Université de Lille

Thèse

pour obtenir le grade de :

Docteur de l'Université de Lille

dans la spécialité

« Informatique »

par

Jarod Vanderlynden

Modélisation multi-agents des comportements clients pour une stratégie marketing efficace.

Thèse soutenue le 25 Novembre 2025 devant le jury composé de :

M.	Nicolas Sabouret	Professeur, LISN, Université Paris-Saclay	(Rapporteur)
M	Sébastien Picault	CR INRAE HDR, BIOEPAR INRAE Nantes	(Rapporteur)
Mme	Suzanne Pinson	Professeur émérite, LAMSADE, Paris-Dauphine PSL	(Examinatrice)
Mme	Laetitia Jourdan	Professeur, CRIStAL, Université de Lille	(Présidente du jury)
M.	Romain Warlop	Directeur data science, fifty-five	(Co-Directeur de Thèse)
M.	Philippe Mathieu	Professeur, CRIStAL, Université de Lille	(Directeur de Thèse)

Ecole Graduée MADIS-631, Laboratoire CRIStAL
Univ. Lille, CNRS, Centrale Lille, UMR 9189, CRIStAL, F-59000 Lille
France

Table des matières

Remerciements

Je tiens tout d'abord à exprimer ma gratitude à l'entreprise fifty-five, qui a rendu cette thèse possible dans le cadre d'une convention CIFRE. Je remercie tout particulièrement Romain Warlop pour son accompagnement, sa disponibilité et ses conseils tout au long de ces années. Sa vision, son exigence et sa bienveillance ont grandement contribué à la qualité et à la pertinence de ce travail. Je remercie également l'ensemble de mes collègues de fifty-five pour leur soutien, leur bonne humeur et les nombreux échanges enrichissants qui ont rythmé ces années de recherche.

Je tiens à exprimer ma profonde reconnaissance à Philippe Mathieu, mon directeur de thèse, pour son encadrement, ses conseils avisés et sa confiance. Son expertise, sa rigueur scientifique et sa bienveillance ont été déterminants dans la conduite et l'aboutissement de ce travail.

Je souhaite également remercier l'équipe SMAC du laboratoire CRIStAL pour son accueil chaleureux et son environnement de recherche stimulant. J'adresse une pensée particulière à Philippe Mathieu, Antoine Nongaillard, Anne-Cécile Carron et Maxime Morge pour leurs conseils et leur accompagnement, ainsi qu'à mes collègues doctorants, Ellie Beauprez, Jules Bompard et Erwan Martin pour leur soutien et leur bonne humeur au quotidien.

Je tiens également à remercier chaleureusement les membres du jury de thèse pour le temps qu'ils ont consacré à l'évaluation de ce travail et pour la richesse de leurs retours. Je remercie tout particulièrement Nicolas Sabouret et Sébastien Picault, pour avoir accepté la tâche exigeante de rapporter cette thèse. Leurs remarques précises, leur bienveillance et leur flexibilité dans le processus d'évaluation ont été très appréciées. Leur lecture attentive et leurs suggestions ont permis d'améliorer la clarté et la portée de ce travail, et je leur en suis sincèrement reconnaissant. J'adresse également mes sincères remerciements à Suzanne Pinson et à Laetitia Jourdan, pour l'intérêt qu'elles ont porté à ce travail et pour avoir accepté de participer à ce jury en tant qu'examinatrices.

Enfin, je souhaite adresser mes remerciements les plus sincères à mes proches, dont le soutien m'a accompagné tout au long de cette thèse. Je remercie mes parents, Éric et Brigitte, ainsi que mon frère Alexis et ma sœur Charlotte, pour leur appui constant et leur confiance en toutes circonstances. Je suis également reconnaissant envers mes amis, pour leur présence, leur bonne humeur et leur capacité à m'aider à prendre du recul lorsque cela était nécessaire. Enfin, je remercie Lya pour sa patience, son écoute et sa présence bienveillante tout au long de ce parcours.

Chapitre 1

Introduction

1.1 Contexte de la thèse

Le travail de recherche présenté dans ce mémoire a été réalisé dans le cadre d'une thèse CIFRE au carrefour des *systèmes-multi agents* (SMA) et du marketing. Cette thèse met en collaboration l'équipe *systèmes multi-agents et comportements* (SMAC) du laboratoire CRIStAL de l'université de Lille et l'entreprise fifty-five un cabinet de conseil marketing centré sur l'exploitation de données. Elle se focalise sur les différentes modélisations informatiques pouvant être utilisées en marketing et propose de nouvelles approches SMA-Marketing.

1.1.1 Introduction aux systèmes multi-agents

Un Système Multi-Agents (SMA) est un système composé de plusieurs entités autonomes ou semi-autonomes, appelées agents, capables d'interagir entre elles et avec leur environnement. Chaque agent est doté de caractéristiques propres, telles que des objectifs individuels, des connaissances locales, des capacités de perception et des mécanismes de décision. Ces entités évoluent dans un environnement potentiellement dynamique, partagent parfois des ressources, et peuvent coopérer, rivaliser ou simplement coexister.

Les SMA s'inscrivent dans la catégorie des modèles de simulation centrés individus. Cette approche privilégie la description fine du comportement d'entités individuelles plutôt que des agrégats globaux. Elle permet de capturer les dynamiques fines d'interactions, souvent inaccessibles par des modèles macroscopiques ou déterministes classiques.

La dynamique globale d'un SMA émerge de la somme des comportements individuels des agents et de leurs interactions, qu'elles soient inter-agents ou entre les agents et leur environnement. Ces interactions peuvent donner lieu à des phénomènes émergents, c'est-à-dire des comportements collectifs ou des structures globales qui ne sont pas explicitement programmés mais qui résultent de la coopération, de la coordination ou de la compétition entre agents. Ces phénomènes peuvent être simples ou complexes, selon les règles locales qui gouvernent les agents. (Goldstein 1999) souligne que l'émergence im-

plique à la fois l'imprévisibilité à partir des éléments de base et la cohérence structurée à l'échelle du système.

L'intérêt principal des SMA réside précisément dans leur capacité à faire émerger des dynamiques complexes à partir de règles simples. Cette propriété les rend particulièrement adaptés à la modélisation de systèmes distribués, adaptatifs, et à forte composante décentralisée.

Exemples de modèles classiques :

— Le modèle de ségrégation de (Schelling 1971) : les agents sont initialement placés sur une grille presque remplie, chacun appartenant à un groupe (par exemple, bleu ou rouge). Chaque agent possède un seuil de tolérance : si la proportion de ses voisins appartenant au même groupe est inférieure à ce seuil, il décide de déménager. Ce modèle illustre comment des préférences individuelles faibles peuvent conduire à une ségrégation spatiale marquée, phénomène typiquement émergent.
— Le modèle Boids de (Reynolds 1987) : il simule le comportement d'un essaim d'oiseaux (ou flocking) à l'aide de règles simples de cohésion, séparation et alignement entre agents. Ce modèle reproduit fidèlement les comportements collectifs observés dans la nature sans qu'un contrôle centralisé soit nécessaire.
— Les automates cellulaires comme le Jeu de la Vie de (Conway 1970) : bien que souvent classés à part, ces systèmes partagent de nombreux traits avec les SMA. Ils montrent comment des règles locales appliquées uniformément peuvent donner lieu à des structures dynamiques complexes.
— Le modèle de la fourmi de (Langton 1986) : un agent unique se déplace sur une grille selon des règles locales simples. Malgré la simplicité du système, un comportement structuré complexe émerge au bout d'un grand nombre d'itérations, illustrant la notion d'ordre émergent.

Le modèle de marché présenté par (Arthur et al. 1999) est un exemple emblématique de système multi-agents appliqué à l'économie. Ce modèle vise à représenter un marché financier artificiel dans lequel des agents hétérogènes interagissent pour prédire et exploiter les fluctuations de prix. Chaque agent possède un ensemble de stratégies prédictives (par exemple, basées sur des tendances passées ou des signaux techniques) qu'il utilise pour estimer les prix futurs. Ces stratégies peuvent être sélectionnées, renforcées ou abandonnées selon leur performance au fil du temps, ce qui introduit un mécanisme d'apprentissage adaptatif. Contrairement aux modèles économiques classiques, les agents ne sont ni omniscients ni parfaitement rationnels : ils apprennent de manière incrémentale à partir de leur environnement. Les comportements émergents qui résultent de ce modèle, tels que la formation de bulles spéculatives, la volatilité auto-renforcée (Volatilité des prix provenant de mécanismes internes et non de chocs externes comme les annonces économiques) ou encore la coordination imparfaite entre les agents, mettent en évidence la capacité des SMA à représenter des dynamiques de marché complexes qui échappent aux équilibres statiques de la théorie néoclassique. Ce modèle a fortement influencé le champ de la finance computationnelle, ainsi que les approches SMA utilisées en économie, sociologie et écologie.

Les SMA sont étroitement liés à la théorie des systèmes complexes, dont ils sont à la fois un outil de modélisation et un exemple concret. Un système complexe désigne un ensemble composé d'entités en interaction, où les relations entre ces entités ne sont ni linéaires ni indépendantes. Autrement dit, les effets d'une interaction ne sont pas proportionnels à ses causes, et l'état global du système ne peut pas être prédit à partir des comportements individuels des composantes.

Les SMA peuvent être vus comme une implémentation concrète de systèmes adaptatifs complexes, tels que définis par (Holland 1992), où l'ordre global émerge de règles locales simples et d'interactions décentralisées. Dans la lignée des travaux de (Epstein et al. 1996), les SMA permettent d'étudier des dynamiques sociales complexes en partant de comportements individuels simples et localisés.

Les SMA modélisent ce type de systèmes en représentant explicitement chaque entité par des agents et en décrivant leurs règles d'action et d'interaction. En retour, les SMA eux-mêmes peuvent être considérés comme des systèmes complexes, car les interactions entre agents peuvent générer des dynamiques collectives difficiles à anticiper.

Parmi les dynamiques que peuvent produire les SMA, on retrouve par exemple :

— Des dynamiques chaotiques, où de petites différences dans les conditions initiales ou dans les paramètres du modèle peuvent conduire à des évolutions très différentes, rendant le comportement global imprévisible à long terme.
— Des phénomènes d'auto-organisation, où un ordre global (comme une structure spatiale, une hiérarchie ou une répartition de tâches) émerge sans qu'il n'y ait de coordination centralisée.
— Des transitions soudaines appelées bifurcations, où une modification progressive d'un paramètre du système entraîne un changement brutal de comportement global (par exemple, le passage d'une phase stable à une phase instable).
— Des formes de résilience (robustesse), où le système parvient à retrouver un fonctionnement stable après une perturbation.

Ces propriétés rendent les SMA particulièrement utiles pour étudier des phénomènes distribués, tels que la propagation d'opinions, la coordination sans autorité centrale, la dynamique de marchés, ou encore l'émergence de normes sociales. En permettant de relier les comportements individuels aux dynamiques collectives, les SMA offrent un cadre puissant pour explorer et comprendre les systèmes complexes dans de nombreux domaines (sciences sociales, économie, écologie, etc.).

Comme le souligne (Bonabeau 2002), les SMA sont particulièrement adaptés à la modélisation de systèmes sociaux complexes du fait de leur capacité à produire des phénomènes émergents difficilement modélisables par des approches top-down.

1.1.2 Le lien avec le marketing

La modélisation du comportement des consommateurs est une thématique interdisciplinaire mettant en lien de nombreux concepts. Nous choisissons de défendre l'intérêt et l'apport des SMA à cette thématique.

Les entreprises commerciales, et en particulier celles de la grande distribution, occupent une place centrale dans l'économie moderne en raison de leur rôle clé dans la chaîne d'approvisionnement, la consommation et la satisfaction des besoins des populations. Ces structures génèrent et collectent quotidiennement une quantité massive de données issues des transactions, des programmes de fidélité, des comportements d'achat, ou encore des interactions en ligne. Depuis plusieurs années, ces entreprises stockent des volumes croissants d'informations concernant leurs clients, leurs produits et leurs processus logistiques. Face à cette explosion de données, elles se tournent de plus en plus vers des technologies de big data et des solutions avancées de traitement de données afin d'optimiser leur gestion, anticiper les tendances du marché et personnaliser leur offre. Cette exploitation intelligente des données devient un levier stratégique majeur pour améliorer la performance commerciale, renforcer la relation client et maintenir leur compétitivité dans un environnement en constante évolution.

L'essor du big data et de la modélisation du comportement client soulève des questions éthiques. Alors que les entreprises commerciales exploitent les données pour anticiper les besoins, personnaliser les offres et influencer les décisions d'achat, il devient essentiel de réfléchir aux limites de ces pratiques. La collecte massive de données personnelles, parfois à l'insu des consommateurs, interroge sur le respect de la vie privée, le consentement éclairé, et l'usage équitable de l'information. Des risques comme la discrimination algorithmique, la manipulation des comportements ou l'exploitation abusive des données sensibles doivent être pris en compte.

Dans un marché de plus en plus concurrentiel, les entreprises ne peuvent plus se contenter d'optimiser leurs opérations internes : elles doivent aussi comprendre finement leur environnement concurrentiel pour conserver ou acquérir un avantage stratégique. Cela implique non seulement d'analyser les comportements et préférences des consommateurs, mais également d'anticiper les actions des marques rivales, qu'il s'agisse de politiques tarifaires, de campagnes promotionnelles ou d'investissements en visibilité. Les décisions prises par un acteur influencent inévitablement les réponses des autres, créant ainsi un système dynamique où chaque mouvement stratégique entraîne des répercussions sur l'ensemble du marché.

