



HAL
open science

Évaluation de la qualité des comportements des agents en simulation

Kévin Darty

► **To cite this version:**

Kévin Darty. Évaluation de la qualité des comportements des agents en simulation : application à un simulateur de conduite en environnement virtuel. Intelligence artificielle [cs.AI]. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066339 . tel-01176764v2

HAL Id: tel-01176764

<https://hal.sorbonne-universite.fr/tel-01176764v2>

Submitted on 15 Jul 2015 (v2), last revised 26 Jan 2016 (v3)

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 PIERRE ET MARIE CURIE**

Spécialité Informatique - Intelligence Artificielle

École doctorale Informatique, Télécommunications et Électronique

Présentée par

Kévin Darty

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

**Évaluation de la qualité des comportements des agents en
simulation : application à un simulateur de conduite en
environnement virtuel**

soutenue le 7 juillet 2015

devant le jury composé de :

Pr. Pierre CHEVAILLIER	Examineur	PR à l'ENIB-CERV – Lab-STICC
Dr. Amélie CORDIER	Examineur	MCF à l'Université Lyon 1 – LIRIS
Pr. Amal EL FALLAH SEGHROUCHNI	Examineur	PR à l'UPMC – LIP6
Pr. Domitile LOURDEAUX	Rapporteur	MCF - HDR à l'UTC – Heudiasyc
Pr. Philippe MATHIEU	Rapporteur	PR à l'Université Lille 1 – LIFL
Pr. Catherine PELACHAUD	Invité	DR à TELECOM ParisTech – LTCI
Pr. Nicolas SABOURET	Directeur	PR à l'Université Paris-Sud – LIMSI
Dr. Julien SAUNIER	Encadrant	MCF à l'INSA – LITIS / IFSTTAR – LEPSIS
Dr. Mahdi ZARGAYOUNA	Examineur	CR à l'IFSTTAR – GRETTIA

Évaluation de la qualité des comportements des agents en simulation : application à un simulateur de conduite en environnement virtuel

Kévin Darty

Résumé

Cette thèse se situe dans le contexte de la Simulation Multi-Agent et s'intéresse à l'évaluation de la capacité des agents à reproduire individuellement des comportements humains. Cette problématique se retrouve dans de nombreux domaines d'application comme la réalité virtuelle, les agents conversationnels animés, les simulateurs immersifs, les jeux...

L'approche dominante pour évaluer les comportements des agents s'appuie sur l'utilisation de questionnaires de Sciences Humaines et Sociales afin d'exploiter le jugement humain. Il existe peu d'approches exploitant l'analyse automatique de données utilisée en Intelligence Artificielle à l'échelle microscopique car cette dernière utilise des données brutes de bas niveau et demande donc des traitements souvent complexes afin de combler le fossé sémantique les séparant des comportements de haut niveau. Ces deux types d'approches permettent chacune d'obtenir des informations différentes sur ce même objet d'étude qu'est le comportement de haut niveau. Nous montrons dans cette thèse que ces informations se complètent et que l'évaluation gagne à exploiter conjointement ces deux approches.

Nous exposons en premier lieu une méthode d'évaluation semi-automatique de la qualité des comportements des agents. Elle s'appuie sur l'extraction automatique de classes de comportement en combinant l'analyse automatique issue de l'Intelligence Artificielle et l'utilisation de questionnaire issue des Sciences Humaines et Sociales. La première étape consiste en une observation objective des traces de simulation des humains pour construire - via des indicateurs données par des experts - des classes servant d'abstractions aux comportements haut niveau. La seconde étape évalue les catégories d'utilisateurs à l'aide d'une annotation des comportements exhibés.

Nous présentons ensuite un algorithme d'agrégation des agents aux classes d'humains qui servent de comportements de référence. Cet algorithme nous permet alors de comparer les agents par rapport aux humains en étudiant la composition de ces classes, afin d'évaluer les capacités, les manques et les erreurs du modèle d'agent. Nous fournissons des métriques basées sur ces types de classe ainsi que des taux de confiance en chaque classe. Puis nous explicitons ces comportements en nous appuyant sur les catégories d'utilisateur.

Enfin, nous exposons un cycle de conception permettant la calibration automatique de la population d'agents par une exploration de l'espace des paramètres. Cette calibration est basée sur les métriques résultantes de notre méthode d'évaluation et sur des scores de représentativité des comportements des agents.

Notre méthode d'évaluation est utilisable dans le but d'analyser un modèle d'agent, de comparer plusieurs modèles d'agent, ou d'évaluer les modifications apportées entre deux versions d'un même modèle.

Nous avons appliqué cette méthodologie sur plusieurs études du comportement de conduite en vue d'analyser la simulation de trafic routier *ARCHISIM* et nous présentons des résultats qui montrent notamment la capacité de son modèle d'agent à reproduire le comportement normatif et le manque de comportements jugés comme dangereux.

Mots-clés : évaluation et explicitation de comportements d'agent, classification et agrégation de comportement, approches issues d'Intelligence Artificielle et de Sciences Humaines et Sociales, calibration de Simulation Multi-Agent.

Abstract

English title: Evaluation of the agents' behaviour quality in simulation: application to a driving simulator in virtual environment

This research work is conducted in the context of Multi-Agents Simulation. It focuses on the evaluation of the agents' ability to reproduce human behaviors. This research question has many application domains such as virtual reality, embodied conversational agents, immersive simulation, games...

The dominant approach for the evaluation of the agents' behaviors is based on human judgement, through Social Science questionnaires. Few approaches are based on Artificial Intelligence and automatic data analysis at the microscopic scale. This is due to the semantic gap between "low-level" raw data used by such algorithms and high-level behavior, which requires complex processing. However, these two approaches can obtain different information on high-level behavior. We show in this thesis that these two types of information complete each other and that the evaluation can benefit from both approaches equally.

We first present a semi-automatic method for evaluating the agents' behavior quality. It relies on the automatic extraction of behavior clusters. The proposed method combines the Artificial Intelligence approach and the Social Science approach. The first one consists in an observation of human's simulation logs in order to build clusters used as high-level behavior abstractions through indicators given by the experts. The second one evaluates users categories by an annotation of the exhibited behaviors.

We then present an algorithm based on the aggregation of the agents to the humans clusters, considered as baseline behaviors. This algorithm allows us to compare agents to humans by studying the clusters composition in order to assess the capacities, the lacks, and the errors in the agent model. We provide metrics based on these cluster types and confidence rates for each cluster. We then make these behaviors explicit based on user categories.

Finally, we present a cycle for automatic calibration of the agents population through exploration of the parameter space. This calibration is based on the metrics of our evaluation method and on agents' behavior representativeness.

Our evaluation method is usable for the analysis of an agent model, for comparing several agent models, and for evaluating the modifications between two versions of the same model.

We applied this methodology on several driver behavior studies to analyse the road traffic simulation *ARCHISIM*, and we present results showing the capacity of its agent model to reproduce the normative behavior and the lack of behaviors considered as dangerous.

Keywords: Evaluation and explicitation of agent's behaviors, behavior clustering and aggregation, approaches from Artificial Intelligence and Social Science, calibration of Multi-Agents Simulation.

*Information is not knowledge.
Knowledge is not wisdom.
Wisdom is not truth.
Truth is not beauty.
Beauty is not love.
Love is not music.
Music is the best.*

Zappa

Table des matières

Liste des figures	ix
Liste des tableaux	x
Liste des annexes	xi
Introduction	2
I États de l’art	5
1 L’évaluation des comportements en <i>SHS</i>	7
1.1 Comportement	7
1.1.1 Définition première	7
1.1.2 Comportements humains	8
1.1.3 Disciplines	8
1.1.4 Niveau de comportement	9
1.1.5 Conclusion	11
1.2 Évaluation en <i>SHS</i>	12
1.2.1 Simulation immersive	12
1.2.2 Moyens d’évaluation	13
1.2.2.a Réactions physiologiques	13
1.2.2.b Questionnaire	13
1.2.2.c Discussion et entretien d’autoconfrontation	14
1.2.3 Évaluation <i>a priori</i>	14
1.2.3.a Propension à l’immersion	15
1.2.3.b Comportement de conduite	16
1.2.4 Évaluation <i>a posteriori</i>	19
1.2.4.a Présence	20
1.2.4.b Dimension comportementale	22
1.2.5 Conclusion	23
2 L’évaluation des comportements en <i>IA</i>	25
2.1 Échelles d’évaluation	25
2.1.1 Validation au niveau macroscopique	25
2.1.2 Validation au niveau microscopique	26
2.1.3 L’humain comme référence	27
2.1.4 Conclusion	29
2.2 Catégoriser pour évaluer	29

2.2.1	Classification non-supervisée	29
2.2.2	Les méthodes de classification non-supervisée	30
2.2.2.a	Classification Ascendante Hiérarchique	30
2.2.2.b	Cartes Auto-Adaptatives	30
2.2.2.c	K -moyennes	31
2.2.3	Nombre de classes	31
2.2.4	Cas des données temporelles	32
2.2.5	Comparaison de classification	33
2.2.6	Conclusion	34
II Méthode		35
3	Méthode d'évaluation des comportements d'agents	37
3.1	Motivations	37
3.2	Approche globale	38
3.2.1	Recueil des données de simulation	39
3.2.2	Recueil des annotations des rejeux vidéo	39
3.2.3	Classification et agrégation	39
3.2.4	Comparaison des classifications	40
3.2.5	Analyse et explicitation des classes	40
3.3	Cas d'étude : la simulation de trafic routier	41
4	Recueil et prétraitements des données	43
4.1	Recueil de traces de simulation	43
4.2	Recueil d'annotations	44
4.2.1	Questionnaire	44
4.2.2	Annotations	45
4.3	Prétraitements	45
4.3.1	Prétraitement des annotations	46
4.3.2	Prétraitement des traces	46
4.3.2.a	Indicateurs haut niveau	47
4.3.2.b	Données temporelles	47
4.3.2.c	Des distances mutuelles aux positions multidimen- sionnelles	48
4.3.2.d	Décorrélacion	49
4.4	Conclusion	49
5	Comparaison entre comportements d'agents et humains	51
5.1	Classification et agrégation	51
5.1.1	Classification des participants	51
5.1.2	Agrégation des agents	53
5.1.3	Traitement des annotations	55
5.1.4	Conclusion	56
5.2	Comparaison et analyse	56
5.2.1	Méthode de comparaison	56
5.2.2	Analyse des classes	58
5.2.2.a	Type de classes	58
5.2.2.b	Taux de confiance	58

5.2.2.c	Scores de type de classe	60
5.3	Conclusion	61
6	Cycle de conception	63
6.1	Calibration	63
6.2	Ensembles de paramètres	65
6.3	Exploration de l'espace des paramètres	66
6.4	Cycle	66
6.5	Conclusion	67
7	Conclusion	69
III	Expérimentations	71
8	Matériel et méthode	73
8.1	Simulation de trafic <i>ARCHISIM</i>	73
8.2	Simulateur de conduite	73
8.3	Scénario	75
8.4	Protocole expérimental	77
8.4.1	Populations	78
8.4.2	Données de simulation	79
8.4.3	Application de la méthode de classification	79
8.5	Questionnaires de SHS	81
8.5.1	Propension à l'immersion	82
8.5.2	Annotations et autoévaluation	83
8.5.3	Évaluation de la crédibilité	84
8.5.3.a	Questionnaire Général de Comportement	84
8.5.3.b	Questionnaire Spécifique de Comportement	85
8.5.4	Questionnaires <i>ad hoc</i>	87
8.6	Implémentation	87
8.7	Conclusion	89
9	Résultats	91
9.1	Étude d'une situation de dépassement	91
9.1.1	Situation de conduite	91
9.1.2	Classification de traces	92
9.1.3	Classification d'annotations	93
9.1.4	Scores	93
9.1.5	Discussions	95
9.1.6	Cycle de conception	95
9.1.7	Conclusion	96
9.2	Étude d'une situation de trafic ralenti	97
9.2.1	Situation de conduite	97
9.2.2	Annotations et habitudes de comportement	98
9.2.3	Classes de traces et classes d'annotations	98
9.2.4	Classes de traces et classes d'habitudes	100
9.2.5	Scores	101
9.2.6	Conclusion	103

9.3	Étude des facteurs de différence dans la conduite	104
9.3.1	Annotations et traces	104
9.3.2	Composition sur trois situations de conduite	105
9.3.3	Discussion	107
9.3.4	Corrélations avec les questionnaires	108
9.3.5	Conclusion	109
9.4	Conclusion	110
	Conclusion et perspectives	114
	Conclusion	115
	Perspectives	117
	Liste des publications	123
	Bibliographie	126
	Annexes	136
	Index	166

Liste des figures

1.1	Modèle de Rasmussen	10
1.2	Modèle de Michon	11
1.3	<i>Game Engagement Questionnaire</i>	17
3.1	Méthode d'évaluation des comportements	39
3.2	Dispositif du simulateur de conduite	41
4.1	Prétraitements, agrégation, et classification	46
5.1	Exemples d'agrégation	55
6.1	Cycle de conception	63
8.1	Simulation de trafic	74
8.2	Dispositif du simulateur de conduite	74
8.3	Circuit du scénario de conduite	76
8.4	Exploitation des questionnaires	82
9.1	Scénario de dépassement en double sens	92
9.2	Comparaison entre traces et annotations	92
9.3	Scénario de dépassement en sens unique	98
9.4	Comparaison entre habitudes et annotations	99
9.5	Comparaison entre traces et annotations	100
9.6	Comparaison entre traces et habitudes	101
9.7	Scénario de trafic ralenti sur autoroute	105
9.8	Comparaison genrée des acteurs principaux	106
9.9	Classifications des comportements pour chaque situation	112
9.10	Évolution des classes de comportements	121

Liste des tableaux

1.1	Échelles du <i>Driving Behaviour Inventory</i>	18
1.2	Échelles du <i>Driving Style Questionnaire</i>	18
1.3	Échelles du <i>Decision-Making Style Questionnaire</i>	19
1.4	Échelles du <i>Driving Behavior Questionnaire</i>	20
1.5	Échelles du <i>Igroup Presence Questionnaire</i>	21
8.1	Paramètres des agents	78
8.2	Variables générales	79
8.3	Variables communes	80
8.4	Variables du véhicule principal	81
8.5	Variables des véhicules simulés	81
8.6	Traduction du <i>Game Engagement Questionnaire</i>	83
8.7	<i>Questionnaire Général de Comportement</i>	85
8.8	Réglette du questionnaire de comportement	85
8.9	<i>Questionnaire Spécifique de Comportement</i>	86
9.1	Corrélations de l'étude de genre pour le <i>DBQ</i>	108
9.2	Corrélations de l'étude de genre	108

Liste des annexes

Annexe A	Questionnaires	137
Annexe A.1	Questionnaire de comportement de conduite (<i>BDQ</i>)	138
Annexe A.2	Questionnaire de jeux video	142
Annexe A.3	Questionnaire d'état civil et conducteur	143
Annexe A.4	Questionnaire d'état de présence (<i>PQ</i>)	145
Annexe A.5	Traduction française du <i>IGroup Presence Questionnaire</i> (<i>IPQ</i>)	148
Annexe A.6	Questionnaire d'implication dans les jeux vidéo (<i>GEQ</i>)	150
Annexe A.7	Questionnaire de propension à l'immersion <i>ITQ</i>	152
Annexe A.8	Questionnaire Général de Comportement (<i>QGC</i>)	155
Annexe A.9	Questionnaire Spécifique de Comportement (<i>QSC</i>)	156
Annexe B	Méthode	161
Annexe B.1	Cycle de calibration	162
Annexe C	Expérimentation	163
Annexe C.1	Circuit	165

Introduction

En Intelligence Artificielle, un agent est une entité autonome au sein d'un environnement qui interagit avec d'autres entités. Une des principales applications du concept d'agent en Intelligence Artificielle se situe dans le cadre de la simulation que ce soit dans le domaine des *Agents Virtuels* [91] incarnés en *Environnement Virtuel* ou au sein des *Simulations Multi-Agents (SMA)*.

Les *Agents Virtuels* sont exploités dans divers domaines tels que les outils d'assistance d'applications, les environnements virtuels participatifs, ou encore les environnements interactifs d'apprentissage humain. La problématique est alors centrée sur sa capacité d'interaction avec l'humain, d'une part au travers divers moyens de communication (*multimodalité*), et d'autre part via l'apparence de l'agent (*incarnation*).

La simulation multi-agents est quant à elle utilisée dans de nombreux domaines comme la simulation économique [98], la simulation sociale [39] et la simulation de trafic [43]. Ces domaines ont en commun de s'intéresser à des *systèmes complexes* qui sont caractérisés par un grand nombre d'entités en interaction qui empêchent l'observateur de prévoir sa rétroaction, son comportement ou son évolution par l'analyse.

Dans nos travaux, nous nous intéressons au cas particulier des simulations immersives dans des environnements de *réalité virtuelle (RV)* pour lesquelles les utilisateurs humains cohabitent avec des agents simulés. Une difficulté apparaît pour la conception de ces simulations : les agents doivent réaliser des tâches similaires à celles des humains. Ils doivent alors exhiber des comportements comparables à un utilisateur humain immergé.

La modélisation et l'évaluation d'agents avec l'humain en référentiel est une tâche complexe pour de nombreuses raisons. Premièrement, le comportement choisi par un humain particulier dépend de ses connaissances qui peuvent induire un comportement non optimal. Ainsi, le fait d'avoir comme référence l'humain implique donc la prise en compte de nombreux comportements avec une forte variabilité. Deuxièmement, la personnalité et l'état mental influent aussi sur le comportement : la fonction d'*utilité* d'une tâche diffère selon les individus. Conséquemment, les humains peuvent choisir de ne pas respecter certaines règles, c'est-à-dire d'exhiber des comportements non normatifs. Troisièmement, il est à noter que le comportement adopté dépend aussi des émotions, ce qui peut entraîner un comportement irrationnel au sens d'incohérent en termes de but. Différents domaines ont ces contraintes comme la simulation de trafic [24], simulation de foule [134] ou la gestion de ressources [68].

Une seconde difficulté émerge pour l'évaluation des comportements de ces agents : le point de vue de l'humain est fortement local et la crédibilité du comportement est jugée essentiellement au niveau microscopique. Cette crédibilité au niveau "*local*" est primordiale pour la simulation : en effet, l'un des principaux objectifs de ces simulations, que ce soit pour des besoins d'entraînement [14] ou pour la simulation participative [73], est que l'utilisateur reste *immergé* dans l'environnement virtuel (*EV*). Si le comportement individuel est incohérent, l'immersion du participant, c'est-à-dire le sentiment d'appartenance à l'*EV*, est rompue [53, 99].

De même, dans l’approche multi-agents, les buts sont définis au niveau local pour les différentes entités : chaque agent va exhiber un comportement local, qui dépend de ses buts propres et des états des agents “*voisins*” avec lesquels il est en interaction. Ces comportements locaux forment l’échelle microscopique de la simulation. Le comportement global du système *émerge* à partir des comportements locaux. Le fonctionnement du système dans son entièreté correspond à l’échelle macroscopique. Il existe donc deux échelles de comportement : le comportement microscopique et macroscopique.

Ainsi, dans la simulation multi-agents comme dans les environnements virtuels, le problème de la validité du comportement individuel est crucial. En effet, le comportement microscopique peut être incohérent (et donc créer une rupture d’immersion) alors même que le comportement macroscopique reste valide. Et réciproquement, le comportement individuel peut rester cohérent alors que le comportement global n’est pas valide [82]. Il faut donc disposer d’outils pour valider le comportement à la fois au niveau macroscopique et au niveau microscopique. Alors qu’il existe déjà de nombreux outils pour la validation au niveau macroscopique, par exemple la comparaison du débit en véhicules par heure entre la simulation et la situation réelle pour la simulation de flux de trafic [116], la littérature propose peu d’outils pour le niveau microscopique.

Plusieurs approches sont envisageables pour évaluer les comportements des agents au niveau microscopique. Premièrement, il est possible de demander à l’utilisateur d’évaluer de manière subjective les comportements à l’aide de questionnaires validés. Cette évaluation subjective se base sur l’expertise humaine. En effet, les Sciences Humaines et Sociales utilisent l’humain comme moyen d’analyse. Cette méthode permet d’obtenir des informations explicites sur les comportements. Cependant elle se voit confrontée à différents problèmes tels que l’attention et l’interprétation du participant, ainsi que le temps nécessaire à l’évaluation lorsque l’expérimentation est longue ou lorsque le nombre de participants est grand. Deuxièmement, il est possible de comparer de manière objective les traces d’exécution des agents avec les actions des humains. En effet, le cadre de la réalité virtuelle nous permet d’obtenir les traces de simulation à la fois des agents virtuels et des participants humains. Cependant, les données obtenues sont fortement bruitées et se résument à des variables bas niveau. Il est néanmoins possible d’effectuer une analyse par similarités basée sur des méthodes de classification. Le regroupement de plusieurs instances analogues sert alors d’abstraction à un archétype de comportement. Cela peut permettre notamment un étalonnage des paramètres d’entrée du modèle d’agent. Toutefois, ce type d’analyse ne permet pas de capturer la sémantique des comportements, elle ne peut aller au-delà des variables utilisées.

C’est la raison pour laquelle nous proposons dans cette thèse de combiner les méthodes d’analyse objective de traces de comportements avec les méthodes d’évaluation subjective issues de Sciences Humaines et Sociales pour évaluer les comportements des agents au niveau microscopique dans ce type de simulations immersives. Nous souhaitons fournir des métriques afin de mesurer la qualité des comportements des agents et être apte à analyser chez les agents leurs capacités à reproduire des comportements humains, pouvoir relever les comportements humains manquants chez les agents, et être en mesure de détecter les comportements erronés des agents.

Nous aspirons à utiliser cette méthodologie au sein d'un cycle de conception afin de calibrer une population d'agents.

La méthodologie proposée s'applique à l'évaluation au niveau microscopique des comportements des agents artificiels ayant pour but de reproduire des comportements humains dans le contexte des simulations immersives. Ces travaux de thèse ont pour cadre applicatif l'évaluation des agents en simulation de trafic routier. En effet, le *Laboratoire Exploitation, Perception, Simulateurs et Simulations (LEPSiS)* de l'*Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR)* est un laboratoire pluridisciplinaire au carrefour de l'Ingénierie, de l'Intelligence Artificielle, et des Sciences Humaines et Sociales qui s'intéresse, entre autre, à la réalité virtuelle appliquée aux simulateurs de conduite. Les *SMA* sont notamment utilisés pour ces simulateurs. Ces derniers peuvent avoir plusieurs buts : aide à la conduite, outils pour l'étude des comportements routiers, jeux vidéo, . . . Le *LEPSiS* a développé l'architecture multi-acteurs de simulation des comportements du conducteur qu'est *ARCHISIM* [24, 45] dont nous devons évaluer les agents.

Dans la *partie I*, nous présentons les travaux sur lesquels nous nous appuyons pour définir notre modèle d'évaluation. Nous présentons les approches subjectives et les approches objectives et nous montrons leur complémentarité. Cela permet de positionner notre contribution dans le domaine. D'une part, nous utiliserons les connaissances en Sciences Humaines et Sociales présentées au *chapitre 1* pour définir la notion de comportement (voir *section 1.1*) ainsi que pour exploiter les moyens d'évaluation du comportement (voir *section 1.2*). D'autre part, nous exposons les méthodes existantes en Intelligence Artificielle pour l'évaluation des comportements des agents au *chapitre 2* que ce soit à l'échelle macroscopique en *sous-section 2.1.1* ou à l'échelle microscopique en *sous-section 2.1.2*. Enfin, nous donnons un aperçu des méthodes de catégorisation applicables à notre problématique en *section 2.2*.

La *partie II* présente la méthode élaborée. Premièrement, le *chapitre 3* explique les motivations nous ayant conduit à notre approche, présente notre méthodologie de manière générale ainsi que ses différents composants, et expose le cas d'étude sur lequel nous l'avons appliquée. Secondement, le *chapitre 4* présente le recueil des données des approches objectives et subjectives en *sections 4.1 et 4.2*, ainsi que les pré-traitements qui leurs sont appliqués et les algorithmes mis en œuvre en *section 4.3*. Troisièmement, le *chapitre 5* expose la méthode de comparaison des comportements des participants humains entre eux par classification, puis avec des agents virtuels par agrégation en *section 5.1*, et l'étude de leur crédibilité par une méthode de comparaison et d'analyse des classes en *section 5.2*. Quatrièmement, nous montrons dans le *chapitre 6* l'utilisation de cette méthode au sein d'un cycle de conception afin de calibrer le modèle d'agent en extrayant des ensembles de paramètres.

La *partie III* présente les applications de notre méthode sur le cas d'étude qu'est le simulateur de conduite immersif *ARCHISIM* de l'*IFSTTAR*. Ce dernier est décrit, ainsi que le protocole expérimental, au *chapitre 8*. Les résultats obtenus lors des différentes expérimentations sont exposés au *chapitre 9* et une discussion de ces derniers constitue la *section 9.4*.

Enfin nous présentons dans la dernière partie une conclusion générale sur les méthodes proposées ainsi que les perspectives de recherche de nos travaux.

Première partie

États de l'art

Dans cette partie, nous présentons les questions de recherche soulevées par notre problématique qu'est l'évaluation de la qualité des comportements des agents. Le *chapitre 1* s'intéresse aux Sciences Humaines et Sociales. Tout d'abord, nous définissons dans la *section 1.1* la notion de comportement qui est au cœur de notre étude. Nous présentons ensuite en *section 1.2* les travaux de la littérature pour l'évaluation des comportements en *SHS*. Le *chapitre 2* s'intéresse à l'évaluation des comportements en Intelligence Artificielle. L'évaluation au niveau macroscopique est étudiée en *sous-section 2.1.1*, puis celle au niveau microscopique en *sous-section 2.1.2*. Enfin, la *section 2.2* présente les méthodes de classification utilisées ou adaptées à cette évaluation. Nous montrons les limites de chaque approche et nous expliquons comment elles se complètent.

1 L'évaluation des comportements en Sciences Humaines et Sociales

Les Sciences Humaines et Sociales se sont intéressées à l'évaluation des comportements afin de comprendre l'être humain. Dans ce chapitre, nous nous intéressons aux méthodes d'évaluation des comportements issues de ce domaine. Pour cela nous allons définir dans la *section 1.1* le terme *comportement* et les sens qu'il peut prendre pour l'être humain. La *section 1.2* présentera les spécificités de la simulation immersive et les moyens d'évaluation existants. Par la suite, nous présenterons les méthodes d'évaluation *a priori* de l'*immersion*, puis dans le cadre de la conduite. Enfin, les méthodes d'évaluation *a posteriori* de la *présence* seront exposés de manière générale, puis plus spécifiquement dans sa dimension comportementale.

1.1 Comportement

Comme nous cherchons à analyser les *comportements* d'agents logiciels, nous devons avant tout préciser la définition de ce terme. Les comportements d'un individu sont l'ensemble des actions et réactions observables de ce dernier dans l'environnement [10]. La notion de comportement englobant différents concepts (individuel ou collectif, comportement moteur ou comportement social, ...), il est nécessaire de préciser ce que nous entendons par ce terme.

1.1.1 Définition première

Le comportement est la manière dont un *objet* fonctionne et évolue dans certaines circonstances. Nous nous concentrerons ici sur les comportements présents dans le vivant et plus précisément ceux des organismes pourvus d'un système nerveux. Le comportement correspond alors aux actions observables d'un organisme, c'est-à-dire sa capacité à agir sur l'environnement extérieur. Le fonctionnement habituel d'un individu d'une espèce est la réponse à un stimulus. Le comportement permet alors une conséquence prévisible et est mis en œuvre dans un but précis [10]. Les comportements peuvent être observés par une étude au niveau physiologique (la maîtrise des interactions entre nerfs, muscles et sens) et par l'analyse des conséquences des comportements sur l'individu dans son milieu. L'éthologie est la branche zoologique qui a pour but de répertorier les unités comportementales animales pour chaque espèce [90].

1.1.2 Comportements humains

L'être humain a la capacité d'attribuer un état mental à un individu, permettant ainsi la prédiction de ses actions et de celles des autres. De plus, il peut adopter une multitude de comportements complexes, notamment lors d'interactions sociales : communication, collaboration, compétition, ... [113]. Ses comportements nécessitent donc une étude particulière. De nombreuses disciplines ont donc pour objet d'étude le comportement humain spécifiquement. C'est le cas de la psychologie (expérimentale, clinique, psychiatrique, sociale), des neurosciences (neuropsychologie, neurologie), de la sociologie, de l'économie, ... Ces disciplines n'étudient cependant pas les mêmes comportements, par exemple la neurologie s'intéresse au système nerveux alors que la sociologie étudie les relations entre les individus.

La complexité des comportements humains vient notamment de la variabilité interindividuelle [74]. Les variations de comportement d'un individu dépendent de ses préférences, ses désirs, ses motivations, ses affects, ses pensées et ses croyances [113]. Ils ne sont ainsi pas nécessairement rationnels (*i.e.* basés uniquement sur un raisonnement) et parfois difficilement explicables. C'est pourquoi l'analyse du comportement humain est une tâche difficile. Nous verrons dans les sous-sections suivantes les disciplines s'y intéressant.

1.1.3 Disciplines

La reproduction de comportements humains est un problème de recherche actuel. Le domaine de la simulation qui modélise ces comportements a un besoin de validation de ces derniers. Afin de valider ces comportements simulés, il est nécessaire d'avoir des outils d'analyse de ces comportements. Pour cela, il est nécessaire de définir le comportement humain et ce qui l'influence. Nous allons donc voir ici les différentes définitions utilisées et leurs implications.

Éthologie : Pour les éthologistes comme Darwin, le comportement est exécuté de manière invariante pour l'ensemble d'une espèce. Il est la réponse préprogrammée à un stimulus exogène (réactif) ou endogène (proactif). Cette réponse se manifeste sous la forme d'une chaîne de réflexes. Cette définition place intégralement le comportement comme instinctif et inné [132]. L'éthologie nous montre ainsi l'importance de l'inné pour les comportements. De plus, elle démontre la nécessité de prendre en compte l'environnement au sens de stimuli-réponses.

Béhaviorisme : Les behavioristes se sont séparés de la psychologie pour devenir une science naturelle. En effet, ils se limitent uniquement aux conséquences observables des comportements évitant toute introspection. Cette théorie se base sur l'association stimulus-réponse, concept clé du modèle de réflexe conditionné *pavlovien* [109]. Ce domaine intègre les comportements au sens éthologique (inné) et y ajoute la notion d'acquis. Cette notion permet une adaptabilité à un environnement dynamique [42]. Dans ce domaine, un comportement est initialement aléatoire ou inné. Par la suite, l'individu visant à s'adapter, son comportement se modifie dans une recherche de maximisation de l'utilité (notion de succès, de plaisir, satisfaction, ...).

L'apprentissage de comportements est nommé *conditionnement opérant* ou *apprentissage skinnerien* [128]. Si la conséquence d'un comportement tend à ce qu'il apparaisse plus souvent (respectivement moins souvent), c'est un *renforcement* (respectivement une *punition*). Si cet apprentissage nécessite l'ajout (respectivement le retrait) d'un stimulus, il est alors dit *positif* (respectivement *néгатif*). Le béhaviorisme apporte donc la notion d'apprentissage et donc d'acquis.

Psychologie : La psychologie cognitive ajoute la présence d'*états mentaux* [107]. Elle propose des modèles qui intègrent la capacité de représentations [28]. Les comportements peuvent alors dépendre des activités internes (pensée, affects, ...) [88]. Un être humain a la capacité de se représenter ses propres états mentaux ainsi que ceux des autres. Ses états mentaux modifient alors sa compréhension des comportements observés ainsi que les comportements qu'il adopte [113]. Cette caractérisation de l'individu nous concerne particulièrement car les agents virtuels sont construits de manière similaire, c'est-à-dire avec une boucle perception-cognition-décision-action. De plus, le fonctionnement cognitif décrit par les chercheurs en Sciences Cognitives est basé sur des modèles qui sont parfois transposables à l'informatique. L'utilisation de cette définition nous permettrait ainsi d'analyser les comportements des agents dans leurs causes et leurs effets en faisant des parallèles avec les comportements humains. Il est à noter que la psychologie sociale montre l'importance de la société et l'influence de la culture sur les comportements adoptés [133]. Cette dimension impacte donc la portée des comportements d'agents modélisés : les cultures, les règles juridiques (ainsi que leur respect) et les normes sociétales peuvent différer par exemple selon la nationalité. Cela devra être pris en compte dans la portée des résultats obtenus qui dépend de la population de participant étudiée.

La définition que nous adopterons pour la suite de ce manuscrit est la suivante : les comportements sont un ensemble de réactions observables d'un individu sur son environnement. Ils dépendent fortement des stimuli de l'environnement extérieur pour réagir de manière rapide et adaptée. Ils ont un état initial provenant de l'inné puis évoluent par apprentissage via l'acquis obtenu lors du vécu. Ces comportements sont sujets à modification en fonction de l'état mental de l'individu. Les pensées et les émotions font varier les comportements habituels et leurs déclencheurs. Aussi, la dimension sociale impacte les comportements.

1.1.4 Niveau de comportement

Le comportement humain peut être analysé selon plusieurs angles. D'une part, considérant les processus cognitifs engagés dans une tâche spécifique, il est possible de séparer différents niveaux de comportements selon la difficulté. Rasmussen [118] propose par exemple une hiérarchisation des comportements en trois niveaux (voir *figure 1.1*) :

- ceux basés sur les *compétences* sensori-motrices. Ils sont automatiques et ne demandent pas d'attention ;
- ceux basés sur des *règles* apprises par l'expérience et le savoir-faire ;

- ceux basés sur les *connaissances* en vue de la satisfaction d'un but, demandant un choix de planification fait sur des prédictions.

D'autre part, il est possible de différencier les comportements en fonction de la difficulté de la tâche. Par exemple, dans le cadre de la conduite, Michon [100] a proposé de diviser les tâches en trois niveaux (voir *figure 1.2*) :

- la tâche de niveau *opérationnel* est principalement inconsciente et de l'ordre de l'automatisme, par exemple rester au centre de la voie ;
- la tâche de niveau *tactique* correspond à des situations nécessitant une prise de décision comme dans le cadre d'un dépassement. Les interactions avec les autres individus font partie de ce niveau de comportement ;
- la tâche de niveau *stratégique* définit la planification d'itinéraire ou encore le choix du véhicule.

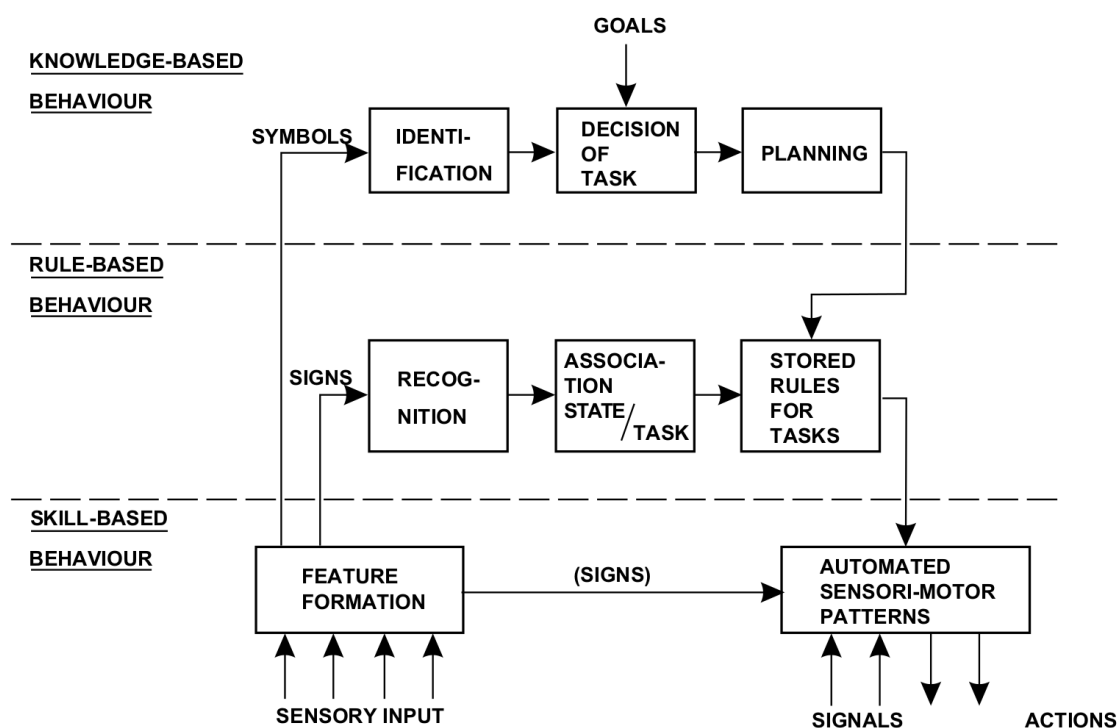


Figure 1.1 – Modèle de Rasmussen.

Nous nous intéressons à la capacité comportementale des agents virtuels en simulation. Des parallèles peuvent être faits entre les niveaux de comportement chez les humains et chez les agents. Nous considérons les différents niveaux de comportements suivants :

- le plus bas niveau est le niveau *opérationnel*. Il correspond aux actions simples de type réflexes comme passer en première vitesse dans une voiture. Ces comportements sont identiques aux opérations élémentaires d'un agent ;
- Un type de comportement de niveau intermédiaire met en place une *tactique* tel le changement de voie d'une voiture sur autoroute. Cela correspond chez l'agent à l'application d'une suite ordonnée de comportements élémentaires ;

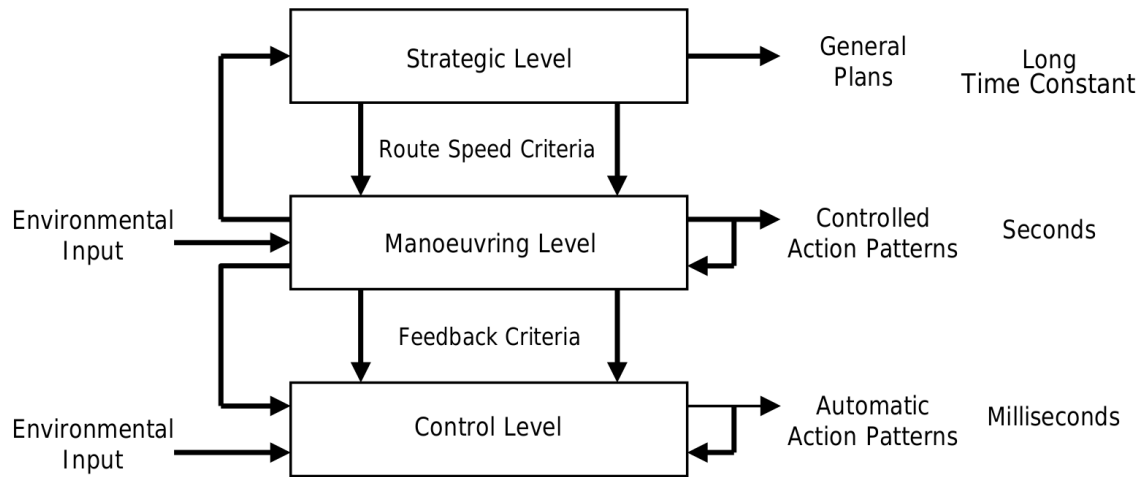


Figure 1.2 – Modèle de Michon.

- Le comportement de niveau plus élevé est le niveau *stratégique* sur le long terme. Il est basé sur des choix et évolue suivant la dynamique de l'environnement, l'état mental de l'individu et ses émotions [50] comme choisir de doubler une file ralentie de camion. Pour un agent, cela représente l'état du modèle cognitif (et possiblement émotionnel) de l'agent. Cet état mental virtuel modifiera la prise de décisions, impactant alors les tactiques adoptées.

Il existe déjà des moyens d'analyse des comportements de niveau *opérationnel* basés sur les *compétences* sensori-motrices, par exemple sur le système moteur [52]. Les travaux s'attardant sur les tâches de niveau *tactique* ou *stratégique* demandant la mise en œuvre de *règles* ou de *compétences* sont rares [19]. Or, comme expliqué précédemment, la "*cohérence*" des comportements adoptés par les agents est importante en simulation. C'est donc à ces niveaux de comportement que nous souhaitons évaluer la crédibilité comportementale des agents.

1.1.5 Conclusion

Dans notre méthode, nous voulons évaluer le comportement (c'est-à-dire les réactions observables) des agents aux niveaux tactique et stratégique : nous considérons que le comportement est adopté selon un choix tactique, qu'il évolue selon la dynamique de l'environnement et qu'il dépend de l'état mental de l'individu, comme proposé par [50]. C'est pour cela que nous distinguons les traces de simulation (c'est-à-dire les données numériques brutes recueillies après simulation) - qui ne sont que le résultat partiel du comportement sur l'environnement - du comportement lui-même tel qu'il peut être examiné et compris par un observateur humain [113]. Cette distinction nous amène à considérer différents types de recherches effectuées sur l'analyse de ce niveau de comportement.

1.2 Évaluation en Sciences Humaines et Sociales

Les Sciences Humaines et Sociales (*SHS*) sont très souvent utilisées afin de modéliser et de valider les Agents Virtuels et les Agents Conversationnels Animés (*ACA*), et ce dans de nombreux domaines tels que les jeux vidéo [22], l'éducation [120], la communication [8], la simulation de foule [12], l'humain virtuel [65], le secteur financier [11], la génération d'expressions faciales [30]), et la décision macro-économique [98].

Les Sciences Humaines et Sociales s'intéressent notamment à l'étude des comportements humains. Elles fournissent un ensemble de méthodes d'évaluation, dont une partie consiste à utiliser des participants pour évaluer l'objet d'étude. Elles peuvent se baser sur l'analyse des réactions physiologiques du participant ou sur diverses méthodes permettant de saisir avec justesse un élément de son état mental : croyance, désir, intention, connaissance, ... Par exemple, Gratch *et al.* [63] utilisent des questionnaires tels que le *Stress and Coping Process Questionnaire (SCPQ)* afin de comparer l'état mental des participants avec celui de ses agents. Cela permet d'obtenir des informations de manières indirectes sur les stratégies d'évaluation et d'adaptation des participants. Ces informations servent alors à évaluer la qualité du modèle d'agent en les confrontant avec les variables internes des agents. De même dans [64], il est question de vérifier la validité des modèles basés sur la théorie de l'évaluation (ou *appraisal*). Pour cela, des questionnaires sur la tendance à la coopération et la compétition sont soumis avant l'expérimentation. D'autres questionnaires sont soumis pendant l'expérimentation sur les émotions ressenties (peur, joie, tristesse, colère, et espoir) et sur l'évaluation/adaptation (avec des échelles sur la perception de l'utilité de gagner, sur la capacité à contrôler les conséquences des actions, sur l'effort fournis pour gagner, ...).

La force de cette approche est de caractériser les comportements adoptés et de pouvoir saisir directement les comportements haut niveau au travers des réponses des participants humains qui ont cette capacité de compréhension et d'analyse du comportement d'autrui [113].

Dans le cadre de notre problème, la simulation immersive ajoute des contraintes spécifiques aux méthodes d'évaluation que nous présentons dans la sous-section suivante.

1.2.1 Simulation immersive

Dans le cadre de la simulation immersive, il est possible d'évaluer la qualité de l'environnement virtuel de manière générale ou détaillée. Witmer [139] sépare cette notion de qualité de l'*EV* appelé *implication (i.e. involvement)*, des conséquences de cette qualité sur l'utilisateur regroupées sous la notion d'*effet de présence (i.e. presence effect)*. Cette dernière est évaluée notamment au travers de questionnaires [84]¹. Cependant, la définition de la notion de présence ne fait pas

1. “Both involvement and immersion are necessary for experiencing presence. Involvement in a VE (virtual environment) depends on focusing one’s attention and energy on a coherent set of VE stimuli... Immersion depends on perceiving oneself as a part of the VE stimulus flow”

encore consensus au sein des communautés de chercheurs en Sciences Humaines et Sociales. Nous présentons donc en *sous-section 1.2.4.a* l'évolution de cette notion.

Pour la validation du comportement simulé en réalité virtuelle, une approche subjective est souvent utilisée. En simulation, c'est le participant plongé dans l'environnement virtuel qui observe son environnement et donc, dans le contexte de l'immersion, qui évalue la crédibilité de la simulation et des agents qui la composent. Cette approche est basée sur l'évaluation de la similarité des comportements avec les êtres humains. Par exemple dans le contexte d'interaction homme-machine, Campano *et al.* [21] proposent une méthode d'évaluation de modèles affectifs, Pelachaud *et al.* [110] évaluent la crédibilité du modèle de comportement affectif des expressions de l'agent *Greta*. Bosse *et al.* [11] analysent la capacité de leur agent (faisant des opérations financières) à reproduire des expressions faciales via un questionnaire proposé aux participants. Prendinger *et al.* [114] incluent des filtres sociaux aux modèles des agents virtuels et vérifient par des questionnaires que les participants ont bien reconnus les rôles sociaux des agents et que ces derniers répondent de manière appropriée.

1.2.2 Moyens d'évaluation

Nous présentons ici les méthodes utilisées en *SHS* pour étudier des comportements humains par la perception du participant, ordonnés selon le niveau des données qu'ils permettent d'obtenir.

1.2.2.a Réactions physiologiques

Il est possible d'obtenir des informations à partir des réactions du corps humain au travers de mesures objectives. Par exemple, l'activité électrodermale est fortement utilisée dans l'analyse des stimuli émotionnels. Elle permet la reconnaissance des émotions via l'analyse des changements de sudation notamment des doigts. En outre, elle a permis d'évaluer le niveau de présence en environnement virtuel sur sa composante visuelle [57]. Autre exemple, l'oculométrie est une technique d'enregistrement des mouvements oculaires via une caméra. L'analyse des fixations permet entre autre de savoir à quel interlocuteur le participant donne son attention [137].

Cependant, nous n'utiliserons pas ces approches dans notre méthode car elles ne sont pas transposables aux agents artificiels.

1.2.2.b Questionnaire

Un questionnaire permet une mesure directe mais subjective de l'objet d'étude (c'est-à-dire de la question soulevée, par exemple l'importance pour l'utilisateur de la qualité des comportements). Il est préférable et coutumier d'utiliser des questionnaires qui ont été validés par des chercheurs en sciences humaines et des experts. Cette validation assure l'absence de dérive par rapport à l'objet d'étude. Cela prend en compte l'ordre des questions, la compréhension des questions par les sujets avec un vocabulaire non ambigu, l'absence d'induction des réponses par les questions,

la pertinence des modalités de réponses et l'utilité de toutes les questions. Afin de caractériser les comportements observés (mais aussi d'éviter le biais dû à l'introspection), ils peuvent être annotés via un questionnaire rempli a posteriori par une tierce personne.

