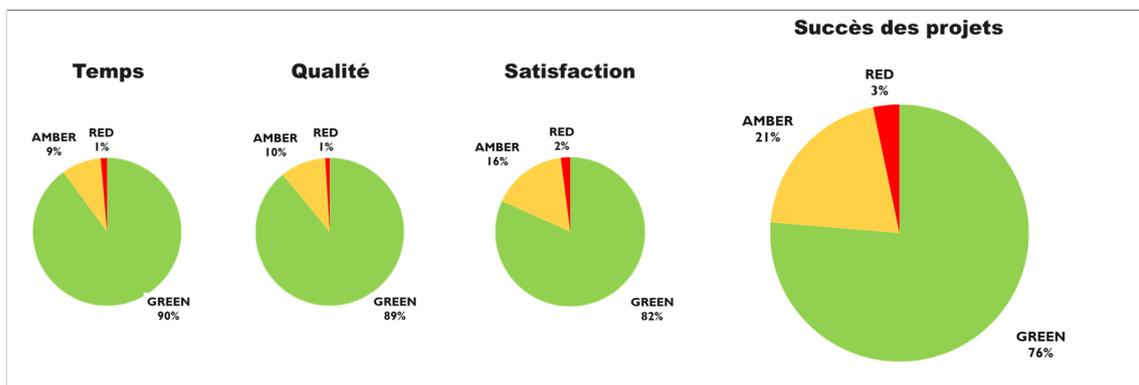


FIGURE 4.1 – Indicateurs de suivi du succès des projets ALTEN

Pourtant, selon le Chaos Report ([Johnson, 2020](#)), un rapport annuel compilant, au niveau mondial, les données issues de plus de 20 000 projets proches de ceux réalisés par ALTEN, le taux de réussite des projets informatiques est particulièrement faible. Ce résultat s’avérait ainsi en claire contradiction avec les taux de réussite des projets tel qu’affiché par ALTEN, comme l’indique la figure 4.2.

La comparaison entre les indicateurs de performance (*Key Performance Indicator*, ou KPI) indiqués par ALTEN et ceux du Chaos Report nous a mis devant une évidence : soit les données d’ALTEN sont biaisées, soit ce sont celles du Chaos Report qui le sont. Or les données du Chaos Report sont fondées sur un large éventail de projets provenant d’un vaste spectre d’industries et liés à des entreprises de tailles variables à travers le monde, ce qui peut laisser penser qu’a priori, celles-ci se rapprochent davantage de la réalité que celles présentes dans le SI d’ALTEN. Il était donc naturel de questionner la qualité des données collectées au sein d’ALTEN.

Notre objectif, ici, est de mettre les participants face à des différences apparentes dans l’évaluation de projets, pourtant similaires, afin de les amener naturellement à aborder l’aspect subjectif de l’évaluation globale des projets.

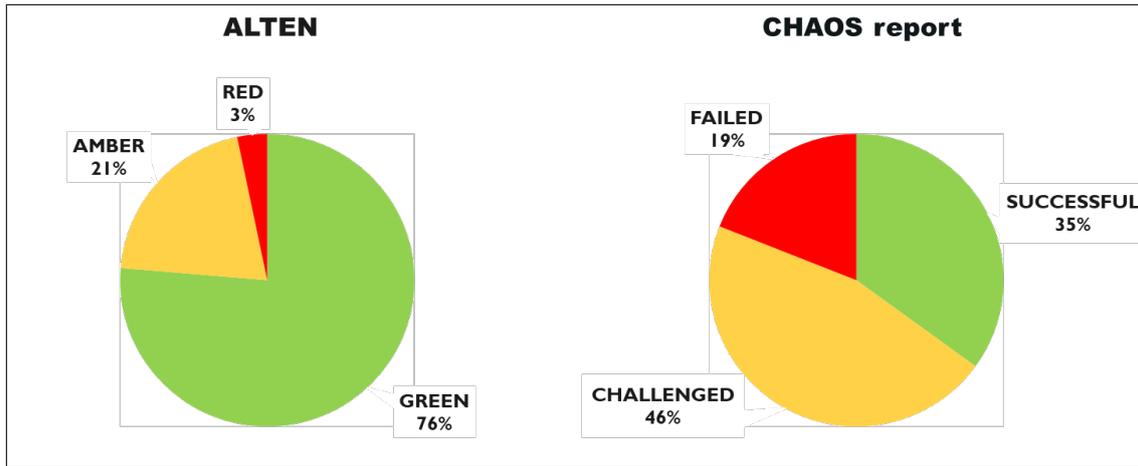


FIGURE 4.2 – Comparaison des KPI de ALTEN et du CHAOS report

## 4.2 Méthode de recherche

Nous présentons dans cette section les éléments essentiels du protocole utilisé lors de cette étude fondée sur des entretiens effectués dans un contexte organisationnel.

### 4.2.1 Sélection de l'échantillon

Les participants sélectionnés pour l'étude étaient 11 chefs ou directeurs de projet, sélectionnés au hasard parmi l'ensemble des 180 chefs et directeurs de projet d'AL-TEN France (voir la sous-section 4.2.4 pour la justification de la taille de l'échantillon).

Les participants ont été invités à participer à l'étude par le biais de courriels et de messages sur l'outil Microsoft Teams de l'entreprise.

Le tableau de la figure 4.3 montre la répartition en termes d'âge, d'années d'ancienneté dans l'entreprise et d'années d'ancienneté dans la direction de l'entreprise de l'ensemble des participants. Tous les participants ont plus de 6 ans d'expérience dans l'entreprise.

Age range	years of practice in the company				Year of practice in the company management			
	1 - 5	6 - 10	11 - 20	Total	0 - 3	4 - 7	7 - 11	Total
30 - 35		2	1	3	3			3
36 - 40		4	1	5	2	2	1	5
41 - 45	1	1	1	3		3		3
	1	7	3	11	5	5	1	11

FIGURE 4.3 – Informations démographiques

### 4.2.2 Ethique

Cette étude a été approuvée par le comité d'éthique d'AL-TEN et la direction de l'entreprise. La personne qui a réalisé les interviews, auteur principal de ce manuscrit,

est employée par ALTEN.

La campagne d'entretiens semi-structurés s'est déroulée entre mai et juillet 2022 et a duré environ une heure par interview. Les entretiens ont eu lieu pendant les heures de travail normales. Tous les participants ont donné leur consentement éclairé et ont accepté d'être enregistrés oralement. Ils n'ont pas été récompensés pour leur participation.

L'étude a produit un total d'environ 20 heures de contenu, qui a été enregistré, transcrit et analysé à l'aide de l'outil NVivo (Myers, 2013). L'auteur a été autorisé par l'entreprise à traiter ces données et à publier les résultats de leur analyse.

### 4.2.3 Procédure d'entretien

Lors de chaque entretien, nous avons suivi le même protocole, conçu à l'avance. Il s'agissait d'abord de montrer à la personne interrogée des données pouvant questionner la qualité et de la subjectivité des données collectées au sein de l'entreprise, puis de lui demander comment, selon elle, elles peuvent être expliquées. En particulier, au cours des entretiens, le chercheur a demandé aux participants de discuter de la mesure dans laquelle l'évaluation des projets qu'ils effectuaient mensuellement était subjective et de suggérer des raisons qui pourraient expliquer leur variabilité.

Les participants n'ont pas été initiés au concept de qualité des données et de l'information telle que présentée précédemment, ni à l'objectif de l'entretien. Nous leur avons simplement présenté des données provenant de diverses sources supposées donner les mêmes résultats, mais pour lesquelles ce n'était pas le cas. De cette manière, nous les avons confrontés à l'aspect subjectif de l'information, pour qu'ils puissent réfléchir par eux-mêmes aux raisons qui pourraient expliquer ces différences, et ce, sans a priori. L'objectif était d'éviter le biais de confirmation bien connu et l'effet d'attente de l'observateur ou de l'expérimentateur. Chaque entretien s'est déroulé en trois phases.

1. Dans la première partie de l'entretien, nous leur avons présenté les conclusions d'une comparaison faite entre l'évaluation des projets telle que produite régulièrement par les chefs et directeurs de projets d'ALTEN et ceux d'une étude globale évaluant un échantillon de 20 000 projets menée par le Standish Group en 2021 (voir la section 4.1). Ces deux études évaluent la réussite des projets sur la base de critères de performance très similaires, mais donnent des résultats très différents ; nous avons demandé aux participants de nous aider à tenter de comprendre ce qui pouvait expliquer ces divergences.
2. Dans la deuxième partie des entretiens, nous leur avons rappelé que nous avons collecté des données sur les projets de l'entreprise par le biais d'un mailing expérimental envoyé aux chefs de projet et qu'ils avaient fait partie de cette expérience (voir le chapitre 5, qui décrit cette étude). Cette évaluation avait été menée en dehors du cadre traditionnel de collecte de données du SI de l'entreprise, et les données ainsi collectées se sont révélées très différentes de celles qui figuraient déjà dans la base de données de l'entreprise. Nous leur avons ensuite montré la comparaison des distributions entre les données de l'entreprise officiellement collectées via le SI et les données collectées en dehors du processus habituel, qui montraient des résultats significativement

différents, puis nous leur avons demandé comment, selon eux, cela pouvait s'expliquer.

3. Dans la dernière partie, nous leur avons expliqué que l'enquête que nous avons menée en dehors du SI et dont ils venaient d'observer les résultats agrégés (voir chapitre 5) avait été réalisée en partitionnant les participants en deux groupes. Nous leur avons expliqué que chaque groupe avait été sollicité de manière différente, l'un en utilisant la méthode de régulation contrôlée et l'autre, celle de régulation non-contrôlée (Ryan & Deci, 2000). Nous leur avons ensuite montré les données de chaque groupe et leur avons demandé, à nouveau, comment ils pouvaient, ou pas, expliquer les différences observées.

#### 4.2.4 Collecte des données

Les données collectées pour cette étude le sont toutes à partir des entretiens réalisés et des notions clés qu'ils contiennent. Après chaque entretien, un résumé des propos tenus a tout d'abord été rédigé à chaud, ainsi qu'une liste des principaux thèmes qui ont émergé, afin d'avoir une vision claire de l'évolution des idées émergentes.

Initialement, un maximum de 30 entretiens était prévu, valeur choisie en raison des contraintes de temps tout en maintenant une taille significative d'échantillon (selon le théorème central limite).

Cependant, dès le 6ème entretien, nous avons commencé à observer une saturation des concepts émergents introduits par les participants ; nous avons donc réalisé cinq entretiens supplémentaires afin d'assurer une saturation totale de l'espace des concepts, ce qui conduit à un échantillon de 11 participants et, donc, entretiens. Cette saturation rapide s'explique par le fait que nous avons réalisé des entretiens approfondis, de plus d'une heure pour la plupart, avec des managers ayant plusieurs années d'expérience dans l'entreprise et dans leur fonction actuelle (voir figure 4.3). Ils avaient donc tous une expérience complète et approfondie de l'entreprise, expérience partagée par tous.

Chaque entretien ayant été enregistré, il a ensuite été transcrit de l'audio en texte manuellement dans un format facilitant l'importation dans l'outil NVivo, un logiciel spécialisé dans l'encodage et l'analyse de données qualitatives. Une fois importées dans NVivo, les transcriptions ont été encodées en utilisant la méthode d'*open coding* (codage ouvert), fondée sur la théorie ancrée, afin de tenter de faire émerger des facteurs pouvant expliquer l'origine des biais présents dans les données (Myers, 2013) (voir section suivante).

La théorie ancrée (*Grounded Theory*) est une méthode de recherche qualitative qui vise à dériver systématiquement des théories du comportement humain à partir de données empiriques (Myers, 2013). Cette méthode de recherche a été choisie ici, parce qu'elle est particulièrement adaptée à l'élaboration de descriptions et d'explications des organisations fondées sur le contexte et orientées vers l'utilisation de processus. Le nom "*Grounded theory*" vise à souligner qu'elle permet de développer des théories "ancrées" dans l'analyse des données recueillies lors d'entretiens et d'expériences, au lieu d'être fondées sur des idées théoriques préconçues. Dans notre cas, cette approche est utile, car la littérature sur la relation entre subjectivité et qualité

de l'information dans les organisations est limitée.

Cette approche "positiviste" vise ici à prendre en compte la nature subjective de l'information et sa relation étroite avec le concept de données, dans l'évaluation de la qualité des données et la caractérisation des facteurs, et biais, affectant celle-ci. L'utilisation de la théorie ancrée est la voie à suivre ici, car il y a un manque important de littérature académique pour répondre à cette préoccupation, et, à notre connaissance, il n'y a pas d'article académique qui utilise la relation entre les concepts de donnée et d'information pour analyser les facteurs responsables de l'émergence de biais. Cette absence dans la littérature est cohérente avec le *gap* (manque) concernant l'étude de cette relation, point aveugle précédemment identifié par des chercheurs en sciences de l'information (Rowley, 2007).

## 4.3 Résultats

Nous décrivons ici l'essentiel des résultats obtenus en suivant le protocole précédemment présenté. Leur analyse est faite dans la section suivante.

### 4.3.1 Codage ouvert

L'encodage, c'est-à-dire l'attribution de termes spécifiques, ou "codes", aux énoncés des personnes interrogées, a été effectué en utilisant la méthode dite "in vivo", ce qui signifie que nous avons choisi les codes en reprenant, autant que possible, les termes utilisés par les personnes interrogées. L'objectif est d'altérer le moins possible le contenu informationnel pendant la phase de codage ouvert, une phase d'affinage venant par la suite.

À la fin de la phase de codage ouvert, 502 verbatims issus des 11 entretiens avaient été codés à l'aide d'un corpus de 60 codes bruts différents. Après une première phase de raffinement, il s'est avéré qu'il existait des codes très similaires, créant ainsi une redondance. Ceux-ci ont alors été fusionnés, réduisant le nombre de codes de 60 à 43 avant structuration par codage axial.

### 4.3.2 Codage axial

Le codage ouvert a été suivi par une phase dite de codage axial, qui conduit à la définition d'une structure sémantique adaptée aux codes, ou propriétés, par le regroupement de ceux-ci selon une hiérarchie enfant/parent, afin de faire émerger des thèmes structurés. À l'issue de cette phase de codage axial, 6 codes de niveau 1, 35 codes de niveau 2 et 8 codes de niveau 3 ont été identifiés (voir cette hiérarchie intermédiaire en annexe A.1).

À l'issue d'une nouvelle phase de raffinement, les 35 codes de niveau 2 et les 8 codes de niveau 3 ont agrégés en 31 codes de niveau 2, et le niveau 3 a été supprimé. On obtient alors la table 4.3<sup>1</sup>.

Les 6 codes de niveau 1, regroupés, avec leurs fréquences d'occurrence, dans la table 4.1, sont considérés ici comme les thèmes émergents qui permettent de classer

---

1. Cette table est présentée en fin de chapitre, puisqu'elle représente l'ensemble des caractéristiques de l'environnement du modèle DBR.

les 31 codes de niveau 2. Ces 31 codes de niveau 2 sont considérés par la suite comme les caractéristiques de l’environnement qui modulent le processus de binding et de réification, et donc, selon le modèle DBR, la production de données et d’informations.

Thème	Fréquence absolue	Fréquence relative
Données et informations	110	0,26
Collecte de données	65	0,15
Projet	63	0,15
Humain	85	0,20
Parties prenantes	67	0,16
Culture d’entreprise	33	0,08

TABLE 4.1 – Fréquences des thèmes

Pour illustrer cette cartographie des catégories émergentes, nous fournissons ci-dessous, pour chacune d’entre elles, quelques phrases significatives extraites des entretiens, que nous les replaçons dans leur contexte. Les thèmes sont décrits dans le tableau 4.1, triés selon leur fréquence relative descendante et groupés en fonction du modèle DBR en deux parties (voir section 4.4). Par exemple, le thème “humain” a été associé à celui des “parties prenantes”, car ils sont étroitement liés et abordent tous deux fréquemment les notions de peur et de transparence.

### Données et informations

Ce thème est celui qui a été le plus discuté ; il concerne la définition de la donnée mais aussi sa pertinence, son adéquation à la réalité du terrain et sa “malléabilité”. Ce qui a été principalement mentionné, c’est que les données, entre le moment où elles sont collectées à l’état brut et le moment où elles sont introduites dans l’outil SI, subissent de nombreuses transformations, adaptations et modifications pour correspondre au format demandé, aux spécificités du projet, mais aussi aux attentes (supposées ou prescrites) de la direction d’ALTEN.

Un exemple typique est celui de l’indicateur de risque lié au retard (anticipé) dans la livraison des livrables d’un projet. Dans un premier temps, la donnée brute est le nombre de jours de retard pour chaque livrable, qu’il faut ensuite agréger pour obtenir un pourcentage pour l’ensemble des livrables. Cependant, si ce critère n’est pas supérieur au seuil conjointement demandé par la direction et le client (et défini de manière contractuelle lors de la signature du contrat du projet), le responsable peut décider de revoir sa manière de définir l’indicateur afin d’atteindre les seuils demandés.