Dans ce contexte, l'intégration des Systèmes Multi-Agents (SMA) dans le domaine du marketing représente une approche innovante et prometteuse pour la modélisation des comportements de consommation. Contrairement aux approches purement statistiques ou centrées sur l'apprentissage automatique les SMA permettent de simuler des comportements individuels en tenant compte de l'autonomie, des interactions sociales, et des prises de décisions des agents (ici, les consommateurs). Cette granularité rend la modélisation plus réaliste, dynamique, et explicable surtout dans des environnements complexes comme ceux de la grande distribution.

De plus la montée en puissance des SMA dans le domaine du marketing est aujourd'hui bien documentée. Une analyse bibliométrique de (Romero et al. 2023) montre que les publications sur ce thème ont connu une croissance continue, notamment autour des problématiques de diffusion de l'innovation, de bouche-à-oreille et de personnalisation

client. Dans cette dynamique, (Liang et al. 2025) proposent une réflexion méthodologique sur l'utilisation des SMA comme outil d'analyse stratégique, notamment pour les petites entreprises. S'appuyant sur les environnements NetLogo et Repast, leur article sert avant tout d'introduction appliquée aux possibilités offertes par la simulation multi-agents.

Dans une perspective éthique et scientifique, un modèle SMA peut devenir un outil de recherche puissant : il permet de tester des hypothèses, de comprendre les dynamiques collectives à partir de comportements individuels, et de visualiser l'impact de certaines stratégies commerciales sur les consommateurs (ex : promotions, changements de prix, personnalisation excessive). Cela permet aussi d'explorer des scénarios dans lesquels la transparence ou la régulation des données peuvent influencer les comportements.

Du point de vue des entreprises commerciales, un tel modèle devient un outil d'aide à la décision. Il peut les aider à interpréter plus finement les données clients, à anticiper les réactions face à de nouvelles offres, ou encore à adapter leur communication en fonction des profils clients simulés. Cela va au-delà de la simple prédiction : c'est une approche explicative et exploratoire, qui offre une meilleure compréhension des mécanismes d'achat.

Enfin, en intégrant des considérations éthiques dans la conception du modèle (ex. : respect du consentement, prise en compte des biais, scénarios de régulation), on montre que la modélisation SMA peut être utilisée non seulement comme un levier économique, mais aussi comme un outil de réflexion autour de l'usage des données dans le marketing.

1.2 Etat de l'art

Dans cette section, nous allons discuter les différents travaux ayant influencé la thèse et servi de base à nos différentes modélisations. Cette section est un état de l'art général sur lequel nous reviendrons plus précisément dans chacun des chapitres.

1.2.1 Modèles de simulation multi-agents : principes et applications au domaine du marketing

Les systèmes multi-agents (SMA) et les modèles classiques, qu'ils soient statistiques, fondés sur l'apprentissage automatique, ou sur les équations différentielles, sont complémentaires selon (Garcia 2005). Leur finalité n'est pas toujours la même : alors que les modèles classiques sont souvent utilisés à des fins prédictives, les SMA visent aussi à comprendre les dynamiques émergentes entre comportements individuels, interactions locales et effets globaux.

Les SMA offrent un niveau de précision et une finesse d'analyse que les approches classiques peinent à atteindre. Là où les modèles statistiques ou les méthodes d'apprentissage automatique se concentrent souvent sur des relations globales entre variables, les SMA permettent d'explorer en détail les conséquences d'actions ciblées sur des individus ou des segments spécifiques. Par exemple, ils peuvent simuler l'impact précis d'une promotion sur différents types de consommateurs, estimer l'effet combiné d'actions de

référencement naturel (SEO) et payant (SEA) sur des profils d'acheteurs distincts, ou encore identifier quelles cibles publicitaires sont les plus susceptibles de réagir positivement à une campagne donnée. Ces questions, qui impliquent de modéliser les comportements, les interactions et les influences au sein d'un marché hétérogène, sont souvent hors de portée des approches plus agrégées. Les SMA, en reproduisant explicitement la diversité des agents et la dynamique de leurs interactions, permettent de répondre à ce type de problématiques de manière plus réaliste et plus adaptée aux besoins opérationnels du marketing.

Bien que les SMA puissent être utilisés pour produire des prévisions chiffrées, leur intérêt ne se limite pas à cet aspect. Ils se distinguent surtout par leur capacité à représenter des comportements individuels, à simuler des interactions locales et à observer comment celles-ci peuvent, par agrégation, générer des phénomènes à l'échelle globale. Cette approche permet non seulement d'estimer des résultats quantitatifs dans certains contextes, mais aussi de mieux comprendre les mécanismes qui sous-tendent ces résultats, en identifiant les dynamiques émergentes et les effets indirects difficiles à capter avec d'autres méthodes.

Par « résultats quantitatifs », on entend des sorties chiffrées issues du modèle, telles que des prévisions de parts de marché, des volumes de ventes, des taux de conversion ou encore de diffusion d'un produit dans une population. Ces valeurs numériques peuvent être directement comparées à des données réelles pour évaluer la précision du modèle ou alimenter des décisions opérationnelles. Dans le cadre des SMA, ces résultats sont généralement obtenus en paramétrant finement le comportement des agents et leurs interactions. Toutefois, leur fiabilité dépend fortement de la justesse des hypothèses et des données utilisées : une approximation ou une erreur sur un paramètre clé peut conduire à des écarts importants par rapport à la réalité.

Il est parfois difficile d'obtenir des résultats quantitatifs avec des SMA. Même s'ils modélisent des comportements individuels, il est souvent impossible ou inutile d'attribuer un comportement distinct à chaque agent (Sauf dans des cas à très peu d'agents). Cela serait non seulement extrêmement coûteux en termes de complexité du fait d'une explosion combinatoire lors de la calibration, mais cela irait aussi à l'encontre de la recherche de régularités comportementales. Il faut trouver des règles communes aux agents.

Dans le cas des agents humains, par exemple, nous partageons des capacités cognitives similaires, des biais communs et des mécanismes biologiques ou psychologique analogues. Il est donc pertinent de définir des règles comportementales globales fondées sur les caractéristiques internes des agents. Dans un tel cadre, toute modification de ces règles peut produire des effets importants à l'échelle macroscopique. Cela signifie aussi qu'un faible écart dans la définition des comportements ou des interactions peut entraîner des dynamiques très différentes, ce qui rend la prédiction difficile : il faut une connaissance fine et précise des règles pour espérer prédire les effets globaux. Dans ce contexte même l'ordre des prises de décisions ou des interactions peut avoir un effet significatif.

Il est possible de différencier les comportements des agents grâce à des caractéristiques internes. On peut alors modifier chaque caractéristique de chaque agent, mais à grande

échelle, cela devient irréaliste. Une telle granularité mène à une spécialisation excessive du système. On peut faire l'analogie avec le surapprentissage (overfitting) en apprentissage automatique. Par exemple d'après théorème d'interpolation de Lagrange il existe toujours un polynôme de degré au plus n qui passe exactement par n+1 points. Mais ce polynôme généralise mal. On ne cherche plus à extraire des lois générales, mais à reproduire à tout prix un phénomène. Cela fait perdre le lien fondamental entre comportements individuels et phénomènes émergents à grande échelle.

(Bonabeau 2002) conclut qu'il est important de connaître le degré de précision et de confiance du modèle (Données, expertise...) afin d'interpréter correctement les résultats et éviter de prendre des décisions sur des résultats quantitatifs lorsque les résultats du modèle sont de nature qualitative.

Il faut retenir que les SMA possèdent des atouts spécifiques, notamment pour l'analyse des comportements. Leur utilisation à des fins prédictives requiert une certaine expertise et impose des précautions dans l'interprétation des résultats. L'état de l'art présenté ici reste volontairement général ; des revues de littérature plus ciblées seront proposées dans chacun des chapitres de la thèse afin d'approfondir les thématiques spécifiques qu'ils abordent.

1.2.2 Interactions sociales et bouche à oreille

Dans le domaine du marketing, le bouche-à-oreille joue un rôle crucial dans la diffusion de l'information et l'adoption de produits ou services. Pour modéliser ces dynamiques complexes d'influence sociale, les SMA se révèlent particulièrement puissants. En effet, ils permettent de représenter explicitement des agents hétérogènes, autonomes et dotés de comportements spécifiques, interagissant dans un environnement donné. Ce cadre offre une grande flexibilité pour capturer les effets d'interaction locale, les influences sociales, et les phénomènes émergents à grande échelle.

Parmi les approches de modélisation, deux grandes familles issues de la recherche en sociologie et en épidémiologie se distinguent : les modèles de propagation et les modèles de seuil. Inspirés respectivement des dynamiques de transmission des maladies et des comportements sociaux, ces modèles fournissent des fondations théoriques pertinentes pour simuler les mécanismes de diffusion du bouche-à-oreille. Intégrés dans des SMA, ils permettent d'explorer finement comment les décisions individuelles et les interactions sociales peuvent façonner la dynamique globale d'un marché.

Pour simuler de manière réaliste ces mécanismes, il est essentiel de s'appuyer sur des réseaux de connexions proches des réseaux sociaux réels. Les réseaux dits « petit monde », tels que ceux générés par le modèle de Watts-Strogatz (Watts et al. 1998) ou Barabasi-Albert (Albert et al. 2002), sont particulièrement adaptés : ils combinent une loi de puissance sur la distribution des connexions et des distances moyennes courtes entre les nœuds. Cela permet une propagation rapide de l'information à travers l'ensemble du réseau tout en conservant une structure sociale plausible. Leur intérêt réside également dans leur facilité de génération et leur capacité à reproduire certaines propriétés structurelles des réseaux sociaux humains.

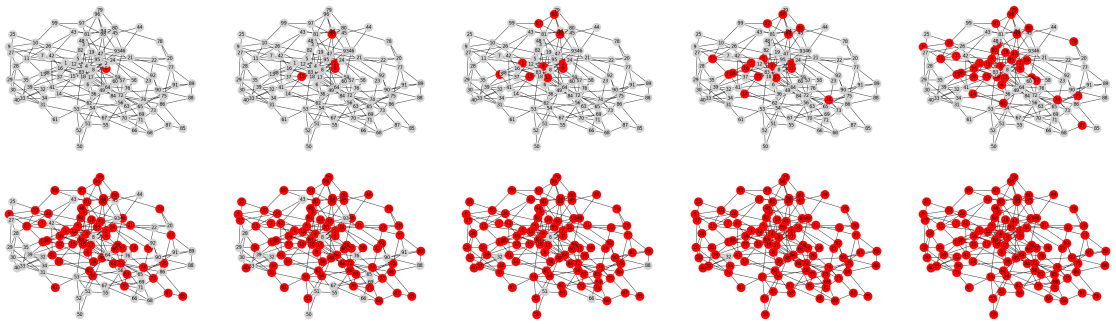


FIGURE 1.1 – *Illustration d'un processus de propagation classique dans un réseau small-world. Chaque image représente un pas de temps. On observe qu'une fois les nœuds les plus connectés atteints, la propagation s'accélère et atteint rapidement l'ensemble du réseau.*

Modèles de propagation

Les modèles de propagation s'inspirent directement des dynamiques épidémiologiques, où l'information (ou l'adoption d'un produit) se transmet d'un individu à l'autre par simple contact. Dans ce cadre, une seule interaction avec un agent informé ou convaincu peut suffire à déclencher un changement de comportement, rendant ces modèles particulièrement adaptés à la modélisation de décisions simples ou peu engageantes.

Chaque individu est connecté à d'autres individus avec lesquels il peut interagir. Au départ, un ou plusieurs individus sont « informés » ou « convaincus », et ils peuvent transmettre cette information à leurs voisins lors d'un contact. Ce processus se répète à chaque étape, créant une chaîne de transmission. La simplicité de la décision (un seul contact suffit) rend la propagation rapide, surtout si le réseau est bien connecté. De plus, on peut même imaginer qu'un individu puisse entrer en contact avec plusieurs autres individus en une seule étape. C'est le cas pour la diffusion d'informations ou de contenu en ligne. Ce type de modèle permet ainsi d'observer comment une idée ou un produit peut se diffuser à grande échelle à partir d'un petit nombre d'initiés.

Dans ce cadre, la dynamique de propagation met naturellement en lumière le rôle des individus les plus connectés nommés « leaders d'opinion ». En supposant qu'un agent transmette l'information à l'ensemble de ses voisins, il devient évident qu'un individu disposant d'un grand nombre de connexions aura un impact plus large sur la diffusion. On observe ainsi une diffusion du centre vers la périphérie : les individus centraux, bien connectés, servent de relais puissants, tandis que les agents périphériques, disposant de peu de liens, sont souvent moins exposés à l'information et peuvent rester en dehors du processus de propagation.