Travaillant avec des sujets français, il nous est nécessaire de soumettre des questionnaires uniquement en français. Avoir une traduction validée du questionnaire est un atout important car cela permet d'avoir une traduction fidèle au niveau littéraire et adaptée à la population (culture, habitude, *etc.*). Cela permet aussi de s'assurer que la traduction a les mêmes propriétés que la version originale (via une analyse factorielle) en termes de normes et de corrélations sur l'échelle globale et ses sous-échelles.

Cependant, il se peut qu'aucun questionnaire existant ne vérifie l'hypothèse posée. Dans ce cas, de nouveaux questionnaires doivent être créés pour répondre à ces besoins spécifiques.

1.2.2.c Discussion et entretien d'autoconfrontation

Les discussions (ou interviews) permettent de récolter des données qualitatives sur les participants. Les phrases dites par les participants peuvent être retranscrites fidèlement grâce à un enregistrement audio. Par exemple, Cahour *et al.* ont utilisé des interviews afin d'analyser les émotions dans le cadre de la conduite [17, 18].

Les entretiens d'autoconfrontation (ou de remise en situation) permettent aussi d'obtenir des données qualitatives. L'introspection permet à l'individu de se remémorer un instant de manière fidèle [136]. Cette méthode est complexe à maîtriser afin de ne pas introduire de biais dans les données récoltées [112].

Ces deux méthodes demandent un traitement important des données récoltées afin d'en extraire des informations utiles. De plus, ce traitement est impossible à automatiser à l'heure actuelle de par la nature qualitative des données. Nous n'utiliserons donc pas ces méthodes.

1.2.3 Évaluation *a priori*

Les comportements des participants peuvent être en partie déterminés *a priori*, c'est-à-dire leur comportement en général ou avant l'expérimentation. En effet, il est possible de prédire le type de comportement qu'adoptera un participant en environnement virtuel. Cela peut permettre de comparer la prédiction faite avec les résultats de l'analyse effectuée. Nous verrons dans les sections suivantes l'analyse a priori de l'immersion d'un individu dans un environnement virtuel et la prédiction du type de comportement de conduite. Cela nous permettra d'introduire des notions et des questionnaires qui nous seront utiles par la suite.

1.2.3.a Propension à l'immersion

Les sujets réagissent différemment face aux techniques d'immersion mises en œuvre. C'est pourquoi ils n'ont pas le même ressenti en environnement virtuel. Leur capacité à se plonger dans l'environnement virtuel est corrélée avec l'effet de présence comme le montre Witmer et Singer [139]. De la même manière, cette capacité joue-t-elle un rôle sur l'appréciation des comportements simulés ? Nous souhaitons donc tester l'hypothèse que la crédibilité des comportements est corrélée positivement avec la capacité à se plonger dans l'environnement virtuel.

Une autre question de recherche nous intéresse : la pratique des jeux vidéo a-t-elle un impact sur la présence ressentie ? Les participants ayant l'habitude des jeux vidéo s'immergent-ils plus facilement en *EV* ? Aussi, par rapport à la problématique de l'évaluation de la qualité des comportements des agents virtuels, nous pensons que l'expérience en jeu vidéo du participant modifie sa capacité à détecter les problèmes du modèle d'agent. Nous nous demandons si son habitude à être immergé augmente sa capacité à repérer des erreurs et si cette habitude facilite cette détection. De plus, un joueur expérimenté ayant de plus fortes espérances, il serait alors plus critique sur les défauts perçus. Il est aussi possible qu'à l'inverse cette habitude rend certaines erreurs *normales/courantes*, *i.e.* le participant passerait outre certains défauts habituels. En conséquence cela diminuerait alors sa capacité de détection. Dans le but de vérifier ces hypothèses, nous présenterons dans la *sous-section 1.2.3.a* un questionnaire permettant d'identifier le type de joueur de jeu vidéo.

Tendance a priori à l'immersion Witmer et Singer valident dans l'article [139] un questionnaire de tendance à l'immersion : l'*Immersive Tendencies Questionnaire (ITQ)*. Ce questionnaire contient trois sous-échelles : *Involvement*, *Focus* et *Games*. Les 18 items de l'*Immersive Tendencies Questionnaire* ont été traduits et validés en 2002 par le laboratoire de Cyberpsychologie de l'université du Québec [121]. Le *Questionnaire de Propension à l'Immersion* contient quatre sous-échelles : *focus*, *implication*, *émotions* et *jeu*. Ce questionnaire permettra une analyse *a priori* de la capacité d'immersion du participant. Cependant, l'échelle *jeu* ne contient que deux items dans le questionnaire et ne permet donc pas une analyse fine du type de joueur de jeu vidéo. C'est pourquoi cette échelle est insuffisante pour analyser le participant en tant que joueur.

Implication dans le jeu vidéo L'article "*Experiencing Presence in Video Games: The Role of Presence Tendencies, Game Experience, Gender, and Time Spent in Play*" écrit par Lachlan et Krcmar en 2011 montre que plus un humain a passé du temps dans un environnement virtuel comme un jeu vidéo interactif, plus le degré d'expérience de présence est élevé, elle est donc modifiée par l'habitude à jouer aux jeux vidéo [80]. Ainsi, l'habitude d'un sujet à jouer aux jeux vidéo modifie leur présence dans l'environnement virtuel. C'est notamment pourquoi l'*Immersive Tendencies Questionnaire* propose deux items en relation avec les jeux vidéo.

Pour évaluer plus finement la typologie de l'engagement du joueur, Brockmyer, Fox, Curtiss, McBroom, Burkhart, Pidruzny ont développé le *Game Engagement*

Questionnaire (GEQ) publié dans l'article "*The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing*" en 2009 [13]. Il contient 19 items permettant de définir le degré d'engagement du sujet et de le caractériser lors d'une expérience dans un jeu vidéo. Il permet de distinguer le niveau d'implication dans un jeu vidéo parmi : *Absorption*, *Flow*, *Presence* et *Immersion* (voir *figure 1.3*). Ce questionnaire est pertinent pour l'évaluation du participant en tant que joueur et permettra la vérification du lien entre habitude aux jeux vidéo et présence ressentie.

Travaillant avec des populations de participants français, il nous est nécessaire d'utiliser des questionnaires en français. Cependant, il n'existe pas à notre connaissance de traduction du *Game Engagement Questionnaire*. Une version en français devra dès lors être faite.

La tendance à l'immersion et la tendance à l'implication dans les jeux vidéo permettent d'évaluer *a priori* l'effet que devrait ressentir le participant lors de l'expérimentation. Des informations sur l'expérience vécue par utilisateur peuvent être obtenus après simulation. Elles peuvent corroborer les résultats obtenus *a priori* mais aussi apporter des données supplémentaires sur l'expérience et les comportements observés.

1.2.3.b Comportement de conduite

Afin d'analyser les comportements dans un domaine particulier, il est possible d'utiliser des méthodes *ad hoc* construites spécifiquement. Le domaine qui nous concerne est celui de la simulation de conducteur. Des outils spécifiques au domaine permettant l'étude du comportement du conducteur dans la vie réelle ou en simulation ont été développés comme dans "*Systematic Approach for Validating Traffic Simulation Models*" [105], "*Systematic verification, validation and calibration of traffic simulation models*" [116], ou encore "*Procedure for validation of microscopic traffic flow simulation models*" [5].

Pour connaître le type de comportement général de notre sujet, il existe des questionnaires créés spécifiquement dans ce but. Plus précisément, le questionnaire de comportement de conduite analyse ce comportement général dans le cas d'un conducteur. Il permet d'évaluer à quel type de conducteur le sujet se rapproche et est effectué avant l'expérimentation pour ne pas être biaisé par la simulation.

Il existe quatre principaux questionnaires sur la conduite qui ont été créés entre 1989 et 1993 :

- Le *Driving Behaviour Inventory (DBI)* publié en 1989 dans l'article "*Dimensions of driver stress*" [67] par Gulian, Matthews, Glendon, Davies et Debney permet de faire des prédictions sur le niveau de stress du conducteur (voir *table 1.1*).
- Le *Driving Style Questionnaire (DSQ)* de French, West, Elander et Wilding ("*Decision-making style, driving style, and self-reported involvement in road traffic accidents*" [54]) publié en 1993 (voir *table 1.2*) montre entre autre une

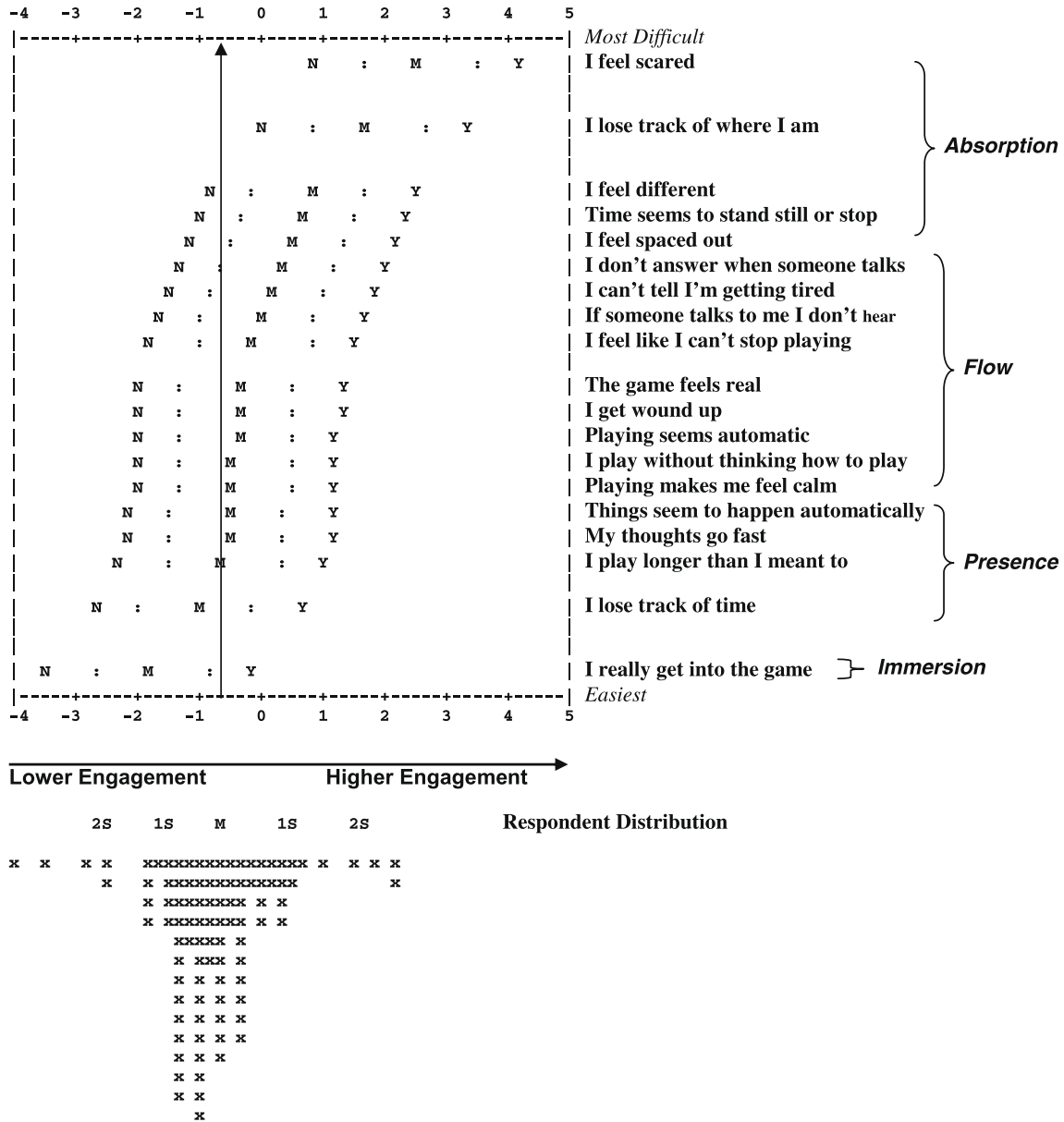


Figure 1.3 – Game Engagement Questionnaire. Scores d’implication des participants représentés par un \times . Le graphique est centré à une moyenne de 0 et un écart-type de 1. Les items sont triés par ordre de difficulté avec les plus simples à être en accord en bas de l’échelle. *N*, *M* et *Y* réfèrent respectivement à “Non”, “Peut-être”, et “Oui”. Le symbole deux-points représente le fait d’être à mi-chemin entre une catégorie ou l’autre. La figure est extraite de [13].

corrélation entre la vitesse et les accidents de la route et une corrélation inverse entre planification et accidents. Ce questionnaire se concentre sur les comportements à risque dans la conduite.

- Le même article [54] présente un questionnaire complémentaire : le *Decision-Making Style Questionnaire (DMSQ)* contenant 30 items (voir *table 1.3*). Il pose des questions sur la fréquence des types de comportement.
- Le *Driving Behavior Questionnaire (DBQ)* contenant 50 items a été créé par Reason, Manstead, Stradling, Baxter, et Campbell en 1990 dans l’article “*Errors and violations on the roads : A real distinction?*” [119]. Il permet de

Table 2. Coefficients alpha for Driving Behaviour Inventory scales on four occasions.

Scale Description	Coefficients alpha			
	Gulian <i>et al.</i> (1989)		Reliability study	
	Study I	Study II	Time 1	Time 2
I Driving aggression	0.799	0.688	0.610	0.583
II Irritation when overtaken	0.792	0.707	0.663	0.629
III Driving alertness	0.595	0.703	0.009	0.352
IV Dislike of driving	0.731	0.667	0.699	0.689
V Frustration in overtaking	0.800	0.661	0.614	0.392
General driving stress scale	0.858	0.819	0.799	0.796

Table 1.1 – Échelles du *Driving Behaviour Inventory* (figure extraite de [61]).

Table 4. Items loading on the dimensions of driving style.

Factor	Loading	Item
1. Speed (18.3% of variance)	0.81	Do you break the motorway speed limit?*
	0.79	Do you drive fast?
	0.78	Do you exceed the speed limit in built up areas?
2. Calmness (12.5% of variance)	0.81	Do you become flustered?*
	-0.69	Do you remain calm?*
	0.68	Do you respond to pressure from other drivers?*
3. Social resistance (9.4% of variance)	-0.82	Are you happy to receive advice from people?*
	0.81	Do you dislike people giving advice?*
4. Focus (8.7% of variance)	0.70	Do you drive cautiously?
	0.69	Do you find it easy to ignore distractions?*
	0.61	Do you ignore passengers?*
5. Planning (7.3% of variance)	-0.82	How often do you set out without looking at a map?*
	0.79	Do you plan long journeys in advance?*
6. Deviance (6.9% of variance)	0.84	Do you overtake on the inside?*
	0.60	Do you ever drive through a red traffic light?*

Table 1.2 – Échelles du *Driving Style Questionnaire* extraites de [54]. Les items avec un astérisque ont été abrégés.

décrire les erreurs et les violations du code de la route effectuées par le sujet et plus particulièrement celles faites volontairement (voir *table 1.4*). Comme la plupart des questionnaires, il est rédigé pour une population anglophone. Nos sujets étant francophones, il ne nous est pas possible d’appliquer directement ce questionnaire, il nous faut alors une version traduite et validée.

À notre connaissance, seul le *Driving Behavior Questionnaire* a été traduit et validé en français. Le questionnaire de Comportement de Conduite est une adaptation du *Driving Behavior Questionnaire* traduit en français en 2010 par deux chercheurs francophones du domaine de la psychologie de la conduite et vérifiée par un anglophone : Gabaude, Marquié et Obriot-Claudiel dans l’article “*Self-regulatory behaviour in the elderly: relationship with aberrant driving behaviours and perceived abilities*” [56].

Table 2. Items loading on the dimensions of decision-making style.

Factor	Loadings	Items
1. Thoroughness (17.9% of variance)	0.79	Do you work out all the pros and cons?*
	-0.70	Do you decide without considering all the implications?*
	0.65	Do you plan well ahead?
	0.64	Is your decision-making deliberate and logical?*
2. Control (9.5% of variance)	0.76	Do you remain calm?*
	0.75	Do you enjoy making decisions?
	0.72	Do you feel in control of things?
	-0.58	Do you avoid making decisions if you can?*
	-0.45	Do you find it difficult to think clearly?*
3. Hesitancy (7.9% of variance)	0.70	Do you favour first one option then another?*
	0.57	Do you change your mind about things?
	0.55	Do you take the safe option if there is one?
4. Social resistance (6.5% of variance)	0.70	Do you avoid taking advice over decisions?
	-0.57	Do you like to consult with others?
	0.55	Do you make up your own mind about things?*
5. Perfectionism (5.8% of variance)	0.70	Do you carry on looking for something better?*
	0.69	Do you settle for an option that will just about do?*
6. Idealism (5.2% of variance)	0.80	Are practicalities more important than principles?*
	-0.69	Are your decisions governed by your ideals?*
7. Instinctiveness (5.0% of variance)	0.67	Do you rely on 'gut feeling' when making decisions?
	0.61	Do you stick by your decisions come what may?

Table 1.3 – Échelles du *Decision-Making Style Questionnaire* extraites de [54].
 Les items avec un astérisque ont été abrégés.

1.2.4 Évaluation *a posteriori*

Après avoir présenté l'analyse possible du comportement du participant *a priori*, nous allons dans cette section présenter les moyens d'évaluer le comportement *a posteriori*, c'est-à-dire via des questionnaires soumis au participant après l'expérimentation.

Nous nous intéressons à l'expérience en *EV* de l'utilisateur et à l'impact des comportements des agents. Nous aborderons donc dans les sections suivantes les notions de *présence* et de *crédibilité comportementale*. Nous posons l'hypothèse que la crédibilité des comportements des agents a un impact sur l'expérience en *EV* vécue par les utilisateurs. De plus, nous pensons que la crédibilité comportementale du système en général est différente de celle spécifique à un comportement ou à un agent. En effet, il nous semble que - comme pour l'évaluation objective macroscopique versus microscopique - les utilisateurs peuvent évaluer de manière positive les comportements des agents de manière générale sans pour autant considérer les comportements individuels comme correctement reproduits.

Si le domaine d'application du système s'est abstenu de faire des études de genre ou si lors de la conception des agents du système la question du genre n'a pas été posée, il est alors possible que seuls les comportements des hommes (respectivement des femmes) soient reproduits et que cela implique des différences dans l'évaluation des

Table 1. Examples of the types of aberrant behaviour used in the driver behaviour questionnaire.

Slips	Misread the signs and exit from a roundabout on the wrong road. Switch on one thing (e.g., headlights), when you intended to activate something else (e.g., wipers).
Lapses	Have no clear recollection of the road just travelled. Attempt to drive away from traffic lights in third gear.
Mistakes	Underestimate the speed of an oncoming vehicle when overtaking. Get into the wrong lane on approaching a roundabout or junction.
Unintended violations	Forget when road tax/insurance expires and drive illegally. Check speedometer and discover that you are unknowingly in excess of the legal speed limit.
Deliberate violations	Become impatient with a slow driver in the outer lane and overtake on the inside. Get involved in unofficial ‘races’ with other drivers.

Table 1.4 – Échelles du Driving Behavior Questionnaire. Table extraite de [119].

comportements en fonction du sexe du participant. Nous posons donc aussi l’hypothèse que le ressenti du participant peut - selon les domaines d’application - différer selon le genre. Le sexe du participant peut avoir une influence sur son ressenti de la qualité des comportements notamment dans le cas où les comportements humains sont genrés (*i.e.* divergent selon le genre) et si les comportements des agents représentent en mauvaise proportion un des deux sexes. Nous souhaitons effectuer une étude de genre afin de vérifier l’hypothèse que la présence ressentie et l’évaluation de la crédibilité comportementale peuvent dépendre du genre des participants.

1.2.4.a Présence

Le terme *présence* vient de *TelePresence* utilisé par Minsky [104]. Il désigne en *RV* le sentiment de réalisme perçu. Pour Lombard [89], c’est le sentiment de ne plus percevoir le “*medium*” entre la réalité et l’environnement virtuel (“*illusion of non-mediation*”) Pour Slater [129], la présence représente un état de conscience, le sentiment d’être dans l’environnement virtuel. Schubert [125] pense que le sentiment de présence n’est pas une conséquence directe de la qualité d’immersion mais quelle dépend des représentations cognitives construites sur la base de stimuli immersifs.

Comme l’ont montré [15, 131], le sentiment de présence dépend de nombreux facteurs dont des éléments sensoriels (rendu visuel, haptique, proprioceptif, *etc.*). Il est à noter que ce sentiment est sensible à des phénomènes psychologiques comme l’irrationalité et l’incapacité des observateurs humains d’expliquer leurs propres jugements [75].

Un questionnaire de présence nous permet d’évaluer l’importance de l’effet de présence dans toutes ses composantes durant l’expérience. Dans nos travaux, nous nous intéressons à l’évaluation des comportements stratégiques des agents au sein d’un dispositif de *RV*.

Plusieurs questionnaires ont été proposés pour mesurer l’effet de présence :

- Le *Presence Questionnaire (PQ)* de Witmer et Singer est publié en 1998 dans l'article "*Measuring Presence in Virtual Environments : a presence questionnaire*" (présentant aussi l'*Immersive Tendencies Questionnaire*) [139]. Il mesure la présence selon quatre facteurs : le niveau de contrôle, la qualité de l'immersion, les distractions perturbant l'expérience et le degré de réalisme. Slater critiquera en 1999 ce questionnaire pour la mauvaise distinction entre immersion et présence. Il critiquera aussi le fait qu'on évalue la perception du système par l'utilisateur et non sa présence. Dans le cadre de notre étude, il semble cependant préférable d'analyser la perception des éléments du dispositif par les sujets que leur état psychologique. De plus, il reste encore aujourd'hui utilisé par la communauté scientifique de la réalité virtuelle (par exemple [122]) et [101]).
- L'*Igroup Presence Questionnaire (IPQ)* fait par Schubert, Friedmann et Regenbrecht en 2001 dans "*The experience of presence: Factor analytic insights*" [125] se concentre sur la présence spatiale, l'implication et le réalisme de l'environnement virtuel (voir *table 1.5*). Plus concis, il contient 14 questions directes. Une traduction² des items en français a été faite par Isabelle Viaud-Delmon (Voir *annexe A.5*).

Table I. Factor Analysis Study I. Factors, Numbers of Items, and Explained Variance

Component	Name	Label	Number of Items	Eigenvalue	% of Variance Explained
1	Spatial presence	SP	14	14.087	20.717
2	Quality of immersion	QI	8	4.574	6.726
3	Involvement	INV	10	3.824	5.624
4	Drama	DRA	7	3.083	4.533
5	Interface awareness	IA	7	2.485	3.655
6	Exploration of VE	EXPL	6	2.262	3.326
7	Predictability & interaction	PRED	6	1.967	2.893
8	Realness	REAL	5	1.901	2.795

Table 1.5 – Échelles du *Igroup Presence Questionnaire* extraites de [125].

- L'*ITC-SOPI* de Lessiter, Freeman, Keogh, Davidoff présenté en 2001 dans "*A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory*" [84]. Les 44 questions de ce questionnaire servent à cerner la présence spatiale, l'engagement du sujet, la perception de l'environnement comme naturel et les effets négatifs de l'environnement. Malheureusement, il n'existe pas à ce jour de traduction en français.
- Le questionnaire *Slater-Usuh-Steed* [129] (*SUS*) publié dans "*Depth of Presence in Virtual Environments*" en 1994 est encore plus court avec uniquement 6 items et se concentre sur la présence spatiale sous 3 échelles : sentiment d'être dans l'environnement virtuel, environnement virtuel comme réalité pour le sujet et mémorisation de l'expérience comme lieu visité. Les auteurs considèrent que soit le sujet est dans le monde réel, soit il est dans le monde virtuel mais

2. La traduction française du *Igroup Presence Questionnaire* est disponible à l'adresse www.igroup.org/pq/ipq/download.php#French

pas dans les deux. Cela se matérialise par une analyse des réponses ayant des scores élevés (au moins 6 sur une échelle à 7 points).

Comme pour les autres questionnaires nous souhaitons utiliser des versions qui ont été traduites et validées. Le *Presence Questionnaire* est particulièrement intéressant car il permet d'obtenir à la fois un score global et des scores sur des sous-échelles ce qui permet de caractériser plus précisément la présence. Il contient notamment les sous-échelles *Involvement/Control*, *Natural* et *Interface Quality* qui permettront d'éviter un biais dans les réponses du questionnaire de crédibilité comportementale : le sujet n'étant pas un expert du domaine, il peut répondre à une question du questionnaire de crédibilité comportementale pensant juger des éléments pertinents alors qu'il note des dimensions faisant partie d'autres composantes de la présence.

La version originale de ce questionnaire a été traduite en 2002 par le laboratoire de Cyberpsychologie de l'université du Québec contrairement à la version de 2005 [121]. Nommé "état de présence", il contient 24 items comme "Dans quelle mesure vos interactions avec l'environnement vous semblaient-elles naturelles?" notés sur une échelle à 7 points (dont les labels diffèrent selon les questions) et 7 sous-échelles : *Réalisme*, *Possibilité d'agir*, *Qualité de l'interface*, *Possibilité d'examiner*, *Auto-évaluation de la performance*, *Auditif* et *Haptique*. Les dimensions *Réalisme*, *Possibilité d'agir* et *Qualité de l'interface* nous intéressent plus particulièrement pour nous permettre de juger la qualité de la trace du véhicule piloté et des réponses au questionnaire de crédibilité comportementale. Il est à noter que les deux questions de la sous-échelle *Haptique* ne font pas partie de l'analyse factorielle et peuvent être supprimées. Ces questions ne s'appliquent pas pour notre expérimentation sur le simulateur à deux tables mais peuvent avoir un intérêt sur le simulateur lacet. C'est pourquoi nous les conserverons.

Ce questionnaire doit être soumis tout de suite après l'expérimentation en simulateur de façon à être au plus proche du moment de la simulation afin d'éviter au mieux une altération de la mémoire.

Pour l'évaluation des comportements, la notion de présence est trop large car elle comprend des éléments autres que comportementaux (qualité visuelle, sonore, etc.). Cependant des sous-parties de l'effet de présence nous intéressent pour l'analyse des comportements des agents et des participants.

1.2.4.b Dimension comportementale

Comme l'ont montré [66, 3], la *crédibilité comportementale* a un impact sur l'immersion du participant. La crédibilité comportementale est donc une des nombreuses composantes de l'effet de présence. Elle évalue dans quelle mesure les utilisateurs interagissant avec l'agent pensent qu'ils observent un être pensant qui a ses propres croyances, des désirs et une véritable personnalité [85].

Quand il s'agit de l'analyse des comportements des agents, les études se concentrent sur la crédibilité comportementale, c'est-à-dire l'évaluation de la similarité ressentie des comportements humains et agents, indépendamment de comment ils ont été produits. Dans cette approche, les agents sont observés par le participant immergé.

À notre connaissance, il n'existe pas actuellement de questionnaire spécialisé dans l'évaluation de la crédibilité comportementale que ce soit pour l'évaluation du comportement global des acteurs de la simulation (*i.e.* du comportement général) ou pour l'évaluation du comportement d'un agent (*i.e.* d'un comportement spécifique).

La fidélité psychologique détermine dans quelle proportion la tâche simulée engendre chez l'utilisateur une activité et des processus psychologiques identiques à ceux de la tâche réelle [108]. Autrement dit, elle évalue dans quelle proportion le simulateur produit un comportement semblable à celui exigé dans la situation réelle [83].

Dans le cadre de l'intelligence artificielle, nous ne souhaitons pas forcément modéliser fidèlement l'activité psychologique des humains dans nos agents simulés : c'est la crédibilité du comportement produit par l'agent qui nous importe le plus, et non la manière par laquelle ce comportement a été produit. De plus, la fidélité psychologique est dépendante de processus internes à l'humain et par conséquent non observable directement. C'est pourquoi nous nous intéressons uniquement à la crédibilité comportementale et à son évaluation quantitative. Afin d'évaluer de manière spécifique la crédibilité comportementale, nous devons nous doter de questionnaires ciblant cette dernière. Afin de cerner la crédibilité comportementale à plusieurs échelles, ces questionnaires devront l'évaluer d'une part de manière globale et d'autre part spécifiquement à une situation.

1.2.5 Conclusion

La force de l'approche subjective pour l'analyse de la crédibilité comportementale est qu'en s'appuyant sur des questionnaires validés par des experts et des psychologues, elle permet d'obtenir des informations de manière indirecte permettant de décrire les différents types de comportements adoptés. Il est alors possible de regrouper les comportements selon ces descriptions haut niveau. Nous parlerons alors de *classes de comportement annoté* ou *classes d'annotations* dans le cas de questionnaires d'annotations.

Cependant, il est difficile de traiter de nombreux participants (et de nombreux agents), ainsi que de tester l'intégralité des différents paramètres des agents avec une telle méthode puisqu'elle requiert l'annotation d'un grand nombre de données. C'est pourquoi nous étudions dans le chapitre suivant les méthodes d'analyse automatique pour l'évaluation des comportements utilisés en Intelligence Artificielle.

Dans le chapitre suivant, nous présentons les méthodes d'évaluations du comportement utilisées en Intelligence Artificielle ainsi que les techniques de catégorisation.

2 L'évaluation des comportements en Intelligence Artificielle

L'analyse de données de simulation pour l'évaluation de la crédibilité des comportements, que nous qualifierons dans la suite d'approche *objective*, est souvent utilisée dans le domaine des *SMA*. Elle consiste à vérifier au travers de données quantitatives que les agents se comportent d'une façon similaire à ce qui est observé dans une situation réelle, comme le font Champion *et al.* [24, 12].

Cette approche a plusieurs atouts. Premièrement, ne demandant pas le jugement humain, elle permet d'effectuer une expérimentation plus rapidement que l'approche de *SHS*. Secondement, elle autorise l'analyse d'un grand nombre d'agents.

Un autre avantage de cette approche est que les données de simulation sont la conséquence directe des paramètres d'entrée du modèle d'agent (ou des actions du participant humain sur le dispositif). En conséquence, une analyse objective des comportements au niveau microscopique peut permettre d'évaluer pour chaque agent l'impact des paramètres d'entrée du modèle. En découle un possible ajustement de ces paramètres de manière automatique dans un cycle de conception.

Ces atouts combinés aident à avoir une évaluation et un cycle de conception efficaces en testant les nouveaux modèles de manière immédiate. Cependant, seules quelques recherches s'appuient sur une analyse *objective* des données de simulation à l'échelle microscopique, comme [40] et [19].

2.1 Échelles d'évaluation

Dans cette section, nous nous intéressons aux différents niveaux de validation. À savoir l'échelle macroscopique qui s'intéresse à la validité du système dans son comportement global et l'échelle microscopique qui se concentre sur la validité des comportements individuels.

2.1.1 Validation au niveau macroscopique

L'évaluation objective se fait principalement au niveau macroscopique car les données sont plus aisées à obtenir [23, 93].

De nombreux outils sont conçus de manière *ad hoc* afin de s'appliquer au mieux à un domaine. Par exemple, dans le domaine de la conduite, la comparaison du débit en véhicules par heure entre la simulation et la situation réelle pour la simulation de

flux de trafic [116] ou la comparaison des courbes du diagramme fondamental entre débit et densité [77] sont courantes.

Certains travaux utilisent des comportements de références, décrits manuellement [81] ou extraits de situations réelles [123], afin de générer les comportements des agents par des méthodes de classification. Le cycle de conception est alors bouclé par une validation à l'échelle macroscopique via par exemple une carte de chaleur [123] ou des variables globales [81]. Soyer *et al.* [130] proposent une méthode d'ingénierie de *SMA* multi-niveaux. La validation proposée se concentre alors sur la cohérence entre les modèles haute et basse résolution.

Il existe donc de nombreuses méthodes objectives de validation macroscopique, mais la littérature ne contient que peu d'outils ou d'approches au niveau microscopique comme fait par [127].

Comme le montrent Drogoul *et al.* [44], la validation d'une simulation passe nécessairement par une évaluation du système à l'échelle *macroscopique*. Le concepteur du *SMA* doit alors disposer d'outils de validation du phénomène collectif. Par exemple, dans la simulation énergétique, on peut utiliser les courbes de charge pour valider le comportement d'un foyer [1].

Cependant, si la validité macroscopique est nécessaire pour la réalité virtuelle (*RV*) ou la simulation participative, elle ne garantit pas une validité au niveau microscopique, c'est-à-dire au niveau du comportement de chaque agent du système. En effet, un comportement collectif valide ne signifie pas que les comportements individuels qui le compose le soient. Par exemple, il est possible dans certains cas d'obtenir une courbe de charge correcte avec des comportements individuels invalides. Vérifier la corrélation entre participants humains et agents de la vitesse moyenne et de son écart-type ne valide qu'en partie les comportements des agents. C'est ce que font Hattori *et al.* [70] pour valider de manière quantitative les comportements de véhicules simulés issus de règles *ad hoc*. En effet, d'autres paramètres du modèle d'agent peuvent amener à des comportements se révélant faux : une mauvaise reproduction des actions sur le volant pourrait entraîner des prises de virages saccadés sans pour autant modifier ces corrélations. Ainsi, il faut une analyse plus fine des comportements des agents en tant qu'entités, c'est-à-dire au niveau microscopique. Or il existe encore peu de travaux sur l'évaluation objective au niveau microscopique.

2.1.2 Validation au niveau microscopique

La validation au niveau microscopique s'intéresse aux comportements individuels des agents. Elle ne valide pas nécessairement le comportement global du système, c'est-à-dire au niveau macroscopique. Cette validation est basée sur les traces de simulation qui sont par essence bas niveau.

Un exemple de validation au niveau microscopique est la méthode proposée par Mathieu et Picault dans [97]. Elle permet d'extraire des profils de comportement de consommateur à partir de traces réelles bas niveau que sont les tickets de caisse. Cette méthode permet ensuite de générer une population d'agents basée sur ces prototypes dont la validation est faite au niveau microscopique. Cette population d'agents adopte alors des comportements qui sont individuellement valides.

Parmi ces travaux qui s'intéresse au niveau microscopique, on peut aussi citer les outils de Serrano [127] qui résument les interactions d'un système multi-agents, c'est-à-dire les envois de messages, pour un débogage manuel. Néanmoins, les données de simulations impliquant des participants ne sont pas uniquement composées de variables issues d'échange de messages de type *FIPA* : il existe aussi des interactions indirectes entre agents comme l'allumage d'un clignotant dans le cadre de la conduite. Ce type d'interaction est plus difficile à analyser car les destinataires de ces interactions ne sont pas directement mentionnés. Ces outils ne sont donc pas directement applicables notamment car les données collectées ne sont pas les interactions elles-mêmes.

Les données bas niveau récoltées lors de la simulation peuvent être proches des comportements bas niveau (*i.e.* comportement moteur / réflexe). Cependant, lorsqu'on s'intéresse à des comportements haut niveau, ces données ne permettent pas directement une analyse de ces comportements. Ainsi, Caillou [19] a montré que l'analyse et la comparaison de données au niveau microscopique ne peuvent être faites directement sur les traces de simulation en raison du fossé sémantique entre les données brutes et les comportements recherchés. Une autre raison complexifiant l'analyse est le fait que ces données sont bruitées de par l'utilisation d'un dispositif. De plus, certaines simulations ne sont pas synchrones. Cela augmente la complexité des données à étudier et ajoute une dimension temporelle à l'analyse. Les travaux de Chiron *et al.* présentés dans [27] utilisent par exemple une approche statistique pour la classification non-supervisée de comportement de fourmis à partir de trajectoire. La difficulté engendrée par la temporalité des données a été écartée en extrayant des caractéristiques scalaires à partir des données brutes : la moyenne et l'écart-type, le minimum et le maximum, ...

Si l'analyse des comportements haut niveau ne peut être faite à partir des traces bas niveau, ces traces bas niveau contiennent néanmoins toutes les informations du comportement haut niveau. La difficulté réside donc dans la transformation complexe de ces données afin de les exploiter. Certains travaux, comme [19], proposent donc de partir des traces (le niveau opérationnel du comportement) et de s'appuyer sur des connaissances d'experts pour construire des abstractions de plus haut niveau (tactique ou stratégique), sur lesquels ils peuvent ensuite travailler. Bien que cette approche nécessite une adaptation spécifique au domaine et une intervention d'experts, elle définit une première technique d'analyse des capacités des agents.

2.1.3 L'humain comme référence

Dans le cadre de nos simulations immersives, l'analyse des comportements haut niveau des agents n'est pas suffisante. Il faut aussi pouvoir comparer les comportements des agents à ceux des humains afin de déterminer quels comportements les agents sont capables de reproduire, mais aussi ceux pour lesquels ils ne sont pas en mesure de les reproduire. Cela nécessite de pouvoir aussi catégoriser les comportements humains.

Une autre solution proche pour l'analyse de données, adoptée par [62] pour les tâches de conduite, consiste à classifier les participants selon des variables objectives

comme la vitesse et la position. Cette solution a été appliquée dans une expérimentation sur 9 participants *recevant l'ordre d'être normatifs* dans un circuit ne comportant pas d'autres véhicules (*i.e.* sans interaction multi-agents). Cependant, le cadre de notre problème nous oblige à traiter à la fois plus de variables, des comportements de plus haut niveau, des agents et des participants libres d'adopter le comportement de leur choix et en interaction les uns avec les autres. Ceci augmente la difficulté de la tâche de classification et rend l'approche de Goncalves insuffisante.

Pour répondre à cette limite, certains travaux proposent des méthodes de classification automatique de participants selon des variables bas niveau comme [62]. D'autres approches (comme sus-cité) s'appuient sur des experts du domaine pour déterminer les variables haut niveau qui décrivent le comportement à analyser. En utilisant ces variables, un algorithme de classification automatique peut être utilisé pour classer les comportements des agents [19].

Il est à noter que, dans le domaine de l'interaction, Delaherche et Chetouani proposent des méthodes d'analyse de données pour l'étude de la synchronie [40] qui utilisent la classification de traces de comportement. Cependant, leur but n'est pas d'évaluer le réalisme d'un agent virtuel.

Georgeon *et al.* [58, 59] cherchent à modéliser les comportements de conducteur. Ils proposent donc une méthode d'analyse de traces d'activité humaine exploitant la connaissance des ergonomes sous la forme d'une base de règles et d'une ontologie de concepts. Cette méthodologie propose conjointement un logiciel permettant de visualiser et d'annoter les traces de simulation du participant afin de servir d'outil interactif d'aide à la modélisation [60]. Cet outil est également utilisé dans le cadre de la conception centrée sur l'utilisateur dans [96] afin de construire de manière itérative des modèles cognitifs partiels. Mathern s'intéresse aussi dans sa thèse [95] à la découverte interactive de connaissance à partir de traces d'activité et à la modélisation à base d'automates.

La limitation principale de l'approche objective est que, si elle permet d'observer la différence entre catégories de comportements extraits à partir des traces (nous parlerons de *classes de trace de comportement* ou *classes de traces*), elle ne fournit pas d'information au-delà des variables utilisées. En particulier, elle ne peut pas donner un sens aux comportements obtenus, c'est-à-dire les rendre explicites. À l'inverse, l'approche subjective (qui se base sur une observation et analyse de plus haut niveau) offre cette possibilité.

Il est à noter que l'exploitation des données issues d'une expérimentation en Environnement Virtuel peut être sujette à un biais. En effet, le comportement observé n'est pas le comportement en situation réelle mais le comportement dans cet *EV*. Il est donc possible que le participant agisse de manière différente en *EV*. Un autre biais peut s'ajouter à ce dernier dans le cas où un dispositif d'action sur le système est utilisé (manette, clavier, volant). Ce biais peut provenir des possibilités d'action proposées au participant via ce dispositif (par exemple des actions simplifiées, schématiques) mais aussi du modèle informatique lié au dispositif (*e.g.* la finesse de la retranscription des mouvements). Cette remarque sur la différence entre comportement en *EV* et en situation réelle est aussi valide pour l'évaluation des comportements en *SHS* via des questionnaires.

2.1.4 Conclusion

Notre approche reprend des idées issues de [19] et [62] : d'une part, nous souhaitons nous placer dans un cadre général et abstraire les traces objectives sous la forme de comportements de plus haut niveau, non définis à l'avance. D'autre part, nous souhaitons mettre en œuvre des algorithmes de classifications temporelle permettant d'analyser les comportements de haut niveau produits par les agents, et donc leurs capacités. Le problème qui demeure est que les classes ainsi obtenues automatiquement ne sont pas nécessairement explicables. Autrement dit, l'approche objective seule ne permet pas de vérifier la cohérence de la classification effectuée, ni de donner un sens aux classes de comportement (c'est-à-dire expliciter les types de comportements détectés).

2.2 Catégoriser pour évaluer

L'objectif de nos travaux est d'obtenir des agents reproduisant des comportements humains. Nous avons donc besoin d'évaluer le comportement des agents par rapport à celui des humains. Les participants humains peuvent adopter un même comportement haut niveau comprenant de nombreuses différences dans les traces de simulation mais aussi dans les variables fournis par des experts. Cela peut être dû à des décalages temporels, à du bruit sur les données des participants de par le dispositif, mais aussi à des choix de comportements bas niveau légèrement différents. Par exemple, des conducteurs choisissant tous d'adopter le comportement de dépasser un véhicule ne le feront pas au même moment, ni avec la même distance intervéhiculaire, ... Il est donc nécessaire d'obtenir des archétypes des divers comportements humains.

L'approche que nous proposons dans cette thèse est d'utiliser un algorithme de classification qui regroupera ensemble les différentes instances d'un même comportement haut niveau en fonction de variables spécifiques. Ces abstractions de comportement pourront ensuite servir de points de comparaison.

L'apprentissage automatique est utilisé ici dans le but de construire des modèles de comportements à partir de traces de conduite en simulation. L'apprentissage automatique nous sert à partitionner nos participants ainsi que nos véhicules simulés afin de combler le fossé sémantique entre les données bas niveau que sont les traces et les comportements haut niveau que nous recherchons.

Nous présentons dans ce chapitre d'une part des techniques d'apprentissage ainsi que les spécificités nécessaires pour notre problème, et d'autre part des techniques de recherche de dimensions décorréées.

2.2.1 Classification non-supervisée

Notre problème, de par son aspect générique, ne nous permet pas d'étiqueter les instances, ni de connaître le nombre de comportements possibles *a priori*. Notre problème exige donc une solution issue de l'apprentissage automatique non-supervisé.

La classification non-supervisée est un problème bien connu en apprentissage. Ce problème de classification non-supervisée étant NP-complet, les algorithmes recherchent donc des optimums locaux. Il est à noter que les algorithmes peuvent alors converger vers différents optimums locaux et donc donner des partitionnements divergents.

L'apprentissage non-supervisé se fait sur des données non-étiquetées de façon à ce que les instances d'une même classe partagent au maximum des caractéristiques communes et que les classes correspondent à des partitions clairement différenciées.

2.2.2 Les méthodes de classification non-supervisée

Il existe de nombreuses familles d'algorithmes d'apprentissage non-supervisé permettant d'obtenir diverses structures de données telles que des partitions et des hiérarchies. Nous expliquerons succinctement trois de ces algorithmes dans les sous-sections suivantes : la classification ascendante hiérarchique, les cartes auto-adaptatives et le K -moyennes.

2.2.2.a Classification Ascendante Hiérarchique

La Classification Ascendante Hiérarchique [102] est une méthode heuristique nécessitant une mesure de dissimilarité entre les instances vue comme une distance.

Initialement, chaque instance est associée à une classe qui correspond à une feuille du dendrogramme. Puis itérativement, les deux classes les moins dissimilaires sont fusionnées pour obtenir une nouvelle classe agrégée et ce jusqu'à obtention d'une unique classe contenant l'ensemble des individus qui correspond à la racine du dendrogramme.

La hiérarchisation dépend fortement du critère d'agrégation. Il en existe un grand nombre dont les plus connus minimisent le saut minimum, le saut maximum, la distance moyenne ou maximisent la variance interclasses (*critère de Ward* [138]).

Le choix du nombre de classes peut se faire visuellement à partir du dendrogramme pour lequel la hauteur des nœuds correspond à la distance interclasses. Il est alors possible de couper l'arbre binaire selon un seuil pour lequel les branches de l'arbre sont plus grandes que la moyenne (*i.e.* les instances sont différentes). Nous verrons dans la *sous-section 2.2.2.c* les critères de sélection automatique du nombre de classes.

2.2.2.b Cartes Auto-Adaptatives

Les Cartes Auto-Adaptatives (*SOM*) ou cartes de Kohonen fonctionne par partitionnement d'instances [78]. L'algorithme est basé sur un réseau de neurones à structure topologique dont dépend la notion de voisinage (par exemple des carrés). Chaque neurone a une localisation fixe sur la carte et une localisation modifiable dans l'espace des données.

L'algorithme commence par initialiser aléatoirement les poids des neurones. Itérativement, une instance est choisie aléatoirement, puis le neurone le plus proche (via une mesure de distance) est sélectionné. Les poids de ce neurone (et dans une moindre mesure ceux de ces voisins) sont alors modifiés selon un coefficient d'apprentissage.

2.2.2.c K -moyennes

L'algorithme des K -moyennes [92] est une méthode de nuées dynamiques partitionnant les instances en K classes. Dans la version originale, ce nombre de classes est déterminé manuellement.

L'algorithme classique initialise aléatoirement les K classes. Jusqu'à convergence, chaque instance est associée à la classe le plus proche selon la distance euclidienne, puis les barycentres des classes sont recalculés en fonction des instances affectées (une description plus détaillée est disponible en *sous-section 5.1.1*).

2.2.3 Nombre de classes

Le choix du nombre de classes est souvent à la charge de l'utilisateur comme le nombre de regroupements pour la méthode des nuées dynamiques ou la profondeur du dendrogramme pour une méthode de regroupement hiérarchique. Il peut cependant être automatisé selon un critère de qualité de la classification.