*"J'ai plusieurs projets où le nombre de sous-projets est très faible sur le mois, et il y a des seuils entre 90% et 95%, et sur 5 projets vous serez à 20, 40, 60, 80, ou 100, jamais 95. Nous essayons donc de trouver une solution pour avoir quelque chose de plus raffiné. Pour ma part, sur chaque élément, je calcule le retard, c'est-à-dire le nombre de jours contractuels sur le nombre total de jours, ce qui me permet*

*d'avoir un résultat OTD<sup>2</sup> de 95% et plus". DP3<sup>3</sup>*

De même, lorsqu'un chef de projet estime que le retard n'est pas dû à un problème de performance de son équipe, mais plutôt à un facteur externe, il renégociera presque systématiquement la date de livraison initiale, annulant ou réduisant le retard initial.

*"Pour nous, si c'est un problème client, on renégocie, donc il n'y a plus de soucis; on n'est pas en retard par rapport à la nouvelle date; on passe au vert." CP1*

Le fait que ces efforts soient faits pour "renégocier" la note à renseigner montre déjà là que, dans une certaine mesure, la saisie des données n'est pas un acte désintéressé ni objectif.

La faible granularité des données à saisir ici par les gestionnaires de projet, dans le cas précédent "rouge", "orange" et "vert", avec la sémantique que l'on devine mais qui n'est pas formellement précisée, a souvent été citée comme un obstacle à l'évaluation exacte ou précise des fluctuations des indicateurs. Le problème est que les fluctuations petites et moyennes ne sont pas suffisamment importantes pour justifier une modification de la valeur de l'indicateur. En outre, compte tenu du fait qu'une diminution de la note n'est pas un acte anodin, dans une situation d'ambiguïté, ne rien changer sera presque toujours perçu comme le choix le plus raisonnable.

*"Nous n'aimons pas cette note de 1 à 4, parce que, lorsque vous mettez 3, c'est bien, mais dès que vous mettez 2, c'est tout de suite délicat. Ce qui serait bien, c'est d'avoir une plus grande granularité; les chefs de projet seraient beaucoup plus enclins à passer de 14 à 12, pour capter des petites variations dans les projets, ce qui reflètera beaucoup plus la réalité. On ne peut pas passer de 3 à 2 aujourd'hui, parce que cela créerait beaucoup plus de problèmes." CP2*

*"Nous effectuons nos mesures projet par projet, et ensuite nous nous contentons d'agréger, et effectivement pour 1 projet sur 40, nous ne voulons pas être pénalisés pour cela. Nous mettrons en place une chaîne d'actions pour résoudre le problème, mais nous n'en parlerons pas." CP1*

D'autres facteurs tels que la pertinence ou la réputation des données ont été souvent mentionnés et liés à l'adéquation entre les données et la réalité du terrain, et à leur réputation en termes d'utilité et de qualité.

## Collecte de données

Le processus de saisie des indicateurs dans le SI a un impact sur la qualité des données, notamment parce qu'il impose, dans le cas d'ALTEN, la saisie des données à un moment déjà très chargé, celui de la clôture financière du budget. De plus, un

---

2. *On-Time Delivery*

3. Pour chaque verbatim, on précise le locuteur, anonymisé, par le label DDP (directeur de projet niveau 2), DP (directeur de projet niveau 1) ou CP (chef de projet) suivi d'un numéro.

certain nombre d'autres données et formulaires sont demandés simultanément aux gestionnaires de projet.

*"Cela ne devrait pas être lié à la clôture; nous ne devrions pas avoir 50 000 formulaires à remplir." CP3*

Ces deux éléments combinés signifient que les gestionnaires de projet ont très peu de temps pour compléter la saisie des données et décident souvent de donner la priorité à d'autres tâches, parfois en négligeant complètement la saisie des données.

*"Les périodes de clôture sont très chargées, [...]; souvent je reviens sur le rapport opérationnel du mois dernier et j'en fais une copie." DP2*

Par ailleurs, l'outil SI n'est pas toujours maîtrisé par les employés, en raison de certains problèmes liés à l'outil lui-même et d'un manque de formation pratique.

*"On a appris beaucoup de choses, mais c'est encore très théorique. Quand on essaie de le mettre en pratique, il y a toujours des obstacles; par exemple, les consultants qui n'ont pas accès, c'est un gros obstacle pour moi." CP1*

Dans l'ensemble, il ne fait aucun doute que l'outil de saisie des données et la manière dont il s'intègre dans le processus de collecte des données, mais aussi dans le processus général de gestion du projet, a une influence significative sur la qualité des données.

*"Il y a eu beaucoup d'évolution, mais on nous demanderait de consacrer la moitié de notre temps à la saisie de données. Ce n'est pas que nous n'aimons pas, mais nous n'avons pas le temps. En tant que gestionnaire de projet, c'est donc un peu lourd, mais cela s'améliore." CP2*

Il est également intéressant de noter qu'une partie significative des personnes interrogées a souligné que, bien que les informations complètes n'aient pas été saisies dans le SI, elles étaient bien connues de la plupart des personnes travaillant et dirigeant le projet, et échangées par le biais d'autres processus informels. C'est un autre indicateur du fait que l'information a potentiellement une plus grande qualité dans son état initial d'information et qu'elle perd en qualité en devenant donnée dans le SI, c'est-à-dire à travers le processus de réification.

## **Projet**

La définition des données et leurs valeurs dépendent fortement des caractéristiques des projets auxquels elles se rapportent.

*"Les paramètres doivent être interprétés différemment selon le type de projet." CP3*

En fonction de ces caractéristiques, une donnée peut avoir une signification ou une importance différente, voire ne pas exister. Les projets avec un engagement de type “temps et matériel” (facturation proportionnelle aux moyens employés) sont soumis au même processus de collecte de données que les projets avec un engagement “prix fixe”, alors que, pour les premiers, les données concernant la qualité et le retard des livrables n’existent parfois tout simplement pas. Ceux-ci sont donc rapportés avec la valeur maximale, indiquant que tout se passe bien, par défaut.

*"Pour les IQP3<sup>4</sup> (Temps et matériel), nous n'avons pas d'OTD et de RFT<sup>5</sup> [pour nos projets]; ils sont rapportés à 100%." CP1*

*"Sur l'IQP4, il est très important de piloter l'OTD [temps] et le RFT [qualité], car sinon nous ne sommes pas payés." CP5*

*"Il faut aussi analyser un projet en fonction de son ancienneté, car un jeune projet peut être plein de difficultés, alors que sur des projets qui ont 2 ou 3 ans d'existence, c'est une machine qui tourne toute seule." CP2*

En général, le problème est que ces caractéristiques structurelles ne sont, dans la plupart des cas, pas prises en compte dans le processus de collecte des données, et que la saisie des données est appliquée de manière uniforme, indépendamment de la taille du projet, de son niveau d’engagement, de sa durée ou autre.

## Humain

L’émotion est le thème qui a été abordé le plus largement par les personnes interviewées ; il a d’ailleurs été abordé par toutes sans exception. L’émotion est parfois mentionnée de manière très générale ou de manière plus spécifique, en expliquant pourquoi elle a un impact sur les données.

*"Il y a le facteur humain ; l'émotion peut parfois être un moteur. Nous essayons d'obtenir des faits à partir de choses qui peuvent parfois être motivées par l'émotion, et donc être biaisées." DP5*

La peur est une émotion qui a été mentionnée à plusieurs reprises ; elle est souvent liée à un autre biais, vu par la suite, qui est l’effet de la réaction de la direction, que les personnes interrogées disent vouloir éviter.

*"La peur, la peur d'être sous les feux de la rampe, sous le radar." DP3*

D’une manière générale, le management exerce presque constamment une forme de coercition sur les interviewés, affectant leur motivation extrinsèque, via ce que l’on appelle la régulation contrôlée (Ryan & Deci, 2000).

---

4. Indicateur de qualité de projet

5. *Ready First Time*

*"Quand c'est demandé par la direction, on sait que ce sera suivi, parce qu'on est dans un modèle où, même quand on est en formation, on est évalué en permanence. Donc la notion de contrôle et d'observation, de comment on réagit, comment on répond, est très présente."* DP5

*"Le groupe contrôle me fait un peu penser aux obligations qu'on a tous les mois [...] on active un peu le mode robot et on le fait. Le groupe autonome, je le vois plus comme quelqu'un qui vient me demander un coup de main, et donc on est dans un processus d'entraide."* CP2

## **Parties prenantes**

La présence de plusieurs parties prenantes telles que la direction, les clients ou le service qualité peut influencer la saisie des données. La majorité des chefs de projet ont clairement exprimé que, selon leurs interlocuteurs et, dans le cas de la saisie des données, selon la personne qui va les consulter, il n'y aura pas le même niveau de transparence. Vis-à-vis de la direction, le discours est généralement bien pensé et filtré pour orienter la communication et faire passer les bonnes informations. Au contraire, face au service qualité, les chefs de projet peuvent avoir tendance à s'auto-flageller, car les actions éventuellement suggérées par l'équipe en charge de la qualité au sein de l'entreprise peuvent les aider.

*"Du côté de la qualité, j'ai tendance à m'auto-flageller pour mettre en évidence le mauvais, et dans l'outil de reporting [pour la direction] on a tendance à se dire qu'on ne va pas alerter pour rien."* CP1

La majorité des personnes interrogées ont également déclaré que la réaction de la direction aux indicateurs négatifs peut être effrayante et contre-productive, raison pour laquelle les chefs de projet ont tendance à vouloir les éviter.

*"Les indicateurs apparaissent à la direction générale, qui est très loin du niveau opérationnel, et lorsque des indicateurs "rouge" ou "orange" sont relevés, [...] cela peut créer beaucoup de problèmes en plus de ceux que l'on a déjà, parce que la direction s'excite, pour une bonne raison, mais du coup, on a besoin d'obtenir des résultats très rapidement, alors qu'on peut être confronté à un problème qui dure 6 mois."* CP2

De la même manière, cette relation et cette peur de la réaction se retrouvent dans la relation avec le client, qui a également accès aux indicateurs de risque et de performance d'ALTEN.

*"Il y a certains risques qu'on ne montre pas aux clients, on ne veut pas montrer qu'on est en difficulté."*

## Culture d'entreprise

L'entreprise a une forte culture de la réussite, de la croissance et de la marge.

*"L'objectif ultime d'ALTEN est d'avoir de grands centres de services qui sont durables, qui croissent en permanence et qui font de la marge." CP4*

Cette même culture de la réussite peut être comparée à une pièce de monnaie à deux faces : d'un côté, cette culture pousse à l'excellence et à la performance, ce qui permet à l'entreprise de s'imposer sur le marché comme un leader ; de l'autre, les managers, fortement poussés à la réussite, peuvent en venir à communiquer des informations en décalage avec la réalité, à la fois pour satisfaire la direction et pour véhiculer une bonne image auprès du client.

*"Cette forte culture de la réussite chez ALTEN peut pousser les chefs de projet à parfois enjoliver un peu la réalité et à mettre de côté le challenge, mais le revers de la médaille est que cette forte culture de la réussite permet à ALTEN d'être très performant par rapport à ses concurrents, et de s'imposer comme un leader sur le marché." CP5*

Cette culture de la réussite a des conséquences sur la stratégie de communication utilisée par les chefs de projet. Cette orientation a un effet sur la "capture" des indicateurs, qui dépassent leur rôle de suivi des risques et de la performance et finissent par être utilisés comme moyen de communication avec les clients et le management.

*"Il y a une sorte de green attitude, car ce que nous collectons en interne est ce que nous validons avec nos clients. Le client compare nos indicateurs de performance avec ceux de nos concurrents. Ainsi, si nous générons un graphique contenant plus de projets en difficulté que nos concurrents, nos clients iront chez ces derniers. Nous voulons avoir des graphiques "verts"; c'est une stratégie." DDP1*

### 4.3.3 Codage sélectif

Le codage sélectif consiste à rechercher quelles sont les catégories qui reflètent les relations entre les concepts identifiés lors du codage ouvert.

Ainsi, après la transcription et l'encodage, chaque verbatim a également été balisé manuellement selon qu'il est considéré comme affectant ou non la qualité des données et/ou de l'information. Le résultat d'un tel codage est résumé dans la table 4.2, qui indique dans quelle proportion chaque thème affecte l'information ou la donnée<sup>6</sup>.

D'après ces résultats, les trois thèmes affectant le plus la qualité des données sont « Données et informations », « Humain », et « Parties prenantes ».

---

6. On observe une différence entre les fréquences absolues de la table précédente 4.1 et la table présente 4.2. Celle-ci vient du fait que la première indique le nombre de références total par thème, tandis que celle-ci indique le nombre d'éléments encodés (verbatim) par thèmes. De fait, un élément encodé peut contenir plusieurs références appartenant au même thème (mais des codes de niveau 2 différent), ce qui explique que la table 4.1 indique des fréquences absolues plus importantes.

	Balise	Fréquence absolue	Fréquence relative
Données et informations	67	84	80%
Collecte de données	29	48	60%
Projet	22	55	40%
Humain	49	57	86%
Partie prenante	48	58	83%
Culture d'entreprise	16	28	57%

TABLE 4.2 – Proportion de verbatims balisés par thème

Notons, cependant, que le thème "Parties prenantes" a été très souvent associé au thème "Humain" lors des interviews. En effet lorsque les intervenants évoquaient un facteur humain, tel que la peur, dans 44% des cas ils évoquaient également une partie prenante (voir annexe A.3).

Il s'avère donc que, dans une grande partie des cas, les parties prenantes sont tenues comme responsables des divers facteurs humains affectant les collecteurs de données, comme c'est le cas par exemple pour un chef de projet qui a peur de saisir des données alarmantes pour ses projets par peur de la réaction de sa direction ou de son client.

Il est donc logique d'émettre l'hypothèse selon laquelle le facteur humain est une variable affectant directement l'information et la donnée, alors que la partie prenante (par exemple un client) est davantage une variable affectant indirectement, tout du moins en partie, l'information et la donnée.

Finalement, le thème affectant le moins l'information et la donnée est "projet", ce qui a du sens sachant que le projet est source de l'information et de la donnée; ses caractéristiques affectent le format des données mais pas l'information dont il est la source.

## 4.4 Alignement avec le modèle DBR

A travers le cycle de vie de la donnée, celle-ci va subir une itération d'opérations de binding et de réification à chaque fois qu'elle va passer d'un opérateur à une base de données et d'une base de données à un opérateur. D'ailleurs le processus de réification de la donnée va souvent concerner un ensemble d'informations, détenues par plusieurs opérateurs, plutôt qu'une information unique. Typiquement l'organisation ciblée par ces recherches, ALTEN, organise régulièrement des comités lors desquels plusieurs parties prenantes partagent leurs informations sur un projet avant de lui donner une note à destination de la direction. L'analyse des données d'entretien va nous permettre d'aligner les thèmes et caractéristiques repérés ci-dessus avec le modèle DBR. Plus précisément, nous allons pouvoir déterminer ceux qui relèvent plutôt du processus de binding que de celui de la réification pour un opérateur générique, qu'il s'agisse d'un individu ou d'un groupe (voir table 4.3).

### 4.4.1 Réification

D'après les responsables de projet ayant été interviewés, les données vont subir une succession de modifications permettant d'arriver à des indicateurs agrégés. Dans le cas d'ALTEN, quand la donnée arrive au niveau du chef de projet, elle est transformée puis agrégée pour créer un pourcentage. Les chefs de projet<sup>7</sup> ont une certaine marge de manœuvre pour créer les indicateurs à leur manière, si la manière recommandée par l'entreprise ne leur convient pas. Ces modifications dans la manière de produire la donnée est justifiée par le fait que la définition de la donnée, et ce qui est attendu d'elle, ne correspond pas toujours à la réalité du terrain, ce qui pourrait, sans cette adaptation, entraîner la production de données non pertinentes.

A ce stade, où le processus de réification est impliqué, des biais importants peuvent affecter la production des données, allant jusqu'à modifier leurs valeurs intrinsèques. D'après les propos des chefs de projet, les biais sont liés à l'agrégation de la donnée, la mauvaise définition de la donnée et sont l'inadéquation (perçue) avec la réalité du terrain. L'origine de ces biais est essentiellement liée au fait que les caractéristiques fondamentales des projets ne sont, d'après les responsables interviewés, pas pris en compte par le système d'information actuel, plus précisément en ce qui concerne la définition des données.

Il semble donc clair qu'il y a un lien fort entre les caractéristiques encodées dans les thèmes "projet", "données et informations" et "processus de collecte de données". En effet, d'après les chefs de projet, si la définition des données et le processus de collecte de données ne sont pas conçus en prenant en compte les caractéristiques concrètes des projets et leur cycle de vie, cela affecte la production de données et entraîne une diminution de la qualité des données produites.

Ainsi, ces trois thèmes font référence à l'environnement dans lequel sont saisies les données par les chefs de projet et son effet sur la qualité de celle-ci. En effet, toutes les caractéristiques encodées dans ces thèmes sont liées à des problèmes relatifs au système de management de l'information, que ce soit au niveau de sa conception, de son intégration dans le processus de vie des projets, ou son (in)adéquation à la réalité du terrain.