Les modèles épidémiologiques comme les modèles SIR (Susceptible-Infected-Recovered) ont longtemps servi de base à ces approches. Toutefois, ces modèles équationnels présentent certaines limites, notamment en supposant une population homogène et des interactions uniformes. À l'inverse, les modèles de simulation multi-agents permettent d'introduire plus de variabilité comportementale, des structures de réseau plus complexes et d'explorer des effets locaux qui ne peuvent émerger dans un cadre purement mathé-

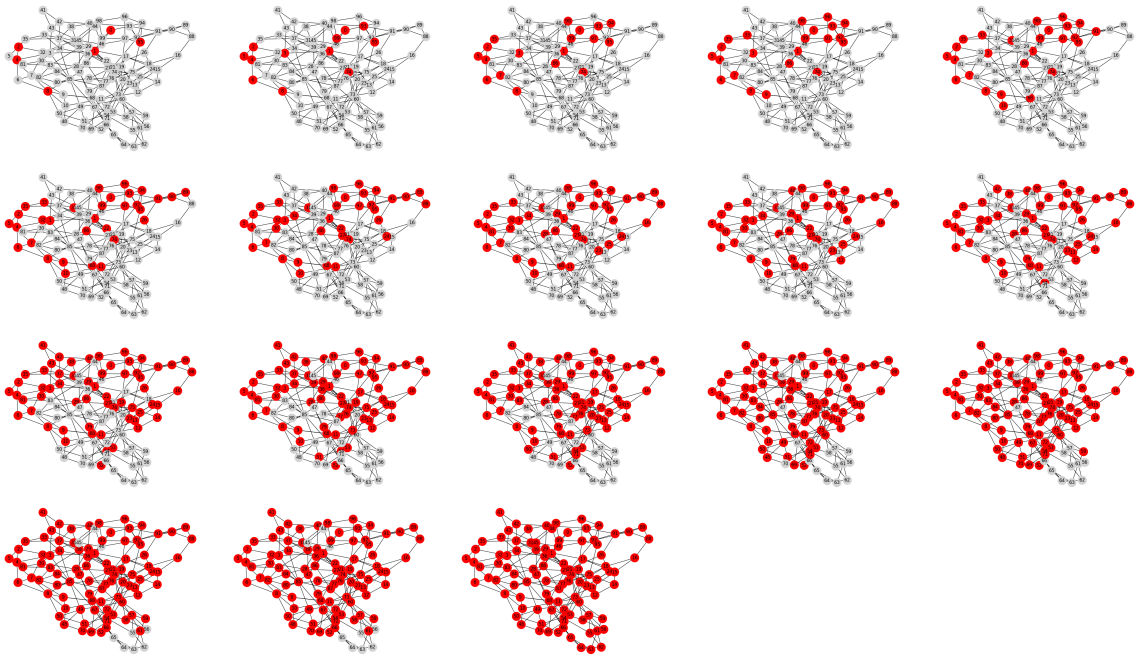


FIGURE 1.2 – *Illustration d'un modèle de seuil dans un réseau small-world. Chaque image correspond à un pas de temps. La simulation commence avec plusieurs nœuds actifs afin d'éviter un arrêt prématuré de la propagation.*

matique. Cela ouvre la voie à des analyses plus fines et contextualisées de la diffusion sociale.

Avec ce type de modélisation, on observe que les SMA sont particulièrement adaptés malgré un cadre limité aux décisions simples ne nécessitant qu'un contact. Cela peut suffire à représenter certaines situations marketing spécifiques, mais dans le cadre du bouche-à-oreille les décisions sont généralement engageantes pour les agents avec une possibilité de perte, économique ou de qualité.

Modèles de seuil

Contrairement aux modèles de propagation simples, les modèles de seuil considèrent que l'adoption d'un comportement ou d'une opinion n'est pas immédiate, mais dépend de la pression sociale exercée par l'entourage. Dans ces modèles, chaque individu possède un seuil interne, une proportion minimale de ses contacts qui doivent avoir adopté un comportement pour qu'il décide lui-même de l'adopter. Tant que ce seuil n'est pas atteint, l'individu reste inchangé. Ce mécanisme rend la diffusion plus lente et plus sélective, et modifie fondamentalement la dynamique : la propagation a souvent lieu de la périphérie du réseau vers son centre, car ce sont les individus faiblement connectés, donc plus exposés à des majorités locales, qui changent en premier, avant que des individus plus centraux, exposés à une diversité d'opinions, ne basculent à leur tour.

Un effet marquant de ces modèles est la formation de clusters d'opinions homogènes. Les agents, influencés par la majorité de leurs voisins directs, ont tendance à s'aligner sur les normes locales, ce qui crée des groupes cohérents dans lesquels l'opinion ou le comportement est partagé, mais qui peuvent coexister avec d'autres groupes aux opinions

différentes. Ces modèles reflètent donc bien les dynamiques d'uniformisation sociale et les résistances locales au changement, en particulier dans le cas de décisions complexes ou engageantes.

Prenons un exemple : si une personne a l'habitude de manger au restaurant le mercredi midi avec ses collègues, il est peu probable qu'il change ses habitudes juste parce qu'un de ses collègues cesse de manger au restaurant. Mais si, au fil des semaines, ils ne sont plus que deux sur douze collègues à manger au restaurant, il devient beaucoup plus probable que ces deux derniers collègues finissent eux aussi par adopter ce nouveau comportement. Le changement s'effectue donc par effet de masse, une fois qu'un certain point de bascule est atteint.

Les modèles de seuil sont particulièrement pertinents dans des contextes où l'adoption implique une certaine hésitation, un coût perçu, ou un besoin de validation sociale. Typiquement ce sont les changements de mode de vie, les adoptions de technologies nouvelles, les engagements politiques, etc. Combinés à des simulations multi-agents, ils permettent de capturer des dynamiques de diffusion plus nuancées, marquées par des phénomènes d'adhésion progressive, de résistance au changement, ou encore de stabilité apparente suivie d'un basculement rapide une fois un seuil critique franchi.

Des modélisations SMA

Dans leur article (Delre et al. 2007) développent, un modèle de seuil multi-agents, appliqué à la diffusion de nouveaux produits. Leur objectif est double : simuler les dynamiques d'adoption selon des interactions sociales locales, et étudier l'impact stratégique de la publicité sur le déclenchement et le développement du processus de diffusion.

Le modèle repose sur des agents connectés dans un réseau social (souvent de type « petit monde »), chacun possédant un seuil d'adoption individuel. Comme dans les modèles de seuil classiques, un agent adopte le produit lorsque la proportion de ses voisins l'ayant déjà adopté dépasse ce seuil. Ce cadre permet de reproduire de manière émergente les courbes en « S » typiques de la diffusion de l'innovation (Rogers et al. 2014), où l'adoption débute lentement, s'accélère ensuite rapidement, puis atteint une saturation progressive. Ce comportement n'est pas imposé par une équation : il résulte des interactions sociales entre agents, ce qui rend le modèle particulièrement réaliste.

Un aspect central de leur étude est l'introduction de la publicité comme levier externe d'influence. La publicité est modélisée comme une probabilité, à chaque étape, de « forcer » certains agents à envisager l'adoption du produit, indépendamment de l'état de leurs voisins. Cette stratégie de ciblage permet d'accélérer ou de déclencher la diffusion en ciblant des individus-clés, souvent en début de simulation. Toutefois, ce mécanisme présente une forme de rigidité : si un agent exposé à la publicité refuse d'adopter le produit, il ne pourra plus jamais changer d'avis, même si plus tard la majorité de ses contacts l'adopte. Ce comportement unidirectionnel rend l'effet de la publicité artificiel par certains aspects, et limite la dynamique adaptative du modèle.

Par ailleurs, le modèle est extrêmement sensible à la structure du réseau social et à cer-

tains paramètres (distribution des seuils, timing de la publicité, densité de connexions, etc.). Dans une large majorité des configurations testées, le système présente une dynamique binaire : soit aucune diffusion ne se produit (les agents restent majoritairement non-adoptants), soit la diffusion devient massive et très rapide, contaminant presque toute la population en quelques étapes. Cette instabilité rend la prédiction du succès d'une innovation difficile à généraliser, et reflète la nature hautement non linéaire de la propagation par seuils sociaux.

Malgré ces limites, ce modèle reste une contribution majeure : il met en évidence le rôle crucial du timing et du ciblage des interventions marketing, tout en soulignant combien la dynamique de bouche-à-oreille peut être dépendante de la structure sociale sous-jacente. Il offre également un cadre expérimental riche pour explorer des politiques de lancement de produit, mais gagnerait à être complété par des mécanismes plus souples.

Ce modèle rapproche les dynamiques d'influence marketing de la logique diffusion/épidémique, tout en tenant compte de la structure du réseau social, du budget de campagne et du profil produit. Il s'inscrit dans la lignée des travaux sur la synchronisation entre structure sociale et stratégie marketing, en actualisant et en spécifiant la question du choix optimal d'influenceur selon le contexte.

(Doshi et al. 2022) explorent, via un modèle SMA, la sélection optimale d'influenceurs, selon le type de produit (luxe ou non-luxe) et le niveau d'intérêt initial des consommateurs. À partir d'un graphe probabiliste, le modèle simule :

- la diffusion de l'information à travers des agents consommateurs,
- l'effet différencié des influenceurs (nano, micro, macro, célébrités),
- des métriques clés : coût d'acquisition client et taux de conversion.

Les principaux résultats montrent que pour les produits non-luxe, les nano-influenceurs produisent un meilleur retour sur investissement (ROI, Return On Investment) en raison de leur coût modéré et d'un engagement ciblé. Ces résultats allant dans le sens des résultats attendus avec un modèle de seuil généralisé. En revanche, pour les produits luxe ou les situations où l'intérêt initial est faible, les célébrités génèrent de meilleurs taux de conversion malgré un coût d'acquisition client plus élevé.

1.2.3 Construire un modèle SMA appliqué au marketing

(Negahban et al. 2014) et (Rand et al. 2011) proposent des lignes directrices sur la manière de construire des SMA appliqués au marketing. Ils mettent en avant la rigueur que nécessite la modélisation de ce type de système. Le développement du commerce omnicanal a donné lieu à de nouveaux modèles SMA, capables d'intégrer simultanément les comportements des clients en ligne et en magasin. Un exemple notable est présenté par (Li et al. 2024), qui montre comment la coordination des canaux peut influencer la répartition des ventes, la fidélité, et les arbitrages des consommateurs entre physique et digital.

1.2.4 Processus de décision

Certains travaux adoptent une approche SMA pour modéliser les facteurs qui influencent la préférence des consommateurs pour des produits locaux ou étrangers. (Danaye et al.

2023) illustrent comment des agents dotés de préférences culturelles, économiques et sociales peuvent simuler des choix complexes en environnement concurrentiel. Ce type de modélisation est particulièrement utile pour comprendre les arbitrages en situation réelle, au-delà des approches économiques standard.

L'approche SMA permet également de modéliser des comportements plus émotionnels, comme l'achat impulsif. (Abbasi Siar et al. 2022) montre que des effets de groupe ou de stimulation contextuelle (promotions, environnement social) peuvent suffire à déclencher des achats non planifiés. Ce travail renforce la capacité des SMA à simuler des processus non rationnels dans des contextes d'influence.

La modélisation SMA nécessite d'adopter le point de vue du consommateur, de comprendre son fonctionnement afin de modéliser son comportement. L'agrégation de ces comportements doit permettre de reproduire des effets observables en magasin, comme les ventes. Plusieurs facteurs, parfois très différents, influencent le processus de décision d'un consommateur. Dans cette thèse, nous analyserons ce processus en décomposant chaque étape menant à la sélection d'un produit. Le premier facteur déterminant est le besoin. Il peut être plus ou moins essentiel et est influencé par les envies du consommateur, le bouche-à-oreille, la publicité et le marketing. Vient ensuite le choix de l'intermédiaire. Quelles options s'offrent au consommateur ? Quels magasins sont disponibles ? Privilégie-t-il un déplacement en magasin ? À quelle distance ? De nombreux paramètres entrent en jeu et varient selon le profil du consommateur. Une habitude d'achat peut, par exemple, reléguer un magasin plus proche au second plan. De plus, des considérations éthiques peuvent orienter le consommateur vers une enseigne ou un type d'intermédiaire spécifique. Enfin, la dernière étape concerne le choix du produit. Face à plusieurs alternatives similaires, le consommateur doit sélectionner celui qu'il achètera. Il prend en compte les caractéristiques des produits, mais notamment lorsqu'il est en magasin, sa capacité d'analyse et de comparaison est limitée. En revanche, pour un achat en ligne, cette étape de choix du produit peut précéder le choix de l'intermédiaire. Ce phénomène peut aussi se produire pour les achats en magasin, bien que cela soit rare.

Cette séparation des décisions permet de réduire le nombre de facteurs exogènes à prendre en compte, facilitant ainsi la modélisation avec un nombre restreint de paramètres. Toutefois, les effets que nous cherchons à reproduire et observer restent complexes. L'agrégation de ces différents choix permet, en bout de chaîne, de modéliser de manière complète le comportement du consommateur.

1.3 Discussion sur la modélisation du mouvement

Les SMA intègrent généralement une composante spatiale forte, qui constitue l'un des atouts majeurs de cette approche. En effet, la capacité à simuler des interactions localisées dans l'espace en fait un outil particulièrement adapté aux analyses spatiales complexes. Toutefois, dans le cadre spécifique du marketing, cette spatialisation nécessite un haut niveau de réalisme (représentation précise des territoires, des réseaux de transport, des

zones de chalandise[1], etc.), ce qui induit des coûts computationnels importants. (Sturley et al. 2018) illustrent bien ce point : leur SMA simule des comportements d'achat individuels dans un environnement spatialement réaliste, mais requiert une calibration fine et des données détaillées sur les clients, les emplacements et les types de magasins. Dans notre approche, nous faisons le choix d'abstraire la dimension spatiale et de ne pas modéliser explicitement les mouvements des agents dans l'espace. Ce choix minimaliste, mais ancré dans des données empiriques, permet de se concentrer sur les dynamiques comportementales sans dépendre de données géographiques difficiles à obtenir et évite de rendre l'interprétation délicate. De plus, ce choix ne constitue pas un rejet définitif de la spatialisation, mais un report en vue d'une intégration ultérieure dans un cadre modulaire maîtrisé. Par ailleurs, une modélisation spatiale innovante impliquerait d'intégrer la concurrence locale de manière crédible, ce qui suppose un accès à des données sensibles, telles que les performances de vente de différents acteurs du secteur, données aujourd'hui inaccessibles. En effet, il existe déjà des travaux concernant la modélisation du mouvement intra-magasin notamment avec la pandémie de COVID-19 : (Plata et al. 2020 ; Ying et O'Clery 2021 ; Ying, Wallis et al. 2019)

1.4 Calibration et Validation : Des étapes cruciales

Dans un SMA, les paramètres désignent les valeurs numériques ou symboliques qui régissent le comportement des agents, leurs interactions et les propriétés de l'environnement. Ces paramètres peuvent être de plusieurs natures :

- Paramètres individuels : caractéristiques propres aux agents (propension à acheter, seuil d'influence, budget disponible).
- Paramètres comportementaux : règles de décision ou biais communs à tous les agents (aversion au risque, effet de leurre).
- Paramètres structurels : configuration du réseau social ou géographique (ex. : degré de connectivité, structure des liens).
- Paramètres globaux : éléments de contexte ou de simulation (ex. : fréquence des promotions, durée d'un pas de temps, durée de simulation).