Contrairement aux classifications ascendantes hiérarchiques pour lesquelles le nombre de classes peut être choisi *a posteriori* via une coupe du dendrogramme, la méthode des K -moyennes demande de choisir *a priori* le nombre de classes. De nombreux critères existent pour évaluer *a posteriori* la qualité de la classification selon le nombre de classes [103] dont :

- Le *Simple Structure Index (SSI)* prend en compte la différence maximale de chaque variable entre les classifications, la taille des classes les plus "contrastées" et l'écart-type des variables dans une classe par rapport à sa moyenne globale.
- Le critère *Calinski-Harabasz (CH)*, aussi appelé *Variance Ratio Criterion*, étudie le rapport entre les sommes des carrés intraclasses et celles interclasses en prenant en compte le nombre d'instances et le nombre de classes [20]. Les classifications intéressantes ont un critère *CH* élevé.

Pour pouvoir choisir le nombre adéquate de classes, la méthode des K -moyennes en cascade consiste à faire la classification N fois avec K variant de x à $x + N - 1$. Puis de sélectionner le K le plus adapté selon un certain critère, déterminant ainsi la meilleure classification. Il est possible de faire en un temps raisonnable les classifications pour un K variant de 2 à N puis de calculer pour chaque classification la valeur du critère. Lorsque la classification correspond à un maximum local de la valeur du critère en fonction du K , ce K est retenu comme solution intéressante. Dans le cas d'un critère strictement croissant (respectivement strictement décroissant), seul $K = N$ (respectivement $K = 2$) est retenu même si la croissance (respectivement décroissance) est très faible. Dans ce cas, le K retenu n'est pas forcément le plus

intéressant. Si le critère pour $K = 3$ est inférieur à celui pour $K = 2$, la classification à deux classes est retenue. De même, si le critère pour $K = N - 1$ est inférieur à celui pour $K = N$, la classification à N classes est retenue. En conséquence, le choix du K_{max} modifie les classifications retenues : si le critère est strictement croissant, la classification à K_{max} classes sera forcément retenue.

2.2.4 Cas des données temporelles

Ces algorithmes de classification non-supervisée prennent pour la plupart des valeurs numériques : pour chaque variable (colonne), pour chaque exemple du jeu d'apprentissage (ligne) est associé un numérique. Or les données de simulation que nous utilisons sont pour la plupart temporelles. Nous avons donc à la place de ce numérique une série temporelle, c'est-à-dire un vecteur de numériques ordonné par le temps. C'est pourquoi les versions originales des algorithmes vus ci-dessus ne peuvent s'appliquer directement à notre problème.

Il existe trois approches pour classifier des données temporelles [86] :

- raw-based : l'utilisation d'algorithmes prenant directement des séries temporelles ou fréquentielles en données brutes,
- feature-based : le prétraitement des données pour revenir à des données numériques à partir de vecteurs ordonnés. Parmi les plus évidentes, nous y retrouvons la moyenne, l'écart-type, la pente, *etc.* comme le font Chiron *et al.* [27]. La distance utilisée – classiquement euclidienne – peut être modifiée pour prendre en compte les relations entre les variables comme :
 - la distance de Mahalanobis qui permet de prendre en compte la corrélation des données en pondérant les axes les plus bruités. Elle est définie comme la similarité entre deux vecteurs \vec{x} et \vec{y} :

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y})} \quad \text{Éq. 2.1}$$

Avec Σ_X la matrice de variance-covariance pour p dimensions :

$$\Sigma_X = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_p} \\ \sigma_{x_2x_1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_px_1} & \cdots & \cdots & \sigma_{x_p}^2 \end{pmatrix} \quad \text{Éq. 2.2}$$

Il faut ensuite calculer à partir de la matrice de distances, les nouvelles coordonnées projetées ; ou encore,

- l'algorithme *Dynamic Time Warping (DTW)* qui permet de calculer la similarité entre deux séries temporelles [7]. *DTW* permet de calculer l'appariement maximal entre deux suites ordonnées de façon à mesurer leur similarité. Cet algorithme est détaillé en *sous-section 4.3.2.b*.
- model-based : la modélisation des séries temporelles brutes en autres données conservant l'aspect temporel comme les chaînes de Markov [115]. Pour prendre en compte la nature temporelle des données, Box et Jenkins ont développé en 1970 le modèle Auto Regressive - Moving Average (ARMA). Cependant, ce

modèle ne permet que l'étude d'une unique variable. Pour palier cela, Sims propose en 1980 le modèle Vector Auto Regressive (VAR). Ces deux modèles s'appliquent pour des variables linéaires avec de faibles variations. Engle publie en 1982 l'algorithme Auto Regressive Conditional Heteroscedasticity (ARCH) qui permet l'étude de telles variables [69].

La présence de données de simulation au caractère temporel complexifie l'étude des comportements. Les algorithmes de classification pouvant être utilisés sont en conséquence plus avancés.

Dans notre méthode d'évaluation, la brique de catégorisation peut être instanciée avec tout algorithme de classification non-supervisée, qu'il utilise des données temporelles brutes, des données prétraitées, ou des transformations conservant l'aspect temporel. Dans nos applications, nous avons choisi d'utiliser des prétraitements simples ainsi que *DTW*.

2.2.5 Comparaison de classification

Dans le cadre de notre problème, nous cherchons à comparer le résultat de l'algorithme de classification sur les sous-échelles d'un questionnaire (*e.g.* de présence, d'immersion, ou encore comportementale) et le résultat du même algorithme appliqué aux traces de simulation (*i.e.* aux traces de conduite). D'autre part, nous souhaitons comparer les résultats obtenus sur les traces de conduite de différentes phases. Nous sommes donc confrontés à un problème de comparaison de classifications.

Il existe différentes approches du problème. Par exemple, en théorie de l'information, l'idée est de représenter une classification par une variable aléatoire [79]. Il est alors possible de calculer l'entropie d'une classification, l'entropie jointe et l'entropie conditionnelle. La similarité peut se calculer par rapport à l'information mutuelle I , c'est-à-dire la quantité d'information que la connaissance de la valeur d'une variable aléatoire nous apporte sur la valeur de l'autre :

$$I(\mathbb{C}_1, \mathbb{C}_2) = H(\mathbb{C}_1) - H(\mathbb{C}_1|\mathbb{C}_2) = H(\mathbb{C}_2) - H(\mathbb{C}_2, \mathbb{C}_1) \quad \text{Éq. 2.3}$$

Une autre approche est celle des treillis de classifications [9] : si une classification \mathbb{C}_1 peut être obtenue en divisant les classes d'une seconde classification \mathbb{C}_2 alors \mathbb{C}_1 est un raffinement de \mathbb{C}_2 . Enfin, l'approche de comptage de paires fournit un ensemble de métriques basées sur l'observation des associations entre paires d'instances. Ces métriques sont des mesures de similarité des classifications basées sur la répartition des paires d'instances selon les deux classifications. L'algorithme compte les paires d'instances pour lesquelles les deux classifications sont en accord (les paires se trouvant dans la même classe dans les deux classifications N_{00} ainsi que celles dissociées dans les deux classifications N_{11}) et celles pour lesquelles ils sont en désaccord (celles dissociées dans une classification mais pas dans l'autre N_{01} et vis-versa N_{10}). Il existe de nombreuses mesures à partir de ces 4 valeurs comme la *Rand Index RI* [117] et sa version ajustée *ARI* [71] très largement utilisé [94], le *Jaccard Index*, le *Wallace Index*, ...

Dans notre méthode, la brique de comparaison de classification peut être instanciée par une de ces techniques au choix. Dans notre expérimentation, nous comparerons des classifications par l'indice *ARI*. Cet indice est détaillé en *sous-section 5.2.1*.

Il est à noter qu'il existe des techniques de *co-clustering* permettant de regrouper des comportements similaires selon deux angles d'approche de manière simultanée [26]. Un angle est représenté par les colonnes d'une matrice, l'autre angle est quant à lui représenté par les lignes de cette matrice. Les comportements semblables sont alors des sous-ensembles de colonnes similaires à des sous-ensembles de lignes. Cette méthode pourrait sembler intéressante afin d'obtenir des regroupements prenant en compte d'une part l'approche *SHS* et d'autre part l'approche *IA*. Cependant, nous obtiendrions alors uniquement des sous-ensembles qui sont à la fois similaires par les réponses aux questionnaires et par les traces de simulation. Or, ces techniques ne tolèrent pas qu'un comportement - extrait par une approche - corresponde à un ensemble de comportements obtenus via une autre approche. Toutefois nous souhaitons conserver de manière inchangée les comportements obtenus avec chacune des deux approches.

2.2.6 Conclusion

Les classes de traces tout comme les classes d'annotations ont pour but d'évaluer les comportements adoptés. L'approche objective et subjective se complètent mutuellement au travers d'une analyse des données de simulation et d'une expertise humaine. La seconde permet de mettre des étiquettes sur les classes obtenues par la première, à partir des données brutes de la simulation. À notre connaissance, il n'existe pas de technique d'analyse de la crédibilité des comportements haut niveau en simulation permettant leur explicitation.

Dans notre méthode, nous proposons de les combiner : nous utilisons une analyse automatique de données et une méthode d'agrégation pour construire les classes de traces des comportements humains et agents, puis des observateurs humains remplissent un questionnaire sur les comportements adoptés afin de construire les classes d'annotations. Leur utilisation conjointe nous permet d'extraire puis d'explicitier les comportements haut niveau adoptés par les humains et les agents.

Deuxième partie

Méthode

Cette partie a pour but de présenter notre méthode d'évaluation des comportements. Nous commençons par décrire notre cas d'étude qu'est la simulation de trafic ainsi que l'approche globale de notre méthode dans le *chapitre 3*. Puis nous détaillons dans les deux chapitres suivants les différentes étapes de notre méthode : premièrement le recueil et le prétraitement des données dans le *chapitre 4*, et deuxièmement la classification des comportements des participants et l'agrégation des comportements des agents ainsi que leur comparaison et analyse dans le *chapitre 5*. Nous présentons ensuite au *chapitre 6* un cycle de conception exploitant cette méthode à des fins de calibration.

3 Méthode d'évaluation des comportements d'agents

Notre objectif est l'évaluation de la qualité des comportements des agents au niveau microscopique dans une simulation multi-agent au sein d'environnements de réalité virtuelle. Notre méthode est basée sur la combinaison de données issues des traces de simulation (approche objective) avec des données issues de questionnaires comportementaux (approche subjective). Les données sont agrégées sous la forme de classes de comportement et permettent d'évaluer sous deux angles les comportements adoptés. C'est par la comparaison de la typologie des sujets issue du questionnaire et de la classification des traces de comportement que nous évaluons la qualité des agents virtuels. Ce chapitre présente nos motivations et nos choix, puis l'approche générale décomposée en étapes, et enfin le cas d'étude.

3.1 Motivations

Notre but est de pouvoir évaluer les comportements des agents, d'être capable d'analyser leurs capacités à reproduire des comportements humains, d'identifier les comportements humains manquant à leur modèle, et de déterminer leurs comportements erronés car différents de ceux des humains. Nous souhaitons de plus être capable de retrouver quels sont les jeux de paramètres ayant aboutis à ces comportements. Enfin, nous souhaitons pouvoir expliciter ces comportements, c'est-à-dire être en capacité de les qualifier de manière compréhensible pour un humain et donc ne pas être restreint au résultat d'une analyse automatique.

Dans le cadre de la simulation, il existe deux types de données disponibles. Un de ces types de données correspond aux traces de simulation. Ces données sont brutes, de bas niveau (*i.e.* des actions simples, des opérations élémentaires, *etc.*), et pour la plupart temporelles. Leur analyse est donc difficile. Cependant, nous pensons qu'elles contiennent toutes les informations nécessaires à la caractérisation des comportements recherchés de haut niveau (tactique ou stratégique).

L'autre type de données est le recueil d'informations auprès d'une population d'humains ayant pour consigne d'évaluer les comportements. En effet, le jugement humain est adapté à l'analyse des comportements de ses semblables, et l'évaluation de la crédibilité peut donc passer par un jugement d'observateurs humains. De plus, un des buts en réalité virtuelle est d'avoir des comportements *crédibles* auprès de ces observateurs humains. L'avantage de ces informations est qu'elles portent directement sur les comportements de haut niveau, évitant ainsi le fossé sémantique existant entre les traces de simulation et les comportements recherchés. Cependant,

ces jugements sont subjectifs et leur justesse est difficilement vérifiable vis-à-vis des traces objectives. Ainsi, il est complexe de les lier directement aux variables du modèle d'agent.

Quel que soit le type de données que sont les traces de simulation et les évaluations faites par des humains, nous devons obtenir - afin d'analyser les comportements des agents - les mêmes informations sur une population d'humains. Les comportements des humains sont alors les points de référence, c'est-à-dire les comportements valides à reproduire. Pour cela, nous devons donc placer les participants dans les mêmes situations que celles soumises aux agents. Les traces de simulation doivent être recueillies à la fois pour les agents et les participants, et ce avec les mêmes variables. Le jugement humain ne peut donc pas être autoévalué, nous devons recourir à une annotation par une autre population observant à la fois les comportements adoptés par les agents et par les participants.

La difficulté dans chacun de ces deux types de données est qu'elles contiennent plusieurs instances des mêmes comportements haut niveau, c'est-à-dire qu'un même comportement peut avoir été adopté par plusieurs participants (ou agents). Cependant, ces instances ne sont pas exactement égales : elles peuvent avoir des différences minimales tout en correspondant néanmoins au même comportement. Afin d'exploiter ces données, nous devons donc regrouper ces instances en une unique abstraction de ce comportement haut niveau. Pour ce faire, nous avons choisi d'utiliser la classification non-supervisée.

Ces classes correspondant aux abstractions de comportement haut niveau des participants peuvent alors servir de point de référence pour la comparaison avec les agents artificiels. Cette comparaison est faite par une agrégation des agents aux classes de comportements de participants les plus proches.

Nous pouvons alors enfin analyser des classes de comportements haut niveau contenant à la fois les agents et les participants.

Afin d'exploiter les avantages des deux approches (traces de simulation et jugements humains), nous analysons conjointement les deux classifications obtenues permettant ainsi de caractériser les comportements des agents par rapport à ceux des participants du point de vue de leur traces de simulation tout en explicitant au sein de ces comportements le jugement humain perçu.

3.2 Approche globale

Notre méthode générale d'évaluation est décrite dans la *figure 3.1* (le traitement des données sera quant à lui détaillé dans les *chapitres 4 et 5* et la *figure 4.1*). Elle ne nécessite pas la connaissance préalable du modèle d'agent, ni de la plateforme de simulation, ni du simulateur utilisé. Nous considérerons donc le modèle d'agent comme une boîte noire.

La méthode est composée de 5 étapes (indiquées par les flèches en haut) : (1) le recueil des données de simulation, (2) l'annotation des rejeux vidéo, (3) la classification et l'agrégation automatique des données, (4) la comparaison des classifications, et (5) l'analyse et l'explicitation des classes par leur composition.

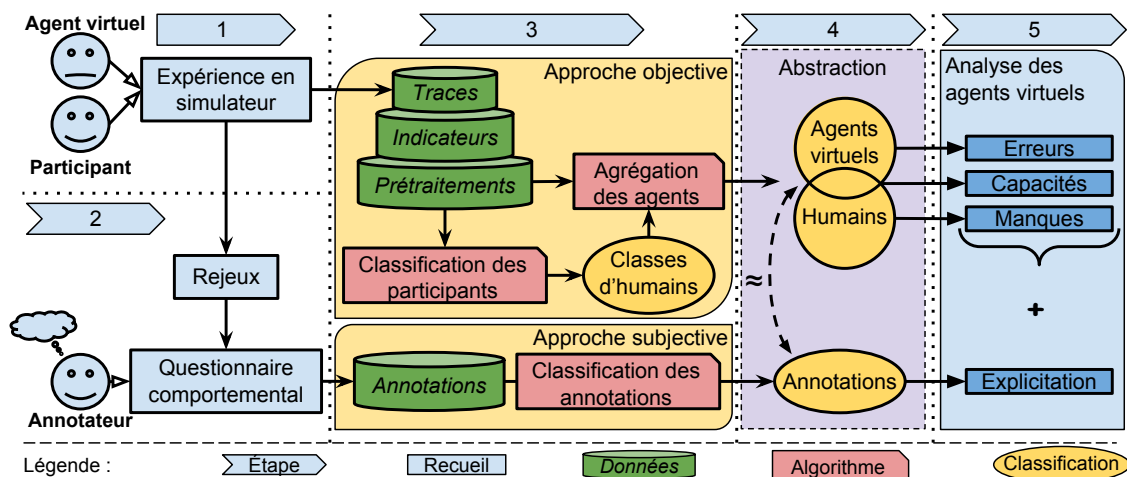


Figure 3.1 – Méthode d'analyse et d'évaluation des comportements d'agents.

3.2.1 Recueil des données de simulation

La première étape de notre méthode est le recueil de données quantitatives sur les comportements des participants. Nous recueillons des données brutes issues d'une simulation immersive en environnement virtuel. Nous générons aussi des données sur les agents. Pour cela, nous produisons de nouvelles simulations dans lesquelles le participant est remplacé par un agent. Nous appelons *acteurs principaux* l'ensemble de ces agents et des participants. Les données brutes des acteurs principaux sont appelées *traces* dans la figure 3.1.

3.2.2 Recueil des annotations des rejeux vidéo

La seconde étape est l'évaluation subjective des simulations effectuées sur l'ensemble des acteurs principaux. Pour cela, des vidéos - appelées *rejeux* - sont enregistrées sur l'ensemble des simulations des acteurs principaux. L'utilisation de l'approche subjective consiste à annoter ces rejeux par une population différente de participants (appelés *annotateurs*) via un questionnaire de comportement. Cette étape produit un ensemble d'annotations.

3.2.3 Classification et agrégation

L'objectif de la troisième étape est double : premièrement, analyser les comportements des agents en les comparant à ceux agents, et secondement expliciter les comportements des agents ainsi que les comportements manquants.

L'analyse des comportements des agents et de leur capacité à simuler des comportements humains est faite par comparaison des traces de comportements entre humains et agents. Cela ne peut pas se faire sur des données brutes. Ces traces, particulièrement dans le cas des participants, sont bruitées : deux traces différentes peuvent représenter le même type de comportement tactique ou stratégique. C'est pourquoi, dans le but de généraliser l'analyse des traces à un plus haut niveau de

comportement, nous proposons d'utiliser des catégories de comportement (appelées abstraction dans la *figure 3.1*). Ces catégories servent d'abstraction aux traces en rassemblant dans une même classe différentes traces représentatives d'un même comportement haut niveau. Cela est fait en utilisant une méthode de classification automatique.

Il serait possible d'utiliser un algorithme de classification automatique supervisée en faisant étiqueter par un expert un nombre important de traces. Cependant, cette solution est coûteuse en temps et en ressource. De plus, elle limite fortement le nombre d'agents analysable et contraint l'efficacité d'un cycle de conception itératif en requérant un nouvel étiquetage pour chaque ensemble de paramètres et à chaque modification du modèle. Pour ces raisons, ces algorithmes seront donc non-supervisés. Aussi, n'ayant pas d'information *a priori* sur le nombre de comportements, le nombre de classes doit être libre. C'est ce qui nous a conduits à utiliser les algorithmes de classification et notre méthode d'agrégation présentés dans la *section 5.1*.

L'explicitation des comportements des agents ainsi que des comportements manquants se fait via l'annotation. Ces annotations permettent une analyse de la typologie des acteurs principaux, et donc une étude des comportements adoptés selon un niveau différent via une approche permettant un autre angle. Dans le même but et de la même manière que pour les traces de simulation, nous utilisons une méthode de classification identique sur le questionnaire d'annotations afin de construire des abstractions, et de les comparer aux abstractions issues des traces.

3.2.4 Comparaison des classifications

Dans le cas où nous analysons pour les deux approches des comportements sous le même angle et de même niveau, nous voulons vérifier que la classification automatique des annotations de comportement observé est liée aux traces de comportements haut niveau dans la tâche donnée. Pour cela, nous évaluons la similarité entre la typologie des participants (par l'annotation) et les classes de comportements observés en simulation (par les traces). S'il y a une forte similarité entre la composition des classes de comportement et la composition des classes d'annotations, cela veut dire que les classes de comportement ont un sens en termes de typologie du participant.

3.2.5 Analyse et explicitation des classes

Dans cette étape, nous analysons la composition de ces classes en termes d'agent et de participant. Cette comparaison conjointe des deux classifications nous permet d'évaluer tout en explicitant les capacités, les manques et les erreurs des agents dans la simulation et d'en extraire des scores et des taux de confiance afin d'évaluer les comportements des agents dans la simulation multi-agents.

Avant de détailler dans le chapitre suivant les différentes étapes de notre méthode, la prochaine section présente brièvement le cas d'application qu'est la simulation de trafic routier.

3.3 Cas d'étude : la simulation de trafic routier

Afin de tester notre méthode, nous proposons d'évaluer le réalisme des comportements des agents du simulateur de trafic routier de l'*IFSTTAR* (dont le dispositif est visible en *figure 3.2* et sa description détaillée en *section 8.2*), au sein du simulateur de conduite *ARCHISIM* dont nous ne détaillerons pas le fonctionnement des agents (les détails du modèle peuvent être trouvés dans [24, 46]). Dans ce simulateur de conduite, le participant pilote un véhicule sur un parcours routier décrit en *annexe C.1*. Lors de la conduite sur ce circuit, le participant est en interaction avec des véhicules simulés générant le trafic. Nous sommes donc dans le cas de la simulation multi-agent immersive pour laquelle notre méthode s'applique.



Figure 3.2 – Dispositif du simulateur de conduite (avec 3 écrans, volant, boîte de vitesse et pédalier) utilisant le simulateur de trafic routier *ARCHISIM*.

Le circuit élaboré à l'occasion a été spécialement conçu afin de fournir de nombreuses situations de conduite indépendantes. Les situations de conduite proposées sont représentatives de la plupart des situations courantes dont certaines considérées comme encore problématiques pour les modèles d'agents actuels. Il est composé de 11 sections de route, appelées *phases* qui comprennent des situations de virages successifs, de carrefours (un prioritaire et un non prioritaire), d'insertion sur autoroute, *etc.* pour une durée totale d'environ 20 minutes de conduite. Le scénario de conduite ainsi que l'ensemble des phases sont détaillées dans les *sections 8.1 et 8.3*.

Dans cette application de notre méthode d'évaluation sur le simulateur de conduite *ARCHISIM*, les données de simulation correspondent aux traces de simulation. Ces traces prennent en compte les informations sur l'environnement de conduite, sur l'acteur principal (c'est-à-dire le conducteur humain ou le conducteur simulé par un agent), et sur les véhicules simulés en interaction. Le détail complet de ces variables est disponible en *sous-section 8.4.2*.

Dans le cadre de la conduite, les données issues de l'annotation correspondent aux réponses à un questionnaire de comportement de conduite. Nous verrons en détail ce dernier en *sous-section 4.2.1*.

Les deux chapitres suivants détailleront les différentes étapes de notre méthode d'évaluation. Puis le *chapitre 6* présentera notre cycle de conception.

4 Recueil et prétraitements des données

Pour la comparaison entre les classes de comportements des participants et celles des agents, nous devons travailler sur des traces recueillies dans des situations similaires. C'est pourquoi les agents sont placés dans une situation initiale identique à celle présentée aux participants.

Nous voulons analyser l'ensemble des capacités des agents, c'est-à-dire un ensemble exhaustif de leurs comportements, afin d'obtenir toute leur variabilité possible. Pour cela, différents types d'agents sont générés en explorant son espace de paramètres. Pour la simulation de conduite, il s'agit de la normativité, de l'expérience, des paramètres décisionnels, *etc.*

4.1 Recueil de traces de simulation

Les données de simulation sont obtenues en recueillant au cours de la simulation les traces de l'acteur principal (participant ou agent). Nous recueillons aussi les traces des autres agents en interaction, ainsi que des informations sur l'environnement telles que le temps de référence, le pas de simulation, le numéro de la phase de conduite, et le numéro d'instruction (déclenchant les éléments scriptés du scénario). Ces traces bas niveau seront combinées afin d'obtenir des abstractions de comportements de plus haut niveau. Il est à noter que ces variables sont pour la plupart temporelles (même si certaines peuvent être des scalaires représentant des agrégations de cet aspect temporel), et d'autres peuvent être des informations constantes.

Dans le cadre de la simulation de conduite, nous recueillons 8 variables pour le véhicule conduit et pour chaque véhicule simulé, ainsi que d'autres variables sur l'environnement. Ces enregistrements sont faits toutes les 300ms. Les variables communes aux véhicules principaux (simulés ou pilotés) sont : le temps, le point kilométrique, la route, le décalage à la voie, l'angle par rapport à la voie, la vitesse, l'accélération et la position du véhicule en trois dimensions. Des variables spécifiques aux véhicules pilotés s'ajoutent : l'angle du volant, la pression sur la pédale d'accélération, de frein et d'embrayage, ainsi que la position de la boîte de vitesse. Cependant, nous devons utiliser les mêmes variables pour les agents que pour les participants, nous n'utiliserons donc pas ces variables spécifiques.

Rappelons que les agents que nous évaluons sont placés dans la même situation, en lieu et place du véhicule piloté par les participants. Ainsi à chaque phase du scénario le véhicule simulé se place dans une situation initiale identique à celle présentée aux participants. Cependant, le comportement des autres agents de la

simulation n'étant pas scénarisé, l'évolution des situations peut plus ou moins diverger en fonction du comportement des acteurs principaux (participants humains ou agents artificiels).

4.2 Recueil d'annotations

Notre méthode d'évaluation utilise un questionnaire d'annotation pour la classification des catégories d'utilisateurs dans la partie "subjective". Notre méthode étant générale, elle ne peut être appliquée que si l'on fournit un questionnaire spécifique au domaine d'application. Ce questionnaire comportemental spécifique au domaine doit être validé pour la population étudiée. Il doit contenir des échelles et sous-échelles qui permettent l'obtention d'informations quantitatives. Cela peut se faire par des échelles de type *Likert* [87] qui sont utilisées dans les questionnaires psychométriques afin d'évaluer le degré d'accord du participant avec une assertion. Un nombre pair de réponses force une prise de décision, alors qu'un nombre impair permet à l'utilisateur de ne pas avoir à trancher en faveur d'une réponse positive ou négative en cas d'avis mitigé (*i.e.* avis neutre).

Dans notre approche, nous avons besoin de ces indicateurs sur l'ensemble des acteurs principaux (humains et agents). Le questionnaire ne peut donc être un questionnaire dont les réponses sont remplies par le participant lui-même sur son propre comportement. De plus, les questions du questionnaire ne doivent pas être relatives aux comportements d'un acteur en général mais spécifiques à une situation. En effet, la réalisation d'un comportement d'un acteur dans une situation précise peut ne pas correspondre à son comportement général habituel, et donc ne pas refléter les scores des échelles du questionnaire. Par exemple, une personne adoptant habituellement une conduite prudente peut exhiber un comportement dangereux lors d'une situation. Le questionnaire doit donc être soumis à une population de participants observant les comportements adoptés par les agents et par les participants étudiés. De plus, cela évite un possible biais dû à l'introspection. Ces observateurs "externes" annotent alors le comportement des participants humains ou agents sur des rejeux de la simulation à la première personne sans connaître la nature de l'acteur observé (agent ou humain). Ils observent chacun l'ensemble des acteurs principaux.

Dans les sous-sections suivantes nous décrivons un questionnaire utilisé dans le domaine de la simulation de conduite, puis notre questionnaire d'annotation basé sur ce dernier.

4.2.1 Questionnaire

Dans le cadre de notre application au comportement de conduite, nous avons choisi le *Driver Behavior Questionnaire (DBQ)* [119] dans sa version française [56]. Il fournit un score général sur les habitudes de comportement de conduite mais aussi des scores sur 5 sous-échelles :

- *contrôle et action* ;
- *mémoire et attention* ;

- *jugement et planification* ;
- *violation involontaire (accidentelle)* ; et
- *violation volontaire (délibérée)*.

De plus, il fournit 3 sous-échelles portant sur le risque d'accident : *aucun risque*, *risque possible*, et *risque certain*.

Cela nous permet de construire un ensemble de 6 indicateurs caractéristiques d'une catégorie d'utilisateur spécifiquement au domaine étudié, ici des catégories de conducteur. Les 5 premiers indicateurs correspondent aux habitudes de comportement de conduite et le dernier évalue le risque d'accident.

4.2.2 Annotations

Le questionnaire d'annotations est défini par l'ensemble des critères issus du questionnaire de comportement spécifique au domaine, avec des valeurs définies sur une échelle de type *Likert* avec neutralité.

Dans notre application aux simulateurs de conduite, le questionnaire contient donc pour chacune des 5 sous-échelles du *DBQ* une question notée sur une échelle à 7 points plus la valeur *sans opinion* (évitant la confusion entre la neutralité et l'absence de réponse), ainsi qu'une question sur *risque d'accident* notée sur une échelle à 3 points (plus la valeur *sans opinion*).

Nous avons aussi ajouté une question relative au contrôle perçu (dans quelle mesure le conducteur observé contrôle son véhicule) notée sur la même échelle à 7 points dans le but d'évaluer le contrôle général des acteurs principaux. Enfin, une dernière question demande si l'acteur principal est un humain ou un agent simulé, ce qui nous permettra de comparer les performances de la classification automatique avec celles d'observateurs humains.

En conclusion, ce questionnaire d'annotation basé sur le *DBQ* permet tout d'abord de définir des types de comportements pour des agents artificiels, contrairement à sa version originelle, et il permet d'obtenir des informations situationnelles (et non générales).

4.3 Prétraitements

Dans cette section, nous présentons - en vue de l'application d'un algorithme de classification non-supervisée - les prétraitements appliqués aux annotations, puis ceux appliqués aux traces de simulation.

4.3.1 Prétraitement des annotations

Le questionnaire d'annotation est conçu de façon à avoir des réponses sur des échelles quantifiables, et en l'occurrence des échelles de *Likert*. Nous obtenons donc les réponses de l'ensemble des annotateurs pour chaque acteur principal. Nous calculons alors le score moyen à chaque échelle du questionnaire. Ces moyennes sont ensuite normalisées puis ramenées entre 0 et 1. Ces seuls prétraitements suffisent à pouvoir appliquer un algorithme de classification pour obtenir une typologie des participants.

4.3.2 Prétraitement des traces

Dans cette section, nous expliquons les différents prétraitements appliqués aux traces de simulation (données brutes) afin d'obtenir des données exploitables par un algorithme de classification non-supervisée en vue d'obtenir les classes de comportement haut niveau que nous recherchons. La *figure 4.1* montre les prétraitements appliqués aux traces dans le but d'obtenir les classes de comportement.

Le recueil des traces des acteurs principaux (voir *section 4.1*) correspond à la partie haute de la figure ; les prétraitements nécessaires expliqués dans cette section sont schématisés au milieu de la figure ; enfin, nous verrons par la suite dans le *chapitre 5* l'application de l'algorithme de classification qui correspond à la partie droite de cette figure.

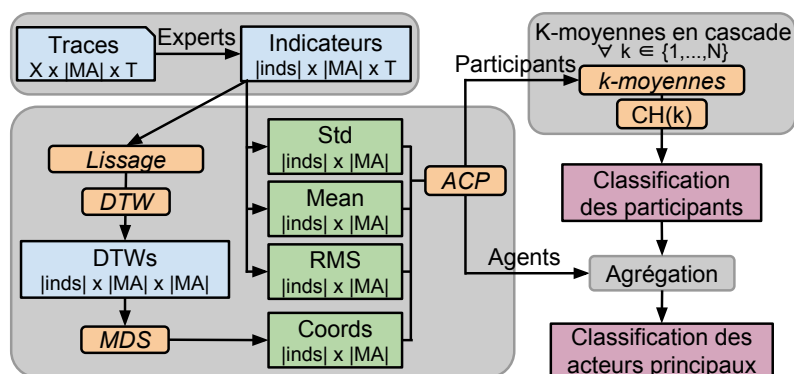


Figure 4.1 – Prétraitements des traces, agrégation, et classification. Les rectangles indiquent le nom des données et leur forme avec le nombre de variables (X), le nombre d'indicateurs ($|inds|$), le nombre d'acteurs principaux ($|MA|$) et le temps (T). Les algorithmes utilisés sont représentés dans des rectangles arrondis au-dessus de flèches. Les indicateurs des experts sont calculés à partir des traces de simulation (en haut à gauche de la figure). Des scalaires (y compris *DTW*) sont calculés à partir de ces données temporelles (à droite de la figure). L'algorithme de classification en cascade est alors appliqué sur les traces des participants (en haut à droite). Enfin, les agents sont agrégés à la classification de participants afin d'obtenir une classification pour l'ensemble des acteurs principaux.

4.3.2.a Indicateurs haut niveau

Nous voulons analyser les comportements haut niveau des acteurs principaux. Cela ne peut pas se faire directement sur les traces de simulations bas niveau à cause du fossé sémantique entre les données brutes et les comportements recherchés [19]. C'est pourquoi les données brutes ne peuvent être utilisées directement : nous utilisons des représentations de plus haut niveau basées sur la connaissance des experts du domaine (appelées *indicateurs*).

Certaines variables significatives, dépendantes du domaine d'application, ne peuvent pas être obtenues directement. Les experts du domaine peuvent identifier les indicateurs manquants importants. Ceux qui n'ont pu être recueillis sont alors calculés à partir des traces de simulation.

Dans notre application, les experts du trafic routier ont choisi les indicateurs suivants : des variables haut niveau telles que le temps avant collision (sous hypothèse que les vitesses des véhicules sont constantes), le temps intervéhiculaire, le *jerk* (dérivée de l'accélération par rapport au temps) et le nombre de changement de voie ; ainsi que des variables de bas niveau telles que le déplacement latéral et l'angle par rapport à l'axiale, la vitesse et l'accélération. Ces indicateurs ainsi que les variables haut niveau sont alors calculés pour tous les acteurs principaux.

4.3.2.b Données temporelles

Dans le cadre d'une simulation dynamique, la plupart des indicateurs sont temporels. Les données à classifier sont donc des suites ordonnées de valeurs pour chaque participant. Deux solutions existent : utiliser un algorithme prenant en entrée des données temporelles, ou étudier les données "à plat", en concaténant les variables temporelles relatives à un acteur principal sur une seule ligne. La première solution augmente fortement la complexité des algorithmes, car ceux-ci doivent prendre en compte les éventuels décalages temporels de comportements similaires. La seconde ignore les décalages temporels, mais permet d'appliquer les algorithmes classiques.

Nous avons choisi une solution hybride de prétraitement des données qui permet à la fois de n'avoir qu'un ensemble d'attributs par acteur principal et de prendre en compte les décalages temporels. Pour ce faire, nous générons autant de vecteurs que d'acteurs principaux. Chaque vecteur contient les informations suivantes extraites des indicateurs identifiés par les experts du domaine : *a*) les moyennes ; *b*) les écarts types ; *c*) les moyennes quadratiques ; et *d*) les agrégations temporelles (*i.e.* un scalaire prenant en compte les différences entre les courbes des indicateurs temporels). Sur ce dernier point, plusieurs acteurs principaux peuvent adopter un même comportement à différents instants, il y a alors un décalage temporel de l'occurrence de ce comportement entre ces acteurs principaux. Afin de prendre cela en compte lors de l'analyse des indicateurs, ils sont comparés avec un algorithme gérant les décalages temporels, c'est-à-dire étant capable de détecter la similitude entre indicateurs sur une fenêtre glissante.

Dynamic Time Warping

Notre brique de prétraitements doit transformer des données temporelles en données scalaires pour l'algorithme de classification non-supervisée tout en conservant au mieux cet aspect temporel. La solution adoptée pour le prétraitement des décalages temporels est d'utiliser un algorithme de recherche de motifs comme *Dynamic Time Warping (DTW)* [7] ou *Longest Common Subsequence (LCS)* [6]. Nous avons choisi d'instancier cette brique par l'algorithme *DTW* qui calcule le coût d'appariement entre deux séquences ordonnées (*e.g.* pour un indicateur *ind*, comparer les acteurs principaux *a* et *b*) dans le but de mesurer leur similarité [7].

La méthode fonctionne de la manière suivante. Soit T le nombre de pas temps de la simulation. L'algorithme calcule une matrice M de taille $T \times T$. Il initialise la première ligne et la première colonne à la valeur *infini* et le premier élément à 0. Il calcule ensuite chaque élément $M_{i,j} \forall (i, j) \in \{2, \dots, T+1\}^2$ de la matrice selon la distance entre les deux séquences au temps t et selon les éléments au voisinage (voir *équation 4.1*). Chaque élément $M_{i,j}$ correspond à la distance au temps t entre le point i de la courbe de l'indicateur sur a et le point j de la courbe de l'indicateur sur b . La dissimilitude (ou coût) entre a et b pour l'indicateur *ind* est $M_{T+1, T+1}$. *DTW* calculant les similarités entre deux instances d'une variable temporelle, moins les instances sont similaires, plus le coût augmente.

$$M_{i,j} \leftarrow distance(ind_i^a, ind_j^b) + \min(M_{i-1,j}, M_{i,j-1}, M_{i-1,j-1}) \quad \text{Éq. 4.1}$$

Étant donné le volume de nos données et que la complexité de *DTW* est en $O(N^2)$, nous utilisons une approximation de cet algorithme : l'algorithme *FastDTW* dont la complexité est de l'ordre de $O(N)$ en temps et en espace pour un taux d'erreur presque nul [124].

Distances mutuelles

Soit *inds* l'ensemble des indicateurs et $|inds|$ le nombre d'indicateurs. Pour chaque indicateur $ind \in inds$, nous calculons la matrice D_{DTW}^{ind} des distances mutuelles de taille $|MA| \times |MA|$, où $|MA|$ est le nombre d'acteurs principaux (agents et participants). Pour l'indicateur *ind*, chaque élément $D_{DTW}^{ind}(a, b)$ de D_{DTW}^{ind} est la dissimilitude entre les acteurs principaux a et b , c'est-à-dire $M_{T+1, T+1}^{ind}(a, b)$ calculé précédemment. Nous avons donc $|inds|$ matrices D_{DTW}^{ind} .

4.3.2.c Des distances mutuelles aux positions multidimensionnelles

Nous avons obtenu avec *DTW* les distances mutuelles entre tous les acteurs principaux pour tous les indicateurs. Nous souhaitons utiliser ces distances comme nouvelles variables décrivant les acteurs principaux au sein d'un algorithme de classification basé sur le partitionnement de l'espace. Nous devons donc transformer ces distances afin de placer chaque acteur principal dans un espace dimensionnel. Pour cela, nous utilisons un algorithme de positionnement multidimensionnel (*MDS*). L'algorithme assigne un point à chaque instance dans un espace multidimensionnel

et essaye de minimiser le nombre de dimensions. Pour chaque indicateur ind , le but est de trouver $|MA|$ coordonnées $(coord_1, \dots, coord_{|MA|}) \in \mathbb{R}$ telles que :

$$\forall (i, j) \in |MA|^2, \|coord_i - coord_j\| \approx D_{DTW}^{ind}(i, j) \quad \text{Éq. 4.2}$$

DTW étant une distance mathématique, l'algorithme MDS appliqué à chaque D_{DTW}^{ind} est capable de minimiser le nombre de dimensions à 1 (*i.e.* un vecteur de coordonnées). Nous avons alors pour chaque indicateur un vecteur de coordonnées de taille $|MA|$, et $|inds|$ vecteurs de coordonnées pour l'ensemble des indicateurs.

4.3.2.d Décorrélation

Les indicateurs peuvent être corrélés, cependant l'algorithme de classification décrit dans la section suivante est basé sur le k -moyennes et utilise donc un espace dimensionnel dont les axes sont orthogonaux entre eux. Afin d'appliquer cet algorithme, nous devons projeter les données de tous les acteurs principaux sur un même hyperplan orthogonal dont les axes sont deux à deux décorrélés.

L'Analyse en Composantes Principales (ACP) cherche des axes décorrélés qui donnent une dispersion maximale des données. Il est alors possible de réduire le nombre de dimensions et d'éviter les informations redondantes en les compressant.

Les données sont représentées par une matrice M de N variables aléatoires X_i contenant chacune K réalisations indépendantes. Cette matrice est centrée sur le centre de gravité $(\bar{X}_1, \dots, \bar{X}_N)$ et réduite par rapport à l'écart type σ des variables aléatoires. Il est alors possible de calculer la matrice de corrélation :

$$\frac{1}{K} \cdot \tilde{M}^T \cdot \tilde{M} \quad \text{Éq. 4.3}$$

$$\tilde{M} = \begin{bmatrix} \frac{X_{1,1} - \bar{X}_1}{\sigma(X_1)} & \dots & \frac{X_{1,N} - \bar{X}_N}{\sigma(X_N)} \\ \vdots & \ddots & \vdots \\ \frac{X_{K,1} - \bar{X}_1}{\sigma(X_1)} & \dots & \frac{X_{K,N} - \bar{X}_N}{\sigma(X_N)} \end{bmatrix} \quad \text{Éq. 4.4}$$

L' ACP cherche alors l'axe maximisant la variance des données par une combinaison linéaire des variables aléatoires pour pouvoir projeter les données sur cet axe : $\pi_u(M) = M \cdot u$. Dans nos expérimentations, nous utiliserons l' ACP lorsque des indicateurs sont corrélés.

4.4 Conclusion

Ce chapitre regroupe les parties 1 (recueil de données) et 2 (prétraitements) de notre méthode comme schématisé dans la *figure 3.1*. Nous avons donc présenté le type de données de simulation devant être recueillies ainsi que la forme du questionnaire d'annotation, puis exposé notre application de cette méthode dans le cadre de la simulation de conduite. Par la suite, nous avons énoncé les prétraitements à effectuer sur les traces de simulation et les annotations en vue de l'application d'un algorithme de classification. Nous exposerons dans le chapitre suivant les parties de

la méthode permettant la classification des participants, l'agrégation des agents, et la comparaison des traces avec les annotations.

5 Comparaison entre comportements d'agents et humains

Notre méthode d'évaluation est composée de 5 étapes : le recueil des données de simulation, l'annotation de ces données, la classification automatique de données, la comparaison des classifications, et l'explicitation des classes. Le chapitre précédent a présenté le recueil des données et leurs prétraitements à la fois pour les traces de simulation et les annotations. Nous allons maintenant nous intéresser aux troisième, quatrième, et cinquième étapes de notre méthode, c'est-à-dire le regroupement des acteurs principaux en classes de comportements, puis la comparaison des classifications obtenues, et enfin l'analyse de ces classes de traces et d'annotations.

5.1 Classification et agrégation

Dans cette section, nous détaillons l'étape de classification automatique des traces ainsi que celle des annotations qui sont au centre de notre modèle d'évaluation de la crédibilité. Nous présentons tout d'abord la partie objective dans les deux sous-sections suivantes, c'est-à-dire la classification des comportements humains à partir des traces de simulation puis l'agrégation des agents à ces classes de participants. Nous exposons ensuite en *sous-section 5.1.3* la classification des catégories de participant à partir du questionnaire d'annotation pour la partie subjective.

5.1.1 Classification des participants

À la fin de l'étape de prétraitement, nous avons obtenu l'ensemble des projections par l'ACP des indicateurs des acteurs principaux. Comme nous l'avons vu dans le *chapitre 3*, l'algorithme de classification appliqué à ces données doit être non-supervisé avec un nombre de classes libre, par exemple le *X-moyennes* basé sur le *Bayesian Information Criterion* [111] ou les *cartes auto-adaptatives* [78].

L'algorithme des *K-moyennes en cascade* qui applique de multiple *k-moyennes* en faisant varier le nombre de classes *k* de 1 à *K* (le nombre maximum de classes) remplit ces conditions nécessaires. Nous avons choisi de l'utiliser pour cette brique de la méthode. Nous décrivons maintenant l'algorithme et les variantes utilisées.

L'algorithme classique des *k-moyennes* utilise *k* centroïdes $\{m_1, \dots, m_k\}$ initialisés aléatoirement. Jusqu'à convergence, l'algorithme applique deux étapes alternativement :

1. l'étape d'assignation qui recherche pour chaque acteur principal ma le centroïde le plus proche. La classe C_i est alors l'ensemble des acteurs qui sont plus proches de m_i que des autres m_j :

$$C_i^{(t)} = \left\{ ma : \forall j \in \{1, \dots, k\}, \left\| ma - m_i^{(t)} \right\|^2 \leq \left\| ma - m_j^{(t)} \right\|^2 \right\} \quad \text{Éq. 5.1}$$

2. l'étape de mise à jour qui calcule les nouveaux centroïdes, en prenant la moyenne des valeurs des acteurs principaux ma de chaque classe :

$$m_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \cdot \left(\sum_{ma_j \in C_i^{(t)}} ma_j \right) \quad \text{Éq. 5.2}$$

La classification obtenue dépend fortement de l'initialisation des classes. L'initialisation des classes de l'algorithme des *k-moyennes* peut être améliorée en utilisant *kmeans++* [2] qui permet une meilleure répartition des centres des classes en fonction des données. Pour ce faire, le centroïde de la première classe est initialisé avec les valeurs des variables d'un participant sélectionné aléatoirement parmi tous les acteurs principaux MA . Jusqu'à obtenir k classes initiales, l'algorithme *kmeans++* calcule la distance de chaque acteur principal au dernier centroïde sélectionné. Ensuite, il sélectionne le centroïde d'une nouvelle classe C_i ayant les valeurs des variables d'un autre acteur principal ma . Cette sélection est faite aléatoirement selon une distribution de probabilité pondérée p proportionnelle au carré de leur distance au centroïde précédemment sélectionné C_{i-1} :

$$p(ma) = \frac{D(ma)^2}{\sum_{ma \in MA} D(ma)^2} \quad \text{Éq. 5.3}$$

$$\text{où } D(ma) = \text{distance}(ma, C_{i-1}) \quad \text{Éq. 5.4}$$

Enfin, pour *K-moyennes en cascade*, le nombre de classes doit être sélectionné judicieusement de façon automatique selon les données. Une "bonne" classification doit produire des classes dont les distances intraclasses sont faibles et les distances interclasses élevées. Nous utilisons pour cela le *Variance Ratio Criterion* qui prend en compte les interdistances et les intradistances des classes [20].

Interdistance : soit $|C_i|$ le nombre d'acteurs principaux dans la classe C_i , $\overline{C_i}$ le barycentre de cette classe, \overline{MA} le barycentre de tous les acteurs principaux (*i.e.* de la classification) et $|MA|$ le nombre d'acteurs principaux. L'interdistance B (voir *équation 5.5*) est la moyenne des carrés des distances entre le barycentre de chaque classe et le barycentre global \overline{MA} . C'est cette valeur que l'on cherche à maximiser en fonction du nombre k de classes.