Étant donné que c'est le processus de réification qui est sollicité au moment de saisir la donnée dans la SI, toutes les caractéristiques de l'environnement encodées dans ces trois thèmes doivent être attribuées au processus de réification (voir la table 4.3). Cependant, cela n'exclut pas le fait que celles-ci peuvent affecter également le processus de binding, puisque ces processus sont en partie intriqués, à l'image de la relation entre les concepts de données et d'information.

### 4.4.2 Binding

Les données du SI sont également discutées au cours de comités lors desquels plusieurs personnes, dont la direction, se réunissent pour créer des indicateurs clés en fonction des données à disposition, mais également des informations de chacun. Certains chefs de projet ont expliqué que les parties prenantes lors de ces réunions

---

7. Nous regroupons ici sous ce terme à la fois les chefs de projet proprement dits et les directeurs de projet.

“s’en foutaient (sic)” de certaines données. Par exemple, certains n’accordent pas d’importance aux indicateurs qualitatifs et se concentrent uniquement sur les résultats financiers, tandis que c’est l’opposé pour d’autres tels que les responsables qualité. Ce phénomène est expliqué par un certain nombre de facteurs affectant le binding tels que les connaissances pré-existantes, l’attention ou encore la mémoire (Zimmer et al., 2006).

Il apparaît également que la présence de différentes parties prenantes affecte le jugement, les décisions et même le choix des méthodes de production de données des chefs de projet en provoquant notamment des émotions relatives à la crainte. Par exemple, les chefs de projet veulent que leurs projets soient bien évalués pour éviter les complications avec la direction, qui souhaite voir des bons résultats. C’est notamment dû à une culture d’entreprise du succès très présente chez ALTEN. D’après les participants, cette culture d’entreprise les pousse à se surpasser, mais le revers de la médaille est que cela peut également les inciter à être parfois dans une logique stratégique en ce qui concerne la production d’indicateurs (“green attitude”). A l’inverse, les responsables qualité, qui souhaitent que les projets respectent un certain nombre de normes de qualité et que les données reflètent au mieux la réalité du terrain, vont essayer de pousser les équipes à être plus critiques et apporter des solutions pratiques pour aller dans ce sens.

Pour renforcer cette déclaration, comme cela a été vu précédemment, le thème “humain” a été très fréquemment (44%) associé au thème “parties prenantes” (voir annexe A.3). Il semble donc que les différentes parties prenantes aient bien une influence forte sur les chefs de projet, en suscitant un certain nombre d’émotions, positives ou négatives telles que la peur, la confiance ou la motivation. La culture de l’entreprise pourrait, quant à elle, avoir une influence sur la manière d’interpréter les données en faveur de ce que les chefs de projet souhaitent mettre en avant.

La question désormais est de savoir si ces thèmes font davantage référence à des caractéristiques de l’environnement dans lequel les individus saisissent la donnée ou dans lequel ils prennent des décisions stratégiques et réfléchissent à l’évaluation qu’ils vont donner à leurs projets, en somme créent de l’information. Répondre à cette question a une implication pratique importante, puisque cela permet de déterminer s’il faut classer ces caractéristiques comme étant liées au processus de réification ou plutôt de binding.

Répondre à cette question n’est pas évident, et il est probable que cela corresponde en partie aux deux. Néanmoins, un certain nombre de caractéristiques encodées dans ces thèmes font écho aux caractéristiques du modèle DBR, telles que les émotions, les contextes sociaux, la mémoire ou encore la motivation (Zimmer et al., 2006). Par ailleurs, même s’ils font référence aux deux, ces trois thèmes sont particulièrement liés à l’homme, à son interprétation des événements et aux biais cognitifs auxquels il fait face pour prendre des décisions. Tous ces éléments nous suggèrent d’attribuer les caractéristiques des thèmes “humain”, “parties prenantes”, et “culture d’entreprise” au processus de binding (voir table 4.3). Cependant, tout comme les caractéristiques de l’environnement attribuées à la saisie de données affectent également l’interprétation de celles-ci en information, l’inverse est également vrai.

C’est d’autant plus vrai qu’il y a également souvent, pour ces thèmes, une pro-

blématique ultime de réification. Par exemple, dans le cadre de ces comités, toutes ces informations, données et avis divergents, doivent converger en un nombre réduit d'indicateurs, i.e., de données, ayant une granularité très réduite (“vert”, “orange”, “rouge”). Ce point a été critiqué par de nombreux chefs de projet estimant qu'il n'était pas possible de montrer des variations dans les projets sans s'attirer des ennuis. Cela les poussait donc à mettre toujours la note la plus élevée en cas de doute, car la diminution de celle-ci serait trop radicale. Cet exemple montre bien ici la relation qui existe entre le processus de binding et de réification, puisque la définition de la donnée, relative à la conception du SI donc, a un impact rétroactif sur la manière de créer l'information au préalable. Cela est représenté dans le modèle DBR par le fait qu'un des facteurs ayant un effet important sur le binding est l'action envisagée, qui, dans ce cas, est la réification<sup>8</sup>. La réification a donc bien un effet indirect sur le processus de binding. Cette interdépendance entre le binding et la réification est importante à prendre en compte, puisque cela signifie que les caractéristiques de l'environnement affectant le processus de réification ont une influence indirecte sur le binding et inversement.

In fine, il semble raisonnable de penser que l'information à la fois s'altère et s'enrichit à chaque itération de binding et de réification. Elle s'enrichit, car elle est le fruit de la combinaison de plusieurs informations et données, provenant parfois de plusieurs personnes différentes, et elle s'altère, car elle subit un certain nombre de biais au moment du binding et de la réification, et de pertes d'informations au moment de la réification dues, entre autres, à l'agrégation qui doit être faite.

L'analyse des résultats des entretiens nous a donc permis d'identifier 31 caractéristiques de l'environnement, réparties en 6 thèmes, ayant un effet sur la saisie de données et la création d'informations, et donc sur les processus de binding et de réification introduits dans le modèle DBR. L'essentiel de cette analyse est résumé dans la table 4.3.

## 4.5 Discussion

Comme nous l'avons déjà vu, il existe un lien très fort entre les caractéristiques de l'environnement dans lequel l'action est réalisée et l'utilisation de raisonnement cognitif impliquant des biais cognitifs et des heuristiques. En effet, le concept de rationalité écologique décrit la correspondance entre une structure et une représentation de l'information dans l'environnement, d'une part, et l'algorithme simple de prise de décision de l'individu, comprenant les heuristiques et les biais cognitifs, d'autre part.

Ainsi, toutes les caractéristiques référencées dans la table 4.1 permettent de conceptualiser la structure de l'environnement dans lequel les individus doivent prendre des décisions et émettre des jugements relatifs à la production de données, en l'occurrence, ici, liés à l'évaluation des performances des projets d'ingénierie. Ces caractéristiques ont donc, comme vu lors de l'analyse des entretiens, un effet non

---

8. A noter qu'en principe, au dernier niveau de la hiérarchie organisationnelle, le processus de réification n'a pas lieu, mais est remplacé par celui de la décision, opération qui affecte directement le réel et non plus le SI, si ce n'est de manière indirecte.

Réification	Binding
Information et donnée	Humain
Agrégation	Peur
Définition	Émotion
Malléabilité	Motivation
Pertinence	Confiance
Granularité	Proximité
Représentativité	Transparence
Réputation	
Projet	Parties prenantes
Type d'engagement	Direction
Type d'activité	Client
Visibilité	Département Qualité
Durée	Autres interlocuteurs
Maîtrise	
Collecte de données	Culture d'entreprise
Nature obligatoire	Succès
Ergonomie	Marge et croissance
Timing	Green attitude
Récurrence	Politique
Chronophage	

TABLE 4.3 – Thèmes et modèle DBR

négligeable sur la qualité des données et des informations produites par les chefs de projet.

Par ailleurs, cette structure d'environnement peut également être considérée comme une architecture de choix au sens de Thaler ([Thaler & Sunstein, 2009](#)), qui peut être mise à profit pour améliorer la prise de décision et le jugement des individus. Ainsi, au-delà de permettre de mieux comprendre et conceptualiser le processus de binding et de réification, ces caractéristiques peuvent être exploitées pour agir sur le processus même de collecte de données via ce que Richard Thaler appellerait le *nudge* (coup de pouce).

Certes, la mise en œuvre d'une méthode liée au *nudge* nécessiterait, dans le cas d'ALTEN, de modifier de nombreux aspects de l'entreprise, y compris les processus organisationnels et le système d'information lui-même. Dans le domaine de la science de l'information, l'adaptation de la partie de l'environnement liée au système d'information pour obtenir de meilleurs résultats a souvent été au cœur des domaines d'étude tels que l'acceptation des utilisateurs ([Venkatesh et al., 2003](#)) ou la gamification ([Koivisto & Hamari, 2019](#)).

Un autre enseignement important à tirer de cette étude de cas est qu'il paraît pertinent de considérer que la saisie de la donnée se produit à travers une série de processus de binding et de réifications successifs, permettant de raffiner le résultat de ce processus jusqu'à ce qu'il soit considéré comme satisfaisant pour l'action envisagée par le chef de projet. D'après les échanges avec les chefs de projet, il s'avère qu'ils

ont un grand nombre d'indicateurs et de données à saisir via le SI, et ce, de manière régulière. Cela signifie que les chefs de projet en charge de cette tâche répètent le processus de binding et de réification associé à cet ensemble de tâches.

Or, quand une action est répétée, elle peut conduire à une forme de conditionnement, le système cognitif de l'individu associant les émotions, positives ou négatives, qu'il ressent lorsqu'il effectue cette tâche à celle-ci. Si le système n'est pas agréable à utiliser ou si les tâches sont perçues comme une corvée par les chefs de projet, cela génère une émotion négative pendant et à l'issue de chaque saisie, ce qui, avec la répétition, cause une diminution de l'efficacité du processus de binding associé à la saisie de données. Les entretiens menés suggèrent que c'est d'ailleurs effectivement le cas : les chefs de projet estiment que la saisie des indicateurs est une perte de temps et que le système n'est ni toujours adapté aux métiers ni ergonomique. Dans ce contexte, la motivation pour effectuer un travail de qualité diminue également, ce qui a une influence sur le processus de binding, la motivation étant un des régulateurs du binding. Le chapitre suivant illustre l'importance de cet aspect motivationnel, aussi bien au niveau individuel qu'organisationnel.

## Chapitre 5

# Effet de la motivation sur la qualité de l'information : application du modèle DBR

Le modèle DBR, introduit dans le chapitre 3, introduit les concepts théoriques de binding et de réification caractérisant les processus de transformation de la donnée en information et de l'information en donnée. L'utilisation de la *Grounded theory* et l'appui sur le domaine de l'économie comportementale et de la psychologie cognitive ont ensuite permis d'identifier, dans le chapitre 4, les caractéristiques de l'environnement qui interviennent dans le processus de binding et de réification et qui modulent ainsi la production de données et d'informations.

Une des retombées majeures d'une bonne compréhension et de l'identification de la structure de l'environnement qui régit la production de données et d'informations est de permettre d'agir sur ses différentes caractéristiques afin d'améliorer la qualité de celles-ci. Cependant, agir simultanément sur toutes les caractéristiques de cet environnement ne permettrait pas d'évaluer l'importance respective de chacun de ces traits. Cette information est pourtant d'importance quand il s'agit de mettre à jour le système complexe qu'est une organisation, en particulier quand il s'agit de prioriser les actions en fonction de leurs impacts financiers (coût, rendement). De plus, modifier des aspects de la culture d'entreprise, la manière dont les données sont saisies dans le système d'information ou la définition des KPI, pour ne donner que quelques exemples, demanderait non seulement des moyens très importants, mais aurait également des répercussions multiples sur d'autres facettes de l'entreprise.

Nous avons donc choisi, pour "challenger" expérimentalement le modèle DBR et son impact sur la qualité des données et informations et les opportunités qu'il offre d'améliorer ces dernières, de nous concentrer sur une unique caractéristique. Nous avons défini et suivi un protocole expérimental, reposant entre autres sur les méthodes d'A/B testing et sur le cadre conceptuel de la qualité des données et des informations proposé précédemment dans ce mémoire, pour tester individuellement les caractéristiques du modèle DBR, en commençant par la motivation.

Le choix de la motivation n'a pas été fait par hasard. Celle-ci a, en effet, été identifiée comme étant un régulateur du processus de binding, étudié dans le chapitre 3, mais également un facteur pouvant avoir un effet sur la saisie de données

et la qualité de celles-ci. Dans la littérature, la motivation est, par ailleurs, un facteur généralement associé à la productivité et l'efficacité (Ryan & Deci, 2000), à la conception de systèmes d'information dans le domaine de la gamification (Koivisto & Hamari, 2019), au partage et à la production d'information (Friedrich et al., 2020) ou, encore, à la prise de décision en économie comportementale (Thaler & Sunstein, 2009).

Dans le domaine de la gamification, les recherches se concentrent sur les facteurs de réussite des systèmes d'information hédoniques pour les exporter dans la conception de systèmes d'information motivationnels (Hamari, Hassan, & Dias, 2018). Dans le domaine des technologies de l'information, de nombreuses recherches analysent l'utilisation et la mise en place de systèmes d'information utilitaristes dans différents contextes organisationnels afin d'identifier les facteurs améliorant leur acceptation (Venkatesh et al., 2003). Dans le domaine de l'économie comportementale et du *nudge*, comme on l'a vu, l'accent est mis sur la mise à disposition intelligente d'informations afin d'améliorer (selon le concepteur du système) la qualité du jugement fait par les individus. Les méthodes de *nudge* peuvent également être utilisées pour influencer les choix des individus en utilisant avantageusement les biais cognitifs de ces derniers.

Ce chapitre propose donc la mise en place d'une expérimentation permettant de tester dans quelle mesure la motivation des personnes en charge de la production de données peut avoir un effet sur la qualité intrinsèque de l'information produite, telle que définie dans le modèle unifié de la qualité présenté dans le chapitre 3. Plus précisément, c'est l'entropie de l'information et sa vraisemblance qui seront évaluées, deux caractéristiques de la qualité intrinsèque de l'information telles qu'identifiées dans le modèle DBR.

## 5.1 Contexte

L'expérience ici présentée est le fruit de la collaboration en cours avec l'entreprise d'ingénierie scientifique ALTEN, qui dispose d'importantes quantités de données et de moyens importants permettant d'étudier le lien entre la motivation des chefs de projets et la qualité intrinsèque des informations collectées, en l'occurrence celles relatives à l'avancement de centaines de projets pour le compte de clients ou en interne.

En effet, rappelons qu'ALTEN est équipée d'un système d'information (SI) avancé qui permet un suivi détaillé de l'évolution des projets en cours. Comme mentionné dans le chapitre 4 (voir section 4.1), lors de l'examen initial des données du projet stockées dans le SI, à savoir divers indicateurs de suivi de la gestion de projet d'ingénierie (coût, délai, service, qualité), nous avons mis en évidence ce qui semblait être un problème lié à la qualité fondamentale de l'information. Les valeurs des indicateurs, fournies chaque mois par les chefs de projet concernant les risques et l'état d'avancement de chaque projet, affichaient une très faible dispersion et donc une faible entropie, ce qui semblait peu en accord avec les valeurs généralement citées dans la littérature (Johnson, 2020).

Nous présentons ici le résultat d'une expérimentation qui entend quantifier, par la mise en place d'une communication différenciée par courrier électronique (email) au

sein de ALTEN, l'impact de la motivation sur la qualité intrinsèque de l'information issue de la collecte des données expérimentales. Cette expérimentation s'est déroulée de juillet à décembre 2021 et a impliqué 177 chefs de projet responsables de plusieurs centaines de projets en cours, avec un taux de participation d'en moyenne 43 %.

Le protocole expérimental choisi, détaillé ci-dessous, repose sur la méthode d'A/B testing et sur les méthodes de régulation de motivation de (Ryan & Deci, 2000) et, enfin, sur le cadre conceptuel de la qualité des données et des informations défini dans le chapitre 3 (voir en particulier la figure 3.4). Conformément aux meilleures pratiques de la méthode d'A/B testing (Kohavi, Longbotham, Sommerfield, & Henne, 2009), deux situations ont été comparées, identiques à l'exception du facteur dont on souhaite observer l'effet, en l'occurrence le type de régulation de la motivation utilisé pour demander aux utilisateurs de saisir les données des projets qui les concernent. Si une différence statistiquement significative est observée entre les deux groupes, pour un critère prédéfini, l'hypothèse selon laquelle le facteur observé a influencé ce critère pourra être validée ; dans le cas contraire, elle sera rejetée. La mesure des effets du type de régulation sur la qualité des données et des informations est effectuée en fonction des caractéristiques du cadre conceptuel de la qualité des données et des informations introduit en chapitre 4.

En pratique, les deux situations sont différenciées par l'envoi de deux courriers distincts sollicitant les collecteurs de données, l'un recourant au langage *corporate* traditionnellement utilisé au sein d'ALTEN pour demander aux collaborateurs de remplir un formulaire, le second faisant de même, mais en régulant la motivation des collaborateurs selon les théories de la motivation mobilisées ici et précisées ci-dessous. Dans le premier cas, le message s'appuie sur le concept de motivation extrinsèque et de régulation contrôlée, mobilisant des incitations telles que l'obligation ou le contrôle de performance. Dans le second cas, le message s'appuie sur le concept de motivation intrinsèque et la régulation autonome ; l'action est cette fois-ci volontaire et repose sur des incitations telles que l'intégration et l'internalisation des valeurs et objectifs de la démarche proposée. Ce dernier message profitera également d'un style de texte plus libre et de graphiques motivants afin de le rendre plus attractif, permettant ainsi de stimuler davantage la motivation intrinsèque.