Ces paramètres peuvent être fixés a priori selon des hypothèses, extraits de données empiriques, ou calibrés automatiquement. Leur choix et leur nombre influencent fortement la complexité du modèle et sa capacité à reproduire des comportements réalistes.

Les termes de calibration et validation ont été débattus dans la communauté et n'ont pas forcément le même sens selon le contexte. Dans cette section nous définirons les différents sens généralement utilisés, ceux les plus admis et notre positionnement.

La mise en œuvre d'un SMA peut s'appuyer sur une grande variété d'outils et de langages de programmation. De nombreux frameworks ont été développés pour faciliter la création, l'exécution et l'analyse de SMA, chacun répondant à des besoins spécifiques du paradigme agent. Parmi les plus utilisés, on peut citer NetLogo, Mesa ou Spade en Python, JADE ou MASON en Java, ou encore GAMA, qui combine une interface de mo-

1. La zone de chalandise d'un commerce est sa zone géographique d'influence, d'où provient la majorité de la clientèle.

délisation propre au framework avec des extensions en Java. Bien que des langages généralistes orientés objet comme Python, Java ou C++ puissent être utilisés pour développer un SMA, ils ne fournissent pas par défaut les abstractions spécifiques aux agents (autonomie, communication, environnement partagé, etc.), ni les outils de simulation et de visualisation intégrés que proposent les frameworks dédiés. L'utilisation de langages bas niveau est possible, mais elle implique un développement coûteux et complexe, souvent sans avantage significatif par rapport aux solutions existantes.

Dans une démarche rigoureuse, deux étapes essentielles doivent accompagner la conception d'un SMA : la vérification et la validation.

La vérification vise à répondre à la question suivante : le modèle informatique correspond-il fidèlement à ce qui a été spécifié ? Il s'agit ici de garantir que l'implémentation logicielle respecte les règles, comportements, et dynamiques attendues du système. Plusieurs moyens peuvent être mobilisés :

— Tests unitaires et fonctionnels : pour vérifier la cohérence du comportement des agents.
— Relecture du code et documentation : un code bien commenté, structuré et transparent facilite l'analyse et la vérification.
— Reproductibilité : un modèle vérifié doit pouvoir être répliqué à partir de la description fournie, garantissant la traçabilité des résultats.
— Visualisation et suivi des agents : permettent de détecter des comportements inattendus ou incohérents.

La validation, quant à elle, vise à déterminer si le modèle est capable de représenter de manière crédible un phénomène réel. Elle répond à la question : est-ce que le modèle est utile pour comprendre ou prédire des comportements observables dans le monde réel ? Plusieurs approches sont possibles :

— Validation empirique : comparer les résultats du modèle avec des données observées, des statistiques ou des tendances connues.
— Validation par expertise : faire intervenir des spécialistes du domaine (ex. : sociologues, économistes, professionnels du marketing) pour évaluer la pertinence des hypothèses, des comportements d'agents et des dynamiques simulées.
— Validation structurelle : vérifier que les mécanismes internes du modèle (règles de décision, interaction, environnement) sont crédibles et bien fondés théoriquement.
— Analyse de sensibilité : tester la robustesse du modèle en faisant varier les paramètres pour observer les effets sur les résultats.

Nous nous positionnons entre la validation classique des SMA et la calibration inspirée de l'apprentissage automatique. L'objectif est de reproduire des faits stylisés caractéristiques du domaine, souvent difficiles à observer ou dont l'origine est mal comprise (par exemple, les « fat tails » en économie).

Un modèle est considéré comme validé s'il satisfait les critères suivants :

— Il intègre des processus de décision plausibles, idéalement validés par des experts ou par la littérature.

— Il reproduit des effets observables caractéristiques du phénomène étudié (par exemple, l'augmentation des ventes lors d'une promotion).
— Il est capable de générer des faits stylisés reconnus, même en l'absence de modélisations théoriques établies.

Lorsqu'un modèle satisfait l'ensemble de ces critères, il peut être considéré comme validé. Cette approche rend la validation difficile mais nous semble la plus rigoureuse. Bien souvent la validation des modèles SMA dans la littérature ne valide que partiellement ces points.

La génération de faits stylisés constitue sans doute l'étape la plus déterminante dans l'évaluation d'un modèle multi-agents. Par « faits stylisés », on entend des régularités empiriques, récurrentes et robustes, qui apparaissent dans de nombreuses observations réelles, mais dont la cause profonde n'est pas toujours clairement identifiée. Leur intérêt est double : ils fournissent une référence concrète pour juger de la pertinence d'un modèle, et ils permettent d'établir un lien entre comportements microscopiques et phénomènes macroscopiques. Dans un SMA, ces faits stylisés ne sont généralement pas « codés » directement dans les règles de comportement. Ils émergent des interactions entre agents, de leur hétérogénéité, et des dynamiques collectives qui en résultent. C'est précisément ce caractère émergent qui rend leur reproduction si précieuse : il témoigne que le modèle capture effectivement des mécanismes sous-jacents réalistes, même si ceux-ci ne sont pas explicitement formulés dans les équations ou les règles initiales.

Cette propriété confère aux SMA un avantage décisif par rapport à d'autres approches plus déterministes ou purement statistiques. Dans de nombreux cas, ces régularités sont impossibles, ou du moins très difficiles, à mettre en évidence par observation directe dans le monde réel, soit parce qu'elles nécessitent des volumes de données considérables, soit parce qu'elles se manifestent sur des échelles de temps ou dans des conditions rarement observées. Les SMA agissent alors comme des « laboratoires virtuels » permettant de reproduire et d'analyser ces phénomènes dans un environnement contrôlé. Cela ouvre la voie à une meilleure compréhension des mécanismes générateurs, à l'exploration de scénarios alternatifs et à l'identification de leviers d'action, en particulier dans des domaines comme le marketing, où la dynamique collective résulte de décisions individuelles interconnectées.

En ce qui concerne la calibration, nous adoptons une approche issue de l'apprentissage automatique. L'objectif est de déterminer un ensemble de paramètres permettant au modèle de s'adapter précisément à divers contextes. Le modèle doit reproduire fidèlement des situations variées, tout en conservant une structure interne cohérente.

Malheureusement, les données nécessaires à ce type de calibration sont parfois difficiles à obtenir. En effet, il faut disposer d'un volume important de données, souvent récentes et adaptées au contexte étudié. Dans notre cadre marketing, il s'agit généralement de données de ventes, indispensables pour vérifier que le modèle reproduit fidèlement la réalité observée.

Ces données existent, car les enseignes commerciales sont capables, depuis plusieurs années déjà, de stocker une grande quantité d'informations sur les achats réalisés. Cepen-

dant, elles sont souvent difficiles d'accès, pas toujours de bonne qualité et peuvent varier considérablement selon le contexte étudié. Par exemple, il ne s'agit pas de collecter les données de n'importe quel magasin si l'objectif est de modéliser des comportements d'achats récurrents. Dans cet exemple, il est nécessaire de se concentrer sur un supermarché afin d'analyser l'impact de la fidélité des consommateurs. Le comportement des acheteurs n'est pas le même dans un magasin d'électronique que dans un supermarché alimentaire.

Par ailleurs, la modélisation des comportements d'achat repose idéalement sur des données granulaires, telles que celles issues des cartes de fidélité. Ces informations, particulièrement riches, permettent de suivre les habitudes de consommation sur le long terme et d'identifier des régularités statistiques. Toutefois, leur accès reste limité, ce qui contraint souvent la précision des calibrations possibles. Dans ce contexte, il est nécessaire d'explorer des approches alternatives ou complémentaires afin de pallier le manque de données directement exploitables.

Une analyse de sensibilité aux paramètres est souvent nécessaire, voire indispensable, pour évaluer leur impact sur le comportement du modèle. Il est également crucial de limiter le nombre de paramètres afin d'éviter le surapprentissage. Un modèle ne devrait pas être capable de reproduire des scénarios radicalement différents de ceux pour lesquels il a été conçu. De plus, pour chaque contexte, seules quelques configurations de paramètres doivent conduire à des résultats satisfaisants. Si des paramètres très différents peuvent générer des résultats similaires dans chaque situation, cela suggère un trop grand nombre de paramètres.

1.4.1 Méthodes de calibration

Il existe de nombreuses méthodes de calibration des paramètres, en particulier parmi celles couramment utilisées en apprentissage automatique.

Les systèmes multi-agents (SMA) sont fréquemment associés aux algorithmes génétiques, parfois eux-mêmes considérés comme des SMA (Bruderer et al. 1996 ; Garcia 2005). Ces algorithmes constituent une classe de méthodes d'optimisation inspirées du processus de sélection naturelle en biologie. Chaque individu représente une solution, c'est-à-dire, dans notre cas, un ensemble de paramètres définissant un SMA. Regroupés en populations, ces solutions sont soumises à des opérations de mutation et de croisement selon des règles définies par le modélisateur. Le processus se décompose généralement comme suit :

- **Initialisation** : génération aléatoire d'une population initiale.
- **Évaluation** : évaluation de chaque individu via une fonction objectif à maximiser ou minimiser.
- **Sélection** : choix des individus destinés à se reproduire (meilleurs individus, individus assurant une diversité maximale, etc.).
- **Croisement (crossover)** : combinaison de deux individus pour en produire de nouveaux, partiellement ou totalement.
- **Mutation** : modification aléatoire d'un individu afin d'introduire de la diversité.
- **Remplacement** : génération d'une nouvelle population, puis répétition du cycle.

Les phases de croisement et de mutation peuvent être inversées selon les besoins.

Bien que puissants, les algorithmes génétiques ne sont pas des solutions clé en main. Leur mise en œuvre nécessite un travail de modélisation des individus ainsi qu'un choix rigoureux des stratégies de sélection et de mutation. Ces choix spécifiques rendent l'approche potentiellement très efficace pour certains problèmes, mais en limitent la capacité de généralisation.

1.4.2 L'étude d'impact des paramètres

L'étude de l'impact des paramètres permet de comprendre quels éléments influencent le comportement global du modèle. En analysant la sensibilité du système à chaque paramètre, on identifie ceux qui ont un effet important sur les résultats. Les autres peuvent être simplifiés ou ignorés sans perte de précision. Cette étape permet de distinguer les paramètres nécessaires au bon fonctionnement du modèle. Cela permet de garantir la robustesse du modèle et d'orienter la calibration. On passera plus de temps sur la calibration d'un paramètre important. Une telle étude permet également de détecter d'éventuelles interactions non triviales entre les paramètres, sources potentielles de dynamiques émergentes inattendues. Elle constitue donc une étape intermédiaire entre la calibration brute et la validation interprétative, en facilitant l'explicabilité du modèle et en guidant les choix de simplification.

1.4.3 Capacité à reproduire le réel

Un modèle n'a de valeur explicative que s'il parvient à reproduire des comportements observés dans le réel, qu'ils soient de nature qualitative (effets stylisés, régularités empiriques) ou quantitative (valeurs de ventes, parts de marché, tendances temporelles). Cette capacité à « coller » au réel constitue un critère fort de validation, et s'observe notamment par la reproduction d'effets connus, comme l'impact d'une promotion ou la fidélité à une marque. Plusieurs travaux dans la littérature montrent que les SMA sont capables de reproduire des dynamiques collectives réalistes à partir de règles locales simples, et peuvent capturer la non-linéarité des comportements agrégés. Cependant, cette fidélité au réel ne doit pas se faire au prix d'une sur-paramétrisation. Un modèle utile est celui qui, avec un nombre réduit de paramètres bien choisis, parvient à générer des comportements cohérents, interprétables et compatibles avec les données réelles disponibles.

En marketing, plusieurs phénomènes empiriques récurrents, appelés faits stylisés, ont été documentés dans la littérature ou observés sur le terrain. Ces faits ne sont pas des lois strictes, mais des régularités qualitatives qui constituent des références comportementales pour juger de la pertinence d'un modèle.

Voici quelques faits stylisés particulièrement importants dans le cadre de cette thèse :

— Effets de contexte : les choix des consommateurs ne dépendent pas uniquement des attributs absolus des produits, mais également des alternatives disponibles au

moment du choix. Par exemple, l'introduction d'un produit « leurre » peut modifier la préférence entre deux autres options.

— Fidélisation : les consommateurs manifestent souvent des comportements répétitifs, privilégiant certaines enseignes ou certaines marques sur le long terme. Cette fidélité peut résulter d'une satisfaction passée, d'un attachement affectif, ou de coûts cognitifs liés au changement. Elle se traduit par une inertie dans les choix, qu'un modèle comportemental doit pouvoir capter.
— Réactions aux promotions : une baisse temporaire de prix peut engendrer une augmentation significative des ventes, souvent disproportionnée par rapport à la réduction effective. Cet effet promotionnel, bien connu, varie selon les profils clients et les types de produits. Il peut également entraîner des effets de stockage ou de report d'achat, qui modifient la dynamique globale.
— Guerre des prix : dans un environnement compétitif, les consommateurs peuvent adapter leurs comportements en fonction des offres concurrentes. Des cycles de baisse de prix entre enseignes peuvent générer des dynamiques inter-enseignes parfois peu rationnelles.