Intradistance : soit $ma_{i,j}$ l'acteur principal numéro j de la classe i . L'intradistance W (voir *équation 5.6*) est la somme des distances au carré de chaque acteur principal $ma_{i,j}$ au barycentre de sa classe C_i . C'est cette valeur que l'on cherche à minimiser en fonction du nombre de classes.

Le *Variance Ratio Criterion* CH pour k classes est alors défini par l'équation 5.7 :

$$B = \sum_{i=1}^k |C_i| \cdot \|\overline{C_i} - \overline{MA}\|^2 \quad \text{Éq. 5.5}$$

$$W = \sum_{i=1}^k \sum_{j=1}^{|C_i|} \|ma_{i,j} - \overline{C_i}\|^2 \quad \text{Éq. 5.6}$$

$$CH(k) = \frac{B/(k-1)}{W/(|MA| - k)} \quad \text{Éq. 5.7}$$

À l'issue de ce traitement, nous obtenons donc entre 1 et K classes de comportement des acteurs principaux comme dans la *figure 4.1*. Le nombre de classes est déjà un indicateur de la qualité de la classification : dans le cas d'une situation pouvant mener l'acteur principal à opter pour différents comportements, une unique classe ou un nombre de classes proche du nombre d'acteurs principaux signifie que les indicateurs ne permettent pas la différenciation en termes de comportement haut niveau.

5.1.2 Agrégation des agents

Pendant le processus de classification, l'ajout d'acteurs principaux modifie les centroïdes des classes et peut par conséquent changer d'autres affectations. Cependant, les agents et les participants ne doivent pas être considérés de la même manière dans ce processus car les comportements humains représentent la référence à laquelle nous voulons comparer les comportements des agents.

Pour cette raison, nous ne souhaitons pas que les indicateurs des agents modifient la classification des comportements humains. Dans le but de garder intactes les classes de participants, l'algorithme des k -moyennes est uniquement appliqué aux participants. Si les agents sont assez proches, ils sont par la suite agrégés à ces classes d'humains fixées. Sinon, ils sont classifiés dans de nouvelles classes d'agents.

Notre méthode fonctionne comme suit (voir l'*algorithme 1*). Nous définissons, à partir de la classification \mathbb{C} , pour chaque classe de participants C_H – composée de participants p – un seuil $s(C_H)$ au-delà duquel l'agent a est considéré comme étant trop éloigné du centroïde m^{C_H} de la classe pour être agrégé. Ce seuil $s(C_H)$ est défini sur chaque dimension (*i.e.* pour chaque indicateur ind) comme la distance entre le centroïde m^{C_H} et le participant de cette classe le plus éloigné :

$$s_{ind}(C_H) = \forall p \in C_H, \max(|m_{ind}^{C_H} - p_{ind}|) \quad \text{Éq. 5.8}$$

Dans le but de permettre l'agrégation des voisins proches, nous élargissons les seuils s par un pourcentage de la moyenne des seuils $\overline{\delta}$, selon un taux de tolérance ϵ : cela permet aux classes singletons (pour lesquelles $\forall ind, s_{ind} = 0$) de pouvoir rassembler des agents.

Chaque agent $a \in \mathcal{A}$ est agrégé à la classe de participants C_H dont le centroïde m^{C_H} est le plus proche parmi celles sous le seuil s pour chaque dimension ind . Les

Algorithme 1 : Algorithme d'agrégation des agents sur les classes de participants

```

pour chaque  $C_H \subset \mathbb{C}$ ,  $ind \in \text{indicateurs}$  faire
    |  $s_{ind}(C_H) \leftarrow \max_{p \in C_H} (|m_{ind}^{C_H} - p_{ind}|)$ 
fin
pour chaque  $ind \in \text{indicateurs}$  faire
    |  $\delta_{ind} \leftarrow \epsilon \cdot \overline{S}_{ind}$ 
fin
pour chaque  $C_H \subset \mathbb{C}$ ,  $ind \in \text{indicateurs}$  faire
    |  $s_{ind}(C_H) \leftarrow s_{ind}(C_H) + \delta_{ind}$ 
fin
pour chaque  $a \in \mathcal{A}$  faire
    |  $\forall C_H \subset \mathbb{C}, dists \leftarrow |a - m^{C_H}|$ 
    | si  $\forall ind \in \text{indicateurs}, dists_{ind} \leq s_{ind}$  alors
    | |  $a_{C_H} \leftarrow \text{argmin}(dists_{\leq s})$  ;  $C_H \leftarrow C_H \cup \{a\}$ 
    | sinon
    | |  $\mathcal{A}_{restant} \leftarrow \mathcal{A}_{restant} \cup \{a\}$ 
    | fin
fin
miseAJourDesSeuils() ;  $\mathbb{C}_A \leftarrow \emptyset$ 
répéter
    | pour chaque  $a \in \mathcal{A}_{restants}$  faire
    | |  $\forall C_A \subset \mathbb{C}_A, dists \leftarrow |a - m^{C_A}|$ 
    | | si  $\forall ind \in \text{indicateurs}, dists_{ind} \leq s_{ind}$  alors
    | | |  $a_{C_A} \leftarrow \text{argmin}(dists_{\leq s})$ 
    | | | sinon
    | | | |  $C_A \leftarrow \{a\}$  ;  $s(C_A) \leftarrow \delta$  ;  $\mathbb{C}_A \leftarrow \mathbb{C}_A \cup \{C_A\}$ 
    | | | fin
    | | | miseAJourDesSeuils()
    | | fin
fin
jusqu'à convergence

```

agents trop éloignés des classes de participants forment l'ensemble d'agents restants $\mathcal{A}_{restant}$.

Si un ou des agents ne se sont pas agrégés aux classes de participants, le premier agent "restant" crée sa propre classe C_A qui est ajoutée à l'ensemble des classes d'agents \mathbb{C}_A afin que d'autres agents restants puissent s'y agréger (voir l'équation 5.9).

Comme pour l'agrégation aux classes de participants, chaque agent restant essaye de s'agréger à une classe d'agents restants selon la même règle de seuil que celle utilisée pour les classes de participants, ou crée sa propre classe sinon (voir l'équation 5.9).

$$a_C = \begin{cases} C_A^{k+1} \leftarrow \{a\}, & \text{si } \forall i \in \{1, \dots, k\}, |a - m^i| < s^i \\ C_A^i, & \forall j \in \{1, \dots, k\}, |a - m^i| \leq |a - m^j|, & \text{sinon} \end{cases} \quad \text{Éq. 5.9}$$

Ainsi, comme schématisé dans la figure 4.1, nous obtenons une classification basée sur les traces de simulation composée de tous les acteurs principaux : participants

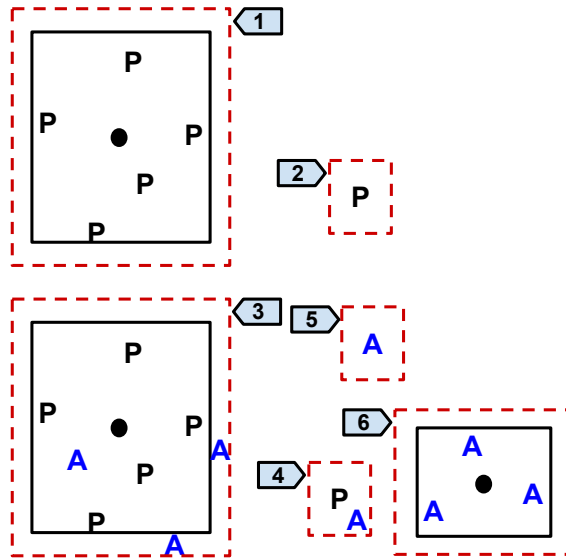


Figure 5.1 – Exemples d’agrégation sur un espace à 2 dimensions. En haut, les participants (P) les plus proches ont formé une classe via l’algorithme des K -moyennes (étiquette 1) et certains participants peuvent avoir formé des classes singletons ne contenant qu’un unique participant (étiquette 2). En bas, de gauche à droite : les agents (A) à l’intérieur de l’espace englobant (rectangle noir) de cette classe (étiquette 3) sont agrégés et ceux suffisamment proches le sont aussi (*i.e.* à l’intérieur de l’espace englobant augmenté du taux de tolérance, représenté par un rectangle rouge pointillé). Ce taux de tolérance permet à un agent de pouvoir s’agréger aux classes singletons comme dans le cas de la classe étiquetée 4. Un agent à l’extérieur de l’espace englobant de toutes les classes forme une classe agent singleton (cas de la classe avec l’étiquette 5). Il est alors possible pour d’autres agents de s’y agréger comme pour la dernière classe (étiquette 6).

humains et agents virtuels. Il est déjà possible de faire une première analyse des comportements d’agents basée sur la composition des classes en termes d’agents et d’humain. Nous détaillerons cette première analyse dans la *section 5.2*.

5.1.3 Traitement des annotations

La méthode de classification que nous utilisons pour la classification des catégories dans la partie “subjective” est indépendante du domaine et du questionnaire utilisé. Cependant, elle ne peut fonctionner que s’il l’on fournit un questionnaire spécifique au domaine d’application. C’est pourquoi nous décrirons par la suite le traitement des annotations au travers de l’application au domaine de la simulation de conduite, et donc via le questionnaire d’annotation décrit en *section 4.2*.

Nous analysons les scores des participants à l’aide d’un algorithme de classification pour obtenir une typologie des participants. Cet algorithme de classification non-supervisée est le même que celui utilisé sur les indicateurs issus des traces de simulation des participants qui a été détaillé en *sous-section 5.1.1*. Nous appliquons

ensuite notre algorithme d'agrégation des agents sur les classes d'annotations des participants comme vue en *sous-section 5.1.2*.

À l'issue de cette étape, nous obtenons donc les typologies des acteurs principaux comme dans la *figure 3.1*.

5.1.4 Conclusion

Dans cette section, nous avons présenté comment passer des données prétraitées aux classes de comportements. Nous avons introduit les caractéristiques nécessaires à l'algorithme de classification des participants ainsi que notre choix de classifieur. Ensuite, nous avons exposé notre méthode d'agrégation des agents sur les classes de participants (*i.e.* sur les abstractions des comportements recherchés). Cet algorithme a été appliqué d'une part sur les indicateurs issus des traces afin d'obtenir une classification des traces sur l'ensemble des acteurs principaux, et d'autre part sur les annotations afin d'obtenir une typologie des acteurs principaux.

Nous avons donc deux classifications des comportements de l'ensemble des acteurs principaux. Il est alors possible d'analyser les comportements des agents par rapport à ceux des participants selon les deux approches (objective et subjective), mais aussi de combiner ces deux analyses afin d'obtenir des informations complémentaires.

5.2 Comparaison et analyse

Dans cette section, nous présentons en premier lieu les méthodes de comparaison existantes et celle retenue, puis nous détaillons la comparaison de la classification de trace avec celle des annotations. Enfin nous présentons notre méthode d'analyse des classes basée sur la typologie des classes en termes de composition, pondérée par des taux de confiance en ces classes, et leur évaluation quantitative via des scores.

5.2.1 Méthode de comparaison

Nous voulons analyser les comportements adoptés par les agents par rapport à ceux adoptés par les participants puis les expliciter. Dans le cas où les indicateurs issus des traces de simulation et les annotations issues du questionnaire analysent le comportement sous le même angle et à un même niveau, les classifications doivent alors être similaires et peuvent donc être comparées avec des méthodes de comptage par paires.

Si nous nous plaçons dans ce cas, nous pouvons alors comparer la classification des traces de comportements et la classification des annotations selon la composition de leurs classes.

La comparaison de classes, dans le cadre de la classification non-supervisée à classe libre, est un problème non trivial car elle dépend à la fois du nombre de classes

et de leur composition. Il est en effet difficile d'associer une classe non étiquetée à une autre quand leurs compositions ne sont pas identiques.

Pour la comparaison de classifications, nous avons besoin d'une mesure de similarité entre deux classifications \mathbb{C}_1 et \mathbb{C}_2 . Les méthodes de comptages par paires permettent d'obtenir des informations sur la similarité entre deux classifications par des métriques basées sur les 4 informations suivantes :

- N_{00} : le nombre de paires qui sont dans la même classe et ce dans les deux classifications (*i.e.* en accord) ;
- N_{11} : le nombre de paires qui sont dans différentes classes pour les deux classifications (*i.e.* en accord) ; et
- N_{01} et N_{10} : le nombre de paires qui sont dans la même classe dans une classification mais dans différentes classes dans l'autre classification (*i.e.* en désaccord).

Parmi les scores calculables, il existe l'indice de Rand [117], ou encore l'indice de Rand ajusté (*ARI*) [71] classiquement utilisé [94] [94].

L'indice de Rand [117] $RI \in [0, 1]$ est décrit dans l'équation 5.10 pour deux classifications \mathbb{C}_1 et \mathbb{C}_2 . L'indice de Rand ajusté $ARI \in [-1, 1]$ (voir équation 5.11) est une version corrigée de RI prenant en compte les accords dus au hasard : l'espérance de la valeur de RI pour deux classifications aléatoires n'est pas constante, alors que l'espérance de la valeur d' ARI est de 0. Une valeur d' ARI proche de 1 signifie donc que les classes obtenues par les deux processus de classification sont similaires en termes de composition de classe, tandis qu'une valeur d' ARI proche de 0 signifie une grande dissimilarité. Il est calculé en utilisant un tableau de contingence $[n_{ij}]$ où n_{ij} est le nombre d'accords (*i.e.* N_{00} et N_{11}) entre les classes i et j (*i.e.*) : $n_{ij} = |\mathbb{C}_1^i \cap \mathbb{C}_2^j|$ (voir équation 5.11).

$$RI(\mathbb{C}_1, \mathbb{C}_2) = \frac{N_{00} + N_{11}}{N_{00} + N_{11} + N_{01} + N_{10}} \quad \text{Éq. 5.10}$$

$$ARI(\mathbb{C}_1, \mathbb{C}_2) = \frac{RI(\mathbb{C}_1, \mathbb{C}_2) - E[RI(\mathbb{C}_1, \mathbb{C}_2)]}{1 - E[RI(\mathbb{C}_1, \mathbb{C}_2)]} \quad \text{Éq. 5.11}$$

$$\text{Où } E[RI(\mathbb{C}_1, \mathbb{C}_2)] = \left[\sum_i \binom{\sum_k n_{ik}}{2} \sum_j \binom{\sum_l n_{lj}}{2} \right] / \binom{n}{2} \quad \text{Éq. 5.12}$$

$$\text{Soit } ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{\sum_k n_{ik}}{2} \sum_j \binom{\sum_l n_{lj}}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{\sum_k n_{ik}}{2} + \sum_j \binom{\sum_l n_{lj}}{2} \right] - \left[\sum_i \binom{\sum_k n_{ik}}{2} \sum_j \binom{\sum_l n_{lj}}{2} \right] / \binom{n}{2}} \quad \text{Éq. 5.13}$$

Dans le cas où le questionnaire d'annotations s'intéresse à des comportements sous un angle différent que celui adopté par les indicateurs issus des traces, les deux classifications apportent des informations complémentaires mais elles ne sont pas nécessairement similaires, leur comparaison n'a donc pas lieu d'être. Plus spécifiquement à notre étude du comportement de conduite, un même type de conducteur défini par les annotations peut être amené à adopter des comportements différents pour une même situation conduisant à deux traces dissimilaires. Pareillement, un même comportement adopté (issu des traces de simulation) peut être effectué par différents types de conducteur en termes d'annotations. Ainsi, il est possible dans

notre problème qu'une classe d'annotations corresponde à plusieurs classes de traces et inversement.

5.2.2 Analyse des classes

Dans cette sous-section, nous présentons l'évaluation - au travers d'une typologie des classes - des capacités des agents à reproduire des comportements humains, des manques dans le modèle d'agent, et des erreurs commises par les agents. Dans les sous-sections suivantes sont présentés les types de classes étudiés, puis les métriques évaluant la composition de ces classes ainsi que la confiance en cette typologie.

5.2.2.a Type de classes

Nous avons vu comment combiner la comparaison des traces de comportements entre participants et agents avec l'analyse de l'annotation des acteurs principaux : les classes de comportements des participants sont corrélées à leurs catégories d'annotations. La composition des classes de comportements - en termes d'agents simulés et de participants - nous permet également d'explicitier des informations sur les comportements des agents. Pour l'étape d'évaluation, il est possible de distinguer trois types de classe :

1. Les classes *mixtes* (une classe *mixte* sera représentée par C_M) contenant à la fois acteurs principaux humains et agents ; elles correspondent à des comportements haut niveau correctement reproduits par les agents et donc à des capacités du modèle d'agent. L'ensemble de ces classes mixtes est noté \mathbb{C}_M .
2. Les classes *agents* (une classe *agent* sera représentée par C_A) constituées uniquement d'agents ; elles correspondent à des comportements produits uniquement par des agents. Dans la plupart des cas cela reflète des erreurs de simulation, au sens où aucun humain n'a adopté ces comportements dans cette situation. Cela peut aussi être dû à un échantillon trop faible de participants si l'hypothèse de représentativité de la population n'est pas respectée. L'ensemble de ces classes agents est noté \mathbb{C}_A .
3. Les classes *humains* (une classe *humain* sera représentée par C_H) constituées uniquement de participants ; elles correspondent à des comportements humains n'ayant pas été reproduits par les agents, et sont donc dus soit à des manques dans le modèle d'agent, soit à un échantillon trop faible d'agents dans l'espace des paramètres. L'ensemble de ces classes humains est noté \mathbb{C}_H .

5.2.2.b Taux de confiance

Les algorithmes de classifications peuvent produire des regroupements erronés parmi les classes de traces de participants, c'est-à-dire des erreurs d'affectation. En conséquence, une classe singleton peut représenter soit un comportement singulier, soit une erreur de classification. De même, l'algorithme d'agglomération des agents peut - par effet de seuil - agréger un agent à une classe d'humain dissimilaire (*i.e.* qui ne représente pas le même comportement). À l'inverse, une instance peut être exclue

d'une classe qui correspond à son comportement. Il est donc nécessaire de fournir des métriques sur la confiance qui peut être attribuée au type d'une classe (erreur, manque, capacité).

Les taux de confiance dépendent du nombre d'acteurs principaux par classe et dans l'ensemble de la classification. Nous définissons donc pour une classe le nombre d'agents, $H(C)$ le nombre d'humains, et $|C| = A(C) + H(C)$ le nombre total d'acteurs principaux. De la même manière, nous définissons $A(\mathbb{C})$ et $H(\mathbb{C})$ pour une classification \mathbb{C} . Aussi, soient $|\mathbb{C}_H|$ le nombre de classes d'humains, $|\mathbb{C}_A|$ le nombre de classes d'agents, $|\mathbb{C}_M|$ le nombre de classes mixtes, et $|\mathbb{C}| = |\mathbb{C}_H| + |\mathbb{C}_A| + |\mathbb{C}_M|$ le nombre total de classes.

Le taux de confiance en une classe dépend de son type. Pour une classe mixte C_M (*i.e.* une capacité), le taux de confiance dépend du ratio d'humains $t_A(C)$ et du ratio d'agents $t_H(C)$ entre la classe étudiée et l'ensemble de la classification :

$$t_A(C_M) = \frac{A(C_M)}{A(\mathbb{C})} \in [0, 1] \quad \text{Éq. 5.14}$$

$$t_H(C_M) = \frac{H(C_M)}{H(\mathbb{C})} \in [0, 1] \quad \text{Éq. 5.15}$$

Pour ces classes mixtes, la confiance est alors faible quand le taux d'humain est "éloigné" du taux d'agents. À l'inverse, nous considérons que la confiance est suffisamment forte quand le taux est supérieur à $\frac{1}{2}$:

$$t_{conf}(C_M) = 1 - |t_A(C) - t_H(C)| \quad \text{Éq. 5.16}$$

$$t_{conf}(C_M) \approx 0 \Rightarrow \text{Confiance faible}$$

$$t_{conf}(C_M) > \frac{1}{2} \Rightarrow \text{Confiance forte}$$

Par exemple, une classe contenant 9 participants sur 12 et seulement 2 agents sur 20 a un taux de confiance de $1 - \left| \frac{9}{12} - \frac{2}{20} \right| = 0,35$. Conséquemment pour cette classe mixte, la confiance en la capacité de l'agent à reproduire ce comportement humain est faible et cette classe mixte représente potentiellement des erreurs de classification. À l'inverse, une classe contenant 2 participants sur 12 et 4 agents sur 20 a un taux de confiance de $1 - \left| \frac{2}{12} - \frac{4}{20} \right| = 0,97$. Il y a donc une forte confiance que cette classe mixte soit réellement une capacité du modèle de l'agent à reproduire le comportement humain adopté, la possibilité d'erreurs de classification est alors écartée.

Pour les classes contenant uniquement des agents, la confiance dépend uniquement du nombre d'agents par rapport au nombre moyen d'agents par classe dans l'ensemble de la classification E_A :

$$E_A(\mathbb{C}) = \frac{A(\mathbb{C})}{|\mathbb{C}_M| + |\mathbb{C}_A|} \quad \text{Éq. 5.17}$$

$$t_{conf}(C_A) = \frac{A(C)}{E_A(\mathbb{C})} \quad \text{Éq. 5.18}$$

De même, la confiance en les classes contenant uniquement des humains dépend du nombre moyen d'humains par classe E_H :

$$E_H(\mathbb{C}) = \frac{H(\mathbb{C})}{|\mathbb{C}_M| + |\mathbb{C}_H|} \quad \text{Éq. 5.19}$$

$$t_{conf}(C_H) = \frac{H(C)}{E_H(\mathbb{C})} \quad \text{Éq. 5.20}$$

Un taux de confiance élevé en une *classe d'humains* (respectivement une *classe d'agents*) signifie que cette classe peut être considérée comme un manque (respectivement une erreur) dans le modèle d'agent avec confiance.

5.2.2.c Scores de type de classe

L'analyse de la composition des différentes classes nous permet de différencier erreurs, capacités, et manques. Il est alors possible de les quantifier. Nous proposons de calculer un score pour chacun de ces trois types :

1. Le score de capacités S_c :

$$S_c = \frac{|\mathbb{C}_M|}{|\mathbb{C}_M| + |\mathbb{C}_H|} \quad \text{Éq. 5.21}$$

2. Le score de manques S_m :

$$S_m = \frac{|\mathbb{C}_H|}{|\mathbb{C}_M| + |\mathbb{C}_H|} \quad \text{Éq. 5.22}$$

$$S_m = 1 - S_c \quad \text{Éq. 5.23}$$

3. Le score d'erreurs S_e :

$$S_e = \frac{|\mathbb{C}_A|}{|\mathbb{C}_M| + |\mathbb{C}_A|} \quad \text{Éq. 5.24}$$

Le score de capacités permet d'analyser quantitativement la capacité du modèle d'agent dans l'ensemble du *SMA* à reproduire des comportements humains en prenant en compte les manques (*i.e.* les classes d'humains). Par exemple pour une classification composée de 2 classes contenant uniquement des humains et 3 classes contenant à la fois des humains et des agents, le score de capacités sera de $\frac{3}{2+3}$. À l'inverse, le score de manques informe sur les manques du modèle en prenant en compte ces capacités. Le score d'erreurs quant à lui quantifie les comportements erronés (*i.e.* le nombre de classes d'agents) relativement aux classes contenant des agents (c'est-à-dire les classes mixtes et les classes d'agents). Plus les comportements adoptés par les agents sont dissimilaires des comportements humains, plus le score d'erreurs est élevé.

5.3 Conclusion

Ce chapitre a présenté comment obtenir une classification de comportement pour l'ensemble des acteurs principaux à partir des données prétraitées via un algorithme de classification et notre méthode d'agrégation. Nous avons montré que ces algorithmes peuvent s'appliquer à la fois aux traces de simulation et aux annotations, et donc qu'il est possible d'obtenir deux classifications de comportement complémentaires. Ce chapitre a ensuite exposé notre méthode d'évaluation basée à la fois sur une analyse de la composition des classes et sur une explicitation des classes par une utilisation conjointe des deux classifications. Comme décrit ci-dessus, cette méthode d'évaluation permet d'analyser les capacités, les manques et les erreurs du modèle d'agent, de jauger de leurs confiances relatives, et d'obtenir des scores sur ces composantes.

Dans le chapitre suivant, nous voyons comment utiliser cette évaluation pour calibrer la population d'agents au sein d'un cycle de conception.

6 Cycle de conception

Dans cette section, nous présentons - au sein d'un cycle de conception - l'utilisation de cette analyse des compositions de classes afin d'évaluer la sur-représentation et la sous-représentation de comportements en se basant sur les scores exposés dans le chapitre précédent. Ce cycle de conception est schématisé dans la *figure 6.1*. Un schéma montrant les parties spécifiques au cycle de conception au sein de la méthode d'évaluation est disponible en *annexe B.1*.

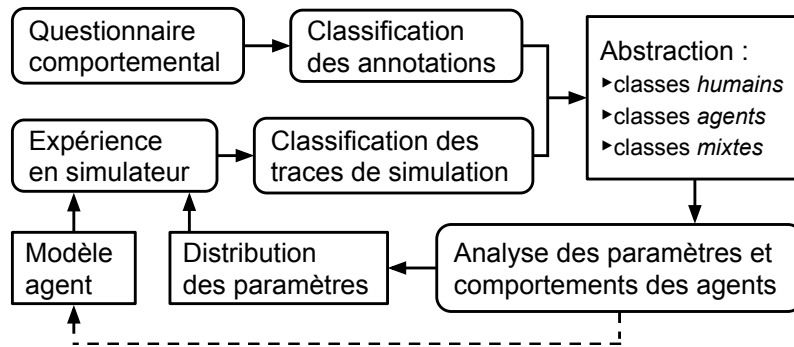


Figure 6.1 – Cycle de conception et détail de la calibration.

6.1 Calibration

Dans la Simulation Multi-Agents, la calibration d'un modèle signifie l'ajustement des paramètres de façon à ce que certains buts globaux ou comportements désirés soient atteints [48, 47], voir *e.g.* [135]. La calibration des paramètres d'un modèle pour des modèles détaillés à base d'agents est un problème pour les techniques standards de calibration dû au large espace des paramètres, au long temps d'exécution de la simulation, et des incertitudes dans la conception du modèle. En simulations participatives, où les agents interagissent avec des participants, il faut de plus que les comportements individuels de ces agents soient "crédibles".

Dans notre cas, nous nous appuyons sur les traces des comportements des participants collectées pendant l'expérimentation. Elles définissent alors un ensemble de comportements valides. En d'autres termes, notre objectif est de réduire la valeur du score d'erreurs S_e et celle du score de manques S_m , et donc d'augmenter la valeur du score de capacité S_c . Considérant le modèle d'agent comme une boîte noire pouvant produire différents comportements en fonction de ses paramètres, le processus de calibration doit s'assurer que :

- le comportement résultant de chaque agent a est crédible, c'est-à-dire que leur ensemble de paramètres $P_a = \{p_1, \dots, p_i\}$ (avec i le nombre de paramètres du modèle) est individuellement valide ; et que
- la population d'agents reproduit bien les comportements humains d'un point de vue global et dans des proportions équivalentes, c'est-à-dire que la distribution des ensembles de paramètres $\mathcal{P} = \{P_1, \dots, P_n\}$ avec n le nombre d'agents est valide globalement.

Concrètement, les agents appartenant aux classes d'erreurs ont montré des comportements non exhibés par les participants humains, et sont donc considérés comme non crédibles. En conséquence, leurs ensembles de paramètres doivent être retirés du groupe d'ensembles de paramètres valides. À l'inverse, les manques (*i.e.* les comportements exhibés par des humains mais par aucun agent) peuvent provenir de paramètres mal choisis, sous l'hypothèse que les agents peuvent produire les comportements identifiés.

Dans le but de proposer des proportions correctes de comportements valides, nous prenons en compte la composition des classes à la fois pour les données des agents et pour celles des humains. Nous pouvons alors raffiner les catégories précédemment présentées dans la *sous-section 5.2.2.a* :

- les comportements erronés lorsque aucun humain n'appartient à la classe, correspondant aux classes d'agents ($H(C) = 0$),
- les capacités humaines basées sur les classes mixtes \mathbb{C}_M , détaillées comme suit :
 - l'ensemble des comportements sur-représentés (\mathbb{C}_{sur}) quand la proportion d'agents adoptant ce comportement est supérieure à la proportion d'humains ($A(C) \gg H(C)$),
 - l'ensemble des comportements représentés de manière correctes (\mathbb{C}_{valide}) quand la proportion d'humains est approximativement égale à celle des agents ($A(C) \approx H(C)$), et
 - l'ensemble des comportements sous-représentés (\mathbb{C}_{sous}) quand la proportion d'agents adoptant ce comportement est inférieure à la proportion d'humains ($A(C) \ll H(C)$).
- les comportements manquants quand aucun agent n'appartient à la classe, correspondant aux classes d'humains ($A(C) = 0$).

Après une première calibration de l'ensemble des paramètres \mathcal{P} , les scores ne dépendent plus du nombre d'agents mais de la proportion d'agent dans la population totale. Généralement, la population d'agents est plus grande que la population humaine pour des raisons expérimentales (principalement le temps de recueil des données). De plus, l'opérateur \approx séparant la sous-représentation et la sur-représentation de la représentation valide d'un comportement conduit à définir un seuil. Nous proposons de prendre en compte la taille de chaque classe (*e.g.* $\delta(C) = \frac{5}{100}|C|$).

Afin d'améliorer la calibration, nous établissons - en plus des deux scores concernant les manques et les erreurs - trois scores pour les classes de comportement mixtes appelés scores de représentativité :

1. Sur-représentation :

$$\mathbb{C}_{sur} = \{C_M \in \mathbb{C}_M, A(C_M) > H(C_M) + \delta|C_M|\} \quad \acute{E}q. 6.1$$

$$S_{sur} = \frac{|\mathbb{C}_{sur}|}{|\mathbb{C}_M|} \quad \acute{E}q. 6.2$$

3. Sous-représentation :

$$\mathbb{C}_{sous} = \{C_M \in \mathbb{C}_M, A(C_M) < H(C_M) - \delta|C_M|\} \quad \acute{E}q. 6.3$$

$$S_{sous} = \frac{|\mathbb{C}_{sous}|}{|\mathbb{C}_M|} \quad \acute{E}q. 6.4$$

4. Représentation valide :

$$\mathbb{C}_{valide} = \{C_M \in \mathbb{C}_M, H(C_M) - \delta|C_M| < A(C_M) < H(C_M) + \delta|C_M|\} \quad \acute{E}q. 6.5$$

$$i.e. \mathbb{C}_{valide} = \mathbb{C}_M \setminus (\mathbb{C}_{sur} \cup \mathbb{C}_{sous}) \quad \acute{E}q. 6.6$$

$$S_{valide} = \frac{|\mathbb{C}_{valide}|}{|\mathbb{C}_M|} \quad \acute{E}q. 6.7$$

Ces scores nous permettent d'évaluer la qualité de la calibration existante et de possiblement la valider. Dans le cas contraire, ils sont utilisés pour modifier les proportions des agents pour chaque ensemble de paramètres \mathcal{P}_i , explorer de nouveaux ensembles de paramètres, et retirer ceux qui sont invalides.

6.2 Ensembles de paramètres

En fonction du modèle d'agent et du nombre de paramètres, il est possible de générer le spectre complet des comportements d'agents possibles. Dans ce cas, un unique cycle via cette méthode est suffisant pour déterminer les paramètres valides et leur proportion dans la population d'agents.

Soit \mathcal{P}_v le groupe des ensembles de paramètres valides correspondant aux comportements valides \mathcal{B}_v , avec $simul(P_i) = b \in \mathcal{B}$ l'ensemble des comportements possibles, et $p(b)$ la proportion de participants humains exhibant ce comportement.

Étant donné que plusieurs ensembles de paramètres peuvent produire le même comportement, la production d'un ensemble de paramètres $P(a_i)$ avec $i \in \{1, \dots, n\}$ pour n agents implique de devoir choisir parmi plusieurs ensembles de paramètres. Nous proposons de sélectionner les ensembles de paramètres de la manière suivante : $P(a_i) = P_i \in \mathcal{P}_v$ avec une probabilité $p(P_i)$ dépendant de la proportion de comportements observés b et du nombre d'ensembles de paramètres P_j amenant au comportement b , d'où :

$$p(P_i) = n \frac{p(b)}{|P_j|} \text{ avec } P_j \in \mathcal{P}_v | simul(P_j) = b \quad \acute{E}q. 6.8$$

De cette manière, les comportements qui étaient sous-représentés ont une probabilité plus forte d'être produits. En effet, la probabilité de sélectionner un ensemble de paramètre compatible avec eux augmente, tandis que l'inverse est vrai pour les comportements sur-représentés.

Il est aussi possible de choisir de manière arbitraire un des ensembles de paramètres $P_j \in \mathcal{P}_v | simul(P_j) = b$ et générer $n \times p(b)$ agents avec cet ensemble de paramètres. Selon la simulation, garder une hétérogénéité contrôlée des agents permet de produire des simulations plus “réalistes”. Cette hétérogénéité est contrôlée puisque tous les ensembles de paramètres (menant aux comportements appartenant à la même classe) produisent des comportements similaires d’après la classification, sinon identiques.

Il est à noter qu’en utilisant uniquement \mathcal{P}_v et non \mathcal{P} , tous les ensembles de paramètres invalides détectés par la classification sont retirés.

6.3 Exploration de l’espace des paramètres

Afin de couvrir tous les ensembles de paramètres, le nombre d’agents devant être générés peut être important. De par l’utilisation de la connaissance d’expert ou des valeurs par défaut incluses dans les modèles, il est aussi possible de ne pas générer tous les comportements lors du premier cycle d’application de la méthode d’évaluation de la Simulation Multi-Agents. Dans le cas où des manques dans le modèle d’agent sont trouvés, nous proposons d’inclure une fonction d’exploration décrite dans l’équation 6.9 qui choisit les ensembles de paramètres dans les zones non-explorées de l’espace des paramètres :

$$P(a_i) = \begin{cases} P_i \in \mathcal{P}_v & \text{si } p > \gamma \\ P_k \notin \mathcal{P} & \text{sinon} \end{cases} \quad \text{Éq. 6.9}$$

Le paramètre d’exploration γ permet de chercher de nouveaux comportements de manière itérative, avec p une variable aléatoire uniforme. Afin de ne pas tester plusieurs fois les mêmes paramètres invalides, P_k ne doit jamais être choisi parmi les étapes précédentes \mathcal{P} . Dans le cas où P_k mène à un comportement valide, il est alors ajouté à \mathcal{P}_v ; dans le cas contraire, il est écarté.

Notons que selon le modèle d’agent considéré, il peut exister une méthode spécifique afin d’explorer de manière efficace l’espace des paramètres. À l’opposé, fixer la valeur de γ à 1 revient à faire uniquement de l’exploration. Dans ce cas \mathcal{P}_v se construit itérativement sur des ensembles différents.

6.4 Cycle

Comme vu dans les sections précédentes, si tous les comportements cibles - *i.e.* déterminés par les classes de traces des participants - sont reproduits, une seule étape de calibration est nécessaire. Quand plusieurs comportements sont manquants, explorer l’espace des paramètres peut permettre de découvrir de nouvelles capacités aux agents.

Les manques et les erreurs peuvent aussi être résolus par une intervention du concepteur du modèle d’agent. Dans ce cas, les informations issues de l’étape de

l'annotation permettent d'identifier sémantiquement les comportements manquants et les erreurs, *i.e.* expliciter ces comportements.

Rappelons que l'annotation et l'expérimentation sur les participants humains n'est nécessaire qu'une unique fois, et ce indépendamment du nombre de cycles puisque le traitement des données des agents est basé sur l'agrégation sur les traces des participants humains. Ainsi, les données des agents ne modifient pas les comportements de référence formés des classes de traces des participants. Une autre application de cette méthode est donc de comparer plusieurs modèles et plusieurs calibrations de modèles par rapport à cette référence.

6.5 Conclusion

Notre méthode introduit des métriques pour mesurer les scores résultants et pour corriger les paramètres des agents selon une telle étude. Les scores d'erreurs, de manques, et de capacités permettent au concepteur de la Simulation Multi-Agents de trouver combien d'archétypes de comportement humain ont été correctement reproduits, combien de comportements d'agents ne devraient pas apparaître, et combien de comportements humains sont manquants. Ensuite, en étudiant uniquement les archétypes des comportements correctement reproduits, les scores de calibration donnent de l'information sur les proportions de chaque comportement et leurs relations à la calibration des agents.

La réalité virtuelle requière que chaque agent adopte des comportements crédibles, *i.e.* des comportements pouvant être produits par des humains. Dans ce contexte, nous retirons premièrement les ensembles de paramètres qui ne produisent pas de comportements valides. Nous calibrons ensuite les proportions d'agents avec les ensembles de paramètres restants selon les données des participants humains. Un unique cycle de notre méthode assure que les comportements valides sont détectés et qu'ils sont produits en des proportions valides, nonobstant les comportements manquants.

Dans le cas d'une "boîte noire" où le modèle d'agent est inconnu et ne peut être modifié, si des comportements sont manquants alors une solution est d'explorer l'espace des paramètres afin de trouver de nouveaux comportements d'agents. Il est à noter que ces nouvelles étapes ne requièrent pas une autre expérimentation avec des participants humains, puisque les données de référence sont déjà disponibles. Chaque nouveau cycle permettra de trouver potentiellement de nouveaux ensembles de paramètres, soit dans des classes déjà *mixtes*, soit dans les classes de *manques* précédentes. Cela permet aussi de trouver quelles zones de l'espace des paramètres produisent des comportements invalides.

Dans le cas d'une "boîte blanche" où le modèle d'agent est connu et peut être potentiellement modifié, les données d'annotations expliquent les comportements manquants et les comportements erronés, permettant ainsi d'améliorer le modèle d'agent [47]. De plus, l'exploration des ensembles de paramètres peut être guidée par la connaissance du modèle [48].

Notre méthode de calibration (schématisée dans la *figure 6.1*) itère l'analyse de la classification afin de raffiner les capacités des agents à reproduire des comportements humains, tout en réduisant les manques et en supprimant les erreurs.

7 Conclusion

Cette partie a présenté notre méthode d'évaluation des comportements des agents dont le but est de reproduire des comportements humains. Cette méthode est basée sur la combinaison d'une classification d'annotations exploitant l'approche subjective et d'une classification d'indicateurs issus des traces de simulation exploitant l'approche objective. Ces classifications portent à la fois sur les comportements des agents et des participants, comparés dans la même situation. Elle utilise des techniques de classification non-supervisée afin d'obtenir des classes servant d'abstraction au comportement des participants humains, puis agrège les agents sur ces classes de participants afin de pouvoir les comparer et en extraire les manques, les capacités, et les comportements erronés du modèle d'agent.

Nos métriques basées sur la composition des classes fournissent des scores et des taux de confiance sur le modèle d'agent qui permettent la comparaison entre deux modèles agents, que ce soit une comparaison entre deux modèles distincts ou entre un modèle et sa version antérieure.

La méthode de calibration des paramètres du modèle présentée suit une approche itérative. Chaque itération permet d'obtenir, à partir de l'ensemble initial de traces valides (*i.e.* celles des participants) servant de points de référence pour la calibration du modèle, les manques et les erreurs du modèle afin de raffiner l'espace des paramètres, puis de générer une nouvelle population d'agents réduisant les comportements erronés, améliorant la représentativité des capacités des agents (*i.e.* des comportements correctement reproduits), et réduisant les comportements manquants de manière exploratoire.

Notre méthode d'évaluation peut être simplifiée pour des cas particuliers :

- Dans le cas où l'expérimentation ne nécessite pas d'annotation ou s'il n'existe pas de questionnaire adapté, il est possible d'utiliser notre méthode uniquement avec la classification issue des traces de simulation. Cette version sans explicitation permet de comparer les comportements d'agents avec ceux des humains par classification puis agrégation, d'obtenir des scores de capacités, de manques et d'erreurs, et d'effectuer des cycles de calibration. Il est à noter que l'absence d'annotation réduit considérablement le temps nécessaire pour une expérimentation.
- Dans le cas où le modèle d'agent n'a pas encore été défini, l'utilisation de notre méthode avec uniquement des participants (*i.e.* sans agents) permet d'obtenir les comportements des participants, de les dénombrer et de les caractériser. Cela peut permettre une conception en conséquence et d'avoir préalablement les comportements de référence.

Dans la partie suivante, nous présenterons l'application de la méthode d'évaluation et de calibration des comportements à travers plusieurs études s'appuyant sur une expérimentation en simulateur de conduite.

Troisième partie
Expérimentations

Nous avons présenté dans la partie précédente notre méthode d'évaluation des comportements des agents et notre cycle de conception. Dans le *chapitre 8*, nous appliquons cette méthode dans le cadre de la simulation de trafic routier. Cela est fait au sein d'une expérimentation avec des participants utilisant un simulateur de conduite exploitant l'architecture *ARCHISIM*. Le *chapitre 9* présente ensuite les résultats obtenus sur diverses études. Enfin, une conclusion discute des résultats obtenus en *section 9.4*.

8 Matériel et méthode

Dans cette partie, nous présentons tout d'abord l'architecture *ARCHISIM* dont le comportement des agents sera évalué et le dispositif de conduite. Ensuite, nous décrivons les scénarios de conduite utilisés pour les expérimentations ainsi que le protocole expérimental détaillé.

8.1 Simulation de trafic *ARCHISIM*

L'application des travaux de cette thèse est l'évaluation des comportements de conduite des agents du simulateur *ARCHISIM*. Cette architecture permet de simuler le trafic routier et a entre autre pour but de pouvoir simuler finement l'ensemble des comportements de conduite humains en toute situation. L'approche adoptée est similaire à celle des systèmes multi-agents : les véhicules sont des entités autonomes avec des capacités de perception agissant et interagissant au sein d'un environnement spatial. Cette particularité permet d'introduire des véhicules pilotés par des humains, notamment via des simulateurs de conduite, de façon transparente pour le système.

Le modèle a été validé à l'échelle macroscopique en termes de flux pour des situations sur autoroute incluant l'entrée et la sortie d'un véhicule par une bretelle, et ce dans le cas de trafic fluide et dense [25]. Le flux de véhicule a aussi été validé dans certaines situations urbaines simples [46].

Un conducteur simulé *ARCHISIM* a un trajet spécifique et des caractéristiques de style de conduite. Les variables d'entrée du modèle d'agent comprennent : la vitesse désirée par le conducteur simulé, la capacité de contrôle du véhicule en fonction des caractéristiques de la voie, la souplesse ou l'agressivité de la conduite, la volonté de respecter le code de la route, la flexibilité de la distance de sécurité (ou distance inter-véhiculaire) acceptée vis-à-vis du trafic, et le fait d'être ou non un conducteur expert. De plus amples explications quant au fonctionnement de cette architecture peuvent être trouvés dans [24] et [45].

8.2 Simulateur de conduite

Comme nous l'avons vu, il est possible d'interchanger agents et humains dans une simulation de trafic. Dans le but de comparer les comportements des agents simulés avec ceux des humains, nous souhaitons récolter des données sur les comportements d'une population de participants dans des tâches de conduite. Ce recueil de données est fait lors d'une expérimentation sur simulateur de conduite.

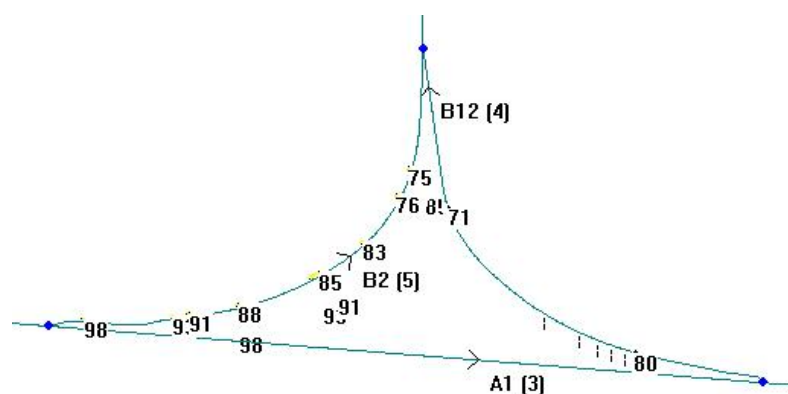


Figure 8.1 – Vue aérienne de la simulation de trafic ARCHISIM : Cas d'une insertion sur autoroute avec des véhicules simulés numérotés. La route B12 rejoint la bretelle B2 qui s'insère sur l'autoroute A1.

Nous nous intéressons à une dimension particulière de la présence qu'est la crédibilité des comportements de haut niveau. Cette dimension comportementale ne demande pas nécessairement une immersion totale (sonore, visuelle, haptique, ...). C'est pourquoi nous avons opté pour un simulateur *léger* sur table. Aussi, l'utilisation d'un tel simulateur - et non d'un simulateur à échelle réelle - facilite la mise en place de l'expérimentation. De plus, ce type de simulateur permet de limiter le taux de participants ayant le mal du simulateur dont les données ne sont pas exploitables. Cependant, les résultats du questionnaire de présence pourraient différer sur un simulateur à échelle réelle.

Notre expérimentation est effectuée sur un dispositif décrit dans la *figure 8.2* comprenant un volant, un pédalier et une boîte de vitesses ainsi que 3 écrans permettant un champ de vision latéral suffisant notamment pour les carrefours. Ces écrans servent aussi à l'intégration des visuels des rétro-viseurs (gauche et central).

Il est à noter que ce simulateur s'est avéré insuffisant pour certaines situations. Cela a été le cas notamment dans lors du premier carrefour pour lequel le conducteur est à l'arrêt à un panneau stop. En effet, le champ de vision du conducteur ne permettait alors pas de prévoir suffisamment à l'avance l'arrivée des véhicules simulés.



Figure 8.2 – Dispositif du simulateur de conduite (avec 3 écrans, volant, boîte de vitesses et pédalier) utilisant le simulateur de trafic routier ARCHISIM.

8.3 Scénario

Le circuit proposé met en jeu un panel de diverses situations que nous voulons étudier. Il a été conçu pour durer environ 20 minutes comprenant 11 sections de route, appelées *phases*, qui comprennent des situations de virages successifs, de carrefours (un prioritaire et un non prioritaire), d'insertion sur autoroute, *etc.*

Le but du circuit est de rejoindre l'autoroute *A75* indiquée par des panneaux puis de conduire sur cette dernière. Un premier passage de test sans véhicule simulé est effectué pour que le participant s'habitue au fonctionnement du simulateur, au pilotage, ainsi qu'au circuit. Ensuite, le participant effectue le scénario dans son intégralité, en interaction avec les véhicules simulés. Les données de simulation sont alors enregistrées pour la phase de traitement et une vidéo de la conduite effectuée (appelée *rejeu*) est conservée pour l'annotation *a posteriori*.