Deux hypothèses sont donc testées à travers cette expérience contrôlée :

- l'utilisation de méthodes de régulation de la motivation a une influence sur la quantité de données produites, caractérisée par le taux de réponse, la vitesse de réponse et la quantité de données collectées (H1) ;
- l'utilisation des mêmes méthodes a une influence sur la qualité intrinsèque de l'information, plus précisément, sur son entropie, mesurée par la tendance centrale et la dispersion des données, et sa vraisemblance, mesurée par un indicateur conçu sur mesure (H2).

L'hypothèse H2, si elle est validée, permettrait de confirmer que la régulation de la motivation, l'une des caractéristiques primaires du binding, a bien une influence significative sur la qualité intrinsèque des informations qui en découlent, caractérisée ici par l'entropie et la vraisemblance de celles-ci. L'hypothèse H1 est testée en complément, car la littérature dans ce domaine suggère, sans pour autant le démontrer, que l'utilisation de méthodes de régulation autonome (reposant sur la motivation

intrinsèque) peut augmenter la qualité de l'information au détriment de la quantité produite, et inversement avec la régulation contrôlée (Friedrich et al., 2020). Si cet effet secondaire est avéré, ce serait un paramètre à prendre en compte lors de l'utilisation de telles méthodes.

Dans la suite de ce chapitre, nous introduisons rapidement les méthodes de régulation de la motivation dans le contexte de la collecte de données en entreprise et les fondements de la méthodologie de l'A/B testing. Nous précisons ensuite le protocole expérimental de la collecte de données, avec le facteur clé du type de courrier électronique utilisé pour cela, avant d'en présenter les résultats, infirmant H1 mais confirmant H2. Nous discutons de ceux-ci avant de conclure.

## 5.2 Outils méthodologiques

Dans cette section, nous présentons la méthodologie des tests A/B et les méthodes de régulation de motivation utilisées, en particulier en les mettant en perspective avec le cadre du monde de l'entreprise.

### 5.2.1 *A/B Testing*

Les expériences dites contrôlées, telles que les tests A/B (*A/B Testing*), constituent le meilleur moyen scientifique d'établir une relation de causalité entre un changement de situation et son effet, dans notre cas, sur le comportement de l'utilisateur. La manière la plus simple de mettre en place une expérience contrôlée consiste à sélectionner au hasard des utilisateurs et à les affecter à l'un de deux groupes :

- le groupe de contrôle, souvent représentatif d'une version de la situation existante ;
- le groupe de traitement, souvent assigné à une nouvelle situation (version), qui est celle évaluée.

Des mesures telles que la performance d'exécution, le comportement de l'utilisateur et les résultats de l'enquête sont collectées. Des critères d'évaluation globale (*Overall Evaluation Criterion*, ou OEC) sont également définis pour comparer les deux groupes (Kohavi et al., 2009). Si l'expérience a été correctement conçue et exécutée, la seule différence entre les deux variantes devrait être le changement dans la variante de traitement, et donc tout changement observé dans les OEC devrait en découler, établissant ainsi un lien de causalité.

Dans la suite du présent document, nous utiliserons les notions classiques suivantes, rappelées ici par souci d'exhaustivité :

- Critère d'évaluation globale (OEC), une mesure quantitative de l'objectif de l'expérience ;
- Variante, l'une des différentes expériences utilisateur testées, ici le contrôle ou le traitement ;
- L'hypothèse nulle, qui signifie qu'il n'y a pas de différence dans l'OEC entre les variantes de contrôle et de traitement ou, s'il y en a une, qu'elle est due à des variations aléatoires ;

- L’hypothèse alternative, qui fait référence à une différence dans l’OEC entre les variantes de contrôle et de traitement (elle est considérée comme acceptée lorsque l’hypothèse nulle est rejetée avec un niveau de confiance suffisant, généralement 5 %) ;
- Niveau de confiance, qui correspond à la probabilité que la valeur réelle d’un indicateur mesuré de la population testée soit effectivement comprise dans un intervalle de confiance fixé.

Le niveau de confiance est généralement fixé à 95 %, c’est-à-dire qu’il y a 5 % de chances que l’on conclue à une différence significative pour la population étudiée, alors qu’il n’y en a pas. En supposant que tout le reste soit constant, l’abaissement de ce niveau diminue le pouvoir discriminatoire de l’expérience.

## 5.2.2 Régulation de la motivation en entreprise

La propension au partage d’informations des utilisateurs dans une organisation et la qualité de l’information partagée peuvent être influencées positivement par un ensemble d’incitations ayant pour objectif de motiver l’utilisateur du système (Friedrich et al., 2020), (Molina, Unsworth, Hodkiewicz, & Adriasola, 2013), (R. Y. Wang & Strong, 1996). Plusieurs approches existent pour la mise en place d’incitation telles que : l’approche par la gamification, qui emprunte des mécanismes utilisés dans la conception de systèmes hédoniques, c’est-à-dire des systèmes utilisés pour le divertissement qu’ils apportent par leur utilisation même, à des fins utilitaristes (Koivisto & Hamari, 2019) ; l’approche de la théorie de l’autodétermination, qui utilise le concept de régulation pour maximiser la motivation des utilisateurs (Ryan & Deci, 2000) ; ou encore UTAUT, fondée sur l’acceptation des utilisateurs, qui s’intéresse aux facteurs favorisant l’acceptation et l’utilisation d’un système dans les organisations (Venkatesh et al., 2003). Pour différencier les deux groupes lors de l’A/B testing sans avoir à recourir au développement d’un système gamifié complet, le courrier motivationnel utilisé ici s’inspire de l’approche de la théorie de l’autodétermination et de la théorie d’acceptation des utilisateurs UTAUT, qui offrent des moyens d’action plus facilement activables dans notre contexte.

Rappelons (voir aussi chapitre 2) que la théorie de l’autodétermination, ou Self-Determination Theory (SDT) (Ryan & Deci, 2000), adopte une approche qui s’inscrit dans le cadre d’une psychologie cognitive qui étudie la pensée humaine sous le jour du traitement de l’information. La SDT opte pour une approche différenciée de la motivation pour permettre la mise en évidence de facteurs déterminants relatifs aux caractéristiques de l’agent, du contexte et de l’action. Dans cette optique de différenciation, deux types de motivation sont distingués :

- la motivation intrinsèque, qui caractérise un engagement libre de l’individu pour ce qu’une activité lui apporte en elle-même en termes de plaisir, satisfaction, détente, etc. ;
- la motivation extrinsèque, qui caractérise un engagement incité par des facteurs externes à l’activité elle-même comme les récompenses, la pression sociale ou la compétition.

Si la SDT s'intéresse aux facteurs qui favorisent respectivement la motivation intrinsèque et la motivation extrinsèque, il est important de comprendre lequel de ces deux types de motivation affecte la collecte de données dans le contexte d'une entreprise. Dans ce cas, on se doit de se demander ce qui motive les utilisateurs à participer à la collecte de données. La théorie unifiée de l'acceptation et de l'utilisation des technologies (UTAUT) analyse les facteurs qui affectent spécifiquement l'intention d'utiliser un système d'information dans une entreprise. Fondée sur une comparaison de plusieurs modèles d'acceptation des utilisateurs existants, elle met en évidence trois facteurs principaux déterminant l'intention d'utilisation : l'espérance d'effort, l'espérance de performance et l'influence sociale (Venkatesh et al., 2003).

Dans l'entreprise ALTEN étudiée, la participation à la collecte des données et l'utilisation du SI sont obligatoires, car imposées par la direction, et la motivation induite par une telle obligation est, par définition, extrinsèque. Les utilisateurs étant extrinsèquement motivés, il pourrait être bénéfique, ne serait-ce que pour améliorer des facteurs tels que la participation au travail ou la qualité de vie sur le lieu de travail, d'agir sur cette motivation extrinsèque afin d'essayer de motiver plus intrinsèquement les utilisateurs impliqués dans la production et la saisie des données. La SDT propose deux processus pour influencer la motivation extrinsèque, appelés régulation contrôlée et régulation autonome.

- La régulation contrôlée est liée à l'obtention d'une récompense, à l'évitement d'une punition ou à la diminution de la pression sociale pour éviter l'anxiété ou satisfaire l'ego, par exemple.
- La régulation autonome, liée à l'intégration des objectifs et à l'intériorisation des valeurs générales sous-jacentes, va plus loin et devrait même permettre d'atteindre un stade de motivation extrinsèque proche de la motivation intrinsèque, appelé "auto-déterminé", et généralement considérée comme plus forte et plus productive.

Ce sont sur ces facteurs que se propose de jouer l'expérience réalisée chez ALTEN selon le protocole décrit ci-dessous.

## 5.3 Protocole expérimental

Les concepts de régulation de la SDT sont utilisés ici pour différencier les groupes dans les tests A/B, en se référant à la régulation contrôlée, pour le groupe qui recevra le courrier *corporate*, et à la régulation autonome, pour le groupe cible du courrier *motivationnel*.

La méthodologie du test A/B utilisée et les OEC étudiées sont détaillées ci-dessous.

### 5.3.1 Méthodologie

Le test A/B de l'impact de la motivation sur la collecte des données du projet a été réalisé selon la méthodologie et les conditions spécifiées par Kohavi et al. dans "Controlled experiments on the web : survey and practical guide" (Kohavi et al.,

2009). Validée par de nombreuses expériences contrôlées chez Amazon, Microsoft et la NASA, cette méthode est bien adaptée à l’environnement des grandes entreprises comme ALTEN.

La première phase de l’expérimentation a consisté à demander à l’ensemble des 177 chefs et directeurs de projets de ALTEN de labelliser tous les 880 projets qu’ils gèrent ou ont géré via l’interface spécialement conçue à cet effet (voir figure 5.1, les traits noirs étant utilisés pour des raisons de confidentialité) dans l’outil de gestion SMART habituellement utilisé par l’entreprise. Pour collecter les données relatives aux projets, l’interface demande à l’utilisateur, pour chaque projet, de répondre à 3 questions à choix fermé (voir la figure 5.2, où les traits noirs sont utilisés pour des raisons de confidentialité).

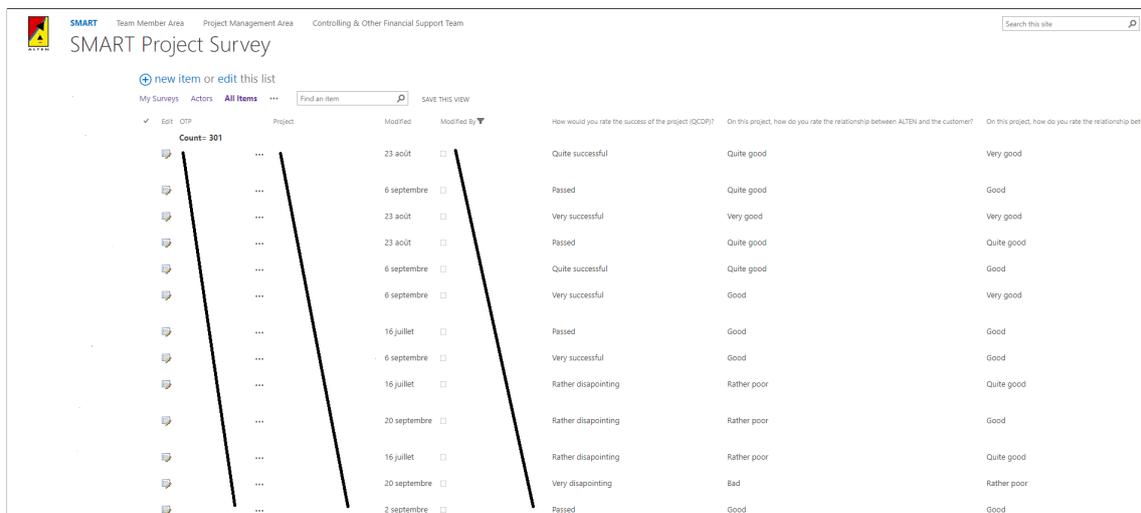


FIGURE 5.1 – Interface SMART de gestion des projets

Les utilisateurs ont accès à une interface dynamique personnalisée comprenant la liste des projets qu’ils ont supervisés dans le passé ou supervisent actuellement. Lorsque l’utilisateur clique sur un projet, une fenêtre s’ouvre lui permettant d’évaluer le projet à travers 3 questions à choix multiple comme le montre la figure 5.2. Le choix d’échelles de Likert évite les réponses neutres et oblige les chefs de projets à prendre parti.

La première question concerne la réussite du projet, définie en termes de Qualité, Coût, Livraison et Performance (QCDP), sur une échelle de Likert de 1 à 6.

La seconde porte sur la qualité de la relation client, également sur une échelle de Likert de 1 à 6, laissant volontairement les responsables libres d’interpréter le sens exact de cette notion, difficile à préciser, en particulier pour les projets en cours.

La dernière question concerne la maturité perçue de l’entreprise par rapport au projet, sur une échelle de Likert de 1 à 10, 1 correspondant à une forte maturité et 10 à une faible maturité. Andersen et Jessen (2003) définissent “mature’ comme étant mûr ou ayant atteint l’état de plein développement naturel ou maximal. La maturité est la qualité ou l’état d’être mature. Si nous appliquons le concept de maturité à une organisation, il peut s’agir d’un état dans lequel l’organisation est en parfaite condition pour atteindre ses objectifs. La maturité d’un projet signifierait alors que

*l'organisation est parfaitement préparée à gérer ses projets.”<sup>1</sup>*

Smart<sub>PS</sub>

OTP \*

Project

How would you rate the success of the project (QCDP)?

Quite successful

On a scale of 1 (Very well known) to 10 (New), do you consider this to be a type of project usually manage by ALTEN?

2

On this project, how do you rate the relationship between ALTEN and the customer?

Quite good

Close

FIGURE 5.2 – Interface SMART pour un projet particulier

La population testée a été divisée de manière aléatoire en deux groupes : un groupe de contrôle de 86 membres et un groupe de traitement de 91 membres<sup>2</sup>. Le groupe de traitement a reçu des messages électroniques initiaux et de suivi conçus selon les principes de la régulation autonome (voir annexes B.1 et B.2), tandis que le groupe de contrôle a reçu des messages électroniques initiaux et de suivi conçus selon les principes de la régulation contrôlée (B.3 et B.4). Le groupe de contrôle a été affecté par une régulation contrôlée, et donc motivé par la simple nature obligatoire de la demande émanant de la direction générale pour cette collecte de données. Le groupe de traitement, qui sera également appelé groupe autonome, a été affecté par une régulation autonome. Le caractère obligatoire de la collecte de données n'est pas mentionné, la présentation est rendue un peu plus attrayante et les cadres sont sensibilisés au fait que leur action sera (jugée) utile et efficace pour eux-mêmes et pour d'autres, même au-delà du champ d'action limité de ALTEN. La raison de cette collecte de données et l'explication de son utilité sont détaillées dans le contenu de l'email, qui est donc plus long que celui du groupe de contrôle et plus illustré.

Afin de s'assurer que le calendrier des envois n'influence pas les résultats, chacun des deux groupes a, à nouveau, été divisés en deux, et les envois ont été effectués en deux étapes, 1 et 2, qui suivront le même processus mais avec une date de lancement à deux mois d'intervalle, comme indiqué sur la figure 5.3.

Le différentiel attendu d'action de ce courrier électronique sur les deux groupes provient du fait que la première raison pour laquelle un individu internalise une activité à laquelle il participe sur la base de justifications externes est le sentiment que celle-ci est perçue comme efficace et utile par les autres personnes de son entourage (Ryan & Deci, 2000). Par ailleurs, donner à un individu les moyens de comprendre le

1. Traduction de l'auteur.

2. La différence de taille des deux groupes est liée au départ de l'entreprise de certains collaborateurs au cours de l'expérimentation.