Ces faits stylisés servent de cibles qualitatives pour la validation de nos modèles. Un SMA crédible ne doit pas seulement produire des résultats chiffrés réalistes, mais aussi être capable de faire émerger ce type de régularités à partir de règles comportementales simples. Ces phénomènes guideront à la fois la construction du modèle, la définition des paramètres clés, et les critères d'évaluation mobilisés lors des phases de calibration.

Il est essentiel que le modèle SMA soit capable de faire émerger ces faits stylisés à partir de règles simples, sans que ces comportements ne soient explicitement codés. C'est précisément cette capacité d'émergence qui confère aux SMA leur pouvoir explicatif. Toutefois, dans certains cas, une partie de ces phénomènes peut être implémentée directement dans les règles de décision des agents, notamment lorsque leur origine est bien identifiée ou qu'une modélisation inductive est justifiée. Même dans ce cas, il demeure pertinent de les observer et d'en évaluer la robustesse : leur persistance dans différents contextes et configurations renforce la crédibilité du modèle, qu'ils soient émergents ou intégrés a priori.

1.4.4 Le nombre de paramètres doit rester faible

L'ensemble de ces considérations souligne l'importance d'une modélisation parcimonieuse. Un modèle trop riche en paramètres devient non seulement difficile à calibrer, mais perd en interprétabilité et en capacité à générer des résultats robustes. La calibration fine, l'analyse de sensibilité et la validation empirique doivent converger vers un objectif commun : élaborer un modèle explicable, réaliste et généralisable. La fidélité à la réalité ne passe pas uniquement par la complexité, mais par la pertinence des mécanismes représentés. En somme, le bon modèle n'est pas celui qui reproduit tout, mais celui qui explique bien, avec peu.

1.5 Objectifs et contributions : vers une modélisation explicable du comportement consommateur en SMA

Dans cette thèse, nous défendons l'idée que la modélisation du comportement des consommateurs dans le cadre des SMA doit être décomposée en plusieurs étapes distinctes. Ce découpage permet de développer des modèles spécialisés, explicables et interprétables, chaque modèle pouvant être calibré et validé à partir de données réelles. L'objectif est de montrer que chaque composante du comportement des consommateurs présente ses propres enjeux et défis, auxquels les SMA peuvent apporter des réponses spécifiques. Ce cadre permet également de mobiliser des concepts et des outils issus d'autres disciplines telles que les sciences sociales, le marketing ou l'économie pour enrichir la modélisation.

En l'absence de ce découpage, il devient difficile, voire impossible, de concevoir un modèle global qui soit à la fois réaliste, interprétable et calibrable. En effet, tenter de représenter simultanément toutes les dimensions du comportement d'achat au sein d'un unique modèle risquerait de produire une structure trop complexe, aux interactions non maîtrisées, et dont les résultats seraient difficiles à interpréter ou à relier à la réalité observée. L'intérêt des SMA réside précisément dans leur capacité à produire des modèles transparents et explicables. Il est donc essentiel d'éviter les approches « boîte noire » comme c'est souvent le cas avec certaines méthodes issues de l'apprentissage automatique, dans lesquelles l'introduction simultanée de multiples facteurs rend l'analyse et la validation délicates.

Adoptant une logique de « diviser pour mieux régner », nous proposons dans cette thèse d'examiner séparément trois grandes dimensions du comportement des consommateurs, avant de chercher à les articuler en vue de construire un modèle unifié. Ces trois dimensions sont les suivantes :

1. Dans le chapitre 2, nous explorons des méthodes, parfois inspirées de l'apprentissage automatique, pour estimer le nombre de clients par catégorie de produits. L'objectif est de répondre à la question suivante : dans un environnement stable, combien de clients peut-on attendre lors du prochain pas de temps ?
2. Dans le chapitre 3 nous nous intéressons à la modélisation du choix effectué par les consommateurs à l'intérieur d'une même catégorie de produits. L'enjeu est ici de comprendre pourquoi un client choisit une marque A plutôt qu'une marque B, et quels facteurs (attributs, produits, préférences individuelles, etc.) influencent ce choix.
3. Dans le chapitre 4 nous proposons une approche complémentaire d'estimation du nombre de clients et de la consommation dans un contexte spécifique : la consommation de produits quotidiens, consommables et périssables. Cette approche est fondée cette fois sur un besoin croissant ressenti par les agents. Cette dernière modélisation permet notamment d'analyser les réactions des agents face à des stimuli tels que les promotions.

Ces trois contributions donnent lieu à la conception de modèles distincts, chacun visant à produire de l'information exploitable pour l'analyse ou la prise de décision, dans leur

domaine respectif. Chacun de ces modèles a été conçu de manière autonome et validé séparément. La construction d'un modèle unifié, capable d'intégrer l'ensemble de ces dimensions du comportement consommateur, demeure un défi important, auquel cette thèse entend contribuer.

Chapitre 2

Modélisation et prédiction du nombre d'acheteurs

Dans ce chapitre, nous présentons la première étape de notre démarche : estimer le nombre d'acheteurs à partir de données de tickets de caisse, dont une part importante n'est rattachée à aucun client. Cette situation rend nécessaire la reconstitution d'une population afin d'obtenir une vision fiable du volume de clients. Nous discutons d'abord du rôle de la granularité, qui oriente le type d'analyse possible, entre une approche globale mettant en évidence des tendances de marché et une approche locale tenant compte de facteurs spécifiques comme les promotions ou la concurrence de proximité. Nous explorons ensuite différentes familles de méthodes pour générer des populations représentatives, en particulier les approches par optimisation combinatoire et celles par extrapolation probabiliste. Dans notre cas, nous retenons cette seconde voie, qui permet de simuler des comportements réalistes à partir de distributions statistiques estimées sur les clients identifiés. Après avoir montré les limites d'approches simples, comme l'assimilation du nombre de clients au nombre de tickets ou l'introduction d'un coefficient de correction, nous proposons un algorithme fondé sur la relation entre montant des achats, fréquence et espacement temporel. Cet algorithme attribue à chaque ticket anonyme un client fictif, en regroupant les transactions selon des profils plausibles. Validée sur des données artificiellement anonymisées, la méthode parvient à reproduire correctement les distributions d'intérêt et à estimer le nombre de clients avec une erreur réduite. Ce résultat permet de transformer des données partielles en un indicateur synthétique, intéressant pour la suite de la thèse. L'étape suivante consistera alors à répartir ces acheteurs estimés entre différentes offres, afin d'analyser plus finement les dynamiques concurrentielles sur le marché.

2.1 Introduction

Pour déterminer ce volume de clients, plusieurs approches méthodologiques sont envisageables. Parmi celles-ci, les méthodes classiques fondées sur l'analyse de séries temporelles offrent un cadre robuste et éprouvé. Dans notre étude, nous bénéficions de données professionnelles issues notamment d'une base de tickets de caisse collectés auprès d'une enseigne de bricolage.

Toutefois, l'exploitation de ces données soulève une difficulté majeure : une proportion

significative des transactions ne peut être rattachée à aucun profil client connu. Pour traiter ce problème, nous proposons dans la seconde partie de ce chapitre un algorithme de reconstruction : il attribue les tickets anonymes à des clients fictifs générés de manière probabiliste. L'objectif est de constituer une population cohérente et représentative, sur laquelle il soit possible d'appliquer des modèles d'analyse.

Avant de présenter ces méthodes, il est nécessaire de préciser à quel niveau d'analyse nous situons notre travail. En effet, les comportements d'achat dépendent de nombreux facteurs, allant des préférences individuelles aux tendances économiques globales. Or, il n'est pas possible de construire un modèle explicable qui intègre toutes ces influences à la fois. Le choix d'une granularité d'analyse devient donc un élément essentiel :

— Une granularité fine (par ex. suivre les comportements d'un client individuel) permet de détecter des régularités locales mais rend la modélisation complexe.
— Une granularité agrégée (par ex. analyser les ventes globales d'une enseigne) met surtout en évidence des tendances collectives, au prix d'une perte de détail.

Le choix de la granularité d'analyse joue un rôle déterminant, car il conditionne les phénomènes que l'on peut observer.

À une échelle globale, par exemple en considérant l'ensemble du marché du bricolage en France, on met surtout en évidence des tendances générales : l'impact d'une crise économique sur la consommation, l'essor du commerce en ligne, ou encore la saisonnalité des travaux de rénovation. Ces facteurs donnent une vision macroscopique du marché, mais ils masquent les comportements individuels.

À une échelle locale, comme celle d'un magasin particulier, apparaissent au contraire des dynamiques beaucoup plus circonstancielles : une campagne promotionnelle qui attire temporairement une clientèle, la concurrence directe d'un magasin voisin ou l'influence de la zone de chalandise. Ces éléments permettent de comprendre finement pourquoi certains produits se vendent mieux dans un point de vente précis, mais ils ne disent pas grand-chose des tendances générales du secteur.

Ainsi, selon le niveau de granularité choisi, on n'obtient pas la même lecture des comportements d'achat : la vision globale révèle des régularités collectives, tandis que la vision locale met en avant des phénomènes spécifiques et contextuels.

Dans notre cas, nous ne cherchons pas à couvrir l'ensemble de ces dimensions. Notre analyse se situe au niveau du magasin, mais sans intégrer les phénomènes très locaux (comme l'impact ponctuel d'une promotion ou la concurrence d'un voisin). L'objectif n'est pas de reproduire toutes les influences, mais de disposer d'une base de clients représentative à partir des tickets disponibles, afin de pouvoir appliquer ensuite des modèles d'analyse. Autrement dit, nous retenons une granularité adaptée aux données : suffisamment détaillée pour distinguer des comportements clients, mais volontairement simplifiée pour rester compatible avec notre algorithme de reconstruction.

La section suivante présentera les méthodes existantes pour générer des clients fictifs et notre choix pour compléter la population observée. Nous montrerons ensuite que ces données enrichies permettent d'alimenter les modèles étudiés dans le reste de la thèse.

2.2 Génération d'une population représentative et cohérente.

La génération d'une population synthétique repose sur l'idée de reproduire artificiellement les caractéristiques d'une population réelle à partir d'un ensemble limité de données disponibles. Pour que cette population simulée soit représentative, il est essentiel de respecter un certain nombre de critères sociodémographiques (âge, sexe, composition des ménages, niveau de revenu, localisation géographique, etc.). Or, l'accès à ces informations constitue souvent un défi : elles sont parfois protégées par des contraintes de confidentialité, ou bien collectées de manière partielle, hétérogène et à des échelles qui ne correspondent pas directement aux besoins de la modélisation. Dans certains cas, les données existent à un niveau agrégé (par exemple au niveau d'une commune ou d'un département), mais ne permettent pas de reconstruire finement les comportements individuels. La rareté des données réelles ainsi que leur fragmentation imposent alors de recourir à des méthodes spécifiques de traitement des données, pour estimer les caractéristiques manquantes et générer une population cohérente. L'enjeu est double : garantir la validité statistique de la population synthétique et préserver sa capacité à alimenter des modèles prédictifs ou de simulation de manière pertinente. (Chapuis et al. 2022) avancent qu'il y a deux grandes familles de génération :

2.2.1 Optimisation Combinatoire

Cette méthode vise à reproduire des enregistrements réels d'individus (généralement issus de données ou d'échantillons) en les combinant et en les pondérant de manière à faire correspondre des statistiques globales cibles (par exemple, des distributions par âge, sexe, emploi, etc.). Elle n'invente pas de nouveaux individus, mais réutilise des profils existants. Le processus cherche la meilleure combinaison d'individus réels pour que les totaux agrégés correspondent aux statistiques connues de la population cible. Les méthodes les plus couramment utilisés sont : Iterative Proportional Fitting (IPF), Simulated Annealing. Cette technique permet une grande cohérence avec les données réelles, mais dépend fortement de la représentativité de ces mêmes données.

Concrètement, le processus peut se résumer ainsi :

- On part d'un échantillon d'individus réels (par exemple, une enquête nationale où l'on connaît l'âge, le sexe, la situation professionnelle de chaque personne).
- On dispose aussi de statistiques agrégées locales (par exemple, la proportion d'hommes et de femmes dans une ville, la répartition par classes d'âge, le taux d'emploi, etc.).
- L'algorithme ajuste des poids sur chaque individu de l'échantillon : certains profils sont « répliqués » plusieurs fois, d'autres sont peu ou pas retenus, jusqu'à ce que les totaux correspondent aux contraintes locales.
- Le résultat est une population synthétique qui reprend exactement la structure statistique observée, tout en étant constituée d'individus plausibles (puisqu'ils proviennent tous de l'échantillon initial).

Par exemple, (Beckman et al. 1996) utilisent l'IPF pour créer des populations locales cohé-

rentes avec les recensements américains. Supposons qu'une petite ville doive être simulée avec 52 % de femmes, 48 % d'hommes, 20 % de retraités et 10 % de chômeurs. L'algorithme va sélectionner et répliquer, à partir d'un échantillon national, le bon mélange de personnes (par ex. davantage de femmes retraitées, moins d'hommes étudiants) jusqu'à ce que la population synthétique respecte exactement ces proportions. Cette approche a ensuite été utilisée dans de nombreux modèles de transport ou d'urbanisme, car elle permet de générer rapidement une population locale réaliste sans inventer de profils artificiels.

Dans le cas français, (Barthelemy et al. 2013) montrent qu'il est possible d'appliquer la même logique en combinant les micro-données de l'INSEE avec les marges locales disponibles (par commune, par catégorie socio-professionnelle, etc.). Le résultat est une population synthétique « ajustée » à chaque territoire, qui conserve la diversité des ménages réels tout en respectant strictement les contraintes statistiques locales.