La *figure 8.3* présente une schématisation des différentes phases de conduite du scénario (voir aussi le circuit seul en *annexe C.1*). Ces phases sont les suivantes :

1. Le circuit commence par une route à sens unique en ligne droite à double voies limitée à 90 Km/h sur laquelle seront analysés le démarrage et la conduite en ligne droite.
2. Sur cette même route se présente un véhicule à vitesse réduite pour lequel les acteurs principaux auront la possibilité d'effectuer une manœuvre de dépassement.
3. S'en suit une série de quatre virages (indiquée via un panneau de signalisation) dont la vitesse est limitée à 70 Km/h servant à observer la conduite en virage au sein d'un trafic fluide.
4. Après une courte ligne droite où le trafic s'intensifie, un rétrécissement à une voie de la route engendre un embouteillage. La limitation de vitesse est rétablie à 90 Km/h . Cela permet d'observer les réactions de conduite en trafic congestionné.
5. Par la suite, vient se greffer une route en sens inverse, la route se transforme alors en deux fois une voie. Cette étape permet également d'évacuer le trafic afin de s'assurer une homogénéité des situations futures avec un acteur principal sans véhicule devant.
6. Cette route de type deux fois une voie en sens inverses contient un panneau de signalisation lui indiquant qu'il n'est pas prioritaire. Un panneau indique que la direction pour l'autoroute *A75* est tout droit. L'acteur principal arrive alors à un stop devant un croisement avec une autre voie similaire mais prioritaire. Cette étape nous informe sur le comportement de l'utilisateur lors de la traversée d'un croisement en étant non prioritaire lorsque le but est de traverser le flux.
7. Le tronçon de route suivant est rectiligne. Il contient un véhicule à vitesse réduite. À son approche, le véhicule principal pourra initier un dépassement mais plusieurs véhicules arrivent dans le sens inverse.
8. Un panneau indique que la route à suivre pour l'autoroute *A75* est celle de gauche. Plus loin, un second croisement se présente pour lequel le participant est prioritaire. C'est alors la voie transversale qui comporte des panneaux stop.

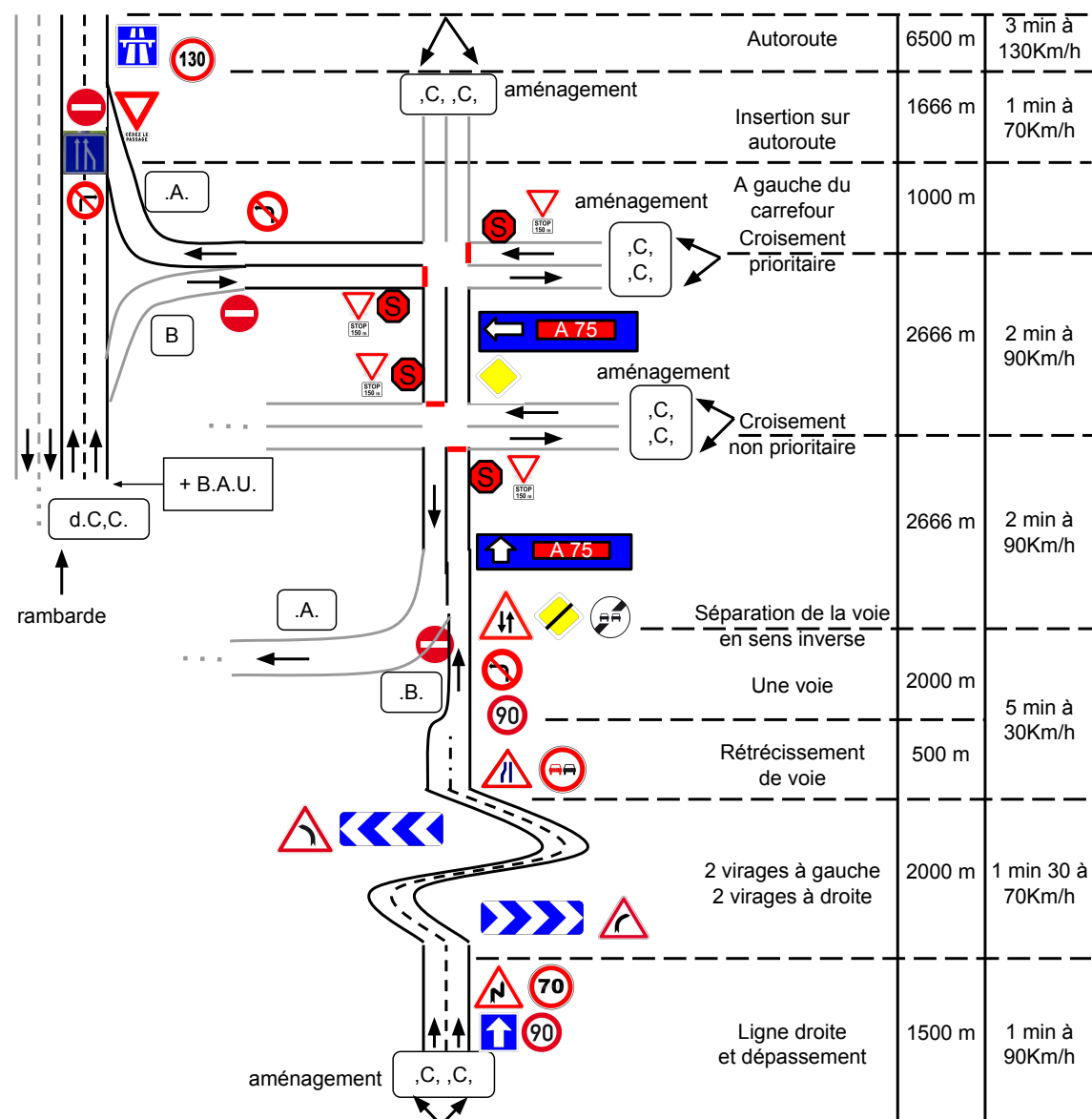


Figure 8.3 – Circuit du scénario de conduite : L'acteur principal commence le circuit sur une route à deux voies dans le mêmes sens (en bas de la figure). S'ensuit un passage en ligne droite, une série de virages, et un rétrécissement de voie. Après qu'une route en sens inverse se soit greffée, le conducteur traverse un premier carrefour puis un second où il doit tourner à gauche sur une route le menant à une bretelle d'insertion. Cette bretelle rejoint l'autoroute sur laquelle le participant doit conduire au sein d'un trafic fluide puis congestionné.

9. Ce trajet mène le participant à devoir s'insérer sur une autoroute, or l'insertion sur autoroute est une situation particulière qui engendre des comportements spécifiques.
10. Nous étudions par la suite la conduite sur autoroute (2 × 2 voies) en trafic fluide à 130 Km/h.
11. Enfin, un ensemble de véhicule à vitesse réduite est présent sur la file de droite de l'autoroute.

Afin de montrer le potentiel de notre méthode, nous cherchons à exploiter des

phases ayant abouti à la fois à la création de classes mixtes, de classes composées uniquement de véhicules simulés, et d'autres uniquement de véhicules pilotés par des participants. En effet, il est possible qu'une phase de conduite demande une manœuvre délicate que l'utilisateur ne peut pas effectuer à cause du dispositif. C'est le cas du second carrefour qui a été accidentogène pour tous les participants. Cependant les véhicules simulés *ARCHISIM* ont réussi à passer ce carrefour sans incident. Les comportements de conduite d'*ARCHISIM* différant intégralement de ceux des participants, la classification de traces ne contient alors aucune classe mixte. Il est aussi possible qu'aucun véhicule simulé n'adopte un comportement *visuellement crédible* qui soit similaire aux participants. Par exemple, lors du suivi en file en rétrécissement de voie pour lequel la vitesse des véhicules simulés était constante à 3 Km/h, les participants ont avancé en *accordéon* (*i.e.* un enchaînement d'accélération afin de rejoindre le véhicule devant et de décélérations jusqu'à l'arrêt du véhicule). La classification de traces montre alors aussi une scission stricte entre les classes de participants et les classes d'agents.

8.4 Protocole expérimental

Avant la simulation, le questionnaire de comportement de conduite *DBQ* est soumis dans sa version traduite en français dans le but de vérifier la cohérence de notre annotation (voir la *sous-section 9.2.2*). La version soumise est disponible en *annexe A.1*.

Ensuite, un premier passage de test sans véhicule simulé est effectué pour que le participant s'habitue au fonctionnement du simulateur ainsi qu'au circuit. Lors de ce passage, les commandes sont expliquées et le participant peut s'habituer à l'immersion physique dans l'environnement virtuel. La connaissance préalable par le participant de la route lui permet de mieux appréhender le circuit avec trafic lors du second passage. Ce premier passage est considéré comme complet à partir du moment où le participant a vu l'intégralité des phases de conduite et qu'il est capable de piloter correctement.

Puis, le participant effectue le scénario dans son intégralité, en interaction avec les véhicules simulés constituant le trafic environnant. Il est à noter que le comportement des véhicules simulés n'étant pas scripté, les situations peuvent légèrement différer selon le comportement de l'acteur principal. Les données sont enregistrées pour l'étape de traitement et une vidéo est faite pour l'étape de rejeu.

Enfin, une autre population de participants remplit le questionnaire d'annotations après visionnage des rejeux de la simulation dans le but d'évaluer le comportement adopté par les acteurs principaux (*i.e.* participants humains et véhicule simulés).

Après la simulation et dans le but d'évaluer le sentiment de présence, le participant remplit le questionnaire de présence (*PQ*) de Witmer et Singer [139] dans sa version française (détaillée en *annexe A.4*) afin de mesurer le niveau de contrôle, la qualité de l'immersion, les distractions perturbant l'expérience et le degré de réalisme. Nous souhaitons aussi obtenir des informations sur la dimension comportementale de la présence (*i.e.* mesurer la qualité des comportements simulés). Pour ce faire,

le participant remplit le *Questionnaire Général de Comportement* que nous avons élaboré. Ce questionnaire sera détaillé en *sous-section 8.5.3.a*.

Ensuite, un rejeu de la simulation est effectué en utilisant la vidéo enregistrée. Lors de ce rejeu le simulateur est mis en pause entre chaque phase. Le *Questionnaire Spécifique de Comportement* (détaillé en *sous-section 8.5.3.b*) est alors remis au participant de façon à évaluer la qualité des comportements simulés pour chaque phase séparément. Les temps de réponse varient en fonction du temps pris par le participant pour répondre et possiblement expliquer ses réponses. Ensuite le rejeu reprend et une nouvelle pause est effectuée pour la phase suivante.

Lorsque l'ensemble du rejeu de la simulation est terminé, les questionnaires nécessitant d'être soumis *a posteriori* (et ceux n'ayant pas de contrainte) sont remplis par le participant. Cela comprend le questionnaire de tendance à l'immersion *ITQ* et le questionnaire d'engagement du joueur *GEQ* détaillés en *sous-section 8.5.1*, ainsi que d'autres questionnaires *ad hoc* décrit en *sous-section 8.5.4*.

8.4.1 Populations

L'échantillon de population a une bonne vue (ou corrigée), a entre 24 et 59 ans (dont 44% de femmes), est automobiliste régulier et détient un permis de conduire voiture depuis plus d'un an. Notre expérimentation est effectuée sur un dispositif comprenant un volant, un pédalier et une boîte de vitesses ainsi que 3 écrans permettant un champ de vision latéral suffisant et l'intégration des visuels des rétro-viseurs gauche et central (voir *section 8.2*). 23 participants ont utilisé ce dispositif.

Les traces de 14 agents ont été générées. Les paramètres d'entrée du modèle d'agent utilisés sont les valeurs moyennes de chaque paramètre avec la modification d'un paramètre. La vitesse est choisie parmi {100, 110, 120, 130, 140}, les paramètres *reglmnt*, *infra*, *trafic* et *soupl* valent par défaut 50 et ont été modifiés à 25 et 75. Le paramètre *exp* quant à lui vaut 0 par défaut, sinon 1. Ces paramètres sont listés dans la *table 8.1*.

Nom	Min.	Défaut	Max.	Description
<i>vitesse</i>	100	130	140	La vitesse désirée par le conducteur simulé. Cette vitesse est bornée par la limitation de vitesse de la route.
<i>infra</i>	25	50	75	La capacité de contrôle du véhicule en fonction des caractéristiques de la voie.
<i>soupl</i>	25	50	75	La souplesse ou l'agressivité de la conduite.
<i>reglmnt</i>	25	50	75	La volonté de respecter le code de la route.
<i>trafic</i>	25	50	75	La flexibilité de la distance de sécurité (ou distance inter-véhiculaire) acceptée vis-à-vis du trafic.
<i>exp</i>	0	0	1	Le fait d'être ou non un conducteur expert.

Table 8.1 – Les paramètres des agents simulés d'*ARCHISIM* avec leurs valeurs utilisées par défaut, en tant que minimum et maximum.

L'ensemble des conducteurs humains et des conducteurs virtuels constitue les acteurs principaux dans cette application.

La population d'annotateurs évaluant les conduites effectuées est composée de 6 participants. Un annotateur a eu le mal du simulateur (*i.e.* des vertiges) et cessa de regarder les rejeux. Ses annotations ont donc été supprimées de l'analyse.

8.4.2 Données de simulation

L'architecture *ARCHISIM* nous permet d'obtenir de nombreuses informations sur l'environnement (voir *table 8.2*), sur les véhicules (voir *table 8.3*), mais aussi des variables spécifique pour le véhicule principal (voir *table 8.4*) et pour les véhicules constituant le trafic alentour (voir *table 8.5*). L'ensemble de ces données correspond aux traces de simulation.

Nom	Type	Description
Pas	UINT	Indice du pas de temps actuel de la simulation.
Temps	DATE	Temps écoulé depuis le début de la simulation au format HH:MM:SS, #####.
IndEssai	UINT	Numéro d'essai courant (de 10 à 120).
NumInstruction	UINT	Numéro d'instruction courant (de 10 à 110).
Rayon de courbure	DOUBLE	Dérivée de l'axiale en X et Y . Donne la courbe de la route.

Table 8.2 – Les variables générales de l'environnement *ARCHISIM*.

Les indicateurs donnés par les experts du domaine (*i.e.* les variables haut niveau) dépendent de la situation : par exemple, le cap est très important en dépassement mais pas en ligne droite sur voie unique. C'est la raison pour laquelle nous ne détaillerons pas dès cette section les indicateurs choisis par les experts mais dans les parties suivantes selon les phases de conduite exploitées.

Il est à noter que la taille des données brutes à analyser est importante : 19 variables pour le véhicule principal, 14 variables pour chaque véhicule simulé, et 5 variables d'environnement. L'ensemble de ses variables est enregistré toutes les 300 *ms* pour un temps total approximatif de 20 minutes, et cela pour chaque participant et chaque agent à analyser.

8.4.3 Application de la méthode de classification

Suite à l'expérimentation, nous obtenons pour chaque véhicule principal (véhicule simulé ou véhicule piloté par un participant) une matrice contenant pour chaque variable un vecteur temporel.

Nous appliquons alors notre méthode (dont les prétraitements sont décrits en *section 4.3* et la classification en *section 5.1*) comme suit. Les variables récoltées sont utilisées pour calculer les indicateurs temporels et ceux non temporels donnés par

Nom	Type	Description
Route	UINT	Numéro de la route sur laquelle le véhicule roule (appartient à $[1 : 5]$).
Voie (mm)	INT	Décalage par rapport à l'axiale (#### négatif ou positif).
Cap (deg)	DOUBLE	Angle de rotation par rapport à la route actuelle. La valeur est proche de -180 si les véhicules vont tout droit mais à contresens du point kilométrique (appartient à $] -180; 180[$).
NoVoie	INT	Indice de la voie. Valeurs positives dans le sens du point kilométrique, négatives à contresens (entre -4 et 4 , 0 exclus car il n'y a pas de voie centrale).
Pk (mm)	UINT	Point kilométrique en millimètre.
NoVh	INT	Numéro du véhicule. 0 pour le véhicule principal piloté, -2 pour le véhicule principal simulé, un entier négatif pour les autres véhicules simulés.
Vit (m/s)	DOUBLE	Vitesse du véhicule (##, ### positif).
Acc (m/s^2)	DOUBLE	Accélération du véhicule (#, ### positif pour l'accélération, négatif pour la décélération).
Topo[x](mm)	DOUBLE	Topographie du circuit sur l'axe X en millimètre (#+, ###).
Topo[y](mm)	DOUBLE	Topographie du circuit sur l'axe Y en millimètre (#+, ###).
Topo[z](mm)	DOUBLE	Topographie du circuit sur l'axe Z en millimètre (#+, ###). Cette variable est inutile dans notre cas car le circuit est sans relief

Table 8.3 – Les variables communes pour le véhicule principal et ceux simulés.

les experts du domaine. Les similitudes entre les véhicules principaux sur ces indicateurs temporels sont calculées par la distance DTW . En résulte une matrice carrée de similitude qui est alors retransformée en coordonnées par positionnement multidimensionnel.

Les nouvelles coordonnées des véhicules des participants sont alors utilisées pour l'algorithme de classification. Les centres des classes des K -moyennes sont initialisés avec K -moyennes++. L'algorithme est exécuté pour K variant de 2 au tiers de la population et la classification retenue est choisie selon le critère *Calinski-Harabasz*.

Nous appliquons alors notre méthode d'agrégation afin d'intégrer les agents proches aux classes de participants. Les agents ne faisant partie d'aucune classe (malgré le seuil de tolérance) créent leurs propres classes. Nous obtenons alors une classification contenant des classes d'agents, des classes de participants, et des classes mixtes.

Nous calculons ensuite les scores de confiance en ces classes qui représentent respectivement les comportements erronés, les comportements humains manquants dans le modèle d'agent, et les capacités du modèle à reproduire le comportement humain.

Nom	Type	Description
Cab.volant	INT	Angle donné par le participant au volant sous la forme ###. À valeurs positives quand le volant est tourné vers la gauche, négatives quand le volant est tourné vers la droite.
Cab.accél	UINT	Degré de la pédale d'accélération en pourcentage (appartient à $\in [0 : 100]$).
Cab.frein	UINT	Degré de la pédale de freinage en pourcentage (appartient à $[0 : 100]$).
Cab.embrayage	UINT	Degré de la pédale d'embrayage en pourcentage.
Cab.BV	UINT	Position de la boîte de vitesse. À valeur dans $[0 : 5]$ avec 0 pour le point mort.
Cab.indics	UINT	Utilisation des boutons du volant et du levier de vitesse. Les valeurs sont gérées par <i>ARCHISIM</i> : 1 par défaut, 104 pour le starter.
Cab.boutons	UINT	Identique à Cab.indics mais géré par <i>Windows</i> . Les valeurs sont 0 par défaut et 00000001 pour le starter.
CibleAr.voie	INT	Décalage à l'axiale de la cible arrière (correspond donc à la variable "Voie (mm)" du véhicule derrière). Cette variable prend la valeur -32768 s'il n'existe aucune cible arrière.

Table 8.4 – Les variables spécifiques au véhicule principal.

Nom	Type	Description
T.I.V.(s)	DOUBLE	Temps inter-véhiculaire, c'est-à-dire le temps en secondes que mettrait le véhicule simulé à rejoindre le véhicule devant (#+, ###). Prend la valeur -1 s'il n'y a pas de véhicule avant.
Jerk (m/s^3)	DOUBLE	Le jerk est la dérivée du vecteur accélération par rapport au temps (#+, ###).
Stratégie	STRING	Stratégie adoptée par le véhicule simulé. Par exemple : "EnCours Doubler Déboit SuiviC InsertC Accélère Adapt.court". Cette variable donne des informations sur le comportement du véhicule à un plus haut niveau. Cependant, elle ne peut être inférée pour les participants. Elle ne sera donc pas exploitée.
Cible.no	INT	Numéro du véhicule cible arrière. Prend la valeur -1 s'il n'y a pas de véhicule arrière.

Table 8.5 – Les variables spécifiques aux véhicules simulant le trafic routier de l'environnement.

8.5 Questionnaires de sciences humaines et sociales

Pour répondre à nos hypothèses sur les facteurs influençant le comportement du participant ou sur les facteurs jouant un rôle dans la perception des comportements, nous soumettons aux participants les questionnaires suivants (voir la *figure 8.4*). Cette partie est réalisée en parallèle des classifications.

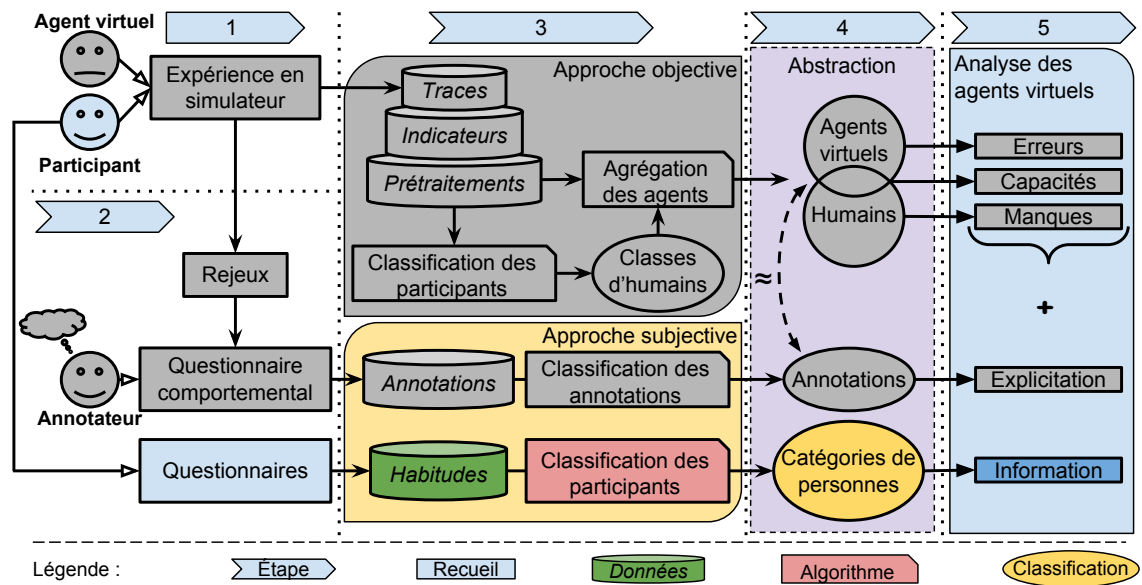


Figure 8.4 – Exploitation des questionnaires au sein de la méthode.

8.5.1 Propension à l’immersion

Une des hypothèses que nous voulons tester est que l’immersion du participant en simulateur a une influence sur son jugement porté sur la crédibilité des comportements observés. Nous souhaitons analyser la capacité d’immersion du participant avant l’expérimentation afin de la comparer avec la présence ressentie lors de l’expérimentation. Cela est fait par une autoévaluation *a priori* de l’immersion via un questionnaire de tendance à l’immersion. Nous utilisons l’*Immersive Tendencies Questionnaire (ITQ)* de Witmer et Singer [139] dans sa version traduite [121]. Cette version est disponible en *annexe A.7*.

Comme vu dans la *sous-section 1.2.3*, nous nous intéressons à la propension à l’immersion via l’*Immersive Tendencies Questionnaire* et aux profils de joueur de jeux vidéo au travers du *Game Engagement Questionnaire (GEQ)*. Il n’existe pas à notre connaissance de traduction du *GEQ*. C’est pourquoi nous avons traduit le questionnaire en français, puis fait vérifier la cohérence de la traduction par un expert-traducteur anglophone (voir *table 8.6*). Le questionnaire dans sa version soumise aux participants est disponible en *annexe A.6*.

Ce questionnaire nous informe sur l’état mental du participant lorsqu’il joue aux jeux vidéo. Cependant, nous souhaitons aussi recueillir des informations plus concrètes sur la relation du participant vis-à-vis du jeu vidéo. Dans ce but, nous avons créé un court questionnaire sur les jeux vidéo (5 items) afin de cibler le participant selon : la première fois où il a joué aux jeux vidéo, sa fréquence de jeu actuelle, le dispositif utilisé (impactant l’immersion), son habitude à voir des comportements simulés (au travers de la présence d’autres joueurs et de personnages non-joueurs), et le type de jeux vidéo auxquels il est habitué à jouer (comme l’expérience des jeux en *3D*, ou des jeux de courses modifiant sa capacité à maîtriser le véhicule piloté). Il nous permet d’obtenir des données concrètes et précises sur le joueur de jeu vidéo de façon complémentaire au *GEQ* qui nous donne la typologie de l’engagement du joueur. Le détail de ce questionnaire est visible en *annexe A.2*.

#	Version anglaise	Version traduite en français
1	I lose track of time	Je perds la notion du temps
2	Things seem to happen automatically	Les événements semblent se dérouler de manière automatique
3	I feel different	Je me sens différent
4	I feel scared	Je me sens effrayé
5	The game feels real	Le jeu semble réel
6	If someone talks to me, I don't hear them	Si quelqu'un me parle, je ne l'entends pas
7	I get wound up	Je deviens irritable
8	Time seems to kind of stand still or stop	Le temps semble se mettre en pause ou s'arrêter
9	I feel spaced out	Je me sens comme déconnecté de la réalité
10	I don't answer when someone talks to me	Je ne réponds pas quand quelqu'un me parle
11	I can't tell that I'm getting tired	Je ne me rends pas compte quand je fatigue
12	Playing seems automatic	Jouer semble automatique
13	My thoughts go fast	Mes pensées vont vite
14	I lose track of where I am	Je perds conscience d'où je suis
15	I play without thinking about how to play	Je joue sans réfléchir à comment jouer
16	Playing makes me feel calm	Jouer me fait me sentir calme
17	I play longer than I meant to	Je joue plus longtemps que ce que j'avais prévu
18	I really get into the game	Je suis vraiment dans le jeu
19	I feel like I just can't stop playing	J'ai l'impression que je ne peux plus m'arrêter de jouer

Table 8.6 – Traduction du *Game Engagement Questionnaire*.

8.5.2 Annotations et autoévaluation

La phase d'annotation, consistant en l'évaluation de l'ensemble des comportements des acteurs principaux par une autre population, requière un questionnaire spécifique au domaine permettant de caractériser les comportements observés au travers d'un rejeu vidéo.

Dans l'application de notre méthode à la simulation de trafic, nous nous basons sur le *Driver Behavior Questionnaire (DBQ)* lors de l'étape d'annotation des rejeux des véhicules principaux. Afin de pouvoir évaluer les agents tout comme les participants, ce questionnaire d'autoévaluation a été modifié de façon à ce qu'une population d'annotateurs puisse le remplir pour des comportements de conduite observés à la première personne. Nous souhaitons néanmoins soumettre le *DBQ* aux participants afin comparer ses résultats avec ceux obtenus via notre questionnaire d'annotation. Ce dernier est soumis avant la conduite en simulateur afin de ne pas être biaisé par l'expérience vécue.

8.5.3 Évaluation de la crédibilité

Lors de l'expérimentation, le participant conduit au sein d'un trafic constitué d'agents simulés dont les comportements sont issus du même modèle que celui des agents analysés (*i.e.* les agents acteurs principaux). Il est intéressant de comparer l'évaluation des annotateurs faite sur les rejoux des comportements des agents analysés (avec une vue à la première personne) à celle faite sur les comportements des véhicules observés au sein du trafic alentour (avec une vue à la troisième personne). C'est pourquoi nous demandons au conducteur lors de la phase de rejou d'évaluer les comportements de certains véhicules du trafic.

Cette évaluation requière un questionnaire comportemental évaluant la crédibilité des comportements sur l'ensemble de la simulation. Afin de cibler l'analyse de certains véhicules au sein du trafic, il est aussi nécessaire d'avoir un questionnaire spécifique de comportement qui s'attarde sur le comportement d'un unique acteur dans une situation précise. Il n'existe cependant pas à notre connaissance de questionnaire permettant au participant d'évaluer spécifiquement la crédibilité des comportements observés dans le cadre d'humains simulés.

Nous proposons donc un *Questionnaire Général de Comportement* évaluant de manière globale les comportements du système. Nous proposons également un *Questionnaire Spécifique de Comportement* afin de répondre à la seconde problématique qu'est l'évaluation d'un comportement ciblé. Dans les sections suivantes, nous présentons les deux questionnaires élaborés.

8.5.3.a Questionnaire Général de Comportement

Nous voulons vérifier que l'immersion du participant influence son jugement des comportements. Nous devons donc obtenir des informations générales sur le ressenti des participants vis-à-vis des comportements observés.

Le questionnaire général de comportement a pour but d'évaluer la qualité ressentie des comportements des agents simulés en général, c'est-à-dire dans l'ensemble de la simulation. Ce questionnaire est soumis après le questionnaire de présence, lui-même étant soumis juste après la phase de simulation de façon à limiter les biais induits par une altération de la mémoire du participant avec le temps.

Les questions demandent pour chacune une unique réponse marquée d'une croix sur une échelle comportant 7 possibilités allant de "Pas du tout" à "Tout à fait" et "Sans opinion". Le participant doit indiquer à quel point il est d'accord avec les affirmations proposées sur une échelle de type Likert [87]. Il contient des affirmations basées sur le modèle suivant : "Un acteur humain peut avoir les comportements effectués par les acteurs simulés dans cette situation". Cette formulation a été pensée pour pouvoir s'appliquer à toute situation. Le fait que la question soit toujours sous la même forme évite d'avoir un biais entre les termes utilisés. Cette question est répétée pour chaque situation évaluée.

La liste de ces questions dans le cadre de l'évaluation de comportements de conduite est décrite dans la *table 8.7*. La version définitive soumise lors de l'expérimentation est en *annexe A.8*.

#	Description
1	en ligne droite sur route nationale (hors situation de dépassement).
2	en virage (hors situation de dépassement).
3	en ligne droite sur autoroute (hors situation de dépassement).
4	face à un véhicule roulant à faible vitesse.
5	en situation de dépassement.
6	en trafic dense.
7	aux carrefours.
8	roulants à faible vitesse.

Table 8.7 – Questionnaire Général de Comportement décrit par situation dans le cadre de l'évaluation de comportements de conduite. Les affirmations sont présentées sous la forme : “Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules simulés...”.

Les corrélations entre notre questionnaire de crédibilité comportementale et le questionnaire de présence seront à observer. Il est important de soumettre le questionnaire de présence avant notre questionnaire de crédibilité comportementale générale pour nous permette de réduire, par une verbalisation du participant, certaines remarques qui concernent d'autres composantes de la présence comme la qualité du graphisme, la qualité de l'interface, *etc.*

8.5.3.b Questionnaire Spécifique de Comportement

De manière à cibler des éléments particuliers du comportement, un questionnaire avec des questions sur des situations précises s'étant présentées a été élaboré. Ce questionnaire contient une ou plusieurs questions par phase pour un total de 32 items. Il est soumis pendant la phase de rejeu. Pour chaque question, une pause est effectuée avant la scène où la consigne d'observer un ou plusieurs véhicules particuliers est donnée. Le rejeu est ensuite lancé pour la durée de la scène nous intéressant. Enfin, une pause est faite après la scène pour poser la question à l'oral. Nous donnons alors une règlette (voir la *table 8.8* ci-dessous) contenant les choix possibles de réponses sur laquelle le participant indique son choix. Suite à cette réponse, nous poursuivons avec une courte discussion servant à expliciter pourquoi le participant a donné cette note.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

Table 8.8 – Règlette du questionnaire de comportement.

Les affirmations sont formées de la même façon que pour le questionnaire général et les réponses sont sur la même échelle. Si la réponse est entre “Modérément” et “Tout à fait”, une seconde affirmation est donnée : “Ce(s) comportement(s) est/sont courant(s)” demandant une réponse sur la même échelle que les autres questions. Cette seconde question permet de savoir si le comportement existe chez les humains mais est rare ou si au contraire c'est un comportement répandu. Tout comme la

première question, sa forme est fixe pour éviter le biais induit par les changements de termes utilisés dans l’affirmation.

La liste des questions posées par phase dans le cadre de l’évaluation de comportements de conduite est donnée dans la *table 8.9*. Une version complète se trouve en *annexe A.9*.

#	Description
1	<ul style="list-style-type: none"> • Le comportement de déboîtement effectué par le conducteur du véhicule devant vous dans cette situation de dépassement. • Le comportement de rabattement effectué par le conducteur du véhicule devant vous dans cette situation de dépassement.
2	<ul style="list-style-type: none"> • Le comportement de conduite effectué par le conducteur du véhicule devant vous dans cette situation.
3	<ul style="list-style-type: none"> • Le comportement de freinage effectué par le conducteur du véhicule devant vous dans cette situation. • Le comportement de suivi en file effectué par le conducteur du véhicule devant vous dans cette situation de bouchon. • Le comportement d’accélération effectué par le conducteur du véhicule devant vous dans cette situation de bouchon.
4	<ul style="list-style-type: none"> • Le comportement de conduite effectué par le conducteur du véhicule devant vous quand il tourne à gauche dans cette situation.
5	<ul style="list-style-type: none"> • Le comportement de conduite effectué par le conducteur du véhicule suivant dans cette situation : Premier véhicule venant de droite et allant tout droit. Second véhicule venant de gauche et allant à droite. Véhicule arrêté au stop venant en face.
6	<ul style="list-style-type: none"> • Le comportement de conduite effectué par le conducteur du véhicule roulant à faible vitesse devant vous dans cette situation de ligne droite.
7	<ul style="list-style-type: none"> • Les comportements de conduite effectués par les conducteurs des véhicules en face de vous dans cette situation au carrefour. • Le comportement de conduite effectué par le conducteur du camion venant de gauche dans cette situation au carrefour.
8	<ul style="list-style-type: none"> • Les comportements de conduite effectués par les conducteurs des véhicules sur l’autoroute dans cette situation d’insertion.
9	<ul style="list-style-type: none"> • Le comportement de conduite effectué par le conducteur du véhicule devant vous sur l’autoroute dans cette situation.
10	<ul style="list-style-type: none"> • Les comportements de dépassement effectués par les conducteurs des véhicules doublant les véhicules roulants à faible vitesse sur l’autoroute dans cette situation.

Table 8.9 – Questionnaire Spécifique de Comportement décrit par étape. Les affirmations sont présentées sous la forme “Un conducteur humain peut avoir...”.

8.5.4 Questionnaires *ad hoc*

Afin de s'assurer de la cohérence de la population étudiée et de la validité du participant pour cette expérimentation, deux autres questionnaires *ad hoc* sont soumis aux participants (voir *annexe A.3*).

- **Questionnaires d'état civil :**

Ce questionnaire de 7 items nous permet de s'assurer d'avoir une population représentative des conducteurs selon leur âge et leur genre. Les autres éléments demandés sont : le statut (actif, étudiant, retraité, . . .), le niveau d'étude et le nombre d'enfants du participant de façon à voir leur impact sur la conduite.

- **Questionnaire conducteur :**

Notre questionnaire sur le conducteur contient 24 items. Certaines questions permettent de s'assurer que le participant remplit bien les conditions nécessaires pour la sélection des candidats : que le participant a bien un permis de type *B* valide et que le participant a bien une vue normale ou corrigée. La date d'obtention du permis est demandée, cela permettra de déduire les participants étant encore en "A" et tester l'hypothèse qu'ils ont un comportement particulier au-delà des changements de limitation de vitesse sur autoroute.

Nous demandons aussi des informations sur les autres permis détenus et les autres véhicules conduits pour mieux juger des réponses. Notamment, avoir l'habitude de conduire un poids lourd peut influencer le comportement de conduite en voiture et le jugement lors de l'observation de véhicules poids lourds. Pour mieux évaluer le niveau de respect du code de la route et celui de prudence du participant, nous souhaitons connaître le nombre d'amendes hors stationnement durant les deux dernières années et le nombre de fois où le participant a été reconnu responsable dans un accident de voiture.

Des informations sur la voiture habituellement conduite par le participant sont demandées (boîte de vitesses automatique, nombre de vitesses, *ABS*, régulateur de vitesse, direction assistée et *GPS*). La dernière partie du questionnaire prend note de la fréquence de conduite globale selon le type de route (routes départementales, routes de ville, voies rapides, *etc.*).

8.6 Implémentation

Notre méthode d'évaluation des comportements demande l'utilisation d'un algorithme de classification non-supervisée.

La première possibilité pour utiliser ces algorithmes est d'implémenter nous-mêmes l'algorithme voulu. Cela permet d'être libre à tous niveaux : sur le choix de l'algorithme à utiliser ; sur le prétraitement des données nécessaires en vue de les utiliser comme entrée des fonctions ; sur le langage utilisé ; sur la justesse de l'implémentation au regard de l'algorithmique. Cependant cette méthode est longue à mettre en œuvre et est inutile si une implémentation existe déjà. De plus, elle ne permet pas de tester rapidement les nombreux algorithmes de classification. Enfin,

le code produit est difficile à optimiser, or la complexité est importante du fait de la taille des données à analyser. C'est pourquoi nous n'utiliserons pas cette méthode.

La seconde possibilité est d'utiliser une bibliothèque contenant l'algorithme désiré. Cela peut nous restreindre sur le choix des algorithmes, nous forcer à effectuer des prétraitements, ou nous obliger à utiliser un certain langage. Parfois, il ne nous est pas possible de vérifier qui a développé la bibliothèque ainsi que comment la bibliothèque a été implémentée et donc sa justesse et son optimisation. C'est pourquoi nous nous sommes concentrés sur les bibliothèques utilisées par les chercheurs du domaine et disponibles en *open source*. De plus, les bibliothèques les plus utilisées contiennent de nombreux algorithmes basés sur un même type de donnée en entrée et donc un unique prétraitement nécessaire.

Par la suite, nous allons présenter quelques bibliothèques d'algorithmes de classification connues ainsi que leur langage souvent adapté à ces traitements :

Matlab est basé sur le calcul matriciel, il permet donc des calculs plus rapides que d'autres langages notamment lors de l'utilisation d'algorithmes d'apprentissage statistique. Il a la capacité de traiter des fichiers sous la forme de *CSV* ce qui évite une partie des prétraitements. Ses bibliothèques s'organisent sous la forme de *toolboxes* et en particulier la *Statistics Toolbox*¹ qui contient des fonctions de classification non-supervisée (comme le K-means, le mélange gaussien, l'analyse en composantes principales), des algorithmes de régression (linéaire, non linéaire) et de classification (Bayes, Knn) et l'analyse de la variance par ANOVA. Il manque cependant des algorithmes que nous aurions souhaité tester comme le SVM (Support Vector Machine) qui est disponible dans une autre *toolbox* : celle de bioinformatique. Il est à noter qu'en plus de l'achat d'une licence *Matlab*, chaque *toolbox* est payante.

Scilab et *GNU Octave* sont les équivalents gratuits et logiciels libres de *Matlab*. Le langage est similaire à celui de *Matlab* et permet aussi la génération de graphiques. Pour Octave, cela aurait été intéressant de l'utiliser dans la continuité de ce qui a été fait pour le prétraitement des données *ARCHISIM* mais il ne propose malheureusement pas directement de bibliothèque de statistique. En revanche, *Scilab* propose des extensions écrites dans des modules externes téléchargeables gratuitement². Il existe entre autre des modules d'analyse de données (*ANN Toolbox*, *Clustering Toolbox*, *etc.*) qui contiennent un nombre intéressant d'algorithmes implémentés³. Mais étant divisé en différents modules, les entrées des fonctions de classification ne sont pas communes ce qui demande donc un prétraitement particulier pour chaque module. De plus, cela demande la recherche des algorithmes voulus dans divers modules et implique donc un coût supplémentaire en temps car la prise en main de chaque module est spécifique (fonctionnement, documentation, *etc.*).

TramineR [55] est un outil d'analyse et de visualisation de données séquentielles basé sur le langage libre *R* [72] (qui sert principalement à l'analyse statistique et à ses représentations graphiques). Il permet d'afficher de nombreuses séquences, de les

1. La bibliothèque statistique peut être achetée à l'adresse www.mathworks.fr/fr/help/stats/

2. atoms.scilab.org

3. Le module correspondant de SciLab est disponible à l'adresse atoms.scilab.org/categories/data_analysis_and_statistics

décrire selon des mesures statistiques, d'appliquer des algorithmes de classification non-supervisée, et d'analyser leurs relations⁴.

WEKA (*Waikato Environment for Knowledge Analysis*) est un logiciel libre *GNU GPL* d'apprentissage automatique. Écrit en Java, il est donc portable sur tout système d'exploitation. Lui aussi contient des outils de visualisation. Il est facile à utiliser notamment grâce à son interface graphique, ses lignes de commandes et son intégration dans une application Java. Il contient un paquetage de prétraitement des données permettant la transformation des fichiers *CSV* en fichiers *ARFF* (*Attribute-Relation File Format*) qui est le format utilisé dans tous les algorithmes de *WEKA*. Souvent mis à jour, il contient le plus grand nombre d'algorithmes parmi les bibliothèques présentées et a une documentation complète à l'aide de sa *Java-Doc*⁵. De plus, il contient une fonction d'analyse en composantes principales (*PCA*). C'est vers cette bibliothèque que notre choix s'est porté. Pour pouvoir utiliser cette bibliothèque, il a été nécessaire de transformer les fichiers de données en fichier *Attribute-Relation File Format* (*ARFF*).

8.7 Conclusion

Dans ce chapitre, nous avons présenté le cadre applicatif de notre méthode qu'est l'évaluation des comportements de conduite de l'architecture de simulation de trafic *ARCHISIM*. Afin d'obtenir des traces de simulation de participant, nous avons utilisé un simulateur de conduite dont nous avons décrit le dispositif ainsi que le scénario de conduite utilisé (*i.e.* le circuit de l'expérimentation).

L'application de notre méthode d'évaluation demande à spécifier des éléments du modèle théorique applicatif. C'est le cas du questionnaire d'annotations de comportement du domaine pour lequel nous avons créé une version spécifique du *Driver Behaviour Questionnaire*. C'est aussi le cas des indicateurs de comportements issus des traces de simulation.

Le protocole expérimental commence par la soumission du *Driver Behaviour Questionnaire*. Le participant effectue ensuite la conduite sur simulateur. Puis les questionnaires de crédibilité comportementale, de propension à l'immersion et de présence lui sont soumis.

Après avoir fait passer l'intégralité de la population de participants, des agents simulés sont placés dans les mêmes situations de conduite. Les données de simulation des participants et des agents sont alors enregistrées pour la classification de traces. Des jeux vidéo des comportements sont enregistrés pour l'évaluation par une autre population d'annotateur afin d'obtenir la classification d'annotations. Les résultats obtenus sur les différentes situations de conduite seront présentés dans le chapitre suivant.

4. Le logiciel *TramineR* est disponible via les paquetages de *R* et aussi à l'adresse mephisto.unige.ch/traminer/install.shtml

5. Le logiciel *WEKA* est disponible à l'adresse weka.sourceforge.net/doc/ et la documentation sur www.cs.waikato.ac.nz/ml/weka/documentation.html

9 Résultats

Dans ce chapitre nous présentons les résultats obtenus par l'application de notre méthode d'évaluation des comportements des agents au travers de l'expérimentation détaillée dans le chapitre précédent.

La première application montre le fonctionnement de notre méthode dans sa globalité avec l'agrégation des agents lors de l'analyse et la calibration de la population d'agents lors d'un cycle de conception. Cette application est décrite dans la *section 9.1* et étudie le comportement dans une situation de dépassement sur une route bidirectionnelle (deux fois une voie).

La deuxième application s'intéresse aux similarités entre questionnaires d'autoévaluation, annotations de comportement, et traces de simulation. Cette application se focalise sur la première partie de la méthode générale. Elle est détaillée dans la *section 9.2*. La situation de conduite étudiée pour cette application est celle du dépassement en trafic ralenti sur route à sens unique.

Enfin dans la *section 9.3*, nous avons appliqué notre méthode sur trois situations de conduite afin d'étudier les facteurs, tels que le genre ou l'âge du conducteur, différenciant l'appréhension de l'environnement virtuel et les comportements de conduite.

9.1 Étude d'une situation de dépassement

Dans cette première étude, nous voulons évaluer la crédibilité des comportements de conduite des agents dans le cadre d'une situation pouvant mener à un possible dépassement de véhicule [35]. Nous appliquons donc notre méthode d'évaluation puis analysons la classification de traces et celle d'annotations. Nous utilisons ensuite ces résultats afin de calibrer la simulation multi-agents.

9.1.1 Situation de conduite

Dans le cadre du simulateur de trafic routier *ARCHISIM* [24] de l'*IFSTTAR* (voir *figure 8.2*), les participants conduisent une voiture sur une route contenant des véhicules simulés.

La phase de conduite (montrée en *figure 9.1*) fournit une situation de conduite sur une route à double sens. Cette phase correspond à environ 1 minute de conduite. L'acteur principal rencontre un véhicule à vitesse réduite sur la voie de droite et plusieurs véhicules en sens inverse sur la voie de gauche. Les distances séparant ces

véhicules vont en augmentant de telle façon qu'il est très difficile de doubler prudemment avant le premier véhicule arrivant en face. Puis, les écarts s'agrandissant, doubler devient de plus en plus facile.

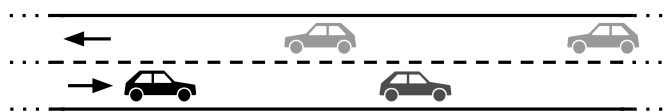


Figure 9.1 – Scénario de dépassement : l'acteur principal (en noir) conduit seul sur une route à deux voies en sens inverse. Un véhicule à faible vitesse (en gris foncé) se trouve sur la voie de l'acteur principal. Aussi, plusieurs véhicules arrivent en face sur la voie opposée. L'espace entre chaque véhicule augmente de façon à ce qu'une manœuvre de dépassement devienne plus facile.

9.1.2 Classification de traces

Les indicateurs sélectionnés pour l'étude de cette situation sont la distance intervéhiculaire, le temps à la collision, le nombre de changements de voies, la vitesse, l'accélération et la distance latérale à l'axiale de la route.

Nous avons comparé la classification des traces avec celle des annotations, leur composition est illustrée dans la figure 9.2. L'indice de Rand entre ces deux classifications est de 0,51.