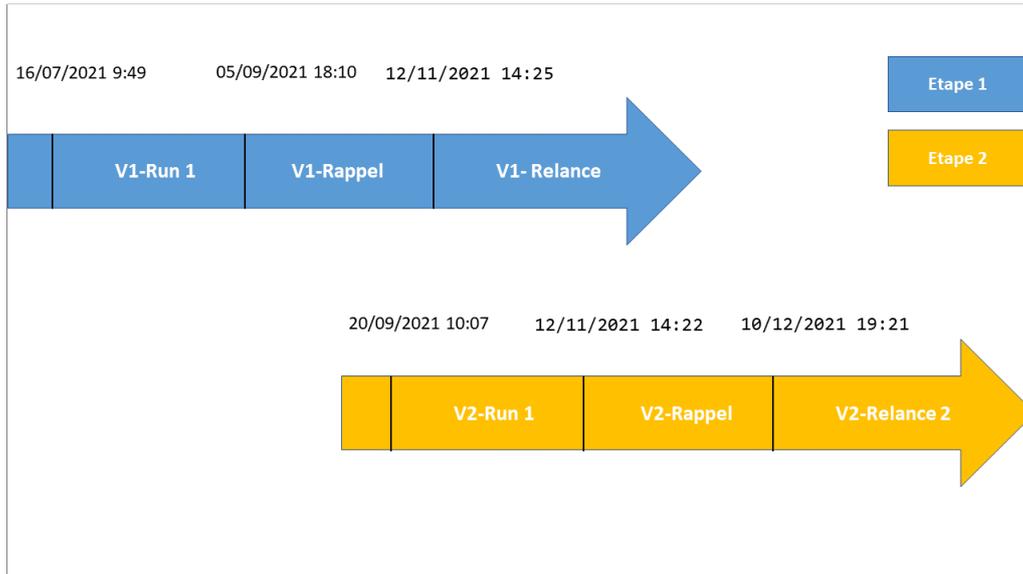


FIGURE 5.3 – Calendrier d’envoi des courriers électroniques

sens de son activité lui permet potentiellement de la fusionner avec ses propres objectifs et valeurs, et donc de favoriser l’intégration et l’internalisation de sa motivation (Kuhl & Fuhrmann, 1998).

### 5.3.2 Définition des critères d’évaluation globaux (OEC)

Pour analyser statistiquement les différences éventuelles entre les deux groupes de l’expérience (groupe de contrôle et groupe autonome), des OEC sont définis pour mesurer, respectivement, la quantité de données produites et la qualité intrinsèque de l’information. La quantité de données produite est mesurée par trois OEC, à savoir :

- le temps de réponse moyen des utilisateurs (en jours) ;
- le taux de participation au processus de collecte ;
- le nombre de projets labellisés, caractérisé par leur nombre et leur proportion.

Pour évaluer la qualité intrinsèque de l’information, les caractéristiques du modèle unifié que sont l’entropie, la vraisemblance et l’objectivité peuvent être utilisées. La compréhensibilité est exclue, car les données de l’expérience ne permettent pas de la mesurer.

L’incertitude est utilisée pour définir une mesure de l’information, que Shannon appelle “entropie”, par analogie avec la notion similaire étudiée en thermodynamique et en physique statistique. De manière intuitive, on peut dire que plus la source émet d’informations différentes, plus l’entropie (ou incertitude sur ce que la source émet) est grande. Cela se traduit notamment par le fait que les données peuvent être caractérisées par une très forte redondance, auquel cas chaque donnée ne contiendra potentiellement pas d’information, si celle-ci est prévisible (Rioul, 2020). Il convient de garder à l’esprit que la mesure de l’entropie de Shannon a été conçue dans un contexte de télécommunication afin d’optimiser le nombre de bits dans un message,

tel que cela peut être le cas pour la compression de données. L’aspect sémantique étant par ailleurs écarté, il ne prend pas en compte les connaissances pré-existantes de l’individu qui reçoit la donnée (voir chapitre 2). Par exemple des données contenant des informations sur la météo des 15 prochains jours pourrait avoir une forte entropie, mais, si le destinataire connaît déjà les prévisions météorologiques, cela ne lui apportera aucune information. Dans un contexte où les destinataires reçoivent régulièrement des données, maximiser l’entropie reviendrait à produire des données éloignées en termes de valeur des données habituellement reçues. En reprenant l’exemple de la météo, si les premières prévisions indiquaient du beau temps, une donnée indiquant le contraire, un orage par exemple, sera susceptible de le surprendre et de représenter une information ayant plus de valeur. Dans le contexte d’ALTEN, nous avons accès à toutes les évaluations produites par les chefs de projet depuis la création du système d’information, ce qui nous donne une indication fiable sur la distribution des données précédemment produites.

La vraisemblance est une caractéristique directement liée à l’interprétation humaine et, donc, au processus de binding. Dans le cas de la vraisemblance, l’interprétation de la donnée est liée aux croyances sur le monde de l’individu et ses connaissances, afin de déterminer si l’information est crédible et vraie à ses yeux (Blitman, 2020). Néanmoins, il n’y a pas de méthode simple pour définir la “croyance sur le monde de l’individu”, mais une approximation peut être utilisée en se référant à la norme dans le domaine auquel se réfère l’information. C’est d’ailleurs la méthode utilisée en économie comportementale et en psychologie cognitive, initiée par Kahneman et Tversky en 1970, et très largement utilisée par la suite (Haselton et al., 2015). Elle consiste à créer un ensemble de réponses théoriques statistiquement normées et de les comparer aux résultats expérimentaux pour déterminer si les individus ont été affectés ou non par des biais. En l’occurrence en ce qui concerne les indicateurs de performance, le Chaos Report semble être une référence solide, que nous postulons représenter la norme statistique (Johnson, 2020).

L’objectivité est difficile à définir, mais elle peut se caractériser comme étant l’inverse de la subjectivité. Sachant que cette expérience suit le protocole de l’A/B testing, la seule différence entre les deux groupes, de taille importante, est supposée être la régulation de la motivation. Or, si l’information était objective, ses valeurs ne devraient pas varier en fonction de la motivation. Montrer que les distributions statistiques entre les deux échantillons sont différentes et que cette différence est statistiquement significative suffit, donc, à suggérer la présence d’une certaine forme de subjectivité dans les informations produites

Ainsi, deux EOC sont respectivement utilisées, dont la définition répond à une observation, faite lors d’une précédente étude d’ALTEN (voir section 4.1), d’un taux très élevé de projets évalués comme ayant d’excellentes performances.

- Le premier critère d’évaluation est fondé sur le test U de Mann-Whitney, en l’occurrence celui présent dans le module python `scipy.stats`, qui évalue la différence entre les distributions empiriques de deux échantillons de données, correspondant respectivement au groupe de contrôle et au groupe autonome. Des distributions statistiques significativement différentes indiquerait que les évaluations de performance des projets ne sont pas objectives, puisque influencées par la motivation. Considérant le taux généralement très élevé de projets

évalués comme ayant d'excellentes performances au sein d'ALTEN, on peut supposer qu'un utilisateur qui consulterait une donnée s'attendrait à voir une évaluation positive. Les valeurs n'indiquant pas une évaluation positive seraient plus surprenantes, et donc une distribution comprenant davantage d'évaluations basses (négatives) indiquerait une entropie plus élevée dans ce contexte. Pour cette raison, c'est un test unilatéral supérieur qui est effectué, c'est-à-dire qu'il teste si la distribution sous-jacente à l'échantillon du groupe contrôle est stochastiquement supérieure à la distribution sous-jacente à l'échantillon du groupe autonome. Si ce test est significatif, alors on peut dire que le groupe autonome a produit des données évaluant moins positivement les projets, ce qui représente une augmentation de l'entropie de l'information.

- Le second critère d'évaluation est la fréquence relative cumulée des deux derniers items de l'échelle de Likert (réponses 5 et 6), qui indiquent une bonne ou excellente performance. La différence entre les deux groupes est testée à l'aide d'un test  $Z$  classique, en l'occurrence celui présent dans le module python `statsmodels`. Cet indicateur spécifique vise à quantifier la tendance (apparente) des chefs de projet à surévaluer les projets, ce qui rentre en opposition avec les valeurs observées dans le Chaos Report, qui produit pourtant ses indicateurs de performance en utilisant les mêmes critères que ALTEN (voir 4.1). Réduire cette tendance reviendrait donc en pratique à améliorer la vraisemblance de ces indicateurs de performance.

Ces indicateurs ne permettent pas de conclure définitivement si les informations du groupe contrôle sont plus ou moins objectives que les données du groupe autonome. Cependant, elles suggèrent, si les données produites par les deux groupes sont différentes, que ces évaluations sont de manière générale subjectives, puisque faire varier la motivation des chefs de projet influencera leur évaluation de la performance des projets. On comprend facilement l'importance d'une telle constatation dans l'écosystème d'une entreprise, et en particulier son SI.

Les projets ayant été labellisés sur trois indicateurs de performance (succès, relations avec les clients et maturité), l'analyse de la qualité intrinsèque de l'information sera donc effectuée respectivement pour chacun de ces indicateurs. L'indicateur de maturité utilise une échelle spécifique ; la comparaison avec le Chaos Report n'est donc pas possible ; la dernière OEC, la vraisemblance, ne sera donc pas utilisée pour cet indicateur.

## 5.4 Résultats

Nous présentons les résultats de l'expérience de collecte de données par envoi différencié de courrier électronique et leur analyse statistique pour les aspects liés à la quantité de données produites et la nature de celles-ci.

### 5.4.1 Quantité de données produites

Nous présentons ici les résultats relatifs à la quantité de données produites, mesurée par temps de réponse, le taux de participation et la quantité de données

produites par personne.

### **Temps de réponse**

Les utilisateurs du groupe contrôle mettent en moyenne 5,01 jours pour répondre, et les utilisateurs du groupe autonome 3,91 jours. Après correction des valeurs aberrantes (temps de réponse supérieur à 1 mois), la moyenne est de 3,48 jours pour le groupe contrôle, et 1,66 pour le groupe traitement. La différence de temps moyen de réponse sans correction entre la version contrôle et la version motivationnelle n'est pas statistiquement significative. L'hypothèse alternative (à l'hypothèse nulle) est largement rejetée avec une p-value de 0,612 pour la différence de moyenne. Cependant, après suppression des valeurs aberrantes (temps de réponse supérieur à 1 mois), la p-value devient 0,096, et on peut donc accepter l'hypothèse alternative avec un niveau de confiance de 10 % ou la rejeter avec un niveau de confiance de 5 % ou moins. Une différence est donc observée entre les deux groupes, tout juste significative après correction des valeurs aberrantes, mais non significative sans correction.

### **Taux de participation**

Sur l'ensemble des deux étapes d'envoi de courrier, 46,5 % des personnes appartenant au groupe contrôle ont répondu à la demande de labellisation des projets et ont labellisé au moins un projet sur l'ensemble de ceux qu'ils gèrent, tandis que 40,7 % des personnes appartenant au groupe autonome ont répondu. Le test donne une p-value de 0,432 ; l'hypothèse nulle ne peut être rejetée, et la différence de taux de réponse entre les deux groupes n'est donc pas significative.

### **Quantité de données labélisées par personne**

Les utilisateurs du groupe contrôle et du groupe autonome ont labellisé en moyenne 5,6 projets par personne. L'hypothèse nulle n'est évidemment pas rejetée, avec une p-value de 0,978. Par ailleurs, le pourcentage de projets labellisés sur le total de projets attribués à chaque personne a été également testé. Les utilisateurs du groupe contrôle ont en moyenne labellisé 62 % de leurs projets, tandis que les utilisateurs du groupe autonome ont en moyenne labellisé 65 % de leurs projets. Cette différence n'est pas statistiquement significative, et l'hypothèse nulle n'est pas rejetée, avec une p-value de 0,744.

## **5.4.2 Qualité intrinsèque de l'information**

Les projets ont été labellisés sur trois indicateurs : le succès, la relation client et la maturité perçue par le chef de projet vis-à-vis du projet. L'analyse de la qualité intrinsèque de l'information sera donc effectuée pour chacun de ces indicateurs avec les deux OEC prévus à cet effet et définis précédemment. Pour rappel, le premier critère, qui évalue l'entropie, est la différence entre les distributions empiriques des deux échantillons, et le second, qui évalue la vraisemblance, est la différence entre les

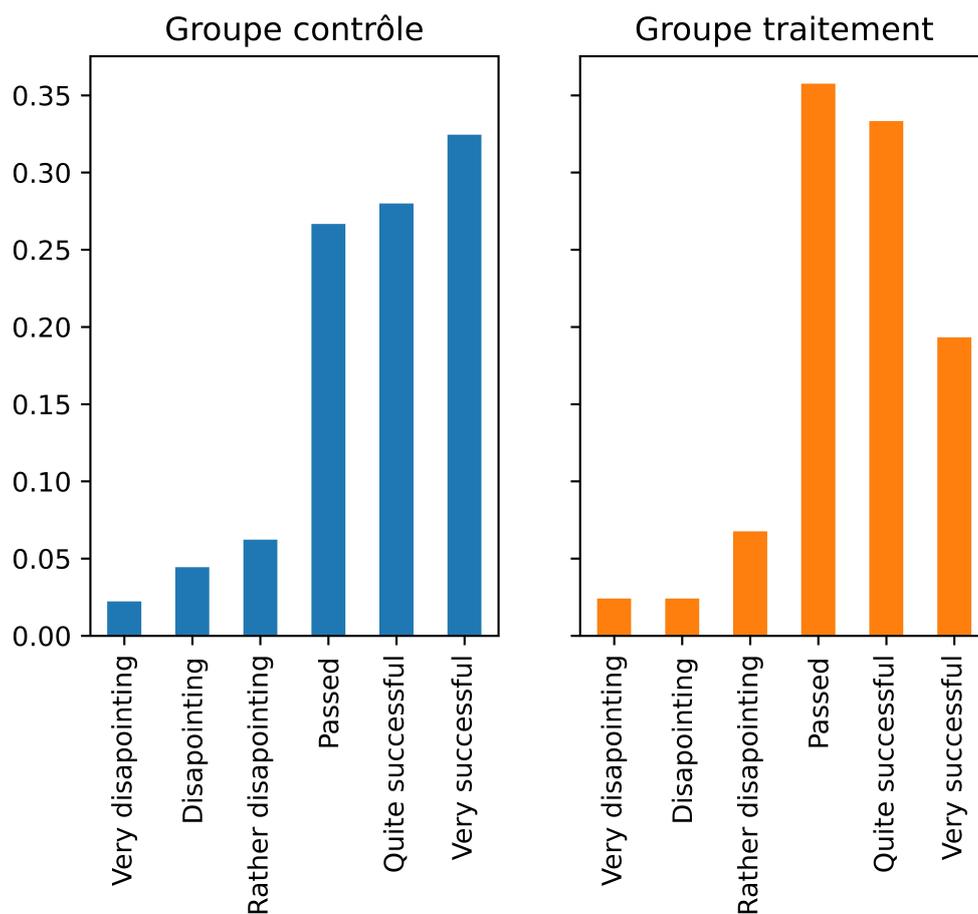


FIGURE 5.4 – Distributions empiriques du succès des projets

deux groupes de la fréquence cumulée des deux derniers items de l'échelle de Likert (réponses 5 et 6).

Les différences entre les deux groupes sont testées avec le test de Mann-Whitney, pour l'entropie, et avec un test de proportion sur la loi normale, pour la vraisemblance (`statsmodels`).

### Succès des projets

Une différence est observée dans la distribution des données entre les deux groupes (voir figure 5.4) avec une tendance à évaluer plus positivement pour le groupe contrôle que traitement. L'indicateur de fréquence cumulée est de 60,4 % pour le groupe contrôle et de 52,7 % pour le groupe traitement. Cette différence est confirmée par le test de Mann-Whitney avec une p-value de 0,015 au test de Mann-Whitney, pour lequel l'hypothèse nulle est donc largement rejetée avec un seuil de confiance de 5%.

Le test de proportion confirme également cette observation, avec une p-value de 0,051 ; cette fois-ci, l'hypothèse nulle est donc rejetée avec un seuil de confiance de 10 % et non rejetée de justesse avec un seuil de confiance de 5 %.

Les utilisateurs du groupe traitement ont donc produit des informations ayant

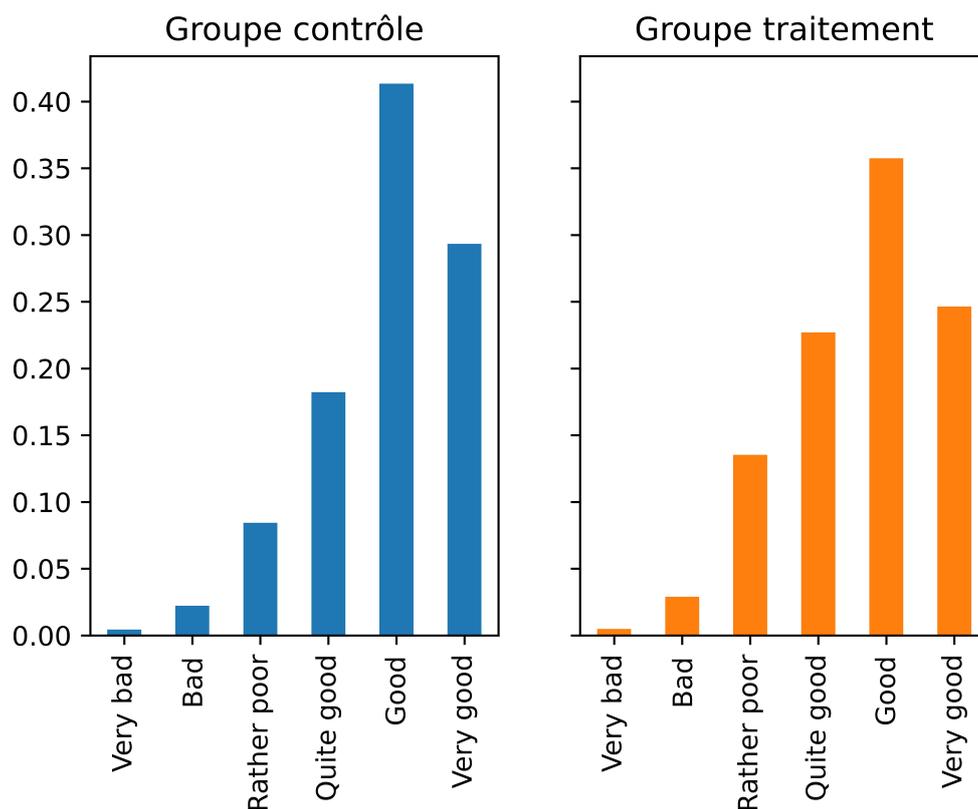


FIGURE 5.5 – Distributions empiriques des indicateurs de qualité de la relation avec le client

une entropie et une vraisemblance plus élevées, ce qui semble indiquer une meilleure qualité intrinsèque. Par ailleurs, les utilisateurs du groupe traitement ont également tendance à évaluer moins positivement les projets que le groupe contrôle, ce qui tend vers la norme que nous avons établie sur la base du Chaos Report, précédemment.