2.2.2 Techniques d'Extrapolation

Ces techniques visent à modéliser une distribution statistique (multivariée) sous-jacente à la population réelle, puis à générer de nouveaux individus synthétiques en échantillonnant cette distribution. Elles permettent de créer des individus entièrement nouveaux qui suivent les tendances statistiques observées. Ces méthodes peuvent impliquer des modèles statistiques ou probabilistes (ex. régressions, réseaux bayésiens, etc.) pour estimer la distribution des caractéristiques. Ces techniques sont d'une grande flexibilité et sont très utiles si la granularité des données n'est pas complète, par exemple s'il manque quelques caractéristiques et informations précises. Cependant, le modèle peut générer des individus peu réalistes s'il est mal calibré.

(Mathieu et Picault 2013) proposent une méthode de reproduction des tickets de caisse en utilisant un indice de Jaccard revisité ainsi qu'un clustering hiérarchique. Cette méthode montre que selon le contexte, la génération de population, ici de tickets de caisse, peut s'appuyer directement sur l'analyse des traces d'activité plutôt que sur des catégories socio-démographiques pré-établies. Elle permet ainsi de reconstruire des comportements réalistes à partir de données brutes, tout en conservant la diversité observée dans les habitudes d'achat. En combinant une mesure de similarité adaptée et une approche non supervisée, cette méthode évite les biais induits par des segmentations a priori et offre une grande souplesse pour explorer différentes hypothèses de comportement dans un cadre de simulation multi-agents.

Dans leur revue systématique, (Hazelbag et al. 2020) analysent 84 études utilisant des modèles individu-centrés en épidémiologie, et identifient une grande hétérogénéité dans les méthodes de calibration. Ils distinguent principalement trois familles : les méthodes manuelles, les heuristiques (comme les algorithmes génétiques ou le recuit simulé), et les approches bayésiennes (notamment l'approximate Bayesian computation). Ici, on parle plutôt de calibration de paramètres et non de génération de population directement, car il s'agit d'ajuster les valeurs de paramètres comportementaux ou biologiques pour que les simulations reproduisent les dynamiques épidémiologiques observées. L'étude révèle que

seuls 38% des travaux évaluent l'incertitude des paramètres calibrés, et que peu décrivent précisément leur processus d'ajustement, limitant la reproductibilité. Cela montre que même dans des domaines autres que le marketing, il existe un besoin clair de transparence et de formalisation concernant la génération des populations et l'évaluation des méthodes utilisées, afin de garantir la fiabilité et la généralisation des résultats.

Un autre exemple notable, (Hörl et al. 2021) proposent une méthode de génération de population synthétique et de demandes de mobilité à l'échelle régionale, en s'appuyant exclusivement sur des données ouvertes (INSEE, Enquêtes, Ménages, Déplacements, etc.). Les auteurs combinent des méthodes de rééchantillonnage, d'optimisation, et de microsimulation pour générer des individus cohérents à la fois aux niveaux agrégés et désagrégés. Ce travail illustre l'applicabilité de plusieurs techniques rigoureuses à grande échelle, même en l'absence de données confidentielles, tout en garantissant la cohérence avec les distributions observées à différents niveaux géographiques.

Concernant les méthodes de validation, les différents articles analysés montrent des approches contrastées. Dans le cas de (Mathieu et Picault 2013), la validation repose sur la similarité des structures de tickets de caisse et la capacité du modèle à reproduire des comportements observés, sans recours à des catégories exogènes. (Hazelbag et al. 2020) soulignent un manque général de validation systématique dans les modèles épidémiologiques, notamment en ce qui concerne l'incertitude et la robustesse des calibrations. (Hörl et al. 2021) propose une validation croisée entre niveaux agrégés (par commune, tranche d'âge, etc.) et profils individuels simulés, en comparant avec les statistiques officielles. Enfin, (Chapuis et al. 2022) valident indirectement leurs résultats via la cohérence des comportements émergents observés dans les simulations. Ces cas montrent que la validation des populations synthétiques reste un défi méthodologique important, qui nécessite à la fois des métriques quantitatives et une évaluation qualitative en lien avec les objectifs de la simulation.

Dans cette thèse et pour les chapitres suivants, nous avons fait le choix d'utiliser des techniques d'extrapolation pour générer des populations synthétiques. En nous appuyant sur des données disponibles publiquement, notamment les statistiques de revenus de l'INSEE et des résultats de sondages réalisés en magasin. Le processus de génération s'appuie sur des méthodes classiques, ce qui le rend adapté à notre cas d'usage centré sur des comportements clients spécifiques. Toutefois, par manque de données empiriques établissant un lien direct entre revenus et comportements d'achat dans notre contexte, nous avons choisi de ne pas croiser directement ces deux sources d'information. Ce choix reflète une certaine forme de compromis entre réalisme statistique et faisabilité, mais il correspond aussi aux limites méthodologiques soulevées dans certains travaux, en particulier concernant la transparence et la formalisation du lien entre données et agents. Si notre méthode peut apparaître artisanale au regard de systèmes plus formels, elle illustre bien les enjeux pratiques liés à la disponibilité des données.

2.3 Notre cas d'application : Les clients d'un supermarché

Dans cette section, nous présentons un cas d'étude portant sur une base de données issue de tickets de caisse en magasin. L'objectif est d'estimer le nombre de clients par unité de temps. On parle de « pas de temps » qui peuvent représenter un jour, plusieurs jours, une semaine, etc. Le nombre de clients n'est pas directement disponible dans les données collectées. Environ 40 % des achats ne sont pas associés à un client identifié, ce qui entraîne une perte d'information significative.

Une première approche, simple et intuitive, consiste à assimiler le nombre de clients quotidiens au nombre total de tickets, en posant l'hypothèse que la proportion de clients effectuant plusieurs visites dans le même magasin au cours d'une même journée est négligeable. Cette hypothèse est acceptable lorsque l'analyse se limite à un pas de temps journalier. En revanche, dès lors que l'on souhaite adopter un pas de temps plus large, par exemple hebdomadaire, elle perd en pertinence, car la probabilité qu'un même client effectue plusieurs achats sur la période augmente sensiblement.

Afin d'éviter de contraindre l'analyse à une granularité journalière et de limiter les biais liés à cette hypothèse, nous proposons dans cette section plusieurs méthodes alternatives pour estimer le nombre de clients, adaptées à différents horizons temporels.

Dans la suite de cette section, nous utiliserons les expressions client identifié et ticket identifié pour désigner respectivement les clients et tickets associés à une carte de fidélité, permettant ainsi de relier chaque achat à un profil client complet. Dans ce cas, l'ensemble des informations nécessaires à l'analyse est disponible. À l'inverse, nous emploierons les termes client anonyme et ticket anonyme pour qualifier les transactions qui ne sont rattachées à aucun client, ainsi que les clients potentiels correspondant à ces tickets.

La base de données à notre disposition contient un ensemble d'achats réalisés en magasin, décrits au niveau du produit. Chaque ligne correspond à un produit acheté, accompagné de ses caractéristiques associées telles que le prix payé, l'éventuelle promotion appliquée, et d'autres informations contextuelles. Les produits sont regroupés en tickets à l'aide d'un identifiant de ticket, ce qui permet de reconstituer les paniers d'achat. Par ailleurs, une partie de ces tickets est associée à un identifiant client, permettant de relier ces transactions à des individus identifiés via leur carte de fidélité. La table 2.1 est un exemple de la table utilisée pour notre analyse.

Table 2.1 – *Exemple de la base de données des achats*

Type de produit	ID Ticket	ID Client	Prix payé (€)	Prix avant remise	Date
DECORATION	380/86/6842	None	23.9	23.90	2023-04-11
ECLAIRAGE	45/4/8360	66833635	16.	18.	2023-04-11
⋮	⋮	⋮	⋮	⋮	⋮

La méthode la plus simple pour estimer le nombre de clients sur une période donnée consiste à calculer, pour cette période, le nombre moyen de tickets par client identifié, puis à diviser le nombre total de tickets par cette moyenne. Cette approche servira de méthode de référence comparative pour la suite de nos analyses.

Cependant, il convient de souligner que le comportement des clients identifiés diffère fondamentalement de celui des clients anonymes, la principale distinction entre ces deux groupes résidant dans la possession d'une carte de fidélité.

2.3.1 Hypothèse d'un coefficient de fréquence d'achat

Cette première méthode est présentée avant tout pour illustrer la difficulté fondamentale du problème : l'absence d'informations directes sur le comportement des clients anonymes. Elle met en évidence qu'il est possible de recourir à des approches simples pour contourner ce manque de données, mais que celles-ci présentent rapidement des limites méthodologiques et introduisent des biais potentiels. Lorsqu'on essaie d'adresser ces biais on se heurte au manque de données. Nous ne retenons donc pas cette méthode dans le cadre de notre analyse principale. Elle constitue néanmoins un exemple utile, car elle permet d'expliciter la nature du problème et de montrer pourquoi il est nécessaire de développer des approches plus élaborées.

Nous postulons que les clients identifiés ont un comportement plus fidèle, se traduisant par une fréquence de visite plus élevée que celle des clients anonymes. Le nombre moyen de tickets par client est plus élevé parmi les clients identifiés.

Soient :
- $\frac{Tickets_{anon}}{Clients_{anon}}$: le nombre moyen de tickets par client anonyme,
- $\frac{Tickets_{id}}{Clients_{id}}$: le nombre moyen de tickets par client identifié.

Nous introduisons un coefficient d'ajustement $\alpha \in (0, 1)$, défini comme suit :

$$\frac{Tickets_{anon}}{Clients_{anon}} = \alpha \times \frac{Tickets_{id}}{Clients_{id}} \tag{2.1}$$

Ce coefficient α peut être estimé à partir de données externes ou historiques pour lesquelles l'identification client est plus complète (tests en environnement contrôlé, magasins pilotes, enseignes comparables, etc.). Il peut également être approché en isolant des clients « semi-identifiés » (par exemple, des clients dont l'identité a été reconstruite sans carte de fidélité), que l'on suppose avoir un comportement similaire à celui des clients anonymes.

Dans ce cas, α peut être exprimé de manière empirique par la formule suivante :

$$\alpha = \frac{Tickets_{anon} \times Clients_{id}}{Clients_{anon} \times Tickets_{id}} \tag{2.2}$$

Une fois la valeur de α déterminée, nous faisons l'hypothèse qu'elle reste stable lorsqu'on change de contexte ou lorsqu'on considère une population élargie de clients anonymes. Le nombre estimé de clients anonymes peut alors être calculé par :

$$Clients_{anon} = \frac{Tickets_{anon}}{\alpha \times \frac{Tickets_{id}}{Clients_{id}}} \tag{2.3}$$

Intuitivement, cela revient à diviser le nombre total de tickets anonymes par un nombre moyen de tickets par client ajusté selon le facteur α.

Nous ne disposons pas des données suffisantes pour estimer précisément α, et celui-ci peut varier selon le contexte d'achat, le type de magasin, ou la période considérée.

Avantages et limites de la méthode

La principale force de cette approche réside dans sa simplicité conceptuelle et sa cohérence avec une observation souvent vérifiée sur le terrain : les clients disposant d'un programme de fidélité ont tendance à fréquenter davantage le magasin. Cette hypothèse permet de construire un modèle interprétable, fondé sur des indicateurs facilement observables (nombre de tickets, nombre de clients identifiés), et pouvant être rapidement mis en œuvre, à condition de disposer d'une estimation raisonnable du coefficient α.

Cependant, cette méthodologie présente également plusieurs limites. Elle repose fortement sur l'hypothèse de stabilité du coefficient α entre différents contextes ou périodes. Or, ce paramètre peut varier en fonction de multiples facteurs : type de magasin, moment de l'année, canal d'achat, ou encore politiques commerciales. En l'absence de données comparables pour estimer α, l'approche peut introduire un biais significatif dans l'estimation du nombre de clients anonymes. Par ailleurs, cette méthode ne prend en compte qu'une dimension du comportement d'achat : la fréquence de visite, et ignore d'autres indicateurs potentiellement discriminants, tels que la valeur moyenne du panier, la diversité des produits achetés, ou les préférences de consommation. Elle doit donc être utilisée avec prudence, idéalement en complément d'autres méthodes ou d'analyses exploratoires.

Pour ces raisons, cette méthode constitue un bon point de départ pour estimer le nombre de clients anonymes, à condition de disposer de données suffisamment riches pour calibrer correctement le coefficient α, par exemple via des clients « semi-identifiés », dont le comportement peut raisonnablement être assimilé à celui des clients anonymes. Cependant, il est nécessaire de définir les clients semi-identifiés comme des clients que nous avons pu reconnaître grâce à une méthode neutre, c'est-à-dire qui ne repose pas sur un type de comportement particulier. Par exemple, l'utilisation d'un numéro de carte bancaire, d'une adresse de livraison ou d'un compte client en ligne ne constitue pas une méthode acceptable, car elle exige un comportement spécifique de la part du client. Une telle méthode permet ainsi de produire un premier estimatif dans un cadre exploratoire ou dans des environnements dans lesquels l'identification client est partielle mais structurée. Néanmoins, cette approche ne saurait être utilisée seule pour des analyses décisionnelles ou stratégiques, sans risquer une erreur d'estimation significative. Elle doit impérativement être complétée par d'autres sources d'information ou méthodes d'inférence permettant de valider, nuancer ou corriger ses résultats.

2.3.2 Modèle de simulation

Étant donné que le comportement des clients identifiés ne peut pas être directement transposé à celui des clients anonymes, il est nécessaire de mettre en place une méthode de contournement. Si les comportements individuels ne peuvent être utilisés pour recons-

truire ceux des clients anonymes, il est possible qu'une relation stable et commune à l'ensemble des clients puisse servir de base à l'estimation.

Nous formulons l'hypothèse suivante : la relation entre le montant moyen des tickets et la fréquence d'achat est constante, indépendamment du type de client.