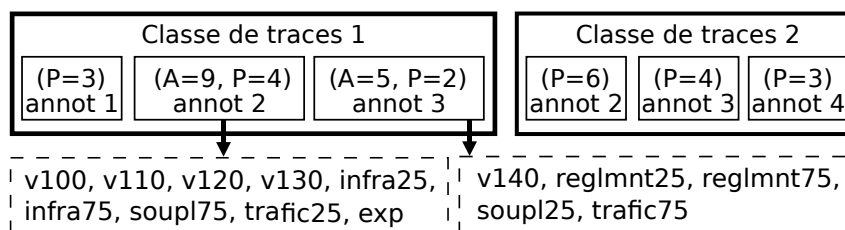


Figure 9.2 – Comparaison des participants entre la classification de traces (dont les classes de traces sont dans des rectangles englobant) et la classification des annotations (dont les classes d'annotations *annot #* sont regroupées avec le nombre de participants *P* et d'agents *A*). Le détail des agents présents dans les classes d'annotations *annot 2* et *annot 3* est indiqué dans les rectangles pointillés (les paramètres correspondants sont détaillés en sous-section 8.4.1).

Il y a 2 classes de comportement dans la classification des traces de simulation :

- la première classe de traces contient 9 participants et l'ensemble des 14 agents. La valeur de l'indicateur *nombre de changements de voies* est 0 signifiant que ces acteurs principaux n'ont pas doublé le véhicule à vitesse réduite et ont préféré le suivre. Étant donné que cette classe est composée à la fois d'agents et de participants, c'est donc une capacité du modèle d'agent à reproduire ce comportement humain qui est de choisir de ne pas dépasser ;
- la seconde classe de traces est uniquement composée de participants. Ces 13 participants doublent le véhicule à vitesse réduite après le deuxième ou le troisième véhicule venant en face. Étant donné qu'il n'y a que des participants

dans cette classe, ce comportement humain peut être considéré comme un manque dans le modèle d'agent : les agents ne peuvent pas choisir de dépasser comme le font certains humains.

9.1.3 Classification d'annotations

La classification d'annotations contient 4 classes d'annotations de comportement (comme schématisé dans la *figure 9.2*) :

- la première classe d'annotations contient 3 participants et aucun agent. Les annotateurs considèrent que ce sont les conducteurs les plus dangereux avec de très hauts scores sur chaque échelle et plus particulièrement sur l'échelle *jugement*. Puisque aucun agent n'a été considéré comme dangereux, et comme le but de ces agents est de reproduire l'intégralité des comportements humains, il y a donc un manque de comportements risqués dans le modèle d'agent ;
- la deuxième classe d'annotations est composée de participants et d'agents. Ils sont annotés comme étant des conducteurs très prudents avec les scores les plus bas sur chaque échelle. L'espace des paramètres de ces agents font qu'ils respectent bien plus le code de la route et adoptent une conduite plus fluide. Le comportement humain normatif peut donc être considéré comme partiellement reproduit ;
- la troisième classe d'annotations est une classe plus petite aussi composée de participants et d'agents. Les annotateurs les ont considérés comme des conducteurs ordinaires avec des scores moyens. Comme pour la classe précédente, le comportement moyen est reproduit dans cette situation ;
- la quatrième classe d'annotations est uniquement composée de participants. Elle a des scores élevés sur les échelles *mémoire* et *jugement*. Ce comportement considéré comme légèrement dangereux n'est pas non plus reproduit par les agents.

9.1.4 Scores

Les scores correspondant à la classification des annotations dans cette expérimentation sont les suivants :

- il n'y a pas de classe contenant uniquement des agents dans cette situation. Le score d'erreur $S_e = \frac{|\mathbb{C}_A|}{|\mathbb{C}_M|+|\mathbb{C}_A|}$ est donc de 0 (car $|\mathbb{C}_A| = 0$),
- le score de manque est $S_m = \frac{|\mathbb{C}_H|}{|\mathbb{C}_M|+|\mathbb{C}_H|} = \frac{1}{1+1} = 1 - S_c = 0,5$. Cela correspond dans la classification de traces à un comportement manquant sur les deux comportements adoptés par les participants,
- le score de capacité est $S_c = \frac{|\mathbb{C}_M|}{|\mathbb{C}_M|+|\mathbb{C}_H|} = \frac{1}{1+1} = 0,5$. Cela correspond dans la classification de traces à un comportement correctement reproduit sur les deux comportements humains.

Le taux de confiance en la classe 1 comme étant une capacité du modèle est de :

$$t_{conf}(C_1) = 1 - \left| \frac{14}{14} - \frac{9}{22} \right| \approx 0,41$$

$t_{conf}(C_1)$ n'étant pas proche de 0, cette classe peut donc être considérée avec confiance comme étant une capacité du modèle. La classe de traces 2 est une classe composée uniquement d'humains, le taux de confiance est alors calculé de la manière suivante :

$$E_H(\mathbb{C}) = \frac{H(\mathbb{C})}{|\mathbb{C}_M| + |\mathbb{C}_H|} = \frac{22}{1 + 1} = 11$$

$$t_{conf}(C_2) = \frac{H(C_2)}{E_H(\mathbb{C})} = \frac{13}{11} \approx 1,18$$

$t_{conf}(C_2)$ est supérieur au seuil défini à 0,5 signifiant que cette classe composée uniquement d'humains peut être traité avec confiance comme étant un comportement manquant réel dans le modèle d'agent.

Ces scores permettent de trouver les nombres de comportements adéquats et inadéquats, et donc de produire l'ensemble des jeux de paramètres qui sont valides : $P \in \mathcal{P}_v = \{v100, v110, v120, v130, infra25, infra75, soupl75, trafic25, exp\} \cup \{v140, reglmnt25, reglmnt75, soupl25, trafic75\}$ pour les deux classes d'annotations mixtes, qui sont dans la même classe de traces.

Il n'y a qu'une classe mixte. Cette dernière est sur-représentée, en effet pour un $\delta = \frac{5}{100}$ (i.e. peu de tolérance aux variations de proportions entre les comportements humains et agents) :

$$A(C_1) > H(C_1) + \delta|C_1| \Leftrightarrow 14 > 9 + \frac{5}{100} \times 23$$

Les scores de représentativité correspondants sont :

- Score de sur-représentativité :

$$S_{sur} = \frac{|\mathbb{C}_{sur}|}{|\mathbb{C}_M|} = \frac{1}{1} = 1$$

- Score de sous-représentativité :

$$S_{sous} = \frac{|\mathbb{C}_{sous}|}{|\mathbb{C}_M|} = \frac{0}{1} = 0$$

- Score de représentativité valide :

$$S_{valide} = \frac{|\mathbb{C}_{valide}|}{|\mathbb{C}_M|} = \frac{0}{1} = 0$$

Les scores de type de classe ont permis de quantifier les erreurs (0), capacités (0,5), et manques (0,5). Les taux de confiance ont permis de s'assurer que la classe mixte est bien une capacité du modèle d'agent ($t_{conf}(C_1) \approx 0,41$) et que la classe d'humains est bien un comportement humain manquant ($t_{conf}(C_2) \approx 1,18$). Les scores de représentativité montrent que la calibration originale sur-représente le comportement humain (correspondant à la classe mixte) dans la population d'agents.

9.1.5 Discussions

Dans la première classe de traces, les indicateurs n'étaient pas capables de distinguer la première classe d'annotations du reste des acteurs principaux. Le score sur l'échelle *judgement* est très élevé et un rejeu des vidéos montre que ces participants ont essayé de doubler à plusieurs reprises sans succès. Pareillement, les participants de la quatrième classe d'annotations n'ont pas été séparés de la seconde classe de traces en une nouvelle classe. Cela peut être dû au fait qu'ils ont essayé de doubler juste après le deuxième véhicule venant en face, ce qui demande alors de se rabattre dans une courte fenêtre temporelle. Cependant, notre algorithme de classification sur ces indicateurs n'a pas pu détecter cette différence.

Nous avons une similarité significative entre la classification des annotations et celle des traces, signifiant que nous sommes capables de classifier nos données de simulation en classes de traces de comportement haut niveau qui ont un sens en termes d'annotations de conduite. Néanmoins, les deux classifications ne sont pas identiques en termes de composition de classes ni en termes de nombre de classes.

Cela peut être dû au nombre restreint d'annotateurs. Ce problème peut aussi provenir de l'algorithme de classification. Cependant, nous avons tenté d'appliquer d'autres algorithmes tels qu'Espérance-Maximisation [41] et la Classification Ascendante Hiérarchique [102] sans obtenir de meilleurs résultats.

Une piste peut être l'utilisation d'algorithmes basés sur les séries temporelles. Les experts du domaine doivent également dans ce cas être consultés pour comprendre quels indicateurs manquants pourraient expliquer ces différences.

Enfin, la classification des comportements issus des traces de simulation est faite sur des indicateurs bruités pour les participants de par l'utilisation d'un dispositif physique, tandis que les agents simulés ont des indicateurs plus lisses. Cette différence peut perturber le résultat de la classification. De façon à ne pas prendre en compte ces différences non significatives en termes de comportement haut niveau, une solution pourrait être de travailler sur des indicateurs lissés.

9.1.6 Cycle de conception

Nous avons trouvé que seule une partie des comportements humains étaient reproduits, le score de capacité étant de $\frac{1}{2}$ et le score de manque étant de $\frac{1}{2}$). Dans ce cas, il existe deux possibilités pour l'utilisateur de la simulation : soit entrer dans un cycle d'exploration de l'espace des paramètres, soit calibrer le système en utilisant la première analyse. Nous nous concentrons sur la seconde possibilité.

Une fois les scores calculés, il est possible de calibrer une nouvelle population d'agents [34]. Nous avons vu que les paramètres valides sont $P \in \mathcal{P}_v = v100, v110, v120, v130, infra25, infra75, soupl75, trafic25, exp, v140, reglmnt25, reglmnt75, soupl25, trafic75$, et qu'il y avait plusieurs comportements manquants. Par conséquent, afin d'obtenir des agents reproduisant le comportement humain, nous calibrons les nouveaux ensembles de paramètres des agents en les choisissant parmi ceux valides dans la première expérimentation (*i.e.* les classes mixtes) et en raffinant les proportions grâce aux classes d'annotations.

La première classe contient 9 traces d'agents et 4 traces de participants, tandis que la seconde classe contient 5 traces d'agents et 2 traces de participants. Nous obtenons donc, en calculant leur probabilité d'apparition dans la prochaine calibration : $p(v100) = p(v110) = p(v120) = p(v130) = p(infra25) = p(infra75) = p(soupl75) = p(traffic25) = p(exp) = \frac{4}{6}$ et $p(v140) = p(reglmt25) = p(reglmt75) = p(soupl25) = p(traffic75) = \frac{2}{5}$.

De cette manière, les agents représentent stochastiquement la densité de comportements humains dans la simulation. Il est à noter que cette nouvelle calibration ne change pas les scores de capacités et de manques, qui sont basés sur la proportion de comportements humains correctement reproduits. Cependant, il réduit dans tous les cas le score d'erreur à 0 en n'utilisant que des ensembles de paramètres valides, et améliore les scores de représentativité en choisissant des ensembles de paramètres pour les agents selon les comportements observés chez les humains.

9.1.7 Conclusion

Cette section a présenté les résultats de la méthode semi-automatique d'analyse des comportements d'agents et de calibration de simulation multi-agents participative basée sur la combinaison de classifications non-supervisées de traces de simulation et d'annotations par des participants.

L'expérimentation permet de définir un ensemble initial de traces valides qui servent de points de référence pour la calibration du modèle multi-agents. La calibration des paramètres du modèle suit une approche itérative. À chaque itération, nous obtenons les manques et les erreurs du modèle afin de raffiner l'espace des paramètres. Nous générons de nouveaux agents que nous agrégeons aux classes précédentes et nous calculons de nouveaux scores pour l'itération suivante.

Dans le cadre d'une manœuvre de dépassement, cette expérimentation a montré qu'il est ainsi possible d'obtenir des comportements haut niveau à partir des traces et des annotations via notre abstraction. En conséquence, les paramètres des agents peuvent être correctement associés à des catégories de comportement.

Le type de classe composé uniquement d'agents n'apparaît pas dans cette situation, le score d'erreur est donc de 0. Nous pouvons donc considérer que dans cette situation, étant donné que chaque comportement d'agent a été agrégé à un comportement de participant, aucun comportement d'agent n'est inexact (*i.e.* est une erreur au sens de dissimilaire des comportements humains).

Une des originalités de ce processus de calibration est le contexte des simulations participatives. La fonction cible du processus de calibration n'est pas comme habituellement [48] au niveau macroscopique mais à un niveau individuel. La réalité virtuelle requiert que chaque agent adopte des comportements crédibles, *i.e.* des comportements pouvant être produits par des humains. Le but est alors, en itérant l'analyse de la classification, de raffiner les capacités des agents à reproduire des comportements humains, tout en réduisant les manques et en supprimant les erreurs. Dans ce contexte, nous retirons premièrement les ensembles de paramètres qui ne produisent pas de comportements valides (*i.e.* les erreurs). Nous calibrons

ensuite les proportions d'agents avec les ensembles de paramètres restants (*i.e.* les capacités) selon les données des participants humains. Un unique cycle de notre méthode assure que les comportements valides sont détectés et qu'ils sont produits en des proportions correctes, nonobstant les comportements manquants.

9.2 Étude d'une situation de trafic ralenti

Dans cette section nous nous intéressons à la conduite en trafic ralenti sur route à sens unique [32]. Nous nous focalisons sur la première partie de la méthode, hors calibration, afin de comparer les résultats entre autoévaluation des participants de leur conduite effectuée et annotation de cette conduite par une autre population. Il est à noter que dans cette étude nous avons effectué la classification sur l'ensemble des acteurs (*i.e.* sans exploiter la méthode d'agrégation).

Le questionnaire d'autoévaluation n'étant rempli que par les participants, la comparaison ne peut se faire que sur les annotations faites sur les comportements des participants (*i.e.* les annotations des agents sont exclues). La soumission du questionnaire de conduite *DBQ* (dans sa version originale traduite en français) permet de vérifier la cohérence de notre annotation basée sur ce questionnaire.

Nous avons comparé la classification des habitudes de conduite (via le *DBQ*), celle des annotations et celle des traces de simulations. La composition des classifications est illustrée par deux graphiques (*figures 9.4 et 9.5*). Les agents sont représentés avec des rectangles et sont nommés *a#*. Les participants sont représentés avec des ellipses et sont nommés *s#*. Les acteurs principaux d'une classification sont regroupés par classe dans des rectangles contenant le numéro de la classe en haut. Les acteurs principaux de l'autre classification sont regroupés par classe selon des couleurs et le numéro de classe et inscrit juste en dessous de l'identifiant des acteurs principaux.

9.2.1 Situation de conduite

Dans cette phase, schématisée dans la *figure 9.3*, l'acteur principal commence la conduite sur une route à sens unique ayant deux voies contenant un trafic fluide. L'acteur principal peut alors observer plus loin sur la route en ligne droite un véhicule roulant à vitesse réduite sur la file de droite. Ce véhicule engendre l'adoption d'un comportement de dépassement chez les véhicules simulés qui sont devant l'acteur principal.

Dans cette situation, les experts du trafic routier ont choisi les indicateurs suivants : des variables haut niveau telles que le temps avant collision (sous hypothèse que les vitesses des véhicules sont constantes), le temps inter-véhiculaire, le *jerk* (dérivée de l'accélération par rapport au temps) et le nombre de changements de voie ; ainsi que des variables de bas niveau telles que le déplacement latéral et l'angle par rapport à l'axiale, la vitesse et l'accélération.

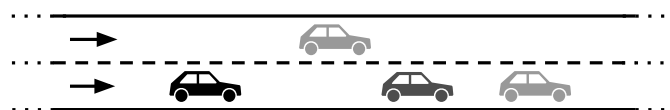


Figure 9.3 – Scénario de dépassement : l’acteur principal (en noir) conduit au sein d’un trafic fluide sur une route à sens unique ayant deux voies. Ensuite, un véhicule roulant à faible vitesse sur la voie de droite (en gris foncé) perturbe le trafic.

9.2.2 Annotations et habitudes de comportement

Nous voulons vérifier si les annotations effectuées sur les participants humains correspondent bien à l’auto-évaluation des participants de leurs habitudes de comportement de conduite par le *DBQ*. Pour cela, nous comparons les résultats obtenus sur le questionnaire d’habitude de conduite *DBQ* et sur les annotations. La *figure 9.4* montre la similarité de la classification des habitudes de conduite (avec des rectangles) et de celle des annotations (regroupé par couleur). Le questionnaire *DBQ* n’étant rempli que par les participants, nous n’affichons pas ici les annotations faites sur les agents.

La classification des habitudes de conduite contient 3 classes, ce qui est proche des 2 classes obtenues pour la classification des annotations. La classe d’habitudes 1 contient presque tous les participants de la classe d’annotations 1 (sauf *s12*). La classe d’habitudes 2 est composée uniquement de participants de la classe d’annotations 2. De même, la classe d’habitudes 3 est principalement composée des participants de la classe d’annotations 2 (sauf *s17*). L’indice de Rand est de 0,71 et l’indice de Rand ajusté de 0,42. Cela signifie donc que notre questionnaire d’annotations basé sur les échelles du questionnaire *DBQ* est cohérent avec les habitudes de comportement de conduite.

9.2.3 Classes de traces et classes d’annotations

Nous voulons analyser les comportements de conduite adoptés par les agents par rapport à ceux adoptés par les participants. Pour cela, nous utilisons la classification de comportement effectuée à partir des traces de conduite. De plus, nous voulons expliciter l’ensemble des comportements de conduite des acteurs principaux à l’aide de la classification des annotations.

Avec la classification de comportement des acteurs principaux (issue des traces), nous avons vu qu’il est possible d’analyser combien de comportements humains sont reproduits par les agents, combien de comportements humains ne sont pas adoptés par les agents, et combien de comportements d’agents ne sont pas adoptés par les participants. Il est également possible de rendre explicite ces comportements par la similarité avec les classes d’annotations si elles sont pertinentes.

La *figure 9.5* présente la classification des comportements (dans des rectangles) et la classification des annotations (regroupé par couleur). Le nombre de classes est similaire dans les deux classifications (3 comportements et 2 types d’annotations). L’indice de Rand est de 0,59 et l’indice de Rand ajusté de 0,17.

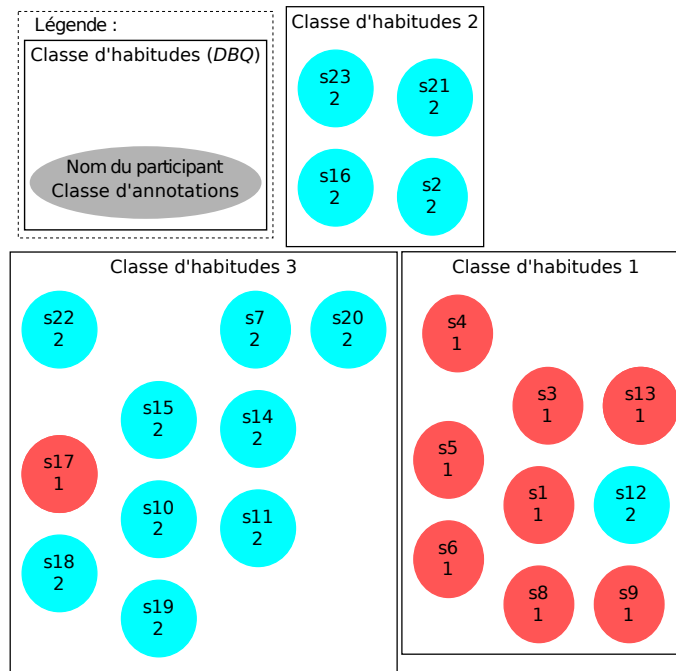


Figure 9.4 – Comparaison des participants entre la classification des habitudes de comportement de conduite par le *DBQ* (dont les classes sont dans des rectangles) et la classification des annotations (dont les classes sont regroupées par couleur).

- La classe de traces 1 contient un participant et presque tous les agents (sauf *a5*). La plupart des acteurs principaux de cette classe de traces sont dans la même classe d'annotations 1. Les acteurs principaux de la classe de traces 1 ont donc adopté un comportement de conduite similaire et ont été annotés de la même façon, c'est-à-dire : le plus haut score sur la question du contrôle perçu et les scores les plus bas sur les autres questions relatives à des problèmes de conduite. En conséquence, ils sont jugés comme étant des conducteurs prudents.
- La classe de traces 2 est composée uniquement de participants qui sont mélangés parmi les deux classes d'annotations. Ces participants n'ont pas doublé le véhicule roulant à faible vitesse sur la voie de droite : certains n'ont pas rejoint la zone de trafic perturbée à cause d'une vitesse faible notamment au démarrage, d'autres ont décidé de ne pas dépasser ce véhicule mais d'adopter un comportement de suivi en file. Ce comportement est caractérisé par une vitesse plus faible et aucun changement de voie.
- La classe de traces 3 est principalement composée de participants (et de l'agent *a5*). Ces participants sont largement annotés dans la même classe d'annotations 2, qui a le plus bas score sur la question du contrôle perçu et les scores les plus hauts sur les autres questions de problèmes de conduite, signifiant qu'ils sont jugés comme étant des conducteurs dangereux.

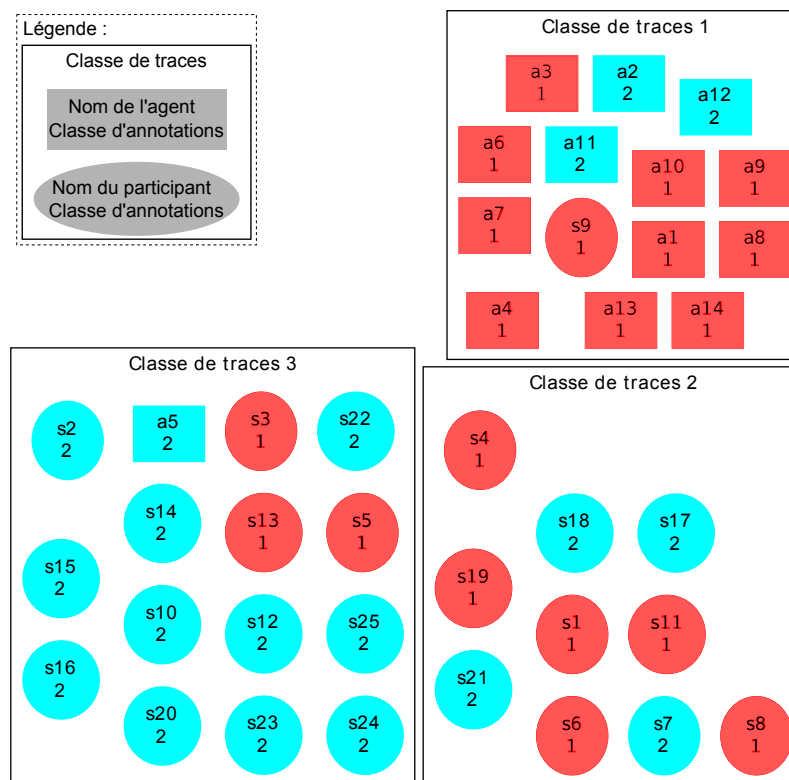


Figure 9.5 – Comparaison des acteurs principaux entre la classification des comportements issus des traces de simulations (dont les acteurs principaux sont regroupés par classe dans des rectangles englobants) et celle des comportements issus des annotations (dont les acteurs principaux sont regroupés par classe par couleur).

9.2.4 Classes de traces et classes d’habitudes

Il n’y a pas de similarité significative entre la classification des habitudes de comportement (*DBQ*) et la classification des comportements issus des traces comme nous pouvons le voir sur la *figure 9.6*. Chacune des classes de la classification des comportements contient un mélange des 3 classes de la classification du *DBQ*. L’indice *RI* est plus faible que pour la classification d’annotations (0,48 contre 0,71) tout comme l’indice *ARI* (0,07 contre 0,42). La classification des comportements observés (via les traces) ne correspond donc pas à la classification des habitudes de comportement de conduite.

Ceci est probablement dû au caractère général du questionnaire *DBQ*. Les habitudes du conducteur peuvent différer des comportements adoptés. Le *DBQ* étant rempli par le conducteur lui-même, le biais d’introspection peut également expliquer en partie ces différences. Ajouté au problème de l’impossibilité d’appliquer des questionnaires aux agents, ceci justifie le recours aux annotations.

Plusieurs précautions doivent être prises cependant dans l’utilisation des questionnaires. Les 3 échelles traitant du risque montrent parfois des incohérences : 8% des participants ont des scores élevés à la fois sur l’échelle *aucun risque* et sur l’échelle *risque certain* mais des scores faibles sur l’échelle *risque possible*.

Un autre problème est que le même type de participant en termes de classes du *DBQ* peut adopter différents comportements dans la même situation, induisant des traces différentes. Pareillement, le même comportement peut être adopté par différents types de participants. C'est problématique pour l'analyse des similarités entre la classification des habitudes de comportement et la classification des traces de comportement avec l'indice de Rand ajusté. Une solution pourrait être de fusionner les sous-ensembles pour lesquels tous les acteurs principaux sont ensemble dans les deux classifications.

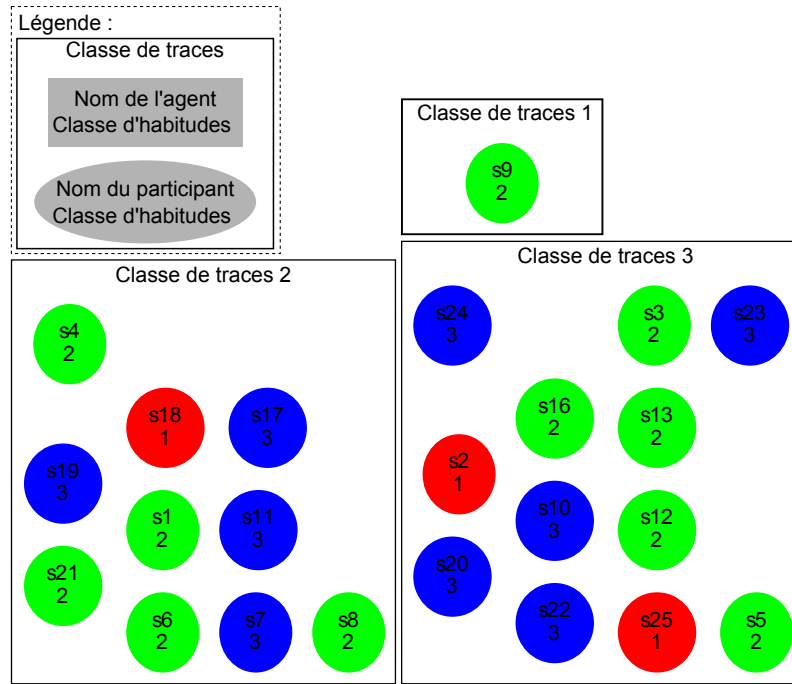


Figure 9.6 – Comparaison des acteurs principaux entre la classification des comportements issus des traces de simulations (dont les acteurs principaux sont regroupés par classe dans des rectangles englobants) et celle des comportements issus des habitudes de conduite autoévaluées (*DBQ*) (dont les acteurs principaux sont regroupés par classe par couleur).

9.2.5 Scores

Les scores de type de classe sont les suivants pour la classification de traces :

- Puisqu'il n'y a pas de classe composée uniquement d'agents, le score d'erreur vaut :

$$S_e = \frac{|\mathcal{C}_A|}{|\mathcal{C}_M| + |\mathcal{C}_A|} = 0$$

- Le score de capacité montre que les agents sont capables de reproduire deux comportements (*i.e.* les classes mixtes) sur les trois comportements exhibés par les participants;

$$S_c = \frac{|\mathcal{C}_M|}{|\mathcal{C}_M| + |\mathcal{C}_H|} = \frac{2}{1 + 2} = \frac{2}{3}$$

- Le score de manque montre qu'il manque un comportement sur les trois comportements humains.

$$S_m = \frac{|\mathbb{C}_H|}{|\mathbb{C}_M| + |\mathbb{C}_H|} = \frac{1}{1 + 2} = \frac{1}{3}$$

Les taux de confiance en chaque classe de la classification de traces sont :

- Pour la classe de traces mixte 1, sachant que le taux d'humain vaut $t_H(C_1) = \frac{1}{25}$ et que le taux d'agents vaut $t_A(C_1) = \frac{13}{14}$, le taux de confiance est de :

$$t_{conf}(C_1) = 1 - |t_A(C_1) - t_H(C_1)| = 1 - \left| \frac{13}{14} - \frac{1}{25} \right| \approx 0,03$$

Le taux de confiance est proche de 0. En conséquence, la confiance dans le fait que cette classe soit une capacité réelle du modèle à reproduire un comportement humain est très faible.

- Pour la classe de traces d'humains 2 :

$$E_H(\mathbb{C}) = \frac{H(\mathbb{C})}{|\mathbb{C}_M| + |\mathbb{C}_H|} = \frac{25}{2 + 1}$$

$$t_{conf}(C_2) = \frac{H(C_2)}{E_H(\mathbb{C})} = \frac{14}{\frac{25}{3}} \approx 1,68$$

Ce taux est supérieur à 0,5 (taux considéré comme suffisamment éloigné de 0) signifiant que cette classe d'humains correspond réellement à un comportement humain, qui de plus est manquant dans le modèle d'agent.

- Pour la classe de traces 3, sachant que le taux d'humain vaut $t_H(C_3) = \frac{10}{25}$ et que le taux d'agents vaut $t_A(C_3) = \frac{1}{14}$, le taux de confiance est de :

$$t_{conf}(C_3) = 1 - |t_A(C_3) - t_H(C_3)| = 1 - \left| \frac{1}{14} - \frac{10}{25} \right| \approx 0,67$$

Le taux de confiance n'est pas proche de 0. En conséquence, cette classe peut être considérée avec confiance comme étant une capacité réelle du modèle à reproduire un comportement humain.

Afin de calculer les scores de représentativité, nous devons définir la représentation des classes mixtes de la classification de traces. La classe de traces 1 est sur-représentée et la classe de traces 3 est sous-représentée :

$$A(C_1) > H(C_1) + \delta|C_1| \Leftrightarrow 13 > 1 + \frac{5}{100} \times 14$$

$$A(C_3) < H(C_3) - \delta|C_3| \Leftrightarrow 1 < 10 - \frac{5}{100} \times 11$$

Il y a donc une classe sur-représentée, une classe sous-représentée, et il n'y a pas de classe dont la représentation est valide. Les scores de représentativité sont donc les suivant :

- Score de sur-représentativité :

$$S_{sur} = \frac{|\mathbb{C}_{sur}|}{|\mathbb{C}_M|} = \frac{1}{2}$$

- Score de sous-représentativité :

$$S_{sous} = \frac{|C_{sous}|}{|C_M|} = \frac{1}{2}$$

- Score de représentativité valide :

$$S_{valide} = \frac{|C_{valide}|}{|C_M|} = \frac{0}{2}$$

9.2.6 Conclusion

Nous avons une similarité significative entre la classification des habitudes de conduite par le *DBQ* et celle des annotations (basées sur le *DBQ*), signifiant que la classification d'annotations a un sens en termes de comportement de conduite.

De plus, il y a aussi une similarité significative entre la classification des annotations et celle des comportements issus des traces, signifiant que nous sommes capable de classifier les traces en des classes de comportement haut niveau ayant un sens en termes de comportement de conduite sur cette situation de trafic ralenti.

Néanmoins, les deux classifications ne sont pas identiques en nombre de classes et en termes de composition. Tout comme pour l'étude de dépassement en *section 9.1*, ces différences peuvent être dues au faible nombre d'annotateurs, à des indicateurs manquants pour l'analyse des traces de simulation, ou à l'algorithme de classification.

Dans la comparaison des annotations et des comportements, un agent était au sein d'une classe principalement composée de participants. Cela veut-il dire que cet agent est capable de simuler le comportement de conduite de cette classe qui est *prudent*? Si oui, nous pouvons alors considérer que ce comportement *prudent* est une capacité des agents. Pour vérifier cette hypothèse, nous aurions besoin d'un test spécifique dans lequel l'ensemble des paramètres qui a été utilisé pour l'agent *a5* serait confronté à différentes situations, et comparé avec les annotations et les traces de participants *prudents*. Cependant, le taux de confiance d'environ 0,03 nous indique que la confiance dans cette capacité est très faible. Dans un cycle de conception, ce comportement est considéré comme sous-représenté. La calibration tentera alors d'ajuster la représentativité de ce comportement en augmentant la population d'agents ayant les paramètres de cet agent afin d'atteindre une proportion similaire à celle des participants (*i.e.* exhiber souvent ce comportement).

Réciproquement, un participant était dans une classe composée principalement d'agents jugés comme dangereux pour la majorité d'entre eux. Cela demande une étude plus approfondie pour comprendre la spécificité dans le comportement de conduite de ce participant qui était similaire aux comportements des agents. Néanmoins, le taux de confiance est moyen avec une valeur d'environ 0,67. Il y a donc une confiance mitigée sur la capacité du modèle. Lors du cycle de conception, ce comportement est considéré comme sur-représenté. Il faut alors diminuer la proportion d'agent ayant ces paramètres afin d'obtenir une population d'agent adoptant ce comportement aussi rarement que les humains.

La classe de traces 2 ne contient aucun agent, signifiant que le modèle d'agent n'est pas capable de reproduire ce comportement de conduite humain (*i.e.* ce comportement est un manque dans le modèle d'agent) qui est de décider de ne pas dépasser et d'adopter un comportement de suivi en file. Le taux de confiance d'environ 1,68 confirme que cette classe de participants correspond bien à un comportement humain manquant.

Il est à noter que dans cette étude également le type de classe composée uniquement d'agents n'apparaît pas (*i.e.* comportement erroné). D'où un score d'erreur de 0 : les agents n'adoptent donc pas de comportements erronés dans cette situation de trafic ralenti.

9.3 Étude des facteurs de différence dans la conduite

Pour cette étude, nous étudions un ensemble de trois situations de conduite, et proposons une analyse de genre menée à l'aide de notre méthode [38]. Nous étendons également la recherche aux autres facteurs susceptibles d'expliquer des différences de conduite ou d'appréhension de l'environnement virtuel.

Nous avons utilisé 3 situations pour lesquelles le participant est en interaction avec d'autres véhicules simulés. Chacune des phases correspond à 1 minute de conduite.

Les deux premières situations sont celles présentées dans les deux expérimentations précédentes. Des différences sont cependant présentes dans les résultats par rapport aux études précédentes à cause d'un découpage des phases légèrement différent.

Dans la première situation (celle présentée dans la *figure 9.3* en *section 9.2*), le véhicule conduit par l'acteur principal est sur une route à sens unique à deux voies et un véhicule à vitesse réduite est sur la voie de droite.

La seconde situation (celle présentée dans la *figure 9.1* en *section 9.1*) met le conducteur principal dans une situation similaire mais sur une route à deux voies en sens opposés avec des véhicules arrivant sur la voie de gauche.

La troisième situation (décrite dans la *figure 9.7*) se situe sur une autoroute à deux fois deux voies. L'acteur principal se retrouve alors confronté à une situation de trafic congestionné dû à plusieurs poids lourds.

9.3.1 Annotations et traces

Avec la classification de traces, nous sommes capables d'analyser les comportements reproduits (et non reproduits) par les agents, et notamment en se focalisant sur un ou plusieurs facteurs. Nous avons choisi de nous concentrer sur le genre¹ du

1. Le sexe d'une personne est biologique et en général déterminé à sa naissance; Le genre est l'ensemble des constructions sociales attribuées aux personnes d'un sexe.

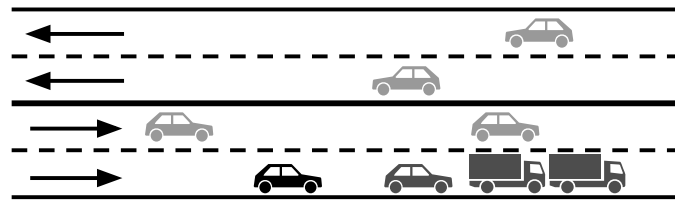


Figure 9.7 – Scénario de trafic ralenti sur autoroute : le véhicule principal roule sur une autoroute à deux fois deux voies quand il rencontre une zone de congestion due à plusieurs véhicules poids lourds.

conducteur, pour tester un ensemble d’hypothèses liées à cette question, et notamment : les comportements de conduite masculins et féminins sont-ils reproduits de la même façon par les agents ? Les hommes et les femmes ont-ils le même rapport au simulateur ?

La figure 9.8 montre les classifications en distinguant les classes de traces de celles des annotations. L’information sur le genre des participants a été ajoutée : m pour les hommes et f pour les femmes. L’indice de rand est de 0,59 et sa version ajustée de 0,18, les scores sont donc légèrement inférieurs de ceux avec le découpage précédent. La classification de traces contient 3 classes :

- la classe de traces 1 contient un homme ($s9(m)$) et presque tous les agents. La plupart sont annotés de la même façon (*i.e.* dans la classe d’annotations 1), c’est-à-dire des scores bas sur toutes les échelles. Cette classe correspond donc aux conducteurs jugés comme prudents. Ils ont dépassé très tôt le véhicule lent qui était encore loin, et ce tout en respectant le code de la route et notamment la première limitation de vitesse de $90km/h$ puis le changement à $70km/h$ dû aux virages annoncés.
- la classe de traces 2 est uniquement composée de participants. Elle contient 6 hommes et 4 femmes répartis entre les classes d’annotations. Ces participants n’ont pas dépassé le véhicule lent ou seulement à la toute fin.
- la classe de traces 3 est principalement composée de participants (6 hommes, 6 femmes, et l’agent $a5$). Ces participants sont largement annotés dans la même classe d’annotations 2, qui a les plus hauts scores, signifiant ainsi que ce sont des conducteurs considérés comme dangereux. Il est à noter que l’agent $a5$ n’est pas regroupé avec les autres agents et qu’il est jugé comme dangereux. Cela est dû au fait qu’il a changé de voie et doublé au dernier moment mais il ne s’est pas rabattu et est donc resté sur la voie de gauche. De plus, il n’a pas respecté la limitation de vitesse avec une vitesse moyenne de $10km/h$ au-dessus de la limite.

9.3.2 Composition sur trois situations de conduite

La figure 9.9 montre la composition des classifications pour chaque situation de conduite. Dans cette représentation, nous avons regroupé les acteurs principaux qui sont classifiés ensemble dans toutes les classifications (*i.e.* dans les trois situations de conduite). La sous-section précédente a présenté en détail la première situation, nous détaillons ici les deux autres situations (au centre et à droite de la figure 9.9).



Figure 9.8 – Comparaison pour la première situation de conduite des acteurs principaux avec le genre pour les participants, regroupés par rectangle pour la classification de traces, et regroupés par couleur pour la classification d’annotations. Les participants (nommés de $s1$ à $s23$) sont représentés par des ovoïdes avec le genre (f pour les femmes et m pour les hommes). Les agents (nommés de $a1$ à $a14$) quant à eux sont représentés par des rectangles.

Pour la seconde situation de conduite, les experts du domaine nous ont indiqué de se concentrer sur le nombre de changements de voie, le numéro de voie, et le temps à la collision. La similitude entre la classification de traces et celle d’annotations donne un indice de Rand de 0,74 et un indice de Rand ajusté de 0,39.

- la *classe* 2.1 contient 8 hommes et 6 femmes. Tous ces participants doublent le véhicule à vitesse réduite après le deuxième ou le troisième véhicule arrivant en face.
- la *classe* 2.2 n’est composée que d’une femme qui a essayé de doubler après le troisième véhicule arrivant en face. Elle a changé de voie mais a reconsidéré la situation en voyant arriver un autre véhicule en face, et décidé alors de ne pas aller au bout de sa manœuvre.
- la *classe* 2.3 est une classe composée uniquement d’agents qui n’ont pas essayé de doubler le véhicule ralenti et qui ont décidé de le suivre avec une vitesse et une distance intervéhiculaire constante.
- la *classe* 2.4 est composée de 5 hommes et 3 femmes qui n’ont pas non plus essayé de dépasser ou seulement après le quatrième véhicule arrivant en face.
- la *classe* 2.5 contient 2 agents dont les comportements sont proches des autres agents (indiscernables en rejeu video).

Pour la troisième situation de conduite, les indicateurs pertinents sont le nombre de changements de voie, le numéro de voie, et la vitesse moyenne. L’indice de rand

entre la première situation de conduite (respectivement la deuxième) est de 0,50 (respectivement 0,54) et la version ajustée de 0,02 (respectivement 0,09).

- la *classe* 3.1 contient 5 agents, 11 hommes, et 7 femmes. Elle contient deux comportements principaux : ne pas dépasser, et dépasser en se rabattant.
- la *classe* 3.2 est composée de 9 agents, 2 hommes et 3 femmes. Les acteurs principaux de cette classe ont de fortes accélérations et décélérations. Ils restent sur la voie de gauche et dépasse. Le participant *s6* ne dépasse pas mais son véhicule a d'importantes oscillations latérales qui ont pu être mal interprétées (*i.e.* comme étant des dépassements) par l'algorithme de classification.

9.3.3 Discussion

Pour la première situation de conduite, la classe de traces 3 contient le même nombre d'hommes que de femmes, signifiant que ce comportement est commun aux deux genres. Puisqu'un agent est dans cette classe, le modèle est capable de simuler ce comportement de conduite jugé comme dangereux. Cependant le taux de confiance est faible et cette population d'agents devrait être augmentée.

À l'inverse, un homme ($s9(M)$) est dans une classe composée d'agents jugés majoritairement comme prudents. Cela requière une étude complémentaire pour comprendre ce qui était spécifique dans le comportement de conduite de ce participant par rapport aux autres conducteurs humains et ce qui était similaire aux comportements des agents. Une explication possible est que les comportements des agents sont plus proche des conducteurs masculins, ou qu'un comportement masculin spécifique est produit par un trop grand nombre d'agents. Une étude sur une population plus large de conducteurs serait nécessaire pour conclure. Cependant, de par la sur-représentativité des agents adoptant ce comportement rare chez les participants, le concepteur de la simulation devrait baisser la proportion de cet agent dans un cycle de conception.

La classe de traces 2 ne contient aucun agent, signifiant ainsi que le modèle d'agent n'est pas capable de reproduire ce comportement de conduite humain (*i.e.* ce comportement est un manque dans le modèle d'agent).

Dans la seconde situation de conduite les agents ne dépassent pas le véhicule à faible vitesse mais le suivent à vitesse constante et à distance intervéhiculaire constante. Le fait qu'aucun participant ne soit dans ces classes agents (*i.e.* aucun participant n'adopte ces comportements) montre que ces comportements sont in-exacts (*i.e.* erronés) et devraient être examinés (ou supprimés dans le cas d'un cycle de conception).

Deux classes contiennent à la fois des hommes et des femmes signifiant que ces comportements sont communs aux deux genres.

La classification contient une classe singleton composée d'une unique femme ($s3(F)$). Si la taille de cette classe n'augmente pas avec une population plus large, cela voudrait dire que le comportement associé est un *outlier* (*i.e.* une donnée aberrante). Dans le cas contraire, *i.e.* si le comportement de cette participante se regroupe avec d'autres comportements adoptés uniquement par des femmes, cela voudrait dire

que ce comportement est spécifiquement féminin et que les agents ne sont pas en capacité de le reproduire.

Dans la troisième situation, les deux classes contiennent des agents, des hommes et des femmes. Cela indique que les agents sont capables de reproduire le comportement adopté sur autoroute. Cela peut être dû au style de conduite spécifique sur autoroute qui est plus facilement reproductible.

9.3.4 Corrélations avec les questionnaires

Nous avons calculé les corrélations de Pearson ($df = 21$) entre les sous-échelles du *Driver Behavior Questionnaire*, de l'*Immersive Tendencies Questionnaires*, du *Game Engagement Questionnaire*, et du *Presence Questionnaire*. Les résultats significatifs (avec $p < 0,05$ comme seuil [51]) sont résumés dans les *tables 9.1 et 9.2* selon l'âge et le genre.

	Violation	Jugement	Contrôle
Genre	-0,45 ($p < 0,05$)	-	-
Âge	-0,42 ($p < 0,05$)	-0,47 ($p < 0,05$)	-0,56 ($p < 0,01$)

Table 9.1 – Corrélations significatives du *Driver Behavior Questionnaire* en fonction du genre et de l'âge (Pearson, $df = 21$).

	PQ - Perf.	PQ - HMI	ITQ - Émotion	ITQ - Implication
Genre	0,52 ($p < 0,02$)	-	-0,47 ($p < 0,05$)	
Âge	-	0,45 ($p < 0,05$)	-	-0,44 ($p < 0,05$)

Table 9.2 – Corrélations significatives des questionnaires en fonction du genre et de l'âge (Pearson, $df = 21$) sur les échelles : *autoévaluation de la performance* et *qualité de l'interface* du questionnaire de présence, et *émotion* et *implication* du questionnaire de tendance à l'immersion.

Nous avons trouvé des corrélations peu significatives entre le genre et les résultats des questionnaires. Les hommes ont des scores plus bas sur l'échelle *violation involontaire* du *DBQ*, des scores plus haut sur la sous-échelle d'*auto-évaluation des performances* du questionnaire de présence, et des scores plus bas sur l'échelle *émotion* de l'*ITQ*.

Concernant les scores résultant d'autoévaluations, il peut y avoir un biais dû au genre puisque les trois sous-échelles sont typiquement chargées de représentations sociales. Les hommes tendent à exposer moins d'émotions, et à surestimer leurs performances dans les tâches (voir *e.g.* [106]).

Il est intéressant de noter que nous avons trouvé plus de corrélations entre l'âge et les résultats des questionnaires, montrant que les générations plus jeunes ont moins de problèmes avec les interfaces homme-machine et dans l'implication générale dans

la simulation. Ces corrélations sont aussi présentes dans les sous-échelles du *DBQ* : jugement, contrôle et violation.

Plusieurs hypothèses n'ont pas été vérifiées. Nous n'avons pas trouvé de corrélation entre le genre et l'habitude aux jeux vidéo ou l'implication dans les jeux vidéo. Concernant les comportements des agents virtuels, il n'y a pas non plus de corrélation entre le genre et les scores de crédibilité des agents virtuels issus des questionnaires *QGC* et *QSC*. La seule corrélation trouvée se situe entre l'âge du conducteur et les scores de crédibilité des conducteurs virtuels.

Il est à noter que les comportements d'agents sont jugés de la même manière par les annotateurs sur les rejeux à la première personne du véhicule principal que par les participants sur le trafic alentour (à la troisième personne) lors du rejeu de leur propre conduite via le Questionnaire Spécifique de Comportement et le Questionnaire Général de Comportement.