### Relation client

Le même phénomène est observé dans le cas de la relation client (voir figure 5.5), de manière plus marquée encore, avec une tendance, donc, à évaluer plus positivement dans le groupe contrôle que dans le groupe traitement. L'indicateur de fréquence cumulée est de 70,7 % pour le groupe contrôle et de 60,3 %, pour le groupe traitement. La signification statistique de cette différence est à nouveau confirmée par le test de Mann-Whitney, avec une p-value de 0,019, pour lequel l'hypothèse nulle est donc largement rejetée avec un seuil de confiance de 5%.

De plus, le test de proportion donne une p-value de 0,012 ; cette fois-ci, l'hypothèse nulle est donc rejetée avec un seuil de confiance de 5 %.

A nouveau les résultats indiquent une qualité intrinsèque des informations produites par le groupe traitement plus élevée, et les utilisateurs du groupe traitement ont tendance à évaluer moins positivement les projets que le groupe contrôle, ce qui rapproche celle-ci de la norme établie.

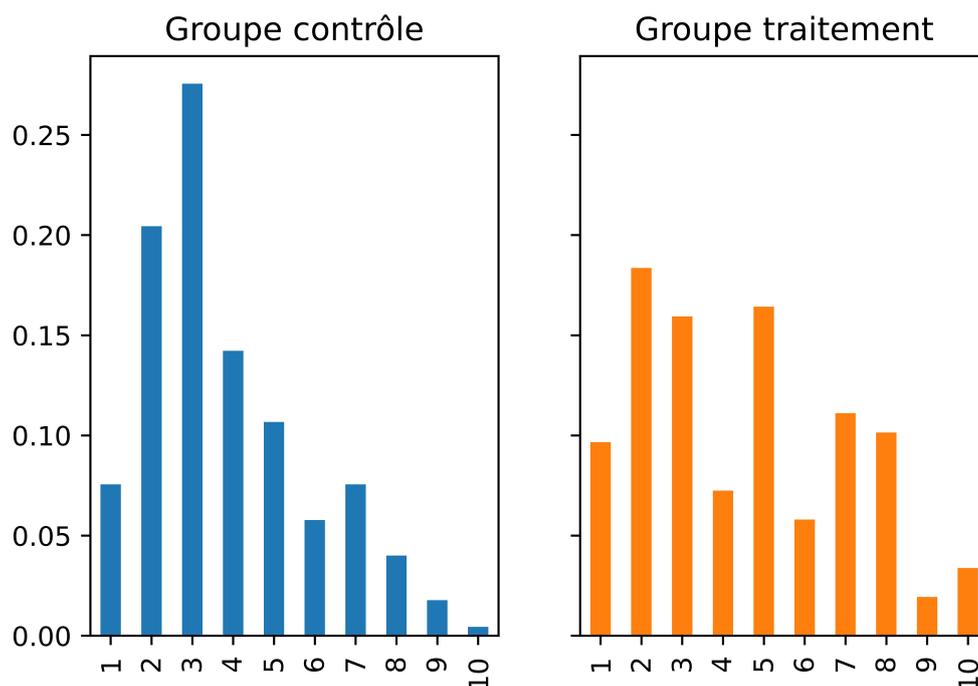


FIGURE 5.6 – Distributions empiriques des indicateurs de maturité des projets

## Maturité

La figure 5.6 montre des dispersions des données de maturité radicalement différentes entre les deux groupes. Le groupe contrôle présente une tendance de réponse concentrée autour de son mode, tandis que le groupe traitement a une distribution beaucoup plus uniforme. Le test de Mann-Whitney confirme cette observation avec une p-value de 0,011 ; l’hypothèse nulle est donc largement rejetée. Les utilisateurs du groupe contrôle ont donc tendance à répondre de manière moins diversifiée que le groupe traitement, avec une plus forte concentration avec les niveaux de maturité faibles pour les projets dont ils sont responsables. Ces résultats indiquent donc bien une entropie plus forte pour le groupe traitement.

Rappelons que la vraisemblance n’est pas testée ici.

## 5.5 Discussion

Au terme de cette expérimentation, les résultats obtenus offrent des pistes intéressantes concernant l’effet de la motivation sur la production de données et d’informations dans le cadre organisationnel. Dans un contexte où la manipulation simultanée de toutes les caractéristiques de l’environnement qui modulent ces processus s’avère impraticable, le choix d’étudier le rôle de la motivation, un facteur déjà reconnu pour son influence dans plusieurs domaines pertinents, s’est révélé instructif.

L’analyse des OEC de mesure de la quantité de données produites montre presque unanimement qu’il n’y a pas de différence significative de taux de participation, de vitesse de collecte et de quantité de données collectées entre les groupes contrôle et traitement. Ce résultat est surprenant, car la littérature sur ce domaine ([Friedrich](#)

et al., 2020) indique que l'utilisation de méthodes de contrôle pour inciter au partage d'information augmente la quantité d'information collectée au détriment de la qualité. Or, nos résultats semblent indiquer que celle-ci n'a pas d'influence sur la quantité de données collectées. D'une manière générale, ce que ces résultats nous indiquent, c'est que l'utilisation de méthodes de contrôle et de méthodes de régulation de la motivation n'a pas d'influence sur la quantité de données produites. Une expérimentation ultérieure avec des échantillons de plus grandes tailles, une autre entreprise, voire une entreprise d'un autre domaine économique, permettraient peut-être de lever la contradiction apparente entre notre étude terrain et l'état de l'art.

En ce qui concerne la qualité intrinsèque de l'information, il y a là bien une différence observée et statistiquement significative : les utilisateurs du groupe contrôle ont davantage cette tendance apparente à surévaluer les projets, comme c'était le cas dans les données produites précédemment par ALTEN. A l'inverse, les utilisateurs du groupe autonome répondent de manière plus diversifiée et évaluent moins positivement leurs projets, ce qui rapproche ces évaluations de celles du Chaos Report. Cette différence interroge bien évidemment quant à la pertinence des données collectées et à leur utilisation en aval pour des processus de décision impactant le management des projets. De fait, l'origine de l'expérimentation présentée ici et des questions de recherche associées était un problème de qualité des données projet de ALTEN (voir section 4.1).

Ce qui avait suscité l'attention était le taux très élevé de projets évalués comme étant des projets ayant des résultats excellents. Cette distribution observée est en contradiction avec la distribution observée dans le Chaos Report (Johnson, 2020), comme mentionné précédemment.

Cette observation et les résultats présentés ici mettent en avant la question de la qualité dite "intrinsèque" de l'information. En l'occurrence, les caractéristiques susceptibles d'avoir fait défaut dans l'analyse initiale des données projet (voir section 4.1) étaient justement la vraisemblance des informations et leur entropie. Plus précisément, l'observation du taux très élevé de projets évalués comme ayant d'excellentes performances était représentative d'une information de très faible entropie. Par ailleurs, cette tendance à surévaluer les projets en comparaison avec les valeurs habituellement observées dans la littérature, et notamment dans le Chaos Report (Johnson, 2020), rendait peu vraisemblables les évaluations faites par les chefs de projet. Or, les données collectées auprès du groupe autonome ont moins tendance à évaluer les projets concernés positivement et indiquent moins de projets comme ayant eu des performances excellentes. En effet, si nous reprenons la comparaison qui avait été faite dans le chapitre 4 de cette thèse, laquelle montrait la différence entre les taux de succès des projets d'ALTEN et ceux du Chaos report, et que nous la comparons aux données collectées auprès du groupe autonome, cette observation est mise en évidence. La distribution observée dans les données du groupe autonome est effectivement nettement plus proche des données du Chaos report que les données initialement présentes dans les bases de données d'ALTEN, comme cela peut être observé dans la figure 5.7. Il est important de noter que, bien que la comparaison entre les données initiales d'ALTEN et celles du Chaos report était triviale, puisque les deux entités utilisent des indicateurs construits de manière très similaire, ce n'est pas le cas des données issues de cette expérience. La différence réside dans le fait

que, pour cette expérience, une échelle comportant 6 modalités a été utilisée, tandis que les précédentes utilisent une échelle comportant 3 modalités. Pour parvenir à réaliser une comparaison, nous avons donc agrégé ces 6 modalités en 3 modalités correspondant à RED, AMBER et GREEN, vues comme équivalentes aux FAILED, CHALLENGED et "SUCCESSFUL du Chaos report.

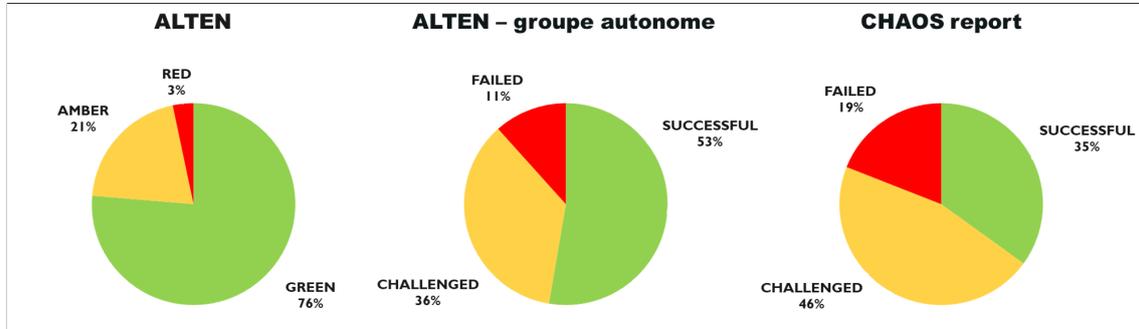


FIGURE 5.7 – Comparaison des indicateurs de suivi du succès des projets

En supposant, encore, que les données du Chaos report font office de norme, les informations produites par le groupe autonome sont donc probablement plus vraisemblables et, en l'occurrence, elle sont également caractérisées par une entropie plus forte. Ces différences pourraient indiquer une amélioration de la qualité intrinsèque de l'information, grâce à l'utilisation de la régulation autonome de la motivation. Cette observation indique que les individus motivés intrinsèquement sont plus susceptibles de produire des informations de qualité, ce qui est cohérent avec les résultats habituellement observés dans la littérature (Friedrich et al., 2020).

Les résultats obtenus par notre expérience comparative semblent confirmer l'effet de la notion de régulation de la motivation sur la collecte de données et la qualité intrinsèque des informations ainsi produites. Un certain nombre d'hypothèses peuvent être formulées quant aux causes de cet écart entre les groupes contrôle et autonome. Étant davantage sensibilisés à l'intérêt de la collecte de données, les utilisateurs du groupe autonome pourraient être plus motivés à répondre de manière réfléchie. Par ailleurs, n'étant soumis à aucune forme de contrôle, il y a peu de chances que leurs réponses soient influencées par des attentes implicites, ou explicites, de la direction. A l'inverse, les utilisateurs du groupe contrôle, ayant été sollicités de manière contrôlée, en indiquant clairement le caractère obligatoire de la collecte de données, ont pu répondre de manière biaisée par appréhension du fait que ces données sont susceptibles d'être observées par la direction. Ces différentes hypothèses font écho aux caractéristiques de l'environnement qui modulent le binding et la réification, selon le modèle DBR. Ces caractéristiques comprennent, par exemple, les réactions des parties prenantes, qui peuvent être liées à la crainte, à la confiance ou encore à la motivation des chefs de projet. En l'occurrence, ces caractéristiques n'ont pas été testées dans le cadre de cette expérience, et donc le lien de causalité ne peut pas être fait ; cependant, les chefs de projet mentionnés dans le chapitre 4 sont soumis au même environnement que celui de l'expérience présente, puisqu'il s'agit de la même entreprise.

En fait, une autre observation, faite en dehors du protocole expérimental, semble indiquer que les données collectées lors de cette expérience, si les échantillons du

groupe contrôle et autonome sont fusionnés, restent très éloignées des données habituellement collectées par l'entreprise. Ces différences sont bien plus importantes que celles observées entre le groupe contrôle et le groupe autonome, ce qui suggère qu'un nombre important d'autres caractéristiques de l'environnement, non maîtrisées dans la présente expérience, ont également affecté cette collecte expérimentale. Cela confirme l'efficacité du protocole de l'A/B testing pour isoler l'effet que l'on souhaite observer, mais cela met en avant également la nécessité qu'il y aurait à tester, par la suite, les autres caractéristiques précédemment identifiées.

## 5.6 Conclusion

Grâce à l'expérience présentée dans ce chapitre, il a été possible d'établir que la régulation de la motivation, et en particulier la régulation autonome, a un effet significatif sur la qualité intrinsèque de l'information produite, spécifiquement en termes d'entropie et de vraisemblance (H2). Cela appuie les théories de l'information selon lesquelles la qualité de l'information collectée peut être améliorée en utilisant des approches qui favorisent la motivation intrinsèque des individus (Friedrich et al., 2020), (Ryan & Deci, 2000). En revanche, l'hypothèse que la régulation de la motivation influe sur la quantité de données produites (H1) n'a pas été validée par l'expérimentation. Ce résultat, bien que contraire à certaines suggestions de la littérature (Friedrich et al., 2020), offre néanmoins des informations précieuses : il semble indiquer que l'adoption de méthodes de régulation autonome n'implique pas nécessairement un compromis entre qualité et quantité d'informations.

Ces résultats ont des implications importantes pour les organisations cherchant à optimiser la qualité de leurs données et informations. Elles suggèrent qu'une révision des approches traditionnelles de collecte de données, souvent fondées sur des motivations extrinsèques, pourrait être bénéfique. En particulier, l'adoption de méthodes de régulation autonome, qui favorisent l'engagement des individus en harmonie avec leurs propres valeurs et objectifs, pourrait améliorer significativement la pertinence des informations collectées pour la gestion de projets.

Il est à noter, toutefois, que si ces résultats sont prometteurs, d'autres études sont nécessaires pour approfondir la compréhension de ces mécanismes et pour explorer d'autres méthodes de régulation de la motivation dans différents contextes organisationnels. La réalisation d'expériences complémentaires et l'application de différentes méthodes de régulation pourraient permettre de généraliser ces conclusions et de développer des recommandations plus précises pour les entreprises.

Plus important encore, ces résultats soulignent de manière significative l'importance du processus de binding dans la création d'informations. En effet, le rôle de la motivation, une des caractéristiques clés du processus de binding, a été mis à l'épreuve et s'est révélé avoir un effet substantiel sur la qualité intrinsèque de l'information produite. Cela suggère également que la manipulation des caractéristiques qui modulent ce processus pourrait être un moyen efficace d'améliorer la qualité des données et informations produites.

Ainsi, ces résultats confirment partiellement la validité du modèle DBR, en montrant que le processus de binding - ou du moins un de ses éléments clés - peut influencer de manière significative la qualité de l'information produite. Par extension,

cela suggère que des expériences similaires pourraient être menées sur les autres caractéristiques du binding et de la réification identifiées précédemment. Ces expérimentations supplémentaires pourraient permettre de confirmer l'intégralité de la pertinence du modèle DBR et de renforcer la confiance dans son utilité pratique pour améliorer la qualité de l'information et des données dans un contexte organisationnel.

En conclusion, cette expérimentation apporte une contribution significative à la compréhension des mécanismes qui sous-tendent la transformation de la donnée en information et vice versa, en soulignant le rôle important du processus de binding. Plus particulièrement, elle confirme l'importance de la motivation, un élément clé de ce processus, comme un levier potentiel pour améliorer la qualité des informations produites au sein des organisations. Finalement, elle ouvre la voie à des recherches futures visant à explorer et à confirmer l'importance des autres caractéristiques du processus de binding et de réification au centre du modèle DBR, en utilisant le même protocole expérimental que celui utilisé ici, de manière fructueuse.

# Chapitre 6

## Conclusion

Ce dernier chapitre met en avant les principales contributions de cette thèse, avant de présenter certaines de limites de ce travail et des perspectives pour de futurs travaux de recherche.

### 6.1 Contributions

Les contributions principales théoriques de cette thèse sont (1) la proposition d'un nouveau modèle, DBR pour *Data Binding and Reification*, qui formalise la relation subtile entre les concepts de données et d'informations, et (2) un cadre conceptuel unifié pour évaluer la qualité des données et informations, en particulier dans les organisations.