L'objectif est donc d'exploiter cette relation *montant d'un ticket / fréquence d'achat* afin d'estimer la fréquence d'achat de clients simulés représentant la population anonyme.

Étude de la relation montant d'un ticket / fréquence

Les observations montrent qu'il n'existe pas de relation linéaire directe entre le montant d'un ticket et la fréquence d'achat. En revanche, on distingue une borne supérieure, visible sur la figure 2.1, qui peut être utilisée comme approximation pour estimer la fréquence d'achat des clients anonymes. Cette approche permet ainsi de relier un montant de ticket à un nombre estimé de tickets par client.

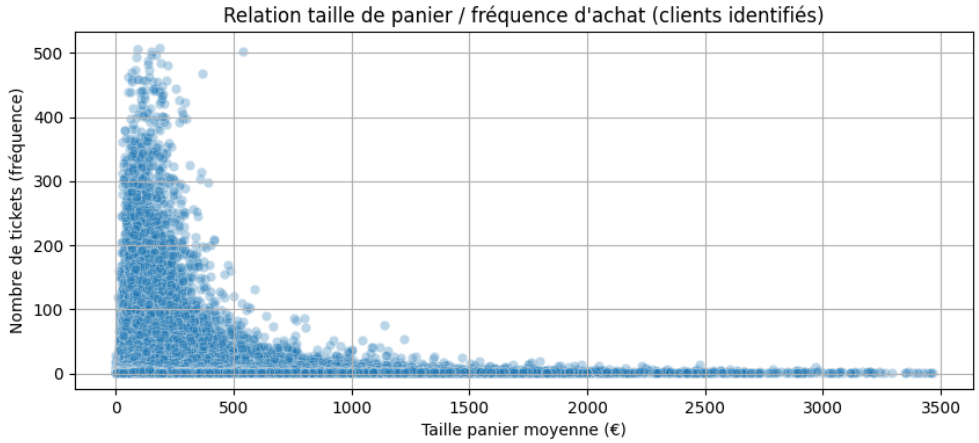


FIGURE 2.1 – *Relation entre la taille de panier moyen et le nombre de tickets par client. Aucun client n'a à la fois une fréquence d'achat élevée et un panier moyen élevé.*

Cependant, cette borne supérieure s'avère très élevée et conduit à des estimations peu satisfaisantes : elle produit, pour les clients anonymes, des fréquences d'achat jusqu'à 100 fois supérieures à celles observées chez les clients identifiés. Dans ces conditions, la borne inférieure calculée pour le nombre total de clients perd toute valeur indicative. Ce constat met en évidence la nécessité de recourir à des méthodes alternatives ou à des contraintes supplémentaires pour affiner l'estimation.

Apprentissage empirique de la relation

Une alternative à l'utilisation d'une borne unique consiste à adopter une approche probabiliste fondée sur l'exploitation empirique des données disponibles. Plutôt que de chercher à apprendre directement une fonction reliant le montant d'un ticket à la fréquence d'achat, nous construisons une table statistique qui associe, pour chaque intervalle de montant, une distribution empirique des fréquences d'achat observées chez les clients identifiés. Cette table peut ensuite être utilisée comme un modèle de correspondance : pour un montant donné, elle fournit un ensemble de fréquences possibles assorties de

leurs probabilités d'occurrence. D'un point de vue méthodologique, cette approche se rapproche des méthodes de modélisation par distributions conditionnelles. Elle présente l'avantage de ne pas imposer de forme fonctionnelle a priori et de pouvoir capturer des relations potentiellement non linéaires et multimodales entre le montant d'un ticket et la fréquence d'achat.

Dans la pratique, on commence par compter, pour chaque tranche de 5 € de montant de ticket, combien de clients ont réalisé chaque nombre possible de tickets sur la période étudiée. On divise ensuite ces effectifs par le total des observations de la tranche pour obtenir une distribution de probabilité : elle indique, pour un montant donné, la probabilité d'observer 1, 2, 3 tickets, etc. La table 2.2 ainsi construite correspond à une estimation non paramétrique d'une distribution conditionnelle par histogrammes (Sart 2017), ce qui permet de décrire la relation montant–fréquence d'achat sans imposer de forme mathématique prédéfinie.

Table 2.2 – *Extrait de la table de probabilités $P(\textit{tickets} \mid \textit{montant})$*

Montant (€) \Nb tickets	1	2	3	4	5	6	7	8	...
0–5	0.258	0.386	0.124	0.057	0.046	0.005	0.012	0.007	...
5–10	0.474	0.200	0.123	0.063	0.032	0.019	0.012	0.007	...
10–15	0.551	0.237	0.082	0.042	0.024	0.009	0.007	0.006	...
15–20	0.534	0.213	0.095	0.048	0.025	0.015	0.012	0.008	...
20–25	0.555	0.228	0.063	0.059	0.026	0.008	0.016	0.009	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Proposition d'algorithme

Nous proposons un algorithme visant à reconstituer les clients de la base de données. L'objectif est d'attribuer à chaque ticket dépourvu d'identifiant client un « client fictif » simulé, en regroupant les tickets anonymes selon des profils plausibles.

Pour ce faire, nous utilisons trois estimations empiriques issues des clients identifiés :

— la distribution conditionnelle $P(\text{tickets} \mid \text{montant})$, construite précédemment par comptage et normalisation pour chaque tranche de montant,

— la distribution conditionnelle $P(\text{montant} \mid \text{tickets})$, obtenue en regroupant les données par nombre de tickets afin d'estimer la probabilité des différents montants observés,

— la distribution $P(\text{fréquence} \mid \text{tickets})$, qui décrit, pour un nombre de tickets donné, la probabilité d'observer différents intervalles de temps entre deux achats. Les intervalles sont discrétisés selon les classes : *0j*, *1j*, *2j*, *3–4j*, *5–7j*, *1–2w*, *2–4w*, *4–8w*, *12–24w*, *24w+*.

La première distribution permet, à partir du montant d'un ticket anonyme, d'estimer un nombre plausible de tickets associés à un même client. La deuxième permet de simuler des montants typiques pour ces tickets. La troisième introduit une contrainte temporelle en

générant des intervalles de temps plausibles entre achats, que nous ajustons ensuite pour les faire tenir dans la période d'observation (ici du 01/01/2023 au 31/12/2024).

Algorithm 1 Reconstruction temporelle et monétaire des achats des clients non identifiés

$T_{\text{non-identifiés}} \leftarrow$ BaseDeTicketsAnonyme
$D^n_{\text{identifiés}} \leftarrow$ DistributionMontantsIdentifiésPourNTickets
$F_{\text{identifiés}} \leftarrow$ DistributionTicketsParMontant
$G_{\text{identifiés}} \leftarrow$ DistributionIntervallesParNTickets
$clients \leftarrow 0$
while $T_{\text{non-identifiés}}$ non vide **do**
    $t \leftarrow \text{randomTicket}(T_{\text{non-identifiés}})$
    $n \leftarrow \text{EstimerNbTickets}(t, F_{\text{identifiés}})$
    $M \leftarrow \text{TirerMontants}(n-1, D^n_{\text{identifiés}})$
    $\Delta T \leftarrow \text{TirerIntervalles}(n-1, G_{\text{identifiés}})$
    $\Delta T_{\text{scalés}} \leftarrow \text{AjusterPourFenêtre}(\Delta T, 01/01/2023, 31/12/2024)$
    $ticketsCompatibles \leftarrow \text{TrouverTickets}(t, M, \Delta T_{\text{scalés}}, T_{\text{non-identifiés}})$
    $T_{\text{non-identifiés}} \leftarrow T_{\text{non-identifiés}} \setminus \{ticketsCompatibles\}$
    $clients \leftarrow clients + 1$
end while
Retourner *clients*

— Sélectionner aléatoirement un ticket t dans la base des tickets anonymes $T_{\text{non-identifiés}}$.
— Estimer le nombre total de tickets n associés au même client que t en tirant aléatoirement selon la distribution $P(\text{tickets} \mid \text{montant})$.
— Générer $n-1$ montants supplémentaires via la distribution $P(\text{montant} \mid \text{tickets})$.
— Générer $n-1$ intervalles de temps entre achats via la distribution $P(\text{fréquence} \mid \text{tickets})$ et les convertir en dates, ajustées pour être comprises entre le 01/01/2023 et le 31/12/2024.
— Identifier dans $T_{\text{non-identifiés}}$ les n tickets dont le montant et la date sont simultanément les plus proches des valeurs générées.
— Répéter jusqu'à épuisement de $T_{\text{non-identifiés}}$.

La figure 2.2 présente le flux de traitement des données et schématise l'algorithme.

Le cœur de l'algorithme repose sur une estimation conjointe montant–fréquence–temps des tickets, dérivée des comportements observés chez les clients identifiés. L'hypothèse centrale reste que la relation entre montant, fréquence et espacement temporel des achats est identique entre clients identifiés et anonymes. Cette hypothèse, bien que moins restrictive qu'une assimilation complète des comportements, impose une structure fixe à cette relation et rend le modèle sensible à toute divergence temporelle ou monétaire entre les deux populations.

Comparée aux autres approches décrites précédemment, notre méthode présente un avantage notable : elle intègre explicitement la dimension temporelle dans la génération des clients fictifs. En modélisant les intervalles entre achats à partir des comportements observés, elle permet de reconstituer des séquences réalistes, sans supposer que les clients

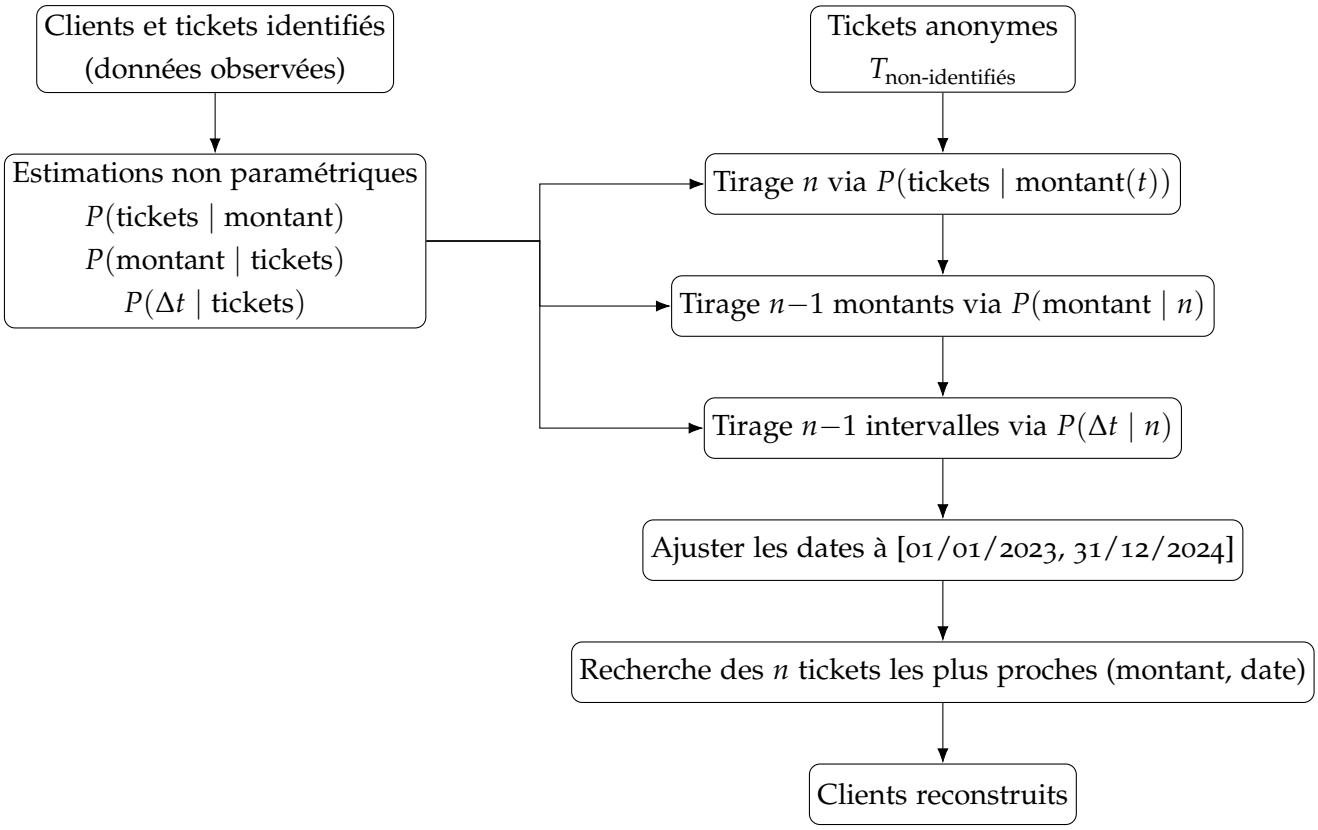


FIGURE 2.2 – *Schéma du flux de traitement pour la reconstitution des clients anonymes*

anonymes adoptent exactement les mêmes comportements que les clients identifiés. Cette absence d'hypothèse forte sur l'équivalence des deux populations rend le modèle plus souple et potentiellement plus robuste aux biais liés à l'identification partielle.

Validation de la méthode

La méthode proposée a été implémentée dans un code dédié, dont la version complète est mise à disposition sur GitHub [1]. Par souci de confidentialité, les données réelles utilisées en production ne peuvent pas être partagées.

À partir des transactions de clients identifiés, nous avons constitué un jeu d'entraînement pour estimer les distributions conditionnelles nécessaires au modèle $P(\text{tickets} \mid \text{montant})$, $P(\text{Montant} \mid \text{ticket})$ et $P(\text{fréquence} \mid \text{ticket})$. En parallèle, un jeu de test a été généré en « anonymisant » artificiellement un sous-ensemble de ces clients identifiés, c'est-à-dire en supprimant leurs identifiants tout en conservant l'intégralité des tickets correspondants.