9.3.5 Conclusion

Bien que les différences de genre dans les comportements de conduite sont bien connues [16], notre étude - via une classification non-supervisée - n'a pas permis de fortement différencier les traces des participants masculins de celles des participants féminins.

Le questionnaire de présence autoévalué est corrélé au genre sur la sous-échelle *performance*, mais aussi au *BDQ* sur les sous-échelles *violations involontaires* et *contrôle*. Cela peut indiquer que les erreurs de *contrôle* faites en situation de conduite réelle sont aussi faites en simulateur. L'assurance joue aussi un rôle dans l'autoévaluation, puisque les annotateurs n'ont pas observé de différence forte non plus.

Bien que les femmes tendent à moins faire confiance aux systèmes automatisés que les hommes [29], nous n'avons pas trouvé de différences significatives dans leur expérience avec le simulateur, et cela ni pour les échelles de présence, ni pour la crédibilité comportementale accordée aux agents virtuels. À travers nos résultats, l'âge est un facteur bien plus fort pour différencier la tendance à l'immersion, la crédibilité, et le comportement général.

Une de nos hypothèses était que l'expérience des participants avec les jeux vidéo impacte les résultats de notre étude. Une fois de plus, aucune corrélation significative n'a été trouvée à partir de nos données. Cependant, l'âge est bien un facteur. Une explication possible est que la présence de la technologie est plus élevée chez les jeunes générations, et puisque l'écart entre les genres diminue, l'expérience du jeu vidéo n'est plus discriminante. Cela est contre-intuitif, notamment par rapport à des études antérieures montrant que le domaine du jeu vidéo n'a pas encore dépassé ce fossé entre les genres [76] (habitudes, types de jeux, et plateformes sont encore largement genrés), bien qu'il ait été montré que jouer à certains types de jeux réduit certaines différences entre les genres comme la cognition spatiale [49].

Concernant la validité des simulateurs de conduite, nous avons trouvé que dans le cas d'*ARCHISIM* il n'y a pas de biais significatif lié au genre ni pour la partie conduite, ni pour le comportement des conducteurs virtuels, bien que des pistes

d'études plus poussées aient été identifiées. Cependant, ce genre d'étude devrait être mené à chaque fois qu'un simulateur est validé pour un usage spécifique dans le but de s'assurer que c'est le cas. Nous avons trouvé plusieurs corrélations significatives avec l'âge demandant des études approfondies sur cette dimension dans la conception et la validation des simulateurs de conduite.

Nous avons aussi trouvé deux résultats contradictoires : les agents virtuels tendent à ne pas reproduire le comportement humain, puisque la classification automatique différencie leurs comportements de ceux des humains dans plusieurs situations. En conséquence, leur réalisme n'est pas élevé. Cependant, la perception de ces comportements par des annotateurs extérieurs a été jugée différemment, puisqu'ils ont annoté les conducteurs virtuels comme plus humain que les rejoueurs humains. En conséquence, l'axe de la crédibilité et celui du réalisme ont des scores opposés. L'une des explications possibles est que les conducteurs humains tendent à être moins normatifs que les conducteurs virtuels qui suivent des règles strictes, tandis que l'attente sociale est que les humains devraient avoir un comportement de conduite normatif, et les déviances à la norme sont considérées comme des erreurs du conducteur virtuel.

Enfin, nous avons trouvé que la crédibilité comportementale des agents n'est pas corrélée aux résultats du questionnaire de présence. Cela est dû à la focalisation sur laquelle la présence est notée dans le questionnaire de présence de Witmer et Singer, qui correspond plus à un score sur la qualité des composants du dispositif qu'au sentiment d'immersion ou qu'à la fidélité psychologique. Ainsi, nous recommandons l'utilisation d'autres questionnaires de présence pour l'étude de la crédibilité comportementale des agents virtuels, tel que [126].

9.4 Conclusion

Nous avons présenté dans ce chapitre les résultats obtenus par l'application de notre méthode d'évaluation et par notre cycle de conception lors de différentes études sur les comportements de conduite. Toutes les études ne portent pas sur l'intégralité de la méthode, mais elles montrent que les environnements virtuels participatifs peuvent être analysés à l'aide de notre méthodologie.

Les situations de conduite de ces études étaient le dépassement, le trafic ralenti, et la conduite sur autoroute. Dans les trois cas, nous avons appliqué un algorithme de classification non-supervisée et étudié la composition des classes en résultant. Puis nous avons calculé des scores permettant l'analyse du modèle d'agent.

Une des originalités de notre algorithme de classification non-supervisée est qu'il tient compte du fait que les classes contenant des humains sont les comportements valides à reproduire et y agrège les agents similaires. Elle permet ainsi d'identifier - au travers des abstractions de comportement haut niveau - les comportements manquants exhibés par des humains, les capacités du modèle à reproduire un comportement humain et les comportements simulés de manière erronés dans le modèle d'agent. Par cette méthode, les paramètres des agents peuvent être associés à des comportements.

Notre méthode contribue à l'analyse quantifiée des simulations multi-agents et des agents virtuels de par les métriques fournies. Les taux de confiance en chaque classe donnent un indicateur de la véracité pour le modèle d'avoir la capacité de reproduire un comportement humain, ou de l'exactitude d'un manque ou d'une erreur dans le modèle. De manière plus générale, les scores de type de classe (capacités, manques, et erreurs) permettent une quantification des corrections et des apports à faire au modèle.

Notre méthode s'intéresse à la réalité virtuelle et aux simulations où les agents reproduisent des comportements humains pour lesquelles la crédibilité comportementale est essentielle. Elle prend donc en compte les comportements simulés individuellement et participe ainsi aux rares méthodes générales d'analyse à l'échelle microscopique.

Notons que nous n'avons pas exploité toutes les phases du scénario. En effet, certaines situations ne présentaient pas de classes mixtes et ne permettaient donc pas de montrer pleinement le potentiel de notre méthode. D'autre part, certaines situations ont posé des problèmes de contrôle du dispositif par les participants. Ce fut le cas lors du second carrefour (où les acteurs principaux devaient tourner à gauche) pour lequel la maniabilité du volant n'était pas évidente et causa des sorties de route. De même, un problème de contrôle du démarrage gêna la phase de rétrécissement de voie embouteillé pour lequel les participants auraient voulu s'arrêter et redémarrer de nombreuses fois. Aussi, l'utilisation du simulateur *léger* empêcha l'exploitation de la phase contenant une traversée de carrefour pour laquelle la visibilité à l'arrêt au stop n'était pas suffisante pour prévoir l'arrivée de véhicule suffisamment en avance, et ce malgré les trois écrans.

Cependant, certaines phases non encore exploitées pourraient servir à l'évaluation des agents d'*ARCHISIM* dans d'autres situations de conduite. Il y a notamment la conduite en virage en interaction avec d'autres véhicules, la conduite en suivi en file sur une route à voie unique, l'insertion sur autoroute à partir d'une bretelle, et la conduite sur autoroute en trafic fluide.

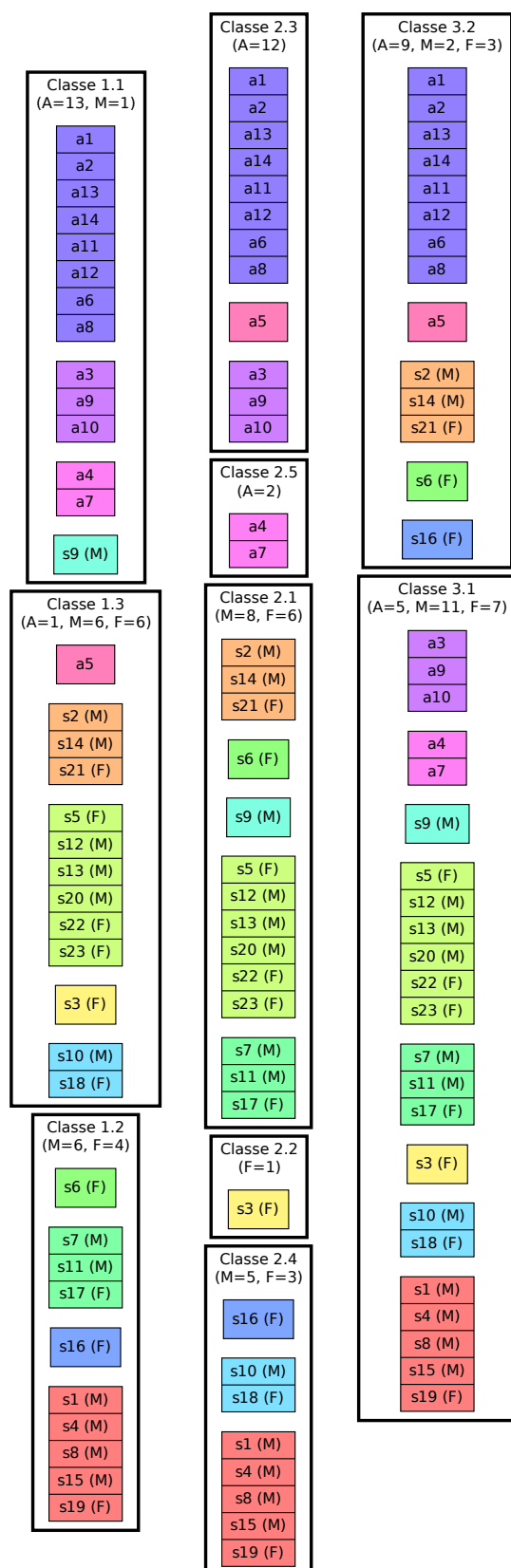


Figure 9.9 – Classifications des comportements issus des traces de simulation pour chaque situation (de gauche à droite). Chaque classe est représentée par un rectangle avec les compositions en termes d’agents (A), d’hommes (M), et de femmes (F). Les acteurs principaux qui sont ensemble dans chacune des classifications sont regroupés au sein d’un rectangle de couleur spécifique.

Conclusion et perspectives

Conclusion

Nous avons présenté dans cette thèse une méthode semi-automatique d'évaluation des comportements des agents ainsi qu'un cycle de conception permettant la calibration d'une population d'agents au sein d'une simulation multi-agents.

Cette méthode se place dans le cadre de la réalité virtuelle et des simulations où les agents ont pour but d'adopter des comportements humains à l'échelle microscopique. L'originalité de notre méthode de classification non-supervisée est de considérer les classes de comportements humains comme les points de référence valides. Les agents sont alors agrégés à ces classes fixes s'ils sont similaires ; ou classés séparément dans le cas contraire, permettant l'identification 1) des capacités du modèle d'agent à simuler un comportement de référence, 2) des comportements exhibés par des humains qui en conséquence manquent dans le modèle d'agent (qui cherche à reproduire le spectre complet des comportements humains), et 3) des comportements qu'aucun humain n'adopte et qui sont donc des erreurs dans le modèle d'agent.

Notre méthode introduit des métriques pour mesurer la qualité de la simulation et pour corriger les paramètres des agents. Les scores de types de classe (c'est-à-dire les scores d'erreurs, de manques, et de capacités) permettent au concepteur de la simulation multi-agents de trouver combien d'archétypes de comportement humain ont été correctement reproduits, combien de comportements d'agents ne devraient pas apparaître, et combien de comportements humains sont manquants. Les taux de confiance en chacune des classes assurent de la justesse du type de classe. Ensuite, en étudiant uniquement les archétypes des comportements correctement reproduits, les scores de représentativité donnent de l'information sur les proportions de chaque comportement et leurs relations à la calibration des agents.

Une deuxième contribution importante de nos travaux est la proposition d'une méthode de calibration de simulations multi-agents participatives via notre cycle de conception. La première itération de ce cycle suffit à supprimer les comportements erronés. L'établissement de classes sur-représentées et sous-représentées ainsi que l'établissement de scores associés autorisent l'ajustement de la population d'agents en accord avec les proportions des comportements humains. Ainsi, notre cycle de conception raffine de manière itérative la calibration des capacités des agents au sein d'une population en prenant en compte leurs paramètres. L'exploration de l'espace des paramètres permet dans le cas d'un modèle d'agent non modifiable de trouver de nouveaux comportements d'agents et d'identifier les parties de l'espace des paramètres produisant des comportements invalides.

Dans plusieurs études, l'application de notre méthode d'évaluation et de calibration des comportements des agents a permis de montrer les capacités du modèle d'agents de l'architecture *ARCHISIM* à reproduire des comportements de conduite humains mais aussi le manque de comportements non normatifs.

Notre méthode permet la comparaison de différents modèles d'agent via les métriques fournies. Cela autorise la comparaison d'une version antérieure d'un modèle d'agent avec la nouvelle afin de pouvoir contrôler et possiblement affirmer les améliorations de cette dernière. Bien que cela n'ait pas été mis en œuvre dans cette thèse, l'utilisation de notre méthode sur différents modèles d'agent permettrait de vérifier - au-delà de l'acquisition de nouvelles capacités de comportements - que le modèle d'agent ne génère pas de comportements erronés supplémentaires et conserve bien les capacités préexistantes évitant ainsi la perte de comportements correctement reproduits à la manière des techniques de tests en génie logiciel. D'autre part, cette comparaison de modèles d'agent peut permettre de d'effectuer un *survey* sur un *benchmark* commun afin d'évaluer les différents modèles de la littérature d'un domaine.

Notre méthode semi-automatique d'évaluation des comportements présente l'originalité de se baser à la fois sur les traces de simulations et sur le jugement d'annotateurs. Elle permet donc de combiner les deux grandes approches de l'étude du comportement que sont l'approche objective (à base de traces) et l'approche subjective (à base de jugement humain). Elle réconcilie ainsi les domaines des Sciences Humaines et Sociales et de l'Intelligence Artificielle qui abordent l'analyse du comportement humain selon deux angles différents en tirant parti des atouts des méthodes proposées dans chacun de ces domaines : l'analyse des indicateurs issus des traces permet une étude objective du comportement et l'annotation par des participants permet d'obtenir directement des informations de haut niveau.

Les perspectives de ces travaux de thèses, notamment sur l'évaluation de l'évolution d'un ensemble de situations, sont présentées dans le chapitre suivant.

Perspectives

Nous présentons dans ce chapitre les perspectives de nos travaux de recherche. Nous pouvons identifier cinq catégories de perspectives en fonction de la méthode de classification, des métriques, de la comparaison de classifications, de l'expérimentation, ou de l'évolution des comportements.

Nous appliquons actuellement notre méthode à travers divers scripts et programmes basés sur de multiples langages. De manière à ce que notre méthode soit utilisable plus facilement par la communauté, il serait avantageux de fournir une application modulaire à code source ouvert fournissant les outils nécessaires à l'ensemble de la méthodologie et laissant ainsi les briques de notre méthode libres d'être instanciées différemment.

Méthode de classification

Nous avons instancié la brique de catégorisation de notre méthode par un algorithme basé sur le K-moyennes. Cependant, l'utilisation d'autres algorithmes de classification peuvent permettre une classification plus juste en fonction des données de simulation. Cela pourrait être le cas des algorithmes conservant la nature temporelle des données. Notre méthode autorise l'exploitation de tout type d'algorithme de classification non-supervisée, si tant est que celui-ci ait un nombre de classes libre déterminé automatiquement par un critère défini (il serait aussi possible de définir le nombre de classes manuellement dans une version plus semi-automatique). Dans le cas où un autre algorithme de classification est choisi, l'algorithme d'agrégation des agents doit être adapté. Par exemple, dans le cas d'un algorithme où l'appartenance à une classe est fonction d'une probabilité, l'agrégation d'un agent à une classe dépendra alors de cette probabilité.

La méthode d'agrégation des agents aux classes d'humains (*i.e.* aux comportements de référence) dépend aussi d'un seuil de tolérance dont la valeur peut impacter la qualité des résultats : cet impact devra être vérifié. De plus, sa valeur est actuellement définie comme un pourcentage de la moyenne des seuils de chaque dimension. Il pourrait être intéressant d'automatiser la définition de ce pourcentage actuellement fixé arbitrairement.

Métriques

Notre méthode fournit après l'expérimentation une classification des comportements issus des traces de simulation et une classification des comportements annotés

sur lesquels nous calculons ensuite des métriques que sont les scores de type de classe, les taux de confiance en chaque classe, et les scores de représentativité des agents. Une limite actuelle des scores de type de classe et des taux de confiance est qu'ils prennent uniquement en compte la classification des comportements issus des traces. S'ils peuvent pareillement être appliqués à la classification d'annotations, nous pourrions envisager d'établir des métriques prenant en compte conjointement les deux classifications comme cela est fait pour les scores de représentativité.

Notre cycle de conception permet la calibration de la population d'agents d'une Simulation Multi-Agents de manière itérative en explorant l'espace des paramètres. Cependant, la convergence du cycle de conception n'a pas encore été étudiée. Notre méthode d'évaluation donne des scores qui pourraient être utilisés afin de stopper le processus, mais une preuve de convergence est nécessaire lorsque le cycle n'est pas utilisé pour explorer de nouveaux paramètres.

Comparaison de classifications

Notre méthode d'évaluation permet d'obtenir deux classifications de comportements : l'une résultant des indicateurs issus des traces et l'autre des réponses au questionnaire d'annotations. Dans le cas où ces indicateurs et ces annotations analysent le comportement sous le même angle et à un même niveau, les classifications doivent alors être similaires et peuvent donc être comparées avec des méthodes de comptage par paires. Dans le cas contraire où les indicateurs et les annotations n'analysent pas le comportement selon le même angle, les classifications ne sont pas nécessairement similaires et la comparaison n'a donc pas lieu d'être.

Si les deux classifications ne se placent pas au même niveau de comportement, alors la classification de plus haut niveau peut avoir regroupée des acteurs principaux qui dans l'autre classification sont séparés en plusieurs classes. Les scores fournis par la méthode de comptage par paires n'ont alors plus le même sens. En effet, ces scores ne prennent pas en compte cette différence de niveau, c'est-à-dire la fusion ou la scission de classes.

Nous pensons qu'il est possible de prendre en compte cette différence de niveau. Pour cela, il est nécessaire d'avoir identifié la classification de comportements de plus haut niveau. Si plusieurs classes de la classification de comportements de plus bas niveau font intégralement partie d'une unique classe de la classification de comportement de plus haut niveau, alors il est possible de les fusionner en une unique classe. Après ce traitement, il est alors de nouveau possible d'appliquer les comptages par paires tels que le *Rand Index*.

Cependant, l'inclusion intégrale de plusieurs classes en une autre est trop stricte. En effet, une unique différence de classification empêcherait la fusion de plusieurs classes. Cette différence implique que tous les acteurs principaux ne se retrouvent plus dans une même classe dans la classification de comportements de haut niveau. Afin d'obtenir un score plus représentatif, nous proposons de minorer l'effet d'une erreur de classification en acceptant la fusion de classes même si l'inclusion n'est pas complète selon un seuil dépendant du nombre d'acteurs principaux.

Actuellement, nous n'avons pas évalué cette proposition de comparaison de classifications de comportements de niveaux différents.

Perspectives expérimentales

Dans nos études, nous exploitons les résultats d'une expérimentation faite sur un simulateur *léger*. L'*IFSTTAR* disposant d'un simulateur à échelle réelle dont la qualité d'immersion est plus grande sur les dimensions sonore, visuelle, et haptique, une expérimentation sur ce dernier permettrait d'étudier le rapport entre la crédibilité comportementale et les autres dimensions de la présence. De plus, cela pourrait résoudre les problèmes de contrôle du dispositif de conduite et de visuel limité afin d'obtenir des traces de simulation pour les participants qui soient exploitables sur certaines phases de notre scénario de conduite (*e.g.* la manœuvre de tourne à gauche et l'arrêt au panneau stop d'un carrefour à 90°, et la conduite suivi en file). Nous préconisons l'utilisation d'un questionnaire de présence intégrant la dimension comportementale contrairement au *Presence Questionnaire*.

Notre méthode permettant la comparaison de modèles d'agent, une comparaison d'*ARCHISIM* avec un autre modèle de la littérature serait pertinente. Par exemple, une comparaison avec *AIMSUN* [4] pourrait être faite dans un premier temps.

Notre méthode a été conçue de manière à pouvoir être appliquée aux agents virtuels et aux simulations multi-agents. Nous l'avons mis en œuvre dans le cadre de la simulation de conduite pour laquelle les interactions sont uniquement indirectes. Il serait intéressant de l'appliquer à d'autres domaines comme la simulation de foule, la simulation économique, les jeux vidéo, ainsi que la programmation orientée émotion [31].

Évolution et ensemble de situations

Notre méthode d'évaluation des comportements permet d'analyser le modèle d'agent sur une situation et notre cycle de conception rend possible la calibration des paramètres d'un ensemble d'agents. Chaque agent adopte alors durant toute la simulation des comportements selon la même "personnalité", *i.e.* les mêmes valeurs des variables d'entrées du modèle d'agent. Chez les humains, pour un même individu, il y a une variabilité de comportements dans le temps suivant le contexte, la fatigue, l'humeur, l'expérience, *etc.* Les agents simulant des comportements humains, leurs comportements devraient alors eux aussi être capables d'évoluer de manière similaire dans un jeu de paramètres fixé (*i.e.* une "personnalité" d'agent). Si des changements locaux chez les participants affectent leurs comportements adoptés, il est alors possible d'analyser l'évolution de ces comportements au cours d'un ensemble ordonné de situations. Nous pouvons alors observer le caractère constant ou changeant du comportement de conduite d'un participant ou d'un agent pour un jeu de paramètres fixé.

Il est alors nécessaire de comparer les classifications faites sur plusieurs situations. Cela permet d'observer si les véhicules ont conservés un comportement similaire et

si ces comportements sont jugés de la même manière. Nous pouvons alors observer le caractère constant ou changeant du comportement de conduite d'un participant ou d'un agent. Si la composition des classes de participants est similaire sur différentes situations, alors les comportements sont donc stables. À l'inverse, des changements de composition (*i.e.* des classes instables) impliqueraient que les comportements des agents doivent aussi évoluer au cours du temps.

Nous avons commencé une réflexion sur l'étude simultanée d'un ensemble de situations ou d'une évolution de comportement au cours de plusieurs situations, et proposons trois types possibles d'évolution de classe comme décrits dans la représentation graphique faite en *figure 9.10*. Les trois types d'évolution des classes sont les suivants : les classes stables dont la composition reste inchangée, les classes issues d'une unique classe (appelées *split*) dans la situation précédente (*i.e.* s'étant séparée en plusieurs classes), et les classes correspondant à un regroupement de plusieurs classes (appelées *merge*) dans la situation précédente.

Cependant, il reste encore beaucoup de travail afin de fournir des outils de visualisation aboutis et d'effectuer une analyse automatique. Cela demanderait dans un premier temps d'étudier l'état de l'art sur l'analyse dynamique de classifications. Plus spécifiquement à notre méthode d'évaluation, il serait ensuite nécessaire de fournir des scores calculables automatiquement mesurant les évolutions décrites précédemment. De plus, dans le but de faire une synthèse complète de l'évolution de cet ensemble de situations, nous pensons qu'un outil de comparaison graphique avancée est indispensable.

Une telle méthode d'évaluation des évolutions des comportements sur un ensemble de classifications permettrait la validation de la dynamique des comportements des agents. Nous pensons que c'est une étape indispensable pour concevoir des agents artificiels qui changent "correctement" de comportements au cours du temps.

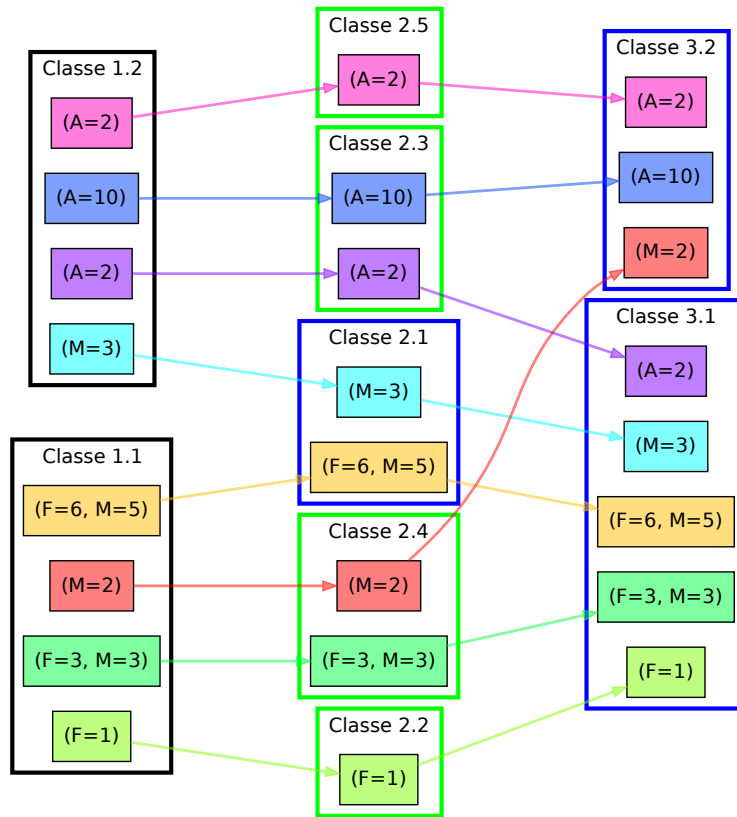


Figure 9.10 – Évolution des classes de comportements issues des traces sur 3 classifications. Les acteurs principaux qui sont toujours classés dans les mêmes classes sont regroupés ensemble dans un rectangle de couleur unique. Il est alors possible de suivre l'évolution de ce groupe au cours des différentes classifications. Les classes (nommées *cluster#_#*) de la première classification sont entourées en noir. Pour les classifications suivantes, les classes sont entourées par un rectangle de couleur : vert dans le cas d'une sous-classe provenant d'une classe s'étant séparée en deux dans la classification précédente (*split*) ; bleu dans le cas d'une classe correspondant à une fusion de plusieurs classes dans la classification précédente (*merge*) ; et rouge (non présent ici) dans le cas d'une classes dont la composition est identique à celle de la classification précédente (*equal*).

Liste des publications

Conférences internationales avec comités de lecture

- [32] K. DARTY, J. SAUNIER et N. SABOURET. « A method for semi-automatic explicitation of agent's behavior: application to the study of an immersive driving simulator ». In : *The 6th International Conference on Agents and Artificial Intelligence (ICAART 2014)*. SciTePress, 2014, p. 81–91.
- [34] K. DARTY, J. SAUNIER et N. SABOURET. « Calibration of Multi-Agent Simulations through a Participatory Experiment ». In : *The International Conference on Autonomous Agents & Multi-Agent Systems (AAMAS 2015)*. 2015, p. 1683–1684.
- [35] K. DARTY, J. SAUNIER et N. SABOURET. « Agents behavior semi-automatic analysis through their comparison to human behavior clustering ». In : *The 14th International Conference on Intelligent Virtual Agents (IVA 2014)*. T. 8637. Springer - LNCS, 2014, p. 154–163.
- [38] K. DARTY, J. SAUNIER et N. SABOURET. « Effects of Gender on Presence and Virtual Driver Perception in Driving Simulators ». In : *The 5th International Conference on Women's Issues in Transportation (WiT 2014)*. 2014.

Conférences nationales avec comités de lecture

- [33] K. DARTY, J. SAUNIER et N. SABOURET. « Extraction de comportements pour l'étude de la crédibilité des agents ». In : *21^{èmes} Journées Francophones des Systèmes Multi-Agents (JFSMA-PFIA)*. T. 21. Cépaduès, 2013.
- [36] K. DARTY, J. SAUNIER et N. SABOURET. « Analyse des comportements agents par agrégation aux comportements humains ». In : *22^{èmes} Journées Francophones sur les Systèmes Multi-Agents (JFSMA 2014)*. Cépaduès, 2014.
- [37] K. DARTY, J. SAUNIER et N. SABOURET. « Calibration de simulations multi-agents à l'aide d'une méthode semi-automatique d'analyse du comportement ». In : *23^{èmes} Journées Francophones sur les Systèmes Multi-Agents (JFSMA 2015)*. 2015.

Bibliographie

- [1] E. AMOUROUX et al. « SMACH : Simuler l'activité humaine pour une meilleure gestion de l'énergie ». In : *JFSMA 2013* (2013).
- [2] D. ARTHUR et S. VASSILVITSKII. « k-means++ : The advantages of careful seeding ». In : *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial et Applied Mathematics. 2007, p. 1027–1035.
- [3] J. N. BAIENSON et al. « The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments ». In : *Presence : Teleoperators and Virtual Environments* 14.4 (2005), p. 379–393.
- [4] J. BARCELÓ et J. CASAS. « Dynamic network simulation with AIMSUN ». In : *Simulation Approaches in Transportation Analysis* (2005), p. 57–98.
- [5] R. F. BENEKOHAL. *Procedure for validation of microscopic traffic flow simulation models*. 1320. 1991.
- [6] L. BERGROTH, H. HAKONEN et T. RAITA. « A survey of longest common subsequence algorithms ». In : *String Processing and Information Retrieval, 2000. SPIRE 2000. Proceedings. Seventh International Symposium on*. IEEE. 2000, p. 39–48.
- [7] D. BERNDT et J. CLIFFORD. « Using dynamic time warping to find patterns in time series ». In : *KDD workshop*. T. 10. 16. 1994, p. 359–370.
- [8] T. W. BICKMORE, L. M. PFEIFER et M. K. PAASCHE-ORLOW. « Health document explanation by virtual agents ». In : *Intelligent Virtual Agents*. Springer. 2007, p. 183–196.
- [9] G. BIRKHOFF. *Lattice theory*. T. 25. American Mathematical Soc., 1967.
- [10] H. BLOCH et al. *Grand dictionnaire de la psychologie*. Larousse, 1992.
- [11] T. BOSSE, G. F. SIDDIQUI et J. TREUR. « An intelligent virtual agent to increase involvement in financial services ». In : *Intelligent virtual agents*. Springer. 2010, p. 378–384.
- [12] T. BOSSE et al. « Agent-based analysis of patterns in crowd behaviour involving contagion of mental states ». In : *Modern Approaches in Applied Intelligence*. Springer, 2011, p. 566–577.
- [13] J. H. BROCKMYER, C. M. FOX et al. « The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing ». In : *Journal of Experimental Social Psychology* 45.4 (2009), p. 624–634.
- [14] C. BUCHE et al. « MASCARET : pedagogical multi-agents systems for virtual environment for training ». In : *Cyberworlds, 2003. Proceedings. 2003 International Conference on*. IEEE. 2003, p. 423–430.
- [15] J. M. BURKHARDT, B. BARDY et D. LOURDEAUX. « Immersion, Réalisme et Présence dans la conception et l'évaluation des Environnements Virtuels ». In : *Psychologie française* 48.2 (2003), p. 35–42.
- [16] J. P. BYRNES, D. C. MILLER et W. D. SCHAFER. « Gender differences in risk taking : A meta-analysis. » In : *Psychological bulletin* 125.3 (1999), p. 367.

-
- [17] B. CAHOUR. « Discomfort, affects and coping strategies in driving activity ». In : *Proceedings of the 15th European conference on Cognitive ergonomics : the ergonomics of cool interaction*. ACM. 2008, p. 22.
- [18] B. CAHOUR, J.-F. FORZY et C. MARTIN. « Feelings and strategies of senior drivers : ways of coping with fear ? » In : *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*. ACM. 2010, p. 161–168.
- [19] P. CAILLOU et J. GIL-QUIJANO. « SimAnalyzer : Automated description of groups dynamics in agent-based simulations ». In : *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*. International Foundation for Autonomous Agents et Multiagent Systems. 2012, p. 1353–1354.
- [20] T. CALIŃSKI et J. HARABASZ. « A dendrite method for cluster analysis ». In : *Communications in Statistics-theory and Methods* 3.1 (1974), p. 1–27.
- [21] S. CAMPANO et al. « An evaluation of the COR-E computational model for affective behaviors ». In : *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents et Multiagent Systems. 2013, p. 745–752.
- [22] M. CAVAZZA, F. CHARLES et S. J. MEAD. « Agents ? interaction in virtual storytelling ». In : *Intelligent Virtual Agents*. Springer. 2001, p. 156–170.
- [23] A. CHAMPION. « Mécanisme de coordination multi-agent fondé sur des jeux : application à la simulation comportementale de trafic routier en situation de carrefour ». In : *PhD thesis, Université de Valenciennes et du Hainaut-Cambrésis* 2.00 (2003), p. 3.
- [24] A. CHAMPION, S. ÉSPIÉ et J. M. AUBERLET. « Behavioral road traffic simulation with ARCHISIM ». In : *Summer Computer Simulation Conference*. Society for Computer Simulation International ; 1998. 2001, p. 359–364.
- [25] A. CHAMPION et al. « Behavioral Simulation : Towards High-Density Network Traffic Studies ». In : ASCE. 2002.
- [26] Y. CHENG et G. M. CHURCH. « Biclustering of expression data. » In : *Ismb*. T. 8. 2000, p. 93–103.
- [27] G. CHIRON, P. GOMEZ-KRÄMER et M. MÉNARD. « Approche Bayésienne non-paramétrique pour la découverte de comportements émergents à partir de trajectoires ». In : *Reconnaissance de formes et intelligence artificielle (RFIA) 2014*. France, juin 2014. URL : <https://hal.archives-ouvertes.fr/hal-00988587>.
- [28] N. CHOMSKY, J. MCGILVRAY et I. EBRARY. *Cartesian linguistics : A chapter in the history of rationalist thought*. T. 116. Harper & Row New YorkLondon, 1966.
- [29] C. D. COTTRILL et P. V. THAKURIAH. « Privacy and Gender : Reviewing Women?s Attitudes Toward Privacy in the Context of Intelligent Transportation Systems and Location-Based Services ». In : *Transportation Research Board Conference Proceedings*. T. 2. 46. 2011.
- [30] M. COURGEON, J.-C. MARTIN et C. JACQUEMIN. « Multimodal affective and reactive character ». In : *1st Workshop on Affective Interaction in Natural Environments*. 2008.
-

- [31] K. DARTY et N. SABOURET. « Emotion Oriented Programming: Computational Abstractions for AI Problem Solving ». In : *The 25th International Conference of the Florida Artificial Intelligence Research Society (FLAIRS)*. AAAI Press, 2012, p. 157–162.
- [32] K. DARTY, J. SAUNIER et N. SABOURET. « A method for semi-automatic explication of agent’s behavior: application to the study of an immersive driving simulator ». In : *The 6th International Conference on Agents and Artificial Intelligence (ICAART 2014)*. SciTePress, 2014, p. 81–91.
- [33] K. DARTY, J. SAUNIER et N. SABOURET. « Extraction de comportements pour l’étude de la crédibilité des agents ». In : *21^{èmes} Journées Francophones des Systèmes Multi-Agents (JFSMA-PFIA)*. T. 21. Cépaduès, 2013.
- [34] K. DARTY, J. SAUNIER et N. SABOURET. « Calibration of Multi-Agent Simulations through a Participatory Experiment ». In : *The International Conference on Autonomous Agents & Multi-Agent Systems (AAMAS 2015)*. 2015, p. 1683–1684.
- [35] K. DARTY, J. SAUNIER et N. SABOURET. « Agents behavior semi-automatic analysis through their comparison to human behavior clustering ». In : *The 14th International Conference on Intelligent Virtual Agents (IVA 2014)*. T. 8637. Springer - LNCS, 2014, p. 154–163.
- [36] K. DARTY, J. SAUNIER et N. SABOURET. « Analyse des comportements agents par agrégation aux comportements humains ». In : *22^{èmes} Journées Francophones sur les Systèmes Multi-Agents (JFSMA 2014)*. Cépaduès, 2014.
- [37] K. DARTY, J. SAUNIER et N. SABOURET. « Calibration de simulations multi-agents à l’aide d’une méthode semi-automatique d’analyse du comportement ». In : *23^{èmes} Journées Francophones sur les Systèmes Multi-Agents (JFSMA 2015)*. 2015.
- [38] K. DARTY, J. SAUNIER et N. SABOURET. « Effects of Gender on Presence and Virtual Driver Perception in Driving Simulators ». In : *The 5th International Conference on Women’s Issues in Transportation (WiT 2014)*. 2014.
- [39] P. DAVIDSSON. « Agent based social simulation : A computer science view ». In : *Journal of artificial societies and social simulation* 5.1 (2002).
- [40] E. DELAHERCHE et al. « Interpersonal synchrony : A survey of evaluation methods across disciplines ». In : *Affective Computing, IEEE Transactions on* 3.3 (2012), p. 349–365.
- [41] A. P. DEMPSTER, N. M. LAIRD et D. B. RUBIN. « Maximum likelihood from incomplete data via the EM algorithm ». In : *Journal of the Royal Statistical Society. Series B (Methodological)* (1977), p. 1–38.
- [42] M. DOMJAN et J. GRAU. *The principles of learning and behavior*. Wadsworth Pub Co, 2009.
- [43] A. DONIEC et al. « A behavioral multi-agent model for road traffic simulation ». In : *Engineering Applications of Artificial Intelligence* 21.8 (2008), p. 1443–1454.
- [44] A. DROGOU, B. CORBARA et D. FRESNEAU. « MANTA : New experimental results on the emergence of (artificial) ant societies ». In : *Artificial Societies : the computer simulation of social life* (1995), p. 190–211.

-
- [45] S. ESPIÉ. « Vehicle-driven simulator versus traffic-driven simulator : the IN-RETS approach ». In : *Driving simulation conference*. 1999, p. 367–376.
- [46] S. ESPIÉ et J. M. AUBERLET. « ARCHISIM : A behavioral multi-actors traffic simulation model for the study of a traffic system including ITS aspects ». In : *International Journal of ITS Research* 1 (2007).
- [47] M. FEHLER, F. KLÜGL et F. PUPPE. « Approaches for resolving the dilemma between model structure refinement and parameter calibration in agent-based simulations ». In : *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*. ACM. 2006, p. 120–122.
- [48] M. FEHLER, F. KLÜGL et F. PUPPE. « Techniques for analysis and calibration of multi-agent simulations ». In : *Engineering Societies in the Agents World V*. Springer, 2005, p. 305–321.
- [49] J. FENG, I. SPENCE et J. PRATT. « Playing an action video game reduces gender differences in spatial cognition ». In : *Psychological science* 18.10 (2007), p. 850–855.
- [50] D. FISHER, M. RIZZO et J. CAIRD. *Handbook of driving simulation for engineering, medicine, and psychology*. CRC Press, 2011.
- [51] S. R. A. FISHER et al. *Statistical methods for research workers*. T. 14. Oliver et Boyd Edinburgh, 1970.
- [52] P. M. FITTS. « The information capacity of the human motor system in controlling the amplitude of movement. » In : *Journal of experimental psychology* 47.6 (1954), p. 381.
- [53] G. FONTAINE. « The experience of a sense of presence in intercultural and international encounters ». In : *Presence : Teleoperators and VEs* 1.4 (1992), p. 482–490.
- [54] D. J. FRENCH et al. « Decision-making style, driving style, and self-reported involvement in road traffic accidents ». In : *Ergonomics* 36.6 (1993), p. 627–644.
- [55] A. GABADINHO et al. « Analyzing and visualizing state sequences in R with TraMineR ». In : *Journal of Statistical Software* 40.4 (2011), p. 1–37.
- [56] C. GABAUDE et al. « Self-regulatory driving behaviour in the elderly : relationships with aberrant driving behaviours and perceived abilities ». In : *Le travail humain* 73.1 (2010), p. 31–52.
- [57] M. GARAU et al. « The responses of people to virtual humans in an immersive virtual environment ». In : *Presence : Teleoperators and Virtual Environments* 14.1 (2005), p. 104–116.
- [58] O. GEORGEON, A. MILLE et T. BELLET. « Abstract : un outil et une méthodologie pour analyser une activité humaine médiée par un artefact technique complexe ». fr. In : *Ingénierie des Connaissances. Semaine de la connaissance* (juin 2006). Sous la dir. de R. L. M. H. N. A.-G. J. CHARLET. URL : http://www.sdc2006.org/cdrom/auteurs.html#Georgeon_Olivier.

- [59] O. GEORGEON et al. « Creating cognitive models from activity analysis : A knowledge engineering approach to car driver modeling ». en. In : *International Conference on Cognitive Modeling*. Taylor & Francis, juil. 2007, p. 43–48. ISBN : 978-1-84169-708-. URL : <http://liris.cnrs.fr/publis/?id=2914>.
- [60] O. L. GEORGEON et al. « Supporting activity modelling from activity traces ». en. In : *Expert Systems* 29.3 (juin 2012), p. 261–275. URL : <http://liris.cnrs.fr/publis/?id=5140>.
- [61] A. GLENDON et al. « Reliability of the driving behaviour inventory ». In : *Ergonomics* 36.6 (1993), p. 719–726.
- [62] J. GONÇALVES et R. J. F. ROSSETTI. « Extending SUMO to support tailored driving styles ». In : *1st SUMO User Conference, DLR, Berlin - Adlershof, Germany* 21 (2013), p. 205–211.
- [63] J. GRATCH et S. MARSELLA. « Evaluating a computational model of emotion ». In : *Autonomous Agents and Multi-Agent Systems* 11.1 (2005), p. 23–43.
- [64] J. GRATCH et al. « Assessing the validity of appraisal-based models of emotion ». In : *Proceedings Affective Computing and Intelligent Interaction*. 2009, p. 1–8.
- [65] J. GRATCH et al. « Creating rapport with virtual agents ». In : *Intelligent Virtual Agents*. Springer. 2007, p. 125–138.
- [66] R. E. GUADAGNO et al. « Virtual humans and persuasion : The effects of agency and behavioral realism ». In : *Media Psychology* 10.1 (2007), p. 1–22.
- [67] E. GULIAN et al. « Dimensions of driver stress ». In : *Ergonomics* 32.6 (1989), p. 585–602.
- [68] P. GUYOT et A. DROGOUL. « Multi-agent based participatory simulations on various scales ». In : *Massively Multi-Agent Systems I*. Springer, 2005, p. 149–160.
- [69] J. D. HAMILTON. *Time series analysis*. T. 2. Princeton university press Princeton, 1994.
- [70] H. HATTORI, Y. NAKAJIMA et T. ISHIDA. « Learning from humans : Agent modeling with individual human behaviors ». In : *Systems, Man and Cybernetics, Part A : Systems and Humans, IEEE Transactions on* 41.1 (2011), p. 1–9.
- [71] L. HUBERT et P. ARABIE. « Comparing partitions ». In : *Journal of classification* 2.1 (1985), p. 193–218.
- [72] R. IHAKA et R. GENTLEMAN. « R : a language for data analysis and graphics ». In : *Journal of computational and graphical statistics* 5.3 (1996), p. 299–314.
- [73] T. ISHIDA et al. « Augmented Experiment : Participatory Design with Multiagent Simulation ». In : *IJCAI*. T. 7. 2007, p. 1341–1346.
- [74] W. JAMES. *The principles of psychology*. 1890.
- [75] C. JAVEAU. *L'enquête par questionnaire : manuel à l'usage du praticien*. Editions de l'Université de Bruxelles, 1978.

-
- [76] Y. B. KAFAI et al. « Preface : pink, purple, casual, or mainstream games : moving beyond the gender divide ». In : *Beyond Barbie and Mortal Combat. New Perspectives on Gender and Gaming*. The MIT Press : London (2008).
- [77] B. S. KERNER. *Introduction to modern traffic flow theory and control*. Springer, 2009.
- [78] T. KOHONEN. « The self-organizing map ». In : *Proceedings of the IEEE* 78.9 (1990), p. 1464–1480.
- [79] S. KULLBACK. *Information theory and statistics*. Courier Corporation, 1997.
- [80] K. LACHLAN et M. KRUMHOLTZ. « Experiencing presence in video games : The role of presence tendencies, game experience, gender, and time spent in play ». In : *Communication Research Reports* 28.1 (2011), p. 27–31.
- [81] B. LACROIX, P. MATHIEU et A. KEMENY. « Formalizing the construction of populations in multi-agent simulations ». In : *Engineering Applications of Artificial Intelligence* (2012).
- [82] L. LECLERCQ. « Modélisation dynamique du trafic et applications à l'estimation du bruit routier ». Thèse de doct. Villeurbanne, INSA de Lyon, 2002.
- [83] J. LEPLAT. « Simulation et simulateur : principes et usages ». In : *Regards sur l'activité en situation de travail : contribution à la psychologie ergonomique* (1997), p. 157–181.
- [84] J. LESSITER et al. « A cross-media presence questionnaire : The ITC-Sense of Presence Inventory ». In : *Presence : Teleoperators & Virtual Environments* 10.3 (2001), p. 282–297.
- [85] J. C. LESTER, S. A. CONVERSE et al. « The persona effect : affective impact of animated pedagogical agents ». In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1997, p. 359–366.
- [86] T. W. LIAO. « Clustering of time series data? a survey ». In : *Pattern recognition* 38.11 (2005), p. 1857–1874.
- [87] R. LIKERT. « A technique for the measurement of attitudes. » In : *Archives of psychology* (1932).
- [88] G. LOEWENSTEIN et J. LERNER. « The role of affect in decision making ». In : *Handbook of affective science* 619 (2003), p. 642.
- [89] M. LOMBARD et T. DITTON. « At the heart of it all : The concept of presence ». In : *Journal of Computer-Mediated Communication* 3.2 (1997).
- [90] K. LORENZ. *The foundations of ethology*. Springer verlag, 1981.
- [91] D. LOURDEAUX et J.-M. BURKHARDT. « Le traité de la réalité virtuelle ». In : t. 5. sciences mathématiques et informatiques, Mines Paris, 3e édition. Les Presses de l'Ecole des Mines de Paris, avr. 2009. Chap. Usages de l'humain virtuel pour la formation, p. 1–10. URL : <http://hal.archives-ouvertes.fr/hal-00462284/en/>.
- [92] J. MACQUEEN et al. « Some methods for classification and analysis of multivariate observations ». In : *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. T. 1. 14. Oakland, CA, USA. 1967, p. 281–297.
-

- [93] P. MAES et R. KOZIEROK. « Learning interface agents ». In : *Proceedings of the National Conference on Artificial Intelligence*. John Wiley & Sons LTD. 1993, p. 459–459.
- [94] C. D. MANNING, P. RAGHAVAN et H. SCHÜTZE. *Introduction to information retrieval*. T. 1. Cambridge University Press Cambridge, 2008.
- [95] B. MATHERN. « Découverte interactive de connaissances à partir de traces d'activité : Synthèse d'automates pour l'analyse et la modélisation de l'activité de conduite automobile ». fr. Thèse de Doctorat en Informatique. Université Claude Bernard-Lyon I, mar. 2012.
- [96] B. MATHERN, T. BELLET et A. MILLE. « An Iterative Approach to Develop a Cognitive Model of the Driver for Human Centred Design of ITS ». en. In : *European Conference on Human Centred Design for Intelligent Transport Systems, Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems*. Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems. HUMANIST publications, avr. 2010, p. 85–95. ISBN : 978-2-9531712-1-1. URL : <http://liris.cnrs.fr/publis/?id=4928>.
- [97] P. MATHIEU et S. PICAULT. « From Real Purchase to Realistic Populations of Simulated Customers ». In : *Advances on Practical Applications of Agents and Multi-Agent Systems*. Springer, 2013, p. 216–227.
- [98] P. MATHIEU et O. BRANDOUY. « A generic architecture for realistic simulations of complex financial dynamics ». In : *Advances in Practical Applications of Agents and Multiagent Systems*. Springer Berlin Heidelberg, 2010, p. 185–197.
- [99] M. W. MCGREEVY. « The presence of field geologists in Mars-like terrain ». In : *Presence : Teleoperators and Virtual Environments* 1.4 (1992), p. 375–403.
- [100] J. A. MICHON. *A critical view of driver behavior models : what do we know, what should we do ?* Springer, 1985.
- [101] H. C. MILES et al. « Efficacy of a Virtual Environment for Training Ball Passing Skills in Rugby ». In : *Transactions on Computational Science XXIII*. Springer, 2014, p. 98–117.
- [102] G. W. MILLIGAN et M. C. COOPER. « A study of the comparability of external criteria for hierarchical cluster analysis ». In : *Multivariate Behavioral Research* 21.4 (1986), p. 441–458.
- [103] G. W. MILLIGAN et M. C. COOPER. « An examination of procedures for determining the number of clusters in a data set ». In : *Psychometrika* 50.2 (1985), p. 159–179.
- [104] M. MINSKY. « Telepresence ». In : (1980).
- [105] D. NI et al. « Systematic approach for validating traffic simulation models ». In : *Transportation Research Record : Journal of the Transportation Research Board* 1876.1 (2004), p. 20–31.
- [106] E. M. O'LAUGHLIN et B. S. BRUBAKER. « Use of landmarks in cognitive mapping : Gender differences in self report versus performance ». In : *Personality and Individual Differences* 24.5 (1998), p. 595–601.