Le modèle DBR est ensuite complété, sur un plan plus pratique, par l'exploration des facteurs environnementaux qui influencent les processus de binding et de réification dans les organisations, l'établissement d'un protocole expérimental pour mettre à l'épreuve sa validité et la démonstration empirique de l'impact significatif de la motivation sur la qualité des informations produites par les organisations.

#### 6.1.1 Modèle DBR

Le modèle DBR répond à un manque clairement identifié dans la littérature en sciences de l'information (Rowley, 2007) (Frické, 2009) de cadre théorique et conceptuel des processus de conversion qui régissent la relation entre les concepts de donnée, d'information, et de connaissance. L'enchevêtrement de ces trois concepts peut être étudié à deux niveaux : la relation entre information et donnée, et la relation entre information et connaissance. Étant donné que ces travaux de recherche ne s'intéresse pas au concept de connaissance, notre étude se concentre sur la relation entre donnée et information, plus proche des besoins des SI au coeur de l'entreprise.

Les fondements théoriques sur lesquels repose le modèle sont issus des sciences cognitives (neurosciences, psychologie cognitive et économie comportementale) appliquées aux sciences de l'information. Elle repose, notamment, sur l'importance de la perception et de l'interprétation dans la définition des processus qui sous-tendent la relation entre les concepts de donnée et d'information. Le processus de conversion de la donnée en information est nommé "binding", terme emprunté au neurosciences

où il caractérise le processus d'interprétation des stimulus par l'homme en expérience subjective, sur laquelle se fonde l'homme pour prendre des décisions et agir. Le processus de conversion de l'information en donnée est nommé "réification", aussi appelé "objectivation", qui désigne, en philosophie, le processus par lequel on transforme quelque chose de mouvant, de dynamique, en être fixe, doté de caractéristiques statiques. Étant donné l'objectif de ces travaux de recherche, il est amusant de remarquer qu'en psychologie la réification désigne un sophisme, celui de traiter une abstraction, un concept ou une formulation comme s'il s'agissait d'un objet réel ou d'une chose matérielle, alors qu'il ne l'est pas.

### 6.1.2 Cadre conceptuel de la qualité des données et des informations

En nous fondant sur le cadre conceptuel de la qualité des données de Wang (R. Y. Wang & Strong, 1996) ainsi que sur sa version adaptée par Khan et Wang en 2002 (Kahn et al., 2002), nous avons conçu un nouveau cadre conceptuel de la qualité des données et des informations. Ce dernier se justifie, notamment, par le fait que Wang ne semble pas faire de différence conceptuelle entre données et informations, puisque son premier cadre conceptuel est dédié à la qualité des données (R. Y. Wang & Strong, 1996) et le deuxième, à la qualité des informations (Kahn et al., 2002). Pourtant, les deux modèles utilisent les résultats de la même expérimentation réalisée en 1996 pour définir leur structure.

Notre proposition consiste donc d'abord à unifier ces deux cadres conceptuels pour n'en faire qu'un qui adresse la qualité des données et des informations, en réutilisant les mêmes données issues de l'expérience de 1996, mais en apportant une nouvelle classification fondée sur une approche neuroscientifique des concepts de donnée et d'information. Ainsi, les quatre nouvelles dimensions sont la qualité intrinsèque de l'information, la qualité contextuelle de l'information, la qualité des données et la qualité du système d'information. Deux nouvelles caractéristiques clés sont également proposées : l'entropie et l'absence de biais. La nécessité de ces dernières provient du fait que deux caractéristiques du modèle initial peuvent concerner simultanément les informations et les données, l'une, relative à la quantité, et l'autre, à la présence ou non de biais.

### 6.1.3 Effet de la rationalité écologique sur le modèle DBR

Au regard du principe de rationalité écologique, l'identification des caractéristiques de l'environnement qui modulent a priori les processus de binding et de réification, en prenant un appui expérimental sur ALTEN, est une des contributions majeures pratiques de ces travaux de recherche. En raison du caractère nouveau des processus de binding et de réification, l'approche de la théorie ancrée - qui privilégie une collecte initiale de données sans préjugés afin de construire, par la suite, une théorie - a semblé particulièrement pertinente et adaptée. Comme disait le philosophe Descartes, "pour atteindre la vérité, il faut une fois dans sa vie savoir se défaire de toutes les opinions reçues, et reconstruire de nouveau tout le système de connaissances" (Descartes & Gilson, 1987).

L'étude approfondie et méthodique de 502 verbatims issus d'environ 20 heures d'interviews auprès de 11 chefs et directeurs de projet de l'entreprise d'ingénierie ALTEN a permis la mise à jour de 60 codes. Après une phase de raffinement théorique, 31 caractéristiques expliquant les processus de binding et de réification, réparties en 6 catégories, ont été identifiées. Ces caractéristiques ont alors été interprétées, au regard des sciences comportementales, comme des caractéristiques de l'environnement modulant la production d'informations et de données.

La structure de l'environnement est étroitement liée à l'utilisation par l'homme de biais cognitifs et heuristiques pour répondre à cet environnement, tel que nous l'avons précédemment évoqué au travers du concept de rationalité écologique. Parmi ces caractéristiques, 14 sont essentiellement liées à l'environnement social et culturel dans lequel les individus produisent l'information, et qui affectent donc le processus de binding. Les 17 autres caractéristiques sont davantage liées à l'environnement organisationnel, au processus de collecte de données et au système d'information elles affectent ainsi la production de données et le processus de réification. L'identification de ces caractéristiques permet ainsi de décliner un cas d'application du modèle DBR dans un contexte organisationnel qui pourrait permettre, à terme et à condition de manipuler les caractéristiques qui modulent les processus de binding et de réification, d'améliorer la qualité des données et des informations produites.

### 6.1.4 Protocole expérimental pour le modèle DBR

Notre travail a permis de proposer le modèle DBR, mettant en lumière les processus de binding et de réification qui transforment les données en informations et inversement. Grâce à l'usage de la théorie ancrée, nous avons aussi identifié les éléments environnementaux qui ont une incidence sur ces processus, et qui régulent donc la génération de données et d'informations.

Comprendre et identifier cette structure environnementale est essentiel. En effet, en modifiant ces éléments, il est possible d'envisager d'améliorer la qualité des données et des informations produites. Néanmoins, une modification simultanée de tous ces éléments serait non seulement très coûteuse, mais pourrait aussi avoir un impact varié sur d'autres aspects organisationnels.

Nous avons donc élaboré un protocole expérimental, reposant en partie sur la méthodologie d'A/B testing, qui vise à isoler et mesurer l'effet de la motivation, l'une de des caractéristiques de l'environnement, sur la qualité des données et informations produites. Nous avons plus particulièrement mesuré l'entropie de l'information et sa vraisemblance, deux indicateurs clés de la qualité intrinsèque de l'information. Ce protocole expérimental à visée partielle repose donc sur le modèle DBR et sur le cadre conceptuel unifié de la qualité des données et des informations que nous proposons pour quantifier les impacts des changements environnementaux. Nos résultats valident en partie le modèle DBR, en montrant une influence notable de la motivation sur la qualité de l'information produite. Il semble donc envisageable d'utiliser les grandes lignes de ce même protocole pour étudier d'autres éléments environnementaux affectant le binding et la réification. Ainsi, de futures expérimentations pourraient confirmer (ou non) la validité globale du modèle DBR et renforcer son utilité pour l'amélioration de la qualité des données et des informations en milieu

organisationnel.

### 6.1.5 Effet de la motivation sur la qualité des informations

Au-delà de la proposition d'un protocole expérimental adapté à la validation du modèle DBR, les résultats pratiques de sa mise en oeuvre chez ALTEN auprès d'une population relativement importante (180 chefs et directeurs de projet) ont permis de confirmer sur le terrain que la motivation a un effet significatif sur la qualité des informations produites dans un contexte organisationnel.

Ces résultats pourraient avoir des répercussions majeures pour les organisations qui cherchent à optimiser la qualité de leurs données et informations. Ils suggèrent, en effet, qu'une remise en question des méthodes traditionnelles de collecte de données, qui s'appuient généralement sur des motivations extrinsèques, pourrait être avantageuse. En prolongation des travaux de Ryan et Deci sur la motivation (Ryan & Deci, 2000), cela suggère que la mise en place de méthodes d'autorégulation, qui encouragent une implication des individus en adéquation avec leurs propres valeurs et objectifs, pourrait améliorer considérablement la pertinence des informations recueillies pour la gestion de projets. Cela ne pourrait qu'être bénéfique à la mise en oeuvre de stratégies organisationnelle fondées sur les données.

### 6.1.6 Mise en perspective

L'objectif général de cette thèse était de fournir des éléments nouveaux permettant une meilleure compréhension des mécanismes qui sous-tendent la création d'informations et de données en utilisant une approche qui place l'homme au cœur de ces processus. L'étude scrupuleuse des concepts d'information et de donnée, leur origine, ce qui les définit, ce qui les différencie et ce qui fait leur valeur, permet déjà certainement de mieux comprendre la thématique de cette recherche.

Fruit d'une réflexion s'appuyant sur de nombreux domaines scientifiques avec une approche largement pluridisciplinaire, le modèle DBR conceptualise les processus qui régissent la création de données et d'informations, en intégrant au cœur de ce modèle les particularités du système cognitif humain et sa manière de percevoir les événements cognitifs. Cette nouvelle perspective a mis en avant l'importance des limitations cognitives de l'homme, étudiées principalement dans le domaine de l'économie comportementale et de la psychologie, dans le processus de création des données et des informations.

L'enjeu principal était alors de parvenir à utiliser les sciences cognitives et comportementales pour identifier dans quelle mesure l'intervention de l'homme et la structure de l'environnement auquel il est soumis affectent le processus de création de données et d'informations, notamment dans un contexte organisationnel. Dans la continuité de nombreux travaux dans les domaines de l'économie comportementale, de la psychologie cognitive et évolutionnaire, il a paru pertinent de prendre en compte les principes de rationalité limitée et de rationalité écologique dans la modélisation des caractéristiques du modèle DBR. Ceux-ci soutiennent que l'homme est rationnel dans les limites de ses capacités cognitives et des informations à sa disposition et que, pour prendre des décisions et agir, il utilise des raccourcis cognitifs

(biais et heuristiques) qui dépendent de la structure de son environnement. Grâce à l'utilisation de la théorie ancrée et de nombreuses interviews avec des chefs de projet en charge de la production de données, il a été possible d'identifier un certain nombre de caractéristiques de l'environnement liées respectivement au processus de binding et au processus de réification. Cette identification permet de savoir quels sont les éléments de l'environnement qui modulent l'utilisation de biais cognitifs et d'heuristiques par l'homme dans le cadre de la production de données et d'informations dans un contexte organisationnel.

L'implication majeure de ces travaux consiste dans la proposition d'outils conceptuels permettant d'améliorer la qualité des données et des informations en limitant les biais cognitifs auxquels l'homme a recours dans le cadre de leur création. Pour des raisons évidentes de moyens et de temps, il n'a pas été possible de tester intégralement le modèle DBR pour intervenir sur l'environnement de saisie afin d'améliorer la qualité des données et des informations. Toutefois, un premier test sur le terrain permet de confirmer en pratique, dans un cadre organisationnel, l'effet de la motivation sur la qualité intrinsèque de l'information ; il s'appuie sur un protocole expérimental reposant sur le modèle DBR, le cadre conceptuel de la qualité des données et des informations et sur les méthodes d'A/B testing. Cette expérimentation, en plus de valider partiellement l'applicabilité du modèle DBR dans un contexte organisationnel, ouvre de nombreuses pistes pour des travaux futurs.

## 6.2 Travaux futurs et limitations

La première piste pour des travaux futurs a déjà été évoquée à plusieurs reprises : il s'agit de tester expérimentalement la validité de chacune des caractéristiques du modèle DBR. En effet, chacune des 31 caractéristiques de l'environnement ayant théoriquement un effet sur les processus de binding et de réification nécessiterait une expérimentation à part entière permettant d'isoler l'effet du dit facteur sur les données et informations produites. Ceci nécessiterait, bien évidemment, un temps et un budget considérables, mais pourrait être le sujet de travaux futurs, ne serait-ce que partiels. Cela permettrait à terme de valider intégralement ou en partie le modèle DBR ou de lui apporter des modifications, et, par extension, cela ouvrirait la porte à la possibilité de concevoir une architecture de choix favorisant la saisie de données et la production d'informations de meilleure qualité. En pratique, cette démarche s'inscrirait dans le prolongement de la théorie du *nudge* de Thaler, qui propose d'optimiser l'environnement relatif à une tâche donnée affectée à un individu pour améliorer son jugement et influencer positivement sa décision.

Dans un autre domaine de recherche, celui de l'économie comportementale et de la psychologie cognitive, ces travaux ouvrent des pistes vers une contribution au domaine de l'identification et de la caractérisation des biais cognitifs. En effet, nous avons identifié un certain nombre de caractéristiques de l'environnement ayant de l'influence sur le modèle DBR. Mais ces caractéristiques font, en réalité, référence au principe de rationalité écologique, qui, pour rappel, indique qu'il existe une forte relation entre la structure de l'environnement et l'utilisation d'heuristiques et biais cognitifs par l'homme pour résoudre les problèmes auxquels il est confronté. Dans cette étude, nous nous sommes contentés d'étudier les caractéristiques de l'en-

vironnement, sans identifier précisément les biais cognitifs et heuristiques qu’elles impliquent. La raison principale à cette omission est que l’identification précise des biais sollicités n’aurait pas eu d’implications majeures sur la manière de conceptualiser et d’utiliser le modèle DBR. Cela dit, leur étude pourrait apporter une contribution importante au domaine des sciences cognitives, et notamment à l’économie comportementale.

Nous avons proposé un modèle unifié de la qualité des données et de l’information, qui repose initialement sur les travaux et matériaux de Wang (1996) mais qui prend en compte les apports du modèle DBR sur la compréhension des mécanismes de production de données et informations. Cependant, la plus grande faiblesse de ce modèle est qu’il repose sur un matériel qui date de 1996 et qui avait été produit à l’origine en ignorant la distinction, dont nous avons souligné l’importance, à faire entre les concepts d’information et de donnée. Il serait donc particulièrement intéressant de concevoir une série d’interviews, en utilisant peut-être la théorie ancrée, pour collecter de nouvelles données plus récentes, et ainsi peut-être apporter des modifications à ce cadre conceptuel.

Le modèle DBR que nous avons conçu permet avant tout de combler un gap dans la littérature en sciences de l’information en ce qui concerne la conceptualisation des processus qui régissent la relation entre donnée, information et connaissance (Rowley, 2007) (Frické, 2009). Les chercheurs ayant identifié ce gap ont estimé que cette relation pouvait être étudiée à deux niveaux distincts : donnée - information ; information - connaissance. Dans ce mémoire, nous avons étudié le premier niveau, donnée - information ; il reste donc le second niveau à explorer.

De manière générale, nos travaux de recherche se sont concentrés exclusivement sur les concepts de données et d’information, en excluant presque complètement le concept de connaissance. Il y a donc clairement un besoin de recherche complémentaire pour intégrer le concept de connaissance à cette analyse et évaluer ce que ces travaux impliqueraient en ce qui concerne la production de connaissances.

Finalement, il apparaît comme pertinent de souligner les nombreuses possibilités de travaux futurs concernant la conception de systèmes d’information, ou plus largement de systèmes de gestion de l’information et de la connaissance, dits “intelligents”. De nombreux efforts sont actuellement engagés dans la conception de systèmes utilisant des méthodes de gamification, d’interface adaptative ou encore d’optimisation de la mise à disposition d’informations, dans l’optique d’influencer positivement l’utilisateur. Suite à la réalisation des pistes de travaux futurs précédemment citées, il serait intéressant d’envisager le développement d’un système de gestion de l’information et de la connaissance intégrant l’ensemble de ces nouvelles connaissances, ainsi que celles déjà existantes sur le sujet.



# Références

- Alomary, A., & Woollard, J. (2015). HOW IS TECHNOLOGY ACCEPTED BY USERS? A REVIEW OF TECHNOLOGY ACCEPTANCE MODELS AND THEORIES.
- Andersen, E. S., & Jessen, S. A. (2003, août). Project maturity in organisations. *International Journal of Project Management*, 21(6), 457–461. doi: 10.1016/S0263-7863(02)00088-1
- APA Dictionary of Psychology - Bias*. (s. d.). <https://dictionary.apa.org/>.
- Ariyaratna, K., & Peter, S. (2019). Business Analytics Maturity Models : A Systematic Review of Literature.
- Attig, C., Rauh, N., Franke, T., & Krems, J. F. (2017). System Latency Guidelines Then and Now – Is Zero Latency Really Considered Necessary? In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics : Cognition and Design* (Vol. 10276, pp. 3–14). Cham : Springer International Publishing. doi: 10.1007/978-3-319-58475-1\_1
- Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, W. W., Ruggeri, F., & Teugels, J. L. (2014). Wiley StatsRef : Statistics reference online.
- Blitman, D. (2020, septembre). 1. Présentation générale de la théorie de Chomsky. In *Le langage est-il inné ? : Une approche philosophique de la théorie de Chomsky sur le langage* (pp. 29–59). Besançon : Presses universitaires de Franche-Comté. doi: 10.4000/books.pufc.14143
- Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014, avril). Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1143–1152). Toronto Ontario Canada : ACM. doi: 10.1145/2556288.2557372
- Colzato, L. S., Barone, H., Sellaro, R., & Hommel, B. (2017, janvier). More attentional focusing through binaural beats : Evidence from the global–local task. *Psychological Research*, 81(1), 271–277. doi: 10.1007/s00426-015-0727-0
- Cours 7 : Biais dans les données : Analyse des métadonnées — Initiation à la Science des Données*. (s. d.). <https://nicolas.thiery.name/Enseignement/IntroScienceDonnees/Devoirs/Semaine7/CM7.html>.
- Delgado-Rodriguez, M. (2004, août). Bias. *Journal of Epidemiology & Community*

*Health*, 58(8), 635–641. doi: 10.1136/jech.2003.008466

- Delone, W. H., & Mclean, E. R. (2003, avril). The DeLone and McLean Model of Information Systems Success : A Ten-Year Update. *Journal of Management Information Systems*, 19(4), 9–30. doi: 10.1080/07421222.2003.11045748
- Denis, G., & Jouvelot, P. (2005). Motivation-Driven Educational Game Design : Applying Best Practices to Music Education. , 4.
- Descartes, R., & Gilson, É. (1987). *Discours de la méthode*. Vrin.
- Ewen, J. G., Armstrong, D. P., Parker, K. A., & Seddon, P. J. (Eds.). (2012). *Reintroduction Biology : Integrating Science and Management* (1<sup>re</sup> éd.). Wiley. doi: 10.1002/9781444355833
- Feldman, J. (2013, février). The neural binding problem(s). *Cognitive Neurodynamics*, 7(1), 1–11. doi: 10.1007/s11571-012-9219-8
- Frické, M. (2009, avril). The knowledge pyramid : A critique of the DIKW hierarchy. *Journal of Information Science*, 35(2), 131–142. doi: 10.1177/0165551508094050
- Friedrich, J., Becker, M., Kramer, F., Wirth, M., & Schneider, M. (2020, janvier). Incentive design and gamification for knowledge management. *Journal of Business Research*, 106, 341–352. doi: 10.1016/j.jbusres.2019.02.009
- Hamari, J., Hassan, L., & Dias, A. (2018, mars). Gamification, quantified-self or social networking ? Matching users' goals with motivational technology. *User Modeling and User-Adapted Interaction*, 28(1), 35–74. doi: 10.1007/s11257-018-9200-2
- Hamari, J., Koivisto, J., & Sarsa, H. (2014, janvier). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). Waikoloa, HI : IEEE. doi: 10.1109/HICSS.2014.377
- Harper, D. (2023). *Etymology of data*. <https://www.etymonline.com/word/data>.
- Haselton, M. G., Nettle, D., & Andrews, P. W. (2015, septembre). The Evolution of Cognitive Bias. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 724–746). Hoboken, NJ, USA : John Wiley & Sons, Inc. doi: 10.1002/9780470939376.ch25
- Johnson, J. (2020). CHAOS 2020 : Beyond Infinity. , 112.
- Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002, avril). Information quality benchmarks : Product and service performance. *Communications of the ACM*, 45(4), 184–192. doi: 10.1145/505248.506007
- Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009, février). Controlled experiments on the web : Survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1), 140–181. doi: 10.1007/s10618-008-0114-1

- Koivisto, J., & Hamari, J. (2019, avril). The rise of motivational information systems : A review of gamification research. *International Journal of Information Management*, *45*, 191–210. doi: 10.1016/j.ijinfomgt.2018.10.013
- Kuhl, J., & Fuhrmann, A. (1998, octobre). Decomposing Self-Regulation and Self-Control : The Volitional Components Inventory. In J. Heckhausen & C. S. Dweck (Eds.), *Motivation and Self-Regulation across the Life Span* (1<sup>re</sup> éd., pp. 15–49). Cambridge University Press. doi: 10.1017/CBO9780511527869.003
- Larousse, É. (s. d.-a). *Définitions : Invariant - Dictionnaire de français Larousse*. <https://www.larousse.fr/dictionnaires/francais/invariant/44070>.
- Larousse, É. (s. d.-b). *Définitions : réification - Dictionnaire de français Larousse*. <https://www.larousse.fr/dictionnaires/francais/r%C3%A9ification/67739>.
- McAlister, V. C. (2016, août). De l’usage du pluriel (ou non) du mot « data ». *Canadian Journal of Surgery*, *59*(4), 221–222. doi: 10.1503/cjs.009816
- Molina, R., Unsworth, K., Hodkiewicz, M., & Adriasola, E. (2013, novembre). Are managerial pressure, technological control and intrinsic motivation effective in improving data quality? *Reliability Engineering & System Safety*, *119*, 26–34. doi: 10.1016/j.res.2013.04.009
- Myers, M. D. (2013). *Qualitative research in business & management* (2nd ed éd.). London : SAGE.
- Olteanu, A., Castillo, C., Diaz, F., & Kiciman, E. (2019, juillet). Social Data : Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*, 13. doi: 10.3389/fdata.2019.00013
- Oxford English Dictionary. (2023, juillet). *Information*, *n*. Oxford University Press.
- Pally, R., & Olds, D. (2018, mars). How the Brain Actively Constructs Perceptions. In *The Mind-Brain Relationship* (1<sup>re</sup> éd., pp. 19–42). Routledge. doi: 10.4324/9780429482465-2
- Raghunathan, R., & Pham, M. T. (1999, juillet). All Negative Moods Are Not Equal : Motivational Influences of Anxiety and Sadness on Decision Making. *Organizational Behavior and Human Decision Processes*, *79*(1), 56–77. doi: 10.1006/obhd.1999.2838
- Rioul, O. (2020). La théorie de l’information. In *Traitement de l’Information (La Main à la Pâte)* (Belin éd.).
- Rowley, J. (2007, avril). The wisdom hierarchy : Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. doi: 10.1177/0165551506070706
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. doi: 10.1037/0003-066X.55.1.68

- Sendrier, N. (s. d.). Introduction à la théorie de l'information.
- Srinivasan, R., & Chander, A. (2021, avril). Biases in AI Systems : A survey for practitioners. *Queue*, 19(2), 45–64. doi: 10.1145/3466132.3466134
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge : Improving decisions about health, wealth, and happiness* (Rev. and expanded ed éd.). New York : Penguin Books.
- Torres, R., & Sidorova, A. (2019, décembre). Reconceptualizing information quality as effective use in the context of business intelligence and analytics. *International Journal of Information Management*, 49, 316–329. doi: 10.1016/j.ijinfomgt.2019.05.028
- Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology : Toward a Unified View. *MIS Quarterly*, 27(3), 425. doi: 10.2307/30036540
- Wang, R., Storey, V., & Firth, C. (1995, août). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4), 623–640. doi: 10.1109/69.404034
- Wang, R. Y., & Strong, D. M. (1996, mars). Beyond Accuracy : What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33. doi: 10.1080/07421222.1996.11518099
- Wilke, A., & Mata, R. (2012). Cognitive Bias. In *Encyclopedia of Human Behavior* (pp. 531–535). Elsevier. doi: 10.1016/B978-0-12-375000-6.00094-X
- Zimmer, H., Mecklinger, A., & Lindenberger, U. (2006). *Handbook of Binding and Memory Perspectives from Cognitive Neuroscience*. Oxford University Press. doi: 10.1093/acprof:oso/9780198529675.001.0001
- Zins, C. (2007, février). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), 479–493. doi: 10.1002/asi.20508



# Annexe A

## Données Nvivo

<b>Information and data</b>	<b>Human</b>
Aggregation Data granularity Definition Malleability New proposal Relevancy Suitability Representativity Reputation Reliability	Fear Feeling Motivation Autonomus Controlled Proximity Transparency Trust
<b>Project</b>	<b>Data collection</b>
Corporate performance Level of commitment Fixed price Time & Material Project managment Time window Type of activity Visibility	Ergonomics Takes over the last value Mandatory nature New proposal Not suited for management Outside the process Recurrence Timing & time-consuming
<b>Company culture</b>	<b>Stakeholder</b>
Culture of success Green Attitude Margin and growth culture Politics	Client response Client Interlocutor Management response Quality department

TABLE A.1 – Hiérarchie intermédiaire du codage axial

	Company culture	Data collection	Data & informations	Human	Project	Stakeholder
Company culture	28	1	3	5	2	7
Data collection	1	53	4	13	4	7
Data and informations	3	4	87	7	9	12
Human	5	13	7	57	4	25
Project	2	4	9	4	55	2
Stakeholder	7	7	12	25	2	58

TABLE A.2 – Croisement thèmes - valeurs

	Company culture	Data collection	Data & informations	Human	Project	Stakeholder
Company culture	100%	2%	3%	9%	4%	12%
Data collection	4%	100%	5%	23%	7%	12%
Data & informations	11%	8%	100%	12%	16%	21%
Human	18%	25%	8%	100%	7%	43%
Project	7%	8%	10%	7%	100%	3%
Stakeholder	25%	13%	14%	44%	4%	100%

TABLE A.3 – Croisement thèmes - ratio

## Annexe B

# Mails utilisés dans le cadre du protocole d'A/B testing



En juin 2020, Alten a initié une collaboration avec l'Ecole des Mines dont l'objectif est de produire un outil d'aide à la décision fondée sur l'Intelligence Artificielle.

L'IA peut constituer un outil essentiel de management et d'accompagnement de projets au sein des entreprises.

Il s'agit pour Alten de mettre au point un système de gestion des données plus fluide, plus proche de la réalité du terrain et plus efficace. Les résultats de ce travail sont déjà très encourageants et nous avons besoin d'aller plus loin dans le process, d'être plus précis et pointus dans la collecte des données.

Pour cela nous avons aujourd'hui besoin de vos compétences, de votre regard et de votre expertise pour que nos outils soient plus robustes scientifiquement, plus proches de vos attentes et de celles de l'entreprise.

Cliquez simplement sur le lien ci-après, vous y trouverez toutes les informations nécessaires.

[SMART Project Survey - My Surveys \(altengroup.com\)](https://www.altengroup.com/SMART-Project-Survey-My-Surveys)

Merci de bien vouloir répondre aux questions portant uniquement sur les projets pour lesquels vous avez une proximité suffisante.

Merci de votre participation.

FIGURE B.1 – Mail groupe autonome



Nous ne sommes plus très loin de notre objectif de labéliser 50 % des projets, soit :

**440 projets**

Nous nous permettons de vous envoyer ce mail de rappel, car nous avons remarqué que vous n'avez pas encore répondu à l'enquête de **labélisation de projets ALTEN**.

L'objectif de cette opération est d'avancer sur la compréhension du potentiel de l'IA et des outils d'aide à la décision dans le cadre de la gestion de projet.

Nous devons disposer d'un échantillon d'une taille suffisamment conséquente pour être statistiquement significatif et, pour cela, nous nous sommes fixé comme **objectif de labéliser 50% des projets présents dans la base**, soit 440 projets (sur un total de 880). Grâce à toutes les personnes qui ont déjà participé à cette opération, **nous ne sommes plus très loin de cet objectif et nous avons besoin d'un dernier effort de votre part**.

Pour rappel, cette opération s'effectue dans d'une collaboration entre ALTEN et l'Ecole des mines de Paris dont l'objectif est de produire un outil d'aide à la décision fondé sur l'Intelligence Artificielle.

Il s'agit pour ALTEN de mettre au point un système de gestion des données plus fluide, plus proche de la réalité du terrain et plus efficace. Les résultats de ce travail sont déjà très encourageants, et nous avons besoin d'aller plus loin dans le process, d'être plus précis et pointus dans la collecte des données.

**Cliquez simplement sur le lien ci-après** ; vous y trouverez toutes les informations nécessaires.

[SMART Project Survey - My Surveys \(altengroup.com\)](https://altengroup.com/SMART-Project-Survey-My-Surveys)

Merci de bien vouloir répondre aux questions portant uniquement sur les projets pour lesquels vous avez une proximité suffisante.

Merci de votre participation.

Cordialement - Kind Regards - mit freundlichen Grüßen

FIGURE B.2 – Mail groupe autonome - relance

Bonjour à toutes et à tous,

Pour donner suite à la présentation faite en comité de direction sur le projet d'intelligence artificielle, destiné à mieux gérer et anticiper les performances des projets, nous avons obtenu le GO des directeurs techniques pour vous solliciter. Nous vous remercions par avance de bien vouloir répondre à un questionnaire pour les projets (en cours et terminés) qui concernent votre périmètre.

Cliquez simplement sur le lien ci-après , vous y trouverez toutes les informations nécessaires.

[SMART Project Survey - My Surveys \(altengroup.com\)](https://my-surveys.altengroup.com)

Merci de bien vouloir répondre aux questions portant uniquement sur les projets pour lesquels vous avez une proximité suffisante.

Merci de votre participation, **et pour celles et ceux qui ont déjà répondu, merci de ne pas tenir compte de cette relance.**

Cordialement - Kind Regards - mit freundlichen Grüßen

FIGURE B.3 – Mail groupe contrôle

Bonjour à toutes et à tous,

Vous n'avez pas encore répondu à l'enquête de caractérisation des projets ALTEN.

Pour Rappel, nous avons obtenu le GO des directeurs techniques pour vous solliciter pour faire avancer le projet d'intelligence artificielle, destiné à mieux gérer et anticiper les performances des projets.

Nous vous remercions par avance de bien vouloir répondre à un questionnaire pour les projets (en cours et terminés) qui concernent votre périmètre.

Cliquez simplement sur le lien ci-après, vous y trouverez toutes les informations nécessaires.

[SMART Project Survey - My Surveys \(altengroup.com\)](https://altengroup.com/SMART-Project-Survey-My-Surveys)

Merci de bien vouloir répondre aux questions portant uniquement sur les projets pour lesquels vous avez une proximité suffisante.

Merci de votre participation.

FIGURE B.4 – Mail groupe contrôle - relance



## RÉSUMÉ

---

Du point de vue du philosophe Friedrich Nietzsche, il n'y a pas de réalité qui existe en soi, pas de fait brut, pas de réalité absolue : tout ce que nous définissons comme étant la réalité n'est, en fait, que le résultat de processus d'interprétation qui nous sont propres. Or, les données stockées sur les systèmes d'information ne sont souvent rien d'autre que la représentation codée de déclarations émises par des êtres humains, relevant donc intrinsèquement de l'interprétation humaine et étant par conséquent affectées par les mêmes biais et limites qui caractérisent la psyché humaine. Cette thèse propose un nouveau cadre conceptuel, le modèle "Data Binding and Reification" (DBR), qui décrit le processus d'interprétation des données, puis de réification de l'information, en utilisant une nouvelle approche qui place les mécanismes de perception humaine au cœur de celui-ci. En mobilisant les sciences cognitives et comportementales, cette approche permet d'identifier dans quelle mesure l'intervention de l'homme et la structure de l'environnement auquel il est soumis conditionnent l'apparition de biais cognitifs affectant ces processus. Les résultats expérimentaux valident partiellement ce modèle en identifiant les caractéristiques de l'environnement affectant, dans un contexte organisationnel, le processus de collecte de données et la qualité des informations produites. Ces travaux ouvrent de nombreuses perspectives, telles que l'élaboration d'une architecture de choix au sens de l'économiste Richard Thaler, pouvant améliorer le processus même de collecte de données en modifiant l'expérience des utilisateurs du système d'information.

## MOTS CLÉS

---

Interaction homme-machine, science de l'information, qualité des données et des informations, biais cognitif, rationalité limitée et écologique, psychologie cognitive, architecture de choix, "Nudge", théorie de l'autodétermination, management de la connaissance, expérience utilisateur, "Binding theory".

## ABSTRACT

---

From the perspective of philosopher Friedrich Nietzsche, there is no reality that exists in itself, no raw fact, no absolute reality: everything that we define as reality is, in fact, only the result of interpretation processes that are unique to us. Moreover, the data stored in information systems is often nothing more than the coded representation of statements made by human beings, thereby inherently involving human interpretation and consequently being affected by the same biases and limitations that characterize the human psyche. This thesis introduces a new conceptual framework, the "Data binding and reification" (DBR) model, that describes the process of data interpretation, and then the reification of information, using a new approach that places human-perception mechanisms at the heart of this process. By mobilizing cognitive and behavioral sciences, this approach allows us to identify to what extent human intervention and the structure of the environment to which one is subjected condition the emergence of cognitive biases affecting these processes. Experimental results partially validate this model by identifying the characteristics of the environment that affect, in an organizational context, the data-collection process and the quality of the information produced. This work opens up numerous perspectives, such as the development of a choice architecture in the sense of the economist Richard Thaler, which could improve the very process of data collection by modifying the experience of users of the information system.

## KEYWORDS

---

Human-Computer Interaction, Information Science, Data and Information Quality, Cognitive Bias, Bounded and Ecological Rationality, Cognitive Psychology, Choice Architecture, Nudge, Auto-determination Theory, Knowledge Management, User Experience, Binding Theory.