-
- [107] A. ORTONY, G. CLORE et A. COLLINS. *The cognitive structure of emotions*. Cambridge Univ Pr, 1990.
- [108] J. PATRICK. *Training : Research and practice*. Academic Press, 1992.
- [109] I. P. PAVLOV et G. V. ANREP. *Conditioned reflexes*. Dover Pubns, 2003.
- [110] C. PELACHAUD. « Modelling multimodal expression of emotion in a virtual agent ». In : *Phil. Trans. R. Soc. B : Biological Sciences* 364.1535 (2009), p. 3539–3548.
- [111] D. PELLEG, A. MOORE et al. « X-means : Extending k-means with efficient estimation of the number of clusters ». In : *Proceedings of the Seventeenth International Conference on Machine Learning*. T. 1. San Francisco. 2000, p. 727–734.
- [112] C. PETITMENGIN et al. « The validity of first-person descriptions as authenticity and coherence ». In : *Journal of Consciousness studies* 16.10-12 (2009), p. 252–284.
- [113] D. PREMACK, G. WOODRUFF et al. « Does the chimpanzee have a theory of mind? » In : *Behavioral and brain sciences* 1.4 (1978), p. 515–526.
- [114] H. PRENDINGER et M. ISHIZUKA. « Social role awareness in animated agents ». In : *Proceedings of the fifth international conference on Autonomous agents*. ACM. 2001, p. 270–277.
- [115] L. RABINER et B.-H. JUANG. « An introduction to hidden Markov models ». In : *ASSP Magazine, IEEE* 3.1 (1986), p. 4–16.
- [116] H. RAKHA et al. « Systematic verification, validation and calibration of traffic simulation models ». In : *75th Annual Meeting of the Transportation Research Board, Washington, DC*. Citeseer. 1996.
- [117] W. M. RAND. « Objective criteria for the evaluation of clustering methods ». In : *Journal of the American Statistical association* 66.336 (1971), p. 846–850.
- [118] J. RASMUSSEN. « Skills, rules, and knowledge ; signals, signs, and symbols, and other distinctions in human performance models ». In : *Systems, Man and Cybernetics, IEEE Transactions on* 3 (1983), p. 257–266.
- [119] J. REASON et al. « Errors and violations on the roads : a real distinction? » In : *Ergonomics* 33.10-11 (1990), p. 1315–1332.
- [120] J. RICKEL et W. L. JOHNSON. « Integrating pedagogical capabilities in a virtual environment agent ». In : *Proceedings of the first international conference on Autonomous agents*. ACM. 1997, p. 30–38.
- [121] G. ROBILLARD et al. « Validation canadienne-française de deux mesures importantes en réalité virtuelle : l’Immersive Tendancies Questionnaire et le Presence Questionnaire ». In : *Poster presented at the 25e congrès annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP), Trois-Rivières, November 1-3 (2002)*.
- [122] J. P. ROWE et al. « Integrating learning, problem solving, and engagement in narrative-centered learning environments ». In : *International Journal of Artificial Intelligence in Education* 21.1 (2011), p. 115–133.
-

- [123] I. SAFFAR et al. « Conception de simulations multi-agents : vers une modélisation basée sur les observations des comportements réels ». In : *Actes de la conférence RFIA 2012*. 2012.
- [124] S. SALVADOR et P. CHAN. « Toward accurate dynamic time warping in linear time and space ». In : *Intelligent Data Analysis* 11.5 (2007), p. 561–580.
- [125] T. SCHUBERT, F. FRIEDMANN et H. REGENBRECHT. « The experience of presence : Factor analytic insights ». In : *Presence* 10.3 (2001), p. 266–281.
- [126] T. W. SCHUBERT. « The sense of presence in virtual environments ». In : *Zeitschrift für Medienpsychologie* 15.2 (2003), p. 69–71.
- [127] E. SERRANO, A. MUÑOZ et J. BOTIA. « An approach to debug interactions in multi-agent system software tests ». In : *Information Sciences* 205 (2012), p. 38–57.
- [128] B. SKINNER. *Science and human behavior*. Free Press, 1953.
- [129] M. SLATER, M. USOH et A. STEED. « Depth of presence in virtual environments ». In : *Presence-Teleoperators and Virtual Environments* 3.2 (1994), p. 130–144.
- [130] J.-B. SOYEZ et al. « A Methodology to Engineer and Validate Dynamic Multi-level Multi-agent Based Simulations ». In : *Multi-agent-based simulation XIII*. T. 7838. LNAI, Springer, 2013, p. 130–142.
- [131] T. A. STOFFREGEN et al. « On the nature and evaluation of fidelity in virtual environments ». In : *Virtual and adaptive environments : Applications, implications, and human performance issues* (2003), p. 111–128.
- [132] N. TINBERGEN. « On aims and methods of ethology ». In : *Zeitschrift für Tierpsychologie* 20.4 (1963), p. 410–433.
- [133] H. C. TRIANDIS. *Culture and social behavior*. McGraw-Hill Book Company, 1994.
- [134] B. ULICNY et D. THALMANN. « Crowd simulation for interactive virtual environments and VR training systems ». In : *Computer Animation and Simulation 2001*. Springer, 2001, p. 163–170.
- [135] A. VEREMME et al. « Evidential calibration process of multi-agent based system : An application to forensic entomology ». In : *Expert Systems with Applications* 39.3 (2012), p. 2361–2374.
- [136] P. VERMERSCH. « Introspection as practice ». In : *Journal of consciousness studies* 6.2-3 (1999), p. 2–3.
- [137] R. VERTEGAAL et al. « Eye gaze patterns in conversations : there is more to conversational agents than meets the eyes ». In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2001, p. 301–308.
- [138] J. H. WARD JR. « Hierarchical grouping to optimize an objective function ». In : *Journal of the American statistical association* 58.301 (1963), p. 236–244.
- [139] B. G. WITMER et M. J. SINGER. « Measuring presence in virtual environments : A presence questionnaire ». In : *Presence* 7.3 (1998), p. 225–240.

Annexes

A] Questionnaires

COMPORTEMENTS DE CONDUITE

Personne n'est parfait. Même les meilleurs conducteurs font des erreurs, des bêtises ou s'arrangent parfois avec les règles. Certains de ces comportements sont banals, mais certains peuvent être potentiellement dangereux.

Pour chacune des questions suivantes, nous vous demandons d'indiquer la fréquence à laquelle vous avez eu ce type de comportement. Basez votre jugement sur ce que vous vous rappelez de votre conduite **depuis une année**. Vous devez pour cela cocher la case correspondante.

Vous ne pourrez, bien entendu, pas donner de réponse précise : seules vos impressions générales nous intéressent. Ne passez donc pas trop de temps à réfléchir à chaque question. Donnez simplement votre impression aussi rapidement que possible en cochant la case qui vous paraît la plus appropriée. Si vous changez d'avis après avoir répondu, grisez simplement votre première réponse, et cochez une autre case.

Lorsque vous aurez terminé, veuillez vérifier que vous avez bien traité chaque question. Nous vous demandons d'être aussi honnête que possible. Vos réponses sont anonymes.

Depuis une année, à quelle fréquence avez-vous eu ces comportements ? Cochez la case correspondante :

1. Vous essayez de démarrer au feu en 3^{ème} :
jamais rarement parfois assez souvent fréquemment presque toujours
2. Vous jetez un coup d'œil sur votre compteur de vitesse et vous vous rendez compte que vous roulez plus vite que la limite autorisée :
jamais rarement parfois assez souvent fréquemment presque toujours
3. Vous avez fermé le véhicule en laissant les clés à l'intérieur :
jamais rarement parfois assez souvent fréquemment presque toujours
4. Vous êtes impatient(e) du fait de la lenteur d'un véhicule sur la voie médiane et vous le doublez par la droite :
jamais rarement parfois assez souvent fréquemment presque toujours
5. Sur une route de campagne, vous conduisez aussi vite avec les codes qu'avec les phares:
jamais rarement parfois assez souvent fréquemment presque toujours
6. Vous tentez de partir sans avoir mis le contact :
jamais rarement parfois assez souvent fréquemment presque toujours
7. Vous collez le véhicule devant vous ou bien vous lui faites des appels de phares pour lui dire d'aller plus vite ou de changer de voie :
jamais rarement parfois assez souvent fréquemment presque toujours
8. Vous ne vous souvenez plus où vous avez garé votre véhicule dans un grand parking :
jamais rarement parfois assez souvent fréquemment presque toujours
9. Vous êtes distrait(e) ou préoccupé(e) et vous réalisez tardivement que le véhicule devant vous a ralenti. Vous devez freiner brutalement pour éviter la collision :
jamais rarement parfois assez souvent fréquemment presque toujours
10. Vous tentez de mettre les essuie-glaces, mais vous allumez les phares ou vice-versa :

jamais rarement parfois assez souvent fréquemment presque toujours

11. Vous tournez à droite pour entrer sur une route principale et vous croisez le chemin d'un véhicule que vous n'aviez pas vu ou dont vous aviez mal estimé la vitesse :

jamais rarement parfois assez souvent fréquemment presque toujours

12. Vous estimez mal vos distances dans un parking et vous percuetez réellement (ou presque) le véhicule voisin :

jamais rarement parfois assez souvent fréquemment presque toujours

13. Vous réalisez soudain que vous n'avez pas de souvenir du trajet que vous venez juste de parcourir :

jamais rarement parfois assez souvent fréquemment presque toujours

14. Vous manquez votre sortie et vous devez faire un long détour :

jamais rarement parfois assez souvent fréquemment presque toujours

15. Vous oubliez en quel rapport de vitesse vous roulez et vous devez vérifier avec votre main :

jamais rarement parfois assez souvent fréquemment presque toujours

16. Sur une route à deux voies (avec ligne blanche continue), vous êtes bloqué(e) derrière un véhicule lent. Frustré(e), vous tentez de le doubler malgré l'interdiction :

jamais rarement parfois assez souvent fréquemment presque toujours

17. En ayant l'intention de vous rendre à la destination A, vous vous retrouvez sur la route pour une destination B qui est votre destination habituelle :

jamais rarement parfois assez souvent fréquemment presque toujours

18. Vous croisez les doigts et passez alors que le feu vient de passer au rouge :

jamais rarement parfois assez souvent fréquemment presque toujours

19. Vous êtes irrité(e) par le comportement d'un autre conducteur et vous décidez de lui donner la chasse afin de lui montrer ce que vous pensez :

jamais rarement parfois assez souvent fréquemment presque toujours

20. Vous essayez de doubler sans avoir regardé dans vos rétroviseurs et vous vous faites klaxonner par l'automobiliste derrière vous qui vient juste d'amorcer sa manœuvre de dépassement :

jamais rarement parfois assez souvent fréquemment presque toujours

21. Délibérément, vous ne respectez pas la limitation de vitesse tard dans la nuit ou très tôt le matin :

jamais rarement parfois assez souvent fréquemment presque toujours

22. Contrairement à la réglementation, vous circulez sans avoir vos papiers (permis de conduire, carte grise, assurance...) :

jamais rarement parfois assez souvent fréquemment presque toujours

23. Perdu(e) dans vos pensées, vous avez oublié que vous roulez en pleins phares jusqu'à ce qu'un autre automobiliste vous le signale :

jamais rarement parfois assez souvent fréquemment presque toujours

24. En tournant à droite, vous manquez de renverser un cycliste qui était sur votre droite :
jamais rarement parfois assez souvent fréquemment presque toujours
25. Dans une file d'attente pour tourner à droite sur une voie principale, vous faites tellement attention au trafic venant de la gauche que vous manquez de percuter le véhicule devant vous :
jamais rarement parfois assez souvent fréquemment presque toujours
26. Vous prenez le volant en sortant d'un restaurant, d'un café ou d'une soirée bien que vous réalisiez que vous êtes peut-être au-dessus du taux légal d'alcoolémie :
jamais rarement parfois assez souvent fréquemment presque toujours
27. Vous avez une aversion pour une catégorie particulière d'usager de la route et/ou une catégorie de comportements et vous indiquez votre hostilité par n'importe quel moyen :
jamais rarement parfois assez souvent fréquemment presque toujours
28. Perdu(e) dans vos pensées, vous ne faites pas attention à la présence d'un piéton qui attend à un passage protégé, ou bien vous ne remarquez pas que le feu vient de passer au rouge :
jamais rarement parfois assez souvent fréquemment presque toujours
29. Vous vous garez près d'un trottoir sur lequel il y a une ligne jaune continue (interdiction de stationner) et vous risquez une amende :
jamais rarement parfois assez souvent fréquemment presque toujours
30. Vous sous-estimez la vitesse d'un véhicule approchant lorsque vous dépassez :
jamais rarement parfois assez souvent fréquemment presque toujours
31. En reculant, vous percuetez quelque chose que vous n'avez pas vu :
jamais rarement parfois assez souvent fréquemment presque toujours
32. Vous ne remarquez pas qu'une personne sort de derrière un bus ou un véhicule stationné jusqu'à ce qu'il soit presque trop tard :
jamais rarement parfois assez souvent fréquemment presque toujours
33. Vous planifiez si mal votre trajet que vous vous retrouvez dans un embouteillage que vous auriez pu éviter :
jamais rarement parfois assez souvent fréquemment presque toujours
34. Vous doublez une file continue de véhicules arrêtés ou au ralenti et vous découvrez qu'ils faisaient la queue :
jamais rarement parfois assez souvent fréquemment presque toujours
35. Sur une voie rapide, vous doublez un véhicule lent par la droite :
jamais rarement parfois assez souvent fréquemment presque toujours
36. Vous coupez un virage à gauche et vous devez faire une violente embardée pour éviter le véhicule qui arrive en face :
jamais rarement parfois assez souvent fréquemment presque toujours
37. Vous pénétrez par une mauvaise voie dans un rond-point ou à l'approche d'une intersection :
jamais rarement parfois assez souvent fréquemment presque toujours

38. Vous lisez mal les panneaux et vous prenez la mauvaise sortie à un rond-point :

jamais rarement parfois assez souvent fréquemment presque toujours

39. Vous refusez la priorité au bus qui vient de signaler le départ de son arrêt :

jamais rarement parfois assez souvent fréquemment presque toujours

40. Vous ignorez les panneaux de priorité et vous évitez de justesse une collision avec un véhicule prioritaire :

jamais rarement parfois assez souvent fréquemment presque toujours

41. Vous oubliez de contrôler dans vos rétroviseurs avant de sortir d'un stationnement, de changer de voie, de tourner... :

jamais rarement parfois assez souvent fréquemment presque toujours

42. Vous tentez de doubler un véhicule et vous n'aviez pas remarqué qu'il avait signalé qu'il allait tourner à gauche :

jamais rarement parfois assez souvent fréquemment presque toujours

43. Vous prenez délibérément une rue déserte en sens interdit :

jamais rarement parfois assez souvent fréquemment presque toujours

44. Lorsque vous conduisez tard dans la nuit sur une route déserte, vous ne respectez pas les feux rouges :

jamais rarement parfois assez souvent fréquemment presque toujours

45. Vous conduisez en jetant un œil sur la route et l'autre sur une carte routière, le réglage de votre radio ou l'insertion d'une cassette :

jamais rarement parfois assez souvent fréquemment presque toujours

46. Vous ne remarquez pas les piétons qui traversent lorsque vous tournez d'une rue principale vers une rue secondaire :

jamais rarement parfois assez souvent fréquemment presque toujours

47. Vous vous trouvez impliqué(e) dans une « course non officielle » avec d'autres conducteurs :

jamais rarement parfois assez souvent fréquemment presque toujours

48. Vous êtes dans une rue étroite dans laquelle les véhicules ne peuvent se croiser. Vous forcez le passage au véhicule qui arrive en face alors qu'il est prioritaire :

jamais rarement parfois assez souvent fréquemment presque toujours

49. Vous freinez trop brutalement sur une route glissante et/ou vous tournez le volant dans le mauvais sens lors d'un dérapage :

jamais rarement parfois assez souvent fréquemment presque toujours

50. Vous estimez mal le temps qu'il vous faut pour tourner à gauche à une intersection et vous manquez d'avoir une collision :

jamais rarement parfois assez souvent fréquemment presque toujours

Veillez vérifier avoir répondu à toutes les questions.

Questionnaire de jeux vidéo

- Quelle est votre fréquence de jeu?
 - Tous les jours
 - Plusieurs fois par semaine
 - Quelques fois par mois
 - Moins fréquemment
 - Jamais
- Depuis quel âge jouez-vous aux jeux vidéo ? _____
- Quels sont les supports que vous utilisez pour jouer aux jeux vidéo ? [Questions à choix multiples]
 - Ordinateur (jeux sur CD ou sur internet)
 - Console de salon (Xbox, Playstation ou Wii par exemple)
 - Console portable (Game Boy, PSP par exemple)
 - Autre support (Téléphone portable par exemple), précisez : _____
- À quels types de jeux jouez-vous? [Questions à choix multiples]
 - Jeux de tir à la première personne (FPS = First Personnal Shooter, Counter-Strike par exemple)
 - Jeux de stratégie (Age Of Empire par exemple)
 - Jeux de plateforme (Rayman par exemple)
 - Jeux d'aventure/action (Assassin's Creed par exemple)
 - Jeux de Rôle (RPG = Role Playing Game, Final Fantasy par exemple)
 - Jeux ludo-éducatifs (Adibou par exemple)
 - Jeux de gestion (Zoo Tycoon par exemple)
 - Jeux de simulation (Sims par exemple)
 - Jeux de sports (Fifa ou PES par exemple)
 - Jeux de courses (Gran Turismo ou Mario Kart par exemple)
 - Jeux de rôle massivement multi-joueurs (MMORPG, World Of Warcraft par exemple)
 - Jeux sur internet (online games, travian par exemple)
 - Jeux d'activité physique ou sportive (EyeToy ou Wii Sport par exemple)
 - Autres, précisez : _____
- Comment jouez-vous ? [Questions à choix multiples]
 - Hors ligne seul
 - Hors ligne à plusieurs
 - En ligne seul
 - En ligne à plusieurs

Questionnaire d'état civil et conducteur

Nom : _____ Prénom : _____ Sexe : __ Date de naissance : __/__/__

- Quel est votre statut ?
 - Étudiant
 - Étudiant salarié
 - Sans activité professionnelle
 - Demandeur d'emploi
 - Actif (exerce un emploi)
 - Retraité
- Quel est le diplôme le plus élevé que vous ayez obtenu ?
 - Brevet
 - CAP, BEP
 - Baccalauréat
 - Diplôme niveau "bac +2" (DUT, BTS par exemple)
 - Diplôme niveau "bac +3" (Licence par exemple)
 - Diplôme niveau "bac +4" (Ancienne Maitrise par exemple)
 - Diplôme niveau "bac +5" (Master par exemple)
 - Diplôme niveau "bac +8" (Doctorat par exemple)
- Combien d'enfants avez-vous ? ____

1. Votre vue est-elle corrigée ? ____
2. Portez-vous des lunettes pour conduire ? ____
3. Portez-vous des lentilles pour conduire ? ____
4. Possédez-vous un permis de conduire (type B) valide ? ____
5. Quelle est l'année d'obtention de votre permis de conduire (type B) ? ____
6. Combien de kilomètres avez-vous parcouru l'année dernière en France ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins de 4000 km	Entre 4000 et 6000 km	Entre 6000 et 8000 km	Plus de 8000 km

7. Possédez-vous un permis de conduire d'autres véhicules ? ____
Si oui, le(s)quel(s) ? _____
8. Possédez-vous une moto ou un scooter ? ____
Si oui, combien de kilomètres faites-vous par an en moto ou en scooter en moyenne ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins de 4000 km	Entre 4000 et 6000 km	Entre 6000 et 8000 km	Plus de 8000 km

9. Maîtrisez-vous des techniques de pilotage avancé ? ____
10. Combien d'amendes pour des infractions de la route avez-vous eu dans les deux dernières années (en dehors des amendes concernant le stationnement) ? ____
11. Combien de fois avez-vous été impliqué dans des accidents dans lesquels vous avez été reconnu responsable alors que vous conduisiez une voiture ? ____

12. Si vous possédez une voiture, est-elle à boîte de vitesse automatique ? ____

13. Si non, combien de vitesse a-t-elle ? ____

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 rapports et marche arrière	5 rapports et marche arrière	6 rapports et marche arrière

14. Si vous possédez une voiture, a-t-elle un système antiblocage des roues (ABS) ? ____

15. Si vous possédez une voiture, a-t-elle un régulateur de vitesse ? ____

16. Si vous possédez une voiture, a-t-elle la direction assistée ? ____

À quelle fréquence conduisez-vous une voiture ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins	1 à 2 fois	3-4 fois	3-4 fois	Plus
	par mois	par mois	par semaine	

À quelle fréquence conduisez-vous en ville ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins	1 à 2 fois	3-4 fois	3-4 fois	Plus
	par mois	par mois	par semaine	

À quelle fréquence conduisez-vous sur des routes de campagne ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins	1 à 2 fois	3-4 fois	3-4 fois	Plus
	par mois	par mois	par semaine	

À quelle fréquence conduisez-vous sur des routes départementales ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins	1 à 2 fois	3-4 fois	3-4 fois	Plus
	par mois	par mois	par semaine	

À quelle fréquence conduisez-vous sur des voies rapides ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins	1 à 2 fois	3-4 fois	3-4 fois	Plus
	par mois	par mois	par semaine	

À quelle fréquence conduisez-vous avec un navigateur GPS ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moins	1 à 2 fois	3-4 fois	3-4 fois	Plus
	par mois	par mois	par semaine	

Questionnaire d'état de présence (QEP)

Décrivez votre expérience dans l'environnement en marquant d'un "X" la case appropriée de l'échelle en 7 points, et ce en accord avec le contenu de la question et les étiquettes descriptives. Veuillez prendre en compte l'échelle en entier lorsque vous inscrivez vos réponses, surtout lorsque des niveaux intermédiaires sont en jeu. Répondez aux questions indépendamment les unes des autres et dans l'ordre dans lequel ils apparaissent. **Ne** sautez **pas** de questions et **ne** retournez **pas** à une question précédente afin de modifier votre réponse.

EN FONCTION DE L'ENVIRONNEMENT DANS LEQUEL VOUS ÉTIEZ

1. Dans quelle mesure étiez-vous capable de contrôler les événements ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

2. Dans quelle mesure l'environnement était-il réactif (sensible) aux actions que vous y faisiez ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas réactif			Modérément réactif			Complètement réactif

3. Dans quelle mesure vos interactions avec l'environnement vous semblaient-elles naturelles ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Extrêmement artificielles			À mi-chemin			Complètement naturelles

4. Dans quelle mesure les aspects visuels de l'environnement vous invitaient-ils à vous y impliquer ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

5. Dans quelle mesure les mécanismes permettant votre mouvement dans l'environnement vous semblaient-ils naturels ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Extrêmement artificiels			À mi-chemin			Complètement naturels

6. Dans quelle mesure vos sens étaient-ils trompés par le réalisme du mouvement des objets à travers l'espace ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout trompés			Modérément			Très trompés

7. Dans quelle mesure les expériences que vous avez vécues dans l'environnement virtuel ressemblaient-elles à celles de l'environnement réel ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas ressemblant			Modérément ressemblant			Très ressemblant

8. Étiez-vous capable d'anticiper les conséquences des mouvements que vous faisiez ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

9. Jusqu'à quel point étiez-vous en mesure d'explorer activement l'environnement de façon visuelle ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

10. Jusqu'à quel point la sensation de déplacement à l'intérieur de l'environnement virtuel était-elle confondante (réaliste) ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas confondante			Modérément confondante			Très confondante

11. À quelle distance pouviez-vous examiner les objets ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout proche			Proche plutôt			Très proche

12. Jusqu'à quel point pouviez-vous examiner les objets sous différents angles ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

13. Jusqu'à quel point étiez-vous impliqué(e) dans l'expérience vécue dans l'environnement virtuel ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout engagé(e)			Moyennement engagé(e)			Complètement absorbé(e)

14. Jusqu'à quel point avez-vous ressenti un délai séparant vos actions de leurs conséquences ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Aucun délai			Délai modéré			Long délai

15. À quel rythme vous êtes-vous adapté(e) à l'expérience vécue dans l'environnement virtuel ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas adapté(e) du tout			Lentement			En moins d'une minute

16. En termes d'interactions et de déplacements dans l'environnement virtuel, jusqu'à quel point vous sentiez-vous compétent(e) à la fin de l'expérience ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas compétent(e)			Raisonnement compétent(e)			Très compétent(e)

17. Jusqu'à quel point la qualité visuelle de l'appareillage graphique vous a-t-elle incommodé(e) dans l'exécution des tâches requises ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez incommodé(e)			Tâches complètement empêchées

18. Dans quelle mesure les mécanismes de contrôle de votre mouvement ont-ils interféré avec l'exécution des tâches requises ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez interféré			Grandement interféré

19. Jusqu'à quel point êtes-vous parvenu(e) à vous concentrer sur les tâches requises plutôt que sur les mécanismes utilisés pour effectuer lesdites tâches ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

20. Dans quelle mesure les aspects auditifs de l'environnement vous invitaient-ils à vous y impliquer ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

21. Dans quelle mesure arriviez-vous à identifier correctement les sons produits dans l'environnement ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

22. Dans quelle mesure arriviez-vous à localiser correctement les sons produits dans l'environnement ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Assez			Complètement

Je n'étais pas conscient de mon environnement réel.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

Dans le monde généré par l'ordinateur, j'ai eu le sentiment "d'y être".

Pas du tout -3 -2 -1 0 +1 +2 +3 Beaucoup

D'une certaine façon, j'ai eu l'impression que le monde virtuel m'entourait.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

Je me suis senti présent dans l'espace virtuel.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

Je faisais toujours attention à l'environnement réel.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

Le monde virtuel semblait plus réaliste que le monde réel.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

J'avais l'impression que j'étais juste en train de percevoir des images.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

J'étais complètement captivé par le monde virtuel.

Pas du tout d'accord -3 -2 -1 0 +1 +2 +3 Tout à fait d'accord

Questionnaire sur l'implication dans le jeu vidéo

Le questionnaire suivant a pour but d'évaluer votre implication dans le jeu vidéo de manière **générale**. Les questions suivantes demandent pour chacune une unique réponse que vous marquerez d'une croix dans la case de votre choix sur une échelle comportant 7 possibilités allant de « Oui » à « Non ».

1. Je perds la notion du temps :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

2. Les événements semblent se dérouler de manière automatique :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

3. Je me sens différent :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

4. Je me sens effrayé :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

5. Le jeu semble réel :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

6. Si quelqu'un me parle, je ne l'entends pas :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

7. Je deviens irritable :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

8. Le temps semble se mettre en pause ou s'arrêter :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

9. Je me sens comme déconnecté de la réalité :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

10. Je ne réponds pas quand quelqu'un me parle :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

11. Je ne me rends pas compte quand je fatigue :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

12. Jouer semble automatique :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

13. Mes pensées vont vites :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

14. Je perds conscience d'où je suis :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

15. Je joue sans réfléchir à comment jouer :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

16. Jouer me fait me sentir calme :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

17. Je joue plus longtemps que ce que j'avais prévu :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

18. Je suis vraiment dans le jeu :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

19. J'ai l'impression que je ne peux plus m'arrêter de jouer :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oui			Plus ou moins			Non

Questionnaire sur la Propension à l'Immersion (QPI)

Indiquer votre réponse en inscrivant un "X" dans la case appropriée de l'échelle en 7 points. Veuillez prendre en compte l'échelle en entier lorsque vous inscrivez vos réponses, surtout lorsque des niveaux intermédiaires sont en jeu. Par exemple, si votre réponse est "une fois ou deux", la deuxième case à partir de la gauche devrait être utilisée. Si votre réponse est "plusieurs fois mais pas extrêmement souvent", alors la sixième case (ou la deuxième à partir de la droite) devrait être utilisée.

1. Devenez-vous facilement et profondément absorbé(e) lorsque vous visionnez des films ou des téléromans ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

2. Vous arrive-t-il d'être tellement absorbé(e) dans une émission de télévision ou un livre que les gens autour de vous ont de la difficulté à vous en tirer ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

3. Jusqu'à quel point vous sentez-vous mentalement éveillé(e) ou vif/vive d'esprit en ce moment même ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas éveillé(e)			Modérément			Complètement éveillé(e)

4. Vous arrive-t-il d'être tellement absorbé(e) dans un film que vous n'êtes pas conscient(e) des choses qui se passent autour de vous ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

5. À quelle fréquence vous arrive-t-il de vous identifier intimement avec les personnages d'une histoire ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

6. Vous arrive-t-il d'être tellement absorbé(e) dans un jeu vidéo que vous avez l'impression d'être à l'intérieur du jeu plutôt qu'en train de manœuvrer des manettes de jeu et de regarder un écran ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

7. Comment vous sentez-vous au plan de la forme physique aujourd'hui ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas en forme			Modérément bien			Extrêmement bien

8. Dans quelle mesure êtes-vous efficace pour vous couper des distractions extérieures lorsque vous êtes en train de faire quelque chose ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas très bon(ne)			Plutôt bon(ne)			Très bon(ne)

9. Lorsque vous assistez à un match sportif, vous arrive-t-il de devenir tellement pris(e) par le match que vous réagissez comme si vous étiez un des joueurs ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

10. Vous arrive-t-il d'être tellement absorbé(e) dans des rêveries, en plein jour, que vous n'êtes pas conscient des choses qui se passent autour de vous ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

11. Vous arrive-t-il d'avoir des rêves qui semblent tellement réels que vous vous sentez désorienté(e) au réveil ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

12. Quand vous faites du sport, vous arrive-t-il d'être à ce point absorbé(e) que vous perdez la notion du temps ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

13. Dans quelle mesure arrivez-vous à vous concentrer lors d'activités plaisantes ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément bien			Très bien

14. À quelle fréquence jouez-vous à des jeux vidéo ? (Souvent devrait correspondre à tous les jours, ou tous les deux jours, en moyenne)

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

15. Vous est-il déjà arrivé(e) d'être excité(e) lorsque vous visionnez une scène de poursuite ou de combat à la télévision ou dans un film ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

16. Vous est-il déjà arrivé(e) d'être apeuré(e) par quelque chose se produisant à la télévision ou dans un film ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

17. Vous est-il déjà arrivé(e) de demeurer apeuré(e) longtemps après le visionnement d'un film d'épouvante ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

18. Vous arrive-t-il de devenir absorbé(e) à un point tel dans une activité donnée que vous en perdez la notion du temps ?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jamais			À l'occasion			Souvent

Questionnaire général de comportement

Le questionnaire suivant a pour but d'évaluer la qualité des comportements de conduite des véhicules simulés **en général**, c'est-à-dire dans l'ensemble de la simulation. Les questions suivantes demandent pour chacune une unique réponse que vous marquerez d'une croix dans la case de votre choix sur une échelle comportant 7 possibilités allant de « Pas du tout » à « Tout à fait ». Indiquez à quel point vous êtes d'accord avec les phrases suivantes :

- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules simulés en ligne droite sur route nationale (hors situation de dépassement) :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules simulés en virage (hors situation de dépassement) :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules simulés en ligne droite sur autoroute (hors situation de dépassement) :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules simulés face à un véhicule roulant à faible vitesse :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules simulés en situation de dépassement :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Des conducteurs humains peuvent avoir les comportements de conduite effectués par les conducteurs des véhicules simulés en trafic dense :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Des conducteurs humains peuvent avoir les comportements de conduite effectués par les conducteurs des véhicules simulés aux carrefours :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Des conducteurs humains peuvent avoir les comportements de conduite effectués par les véhicules roulants à faible vitesse :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

Questionnaire spécifique de comportement par phase

Le questionnaire suivant a pour but d'évaluer la qualité des comportements de conduite des véhicules simulés **étape par étape**. Un rejeu de votre trajet va vous être diffusé et des questions vous seront posées. Ces questions demandent pour chacune une unique réponse que vous marquerez d'une croix dans la case de votre choix sur une échelle comportant 7 possibilités allant de « Pas du tout » à « Tout à fait ». Indiquez à quel point vous êtes d'accord avec les phrases suivantes :

▪ Étape 1

- Consignes : observer le comportement du véhicule devant vous.
- Un conducteur humain peut avoir le comportement de déboîtement effectué par le conducteur du véhicule devant vous dans cette situation de dépassement :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir le comportement de rabattement effectué par le conducteur du véhicule devant vous dans cette situation de dépassement :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ Étape 2

- Consignes : observer le comportement du véhicule devant.
- Un conducteur humain peut avoir le comportement de conduite effectué par le conducteur du véhicule devant vous dans cette situation :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ Étape 3

- Consignes : observer le véhicule devant.
- Un conducteur humain peut avoir le comportement de freinage effectué par le conducteur du véhicule devant vous dans cette situation :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir le comportement de suivi en file effectué par le conducteur du véhicule devant vous dans cette situation de bouchon :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir le comportement d'accélération effectué par le conducteur du véhicule devant vous dans cette situation de bouchon :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ Étape 4

- Consignes : observer le véhicule devant.
- Un conducteur humain peut avoir le comportement de conduite effectué par le conducteur du véhicule devant vous quand il tourne à gauche dans cette situation :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ **Étape 5**

- Consignes : observer les véhicules venant de gauche et de droite ainsi qu'en face de vous au carrefour.
- Un conducteur humain peut avoir le comportement de conduite effectué par le conducteur du véhicule suivant dans cette situation :
 - Premier véhicule venant de droite et allant tout droit :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Second véhicule venant de gauche et allant à droite :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Véhicule arrêté au stop venant en face :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ **Étape 6**

- Consignes : observer le véhicule devant.
- Un conducteur humain peut avoir le comportement de conduite effectué par le conducteur du véhicule roulant à faible vitesse devant vous dans cette situation de ligne droite :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ **Étape 7**

- Consignes : observer les comportements des véhicules venant en face et de celui venant de gauche.
- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules en face de vous dans cette situation au carrefour :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ces comportements sont courants :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Un conducteur humain peut avoir le comportement de conduite effectué par le conducteur du camion venant de gauche dans cette situation au carrefour :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ **Étape 8**

- Consignes : regarder les véhicules sur l'autoroute.
- Un conducteur humain peut avoir les comportements de conduite effectués par les conducteurs des véhicules sur l'autoroute dans cette situation d'insertion :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ces comportements sont courants :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

▪ **Étape 9**

- Consignes : observer le véhicule devant.
- Un conducteur humain peut avoir le comportement de conduite effectué par le conducteur du véhicule devant vous sur l'autoroute dans cette situation :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ce comportement est courant :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

<input type="checkbox"/> Pas du tout	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Modérément	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Tout à fait	<input type="checkbox"/> Sans opinion
--------------------------------------	--------------------------	--------------------------	-------------------------------------	--------------------------	--------------------------	--------------------------------------	---------------------------------------

▪ **Étape 10**

- Consignes : observer les véhicules doublant les véhicules roulants à faible vitesse.
- Un conducteur humain peut avoir les comportements de dépassement effectués par les conducteurs des véhicules doublant les véhicules roulants à faible vitesse sur l'autoroute dans cette situation :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

- Si vous avez coché entre «Modérément» et «Tout à fait» : ces comportements sont courants :

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pas du tout			Modérément			Tout à fait	Sans opinion

B] Méthode

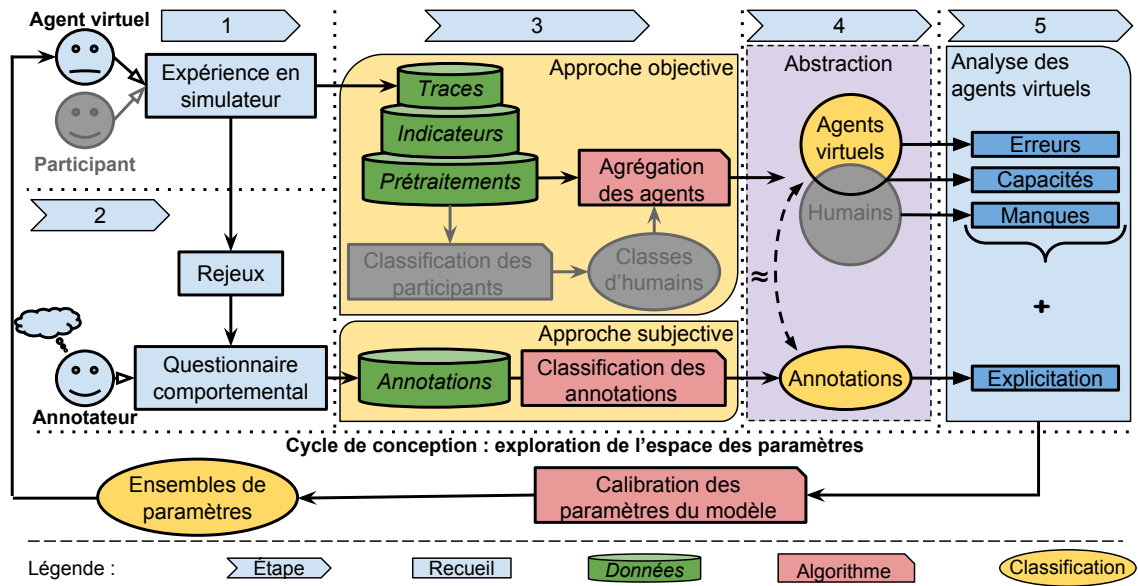


Figure .11 – Cycle de calibration au sein de la méthode générale. Les parties grisées correspondent aux traitements des données non répétés lors du cycle (*i.e.* déjà effectués lors d'une première évaluation).

C] Expérimentation

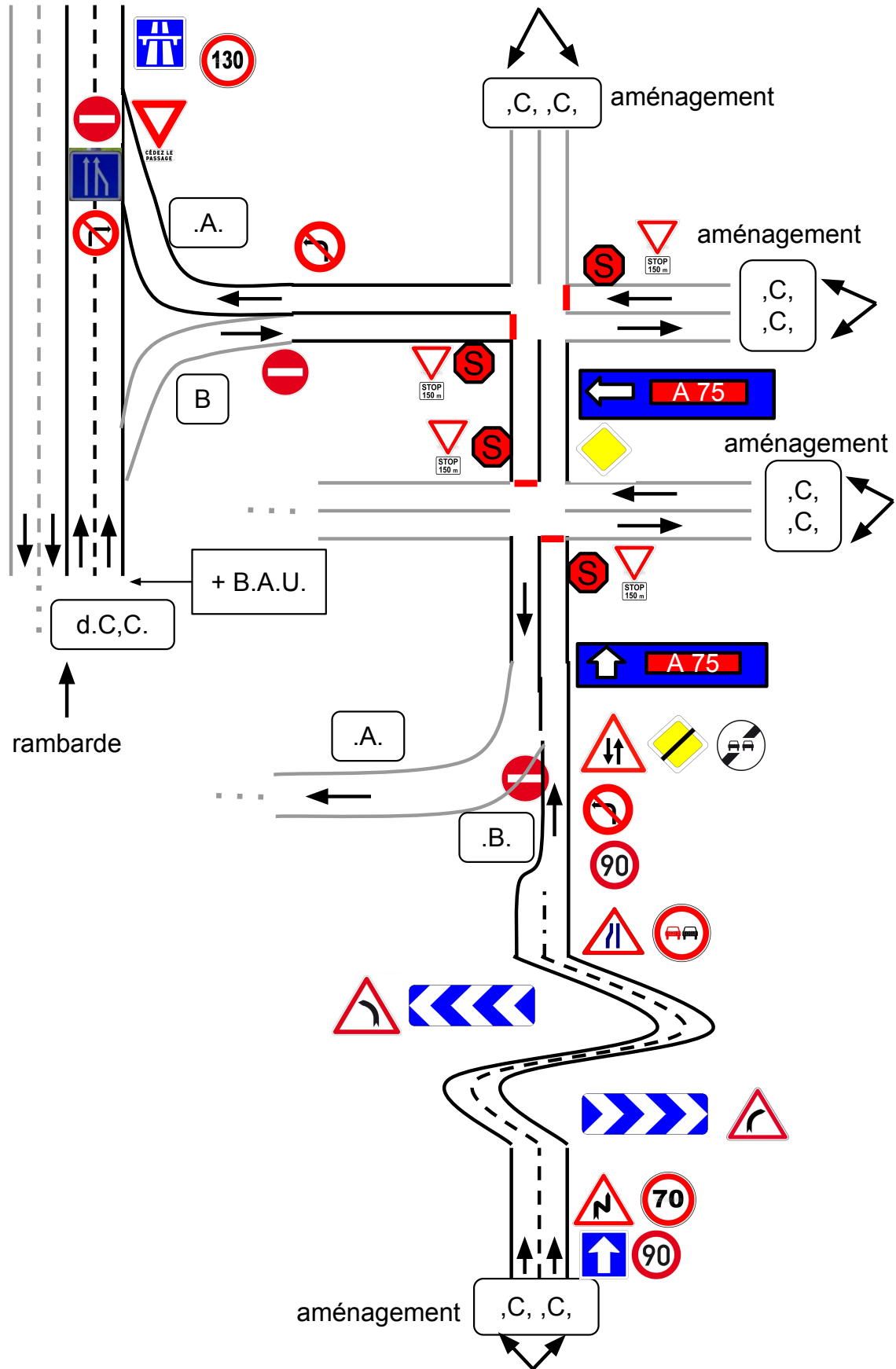


Figure .12 – Circuit détaillé par phase

Index

A	
Acteurs principaux	39, 78
Annotations	14, 39, 43–46, 83, 98
ARCHISIM	4, 41, 73, 79

C	
Calibration	63, 95
Circuit	41, 75, 162
Classification	29
Agrégation	46, 53, 80
Analyse en Composantes Principales (ACP)	49
Classification non-supervisée	30
Calinski-Harabasz (CH)	31
K-moyennes	31
K-moyennes++	52
Simple Structure Index (SSI)	31
Comparaison de classification	33, 118
Adjusted Rand Index (ARI)	33
Comptage par paires	33
Rand Index (RI)	33
Distance	
Distance de Mahalanobis	32
Dynamic Time Warping (DTW)	32, 48
Classification d’annotations	55, 93, 98, 104
Classification de traces	51, 79, 92, 98, 100, 104
Comportement	7
Abstraction	27, 39, 43
Comportement humain	8
Niveaux de comportement	9, 56, 118
Comportement de conduite	16
Decision-Making Style Questionnaire (DMSQ)	17
Driving Behavior Questionnaire (DBQ)	17
Driving Behaviour Inventory (DBI)	16
Driving Style Questionnaire (DSQ)	16
Cycle de conception	63, 66, 95, 160

E	
Échelle	
Échelle macroscopique	3
Échelle microscopique	3
Évaluation	
Évaluation objective	3
Évaluation subjective	3

L

Langage

Matlab	88
Octave GNU	88
R	88
Scilab	88
TramineR	88
Waikato Environment for Knowledge Analysis (WEKA)	89

P

Phase de conduite	41, 75, 76
Propension à l'immersion	15
Implication dans les jeux vidéo	15, 82
Tendance a priori à l'immersion	15, 82
Présence	20
Crédibilité comportementale	22
Questionnaire Général de Comportement (QGC)	78, 84, 109
Questionnaire Spécifique de Comportement (QSC)	78, 85, 109
Fidélité psychologique	23
Igroup Presence Questionnaire (IPQ)	21
ITC-Sense of Presence Inventory (ITC-SOPI)	21
Presence Questionnaire (PQ)	21, 77
Slater-Usuh-Steed Questionnaire (SUS)	21

R

Rejeux	39, 44, 77, 83
Réalité virtuelle (RV)	2

S

Sciences Humaines et Sociales (SHS)

Entretiens d'autoconfrontation	14
Sciences Humaines et Sociales (SHS)	12
Discussion	14
Effet de présence	12
Qualité d'immersion	12
Score de représentativité	64, 94, 102
Score de type de classe	60, 93, 101
Situation de conduite	41, 75, 91, 97, 105

T

Taux de confiance	58, 94, 102
Traces de simulation	11, 29, 37, 39, 41, 43, 46, 79, 92
Type de classes	58

V

Variables

Paramètres des agents	78
Variables communes	80
Variables du véhicule principal	81
Variables du véhicule simulé	81
Variables générales	79