La procédure de reconstruction a ensuite été appliquée sur ce jeu de test en utilisant les distributions estimées à partir du jeu d'entraînement. Les regroupements obtenus par l'algorithme ont été comparés aux regroupements réels connus. En particulier, nous avons analysé :

— la distribution du nombre de tickets par client,
— la fréquence moyenne d'achat,

1. Github CRIStAL-SMAC : https://github.com/cristal-smac/retail

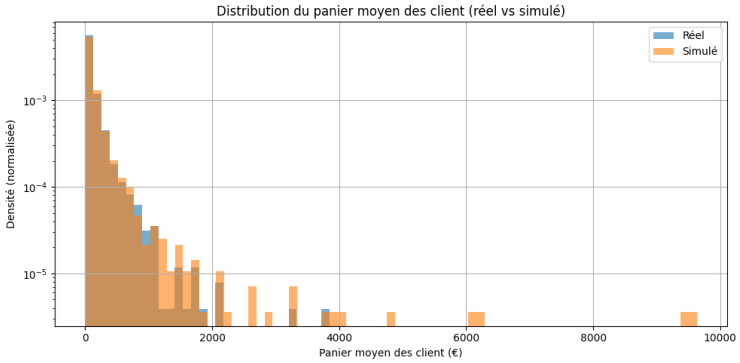


FIGURE 2.3 – *Comparaison du panier moyen des clients : données réelles vs données simulées. L'axe des ordonnées est en échelle logarithmique.*

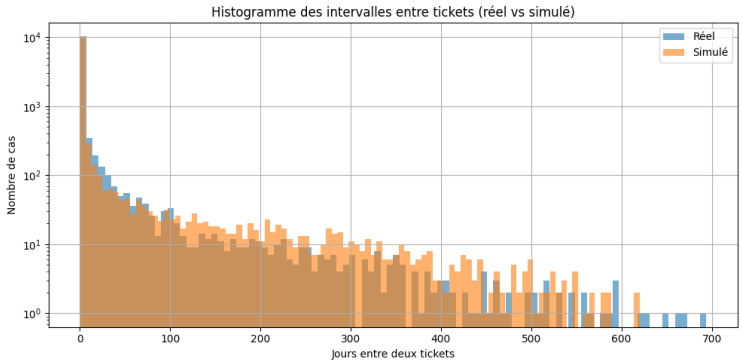


FIGURE 2.4 – *Comparaison des intervalles de temps entre les tickets d'un même client : données réelles vs données simulées. L'axe des ordonnées est en échelle logarithmique.*

— le nombre total de clients reconstitués.

Par ailleurs, les possibilités d'évaluation du modèle sont intrinsèquement limitées. En effet, l'algorithme ne modifie ni le contenu des tickets ni les informations associées aux produits : il ne fait que regrouper des tickets existants pour former des clients fictifs. Ainsi, toute métrique définie au niveau du ticket ou du produit (par exemple la composition des paniers ou les ventes unitaires par référence) reste identique entre les données simulées et réelles, rendant ce type de comparaison non pertinent pour juger de la qualité du modèle. L'évaluation doit donc se concentrer sur des indicateurs agrégés au niveau client, tels que la distribution du nombre de tickets, la fréquence d'achat moyenne ou la volumétrie globale de clients.

Les résultats de cette comparaison montrent que la méthode parvient à reproduire correctement les distributions agrégées d'intérêt, tant pour la fréquence d'achat que pour la volumétrie globale de clients. De plus le nombre de clients reconstruit correspond au nombre de client anonymisés avec un erreur de 5%. Cette erreur a tendance a diminuer a mesure que la taille de l'échantillon augmente.

Il convient de souligner qu'il n'existe aucune correspondance directe entre les regroupements de tickets simulés et les regroupements réels : l'algorithme ne vise pas à reconstruire fidèlement les paniers individuels des clients, mais à générer une structure synthétique respectant les distributions statistiques observées. En conséquence, l'information

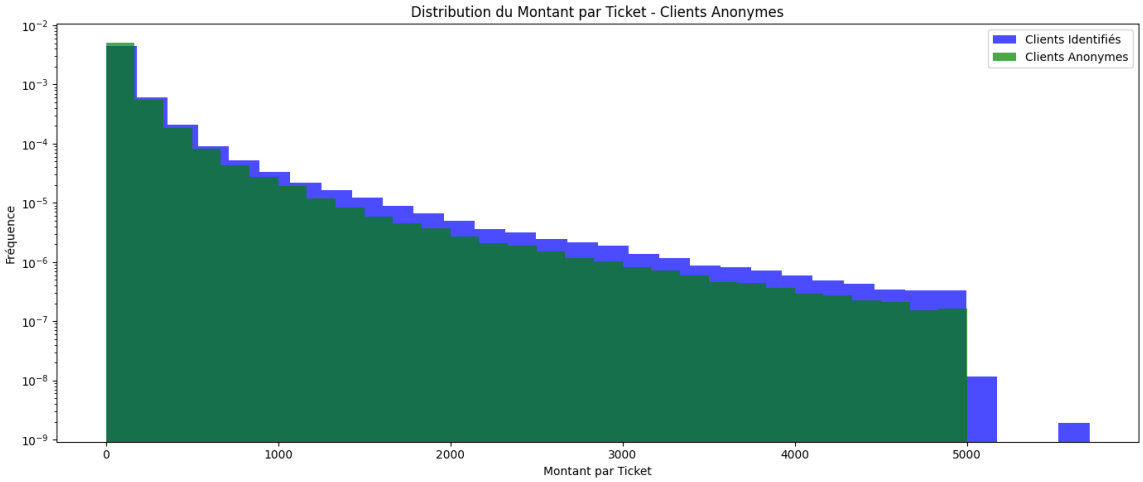


FIGURE 2.5 – *Comparaison de la distribution du montant par ticket entre les tickets identifiés et les tickets anonymes. L'échelle est logarithmique pour mieux visualiser la différence, même si cela rend la différence des clients à petits montants moins visible. On observe que les tickets anonymes ont un montant moins élevé que les tickets identifiés. La médiane des clients identifiés : 54.23€ avec un écart-type de 280. La médiane des clients Anonymes : 42.5€, avec un écart-type de 226.50*

transactionnelle fine (séquence exacte des achats d'un client réel) est perdue. L'objectif est avant tout d'obtenir une estimation cohérente de la fréquentation client et du volume potentiel d'acheteurs par catégorie, plutôt que de reconstituer des profils clients exacts.

2.3.3 Analyse des résultats

Nous avons ensuite appliqué notre méthodologie aux données qui nous intéressent directement : les tickets anonymes. La première étape consiste à comparer statistiquement la distribution des montants de tickets entre clients identifiés et anonymes. La figure 2.5 met en évidence une différence notable : les clients anonymes effectuent en moyenne des achats d'un montant inférieur, et ils sont proportionnellement plus nombreux à réaliser de très petits achats (inférieurs à 20 €). Cette observation suggère que les comportements d'achat pourraient différer entre les deux populations, ce qui justifie la prudence dans toute tentative d'extrapolation.

Dans un second temps, nous avons testé notre algorithme de reconstruction sur un ensemble contrôlé de tickets. L'idée de cette expérience est de vérifier la cohérence du modèle : si l'on utilise 5000 tickets identifiés et que l'algorithme reconstitue environ 2000 clients, alors, en appliquant la même procédure à 5000 tickets anonymes, on s'attend à obtenir un ordre de grandeur similaire pour le nombre de clients reconstruits, éventuellement avec un léger surplus si notre hypothèse, selon laquelle les clients identifiés présentent une régularité de visites plus forte que les clients anonymes, se vérifie.

Les résultats confirment cette intuition : le nombre de clients estimés à partir des tickets anonymes est proche de celui obtenu à partir des tickets identifiés. Une analyse plus fine de la table $P(\text{tickets} \mid \text{montant})$ permet de mieux comprendre ce résultat. On observe en effet que les clients anonymes sont surreprésentés parmi ceux qui réalisent de très petits achats (inférieurs à 20 €). Or, selon la distribution empirique, ces petits montants sont

associés à une probabilité relativement élevée de visites fréquentes. Ce mécanisme vient partiellement compenser la tendance globale des clients anonymes à effectuer des paniers de moindre valeur. Concrètement, plus le montant augmente, plus la fréquence estimée de visites tend à croître, sauf dans le cas particulier des achats très faibles (< 20 €), pour lesquels la probabilité de visites répétées est également forte.

Il est important de souligner que ces résultats reposent sur des distributions probabilistes et non sur des fréquences fixes : notre modèle capture une variabilité intrinsèque et reflète donc des comportements plausibles plutôt qu'une règle déterministe. Cela signifie que l'interprétation doit rester prudente : l'algorithme suggère que, si l'hypothèse de régularité entre montant et fréquence est valide, alors le nombre moyen de tickets par client anonyme est globalement similaire à celui des clients identifiés.

Néanmoins, ce constat soulève une interrogation. Il peut sembler contre-intuitif que des clients anonymes réalisant de très petits achats présentent effectivement une fréquence de visite élevée. Ce point met en lumière une limite possible de notre hypothèse de correspondance entre comportements identifiés et anonymes : certains effets propres à la population anonyme pourraient ne pas être correctement capturés par la distribution conditionnelle estimée à partir des clients identifiés. Cette question appelle donc à des recherches complémentaires, afin de mieux cerner les différences structurelles entre les deux populations.

2.3.4 Discussion

À première vue, on pourrait croire que ce type de problème est relativement simple à résoudre. Notamment après la mise en avant des premières méthodes. Pourtant, l'expérience montre qu'il n'en est rien. Obtenir une estimation fiable du nombre de clients à partir de données incomplètes nécessite des actions de terrain : enquêtes, observations complémentaires ou accès à des bases de données plus détaillées. Sans ces apports, il est difficile d'atteindre un haut niveau de précision.

La distinction entre clients identifiés (via carte de fidélité) et clients anonymes n'est pas uniforme. Elle peut varier fortement selon les enseignes, le taux d'utilisation des cartes de fidélité, le type de magasin, ou encore la zone géographique. Autrement dit, un modèle efficace dans un contexte ne le sera pas forcément dans un autre.

C'est précisément pour cette raison que nous avons proposé ici une méthode alternative. Elle ne prétend pas être la seule possible, ni la plus générale, mais elle illustre une approche différente pour aborder ce problème. Chaque méthode apporte une réponse partielle au problème, en fonction de ses hypothèses de départ. Il est donc essentiel de choisir l'approche la plus adaptée au contexte d'étude et aux contraintes de terrain.

Contrairement aux méthodes reposant sur l'hypothèse d'un comportement homogène, notre approche propose un cadre plus souple, mieux adapté aux contextes où les clients identifiés ne sont pas représentatifs de l'ensemble de la population. Même si cette prudence rend la validation plus difficile, elle constitue un gage de robustesse dans des contextes où peu d'informations sont disponibles sur les clients anonymes.

Face à un problème aussi variable et contextuel, la plupart des méthodes doivent être adaptées au cas étudié. En ce sens, les approches proposées sont souvent construites sur mesure pour répondre aux contraintes spécifiques du terrain et des données disponibles.

2.4 Synthèse

L'algorithme de reconstruction de client présenté dans ce chapitre joue un rôle central dans la chaîne de modélisation que nous proposons. Il permet de reconstituer, à partir de données partielles ou bruitées, une estimation fiable du nombre d'acheteurs potentiels dans une catégorie donnée. Cette étape est essentielle car elle transforme des données brutes difficilement exploitables en un indicateur synthétique, directement mobilisable dans un modèle prédictif.

Les sorties de l'algorithme peuvent alimenter des approches de machine learning visant à anticiper l'évolution future du nombre d'acheteurs. L'idée est de coupler les données reconstruites avec des séries temporelles ou des variables explicatives externes (par exemple l'évolution des prix, les campagnes marketing, ou des indicateurs macroéconomiques) afin d'entraîner un modèle prédictif. Celui-ci permettrait non seulement de décrire la situation actuelle, mais aussi de projeter le nombre d'acheteurs attendus dans différents scénarios futurs. Ainsi, on obtient un nombre d'acheteurs sur différents produits.

Notre résultat principal est un algorithme qui donne le volume global de clients potentiels sur différentes catégories de produits. Ce volume peut ensuite être introduit comme donnée d'entrée dans le modèle que nous développerons au chapitre suivant, lequel cherche à expliquer et à simuler comment ces clients se répartissent sur différents produits en fonction de leurs préférences, des caractéristiques des produits et des interactions concurrentielles. Par exemple, si l'algorithme de reconstruction permet d'identifier qu'environ 10 000 acheteurs appartiennent à une catégorie particulière, le modèle de répartition pourra ensuite donner la manière dont ces 10 000 acheteurs se distribueront entre plusieurs produits concurrents. On fait alors le lien entre les deux premières dimensions du comportement des consommateurs étudiés dans cette thèse.

Chapitre 3

Modélisation de choix, biais et utilité

Dans ce chapitre, nous discuterons de la modélisation du choix d'un produit parmi un ensemble de produits similaires. Nous souhaitons répondre à la question : « Comment un consommateur choisit son produit une fois qu'il sait ce qu'il souhaite acheter ? ». Pour cela nous modélisons le comportement de consommateurs au travers d'un biais cognitif : l'aversion à la perte. Ensuite, nous montrons que notre modélisation permet l'émergence de différents effets, notamment les effets de contexte et nous observons la capacité de notre modèle à reproduire le réel.

*Parmi ces biais, nous mettons particulièrement l'accent sur l'aversion à la perte, décrite initialement dans le cadre de la théorie des perspectives (*Kahneman *et al. 1979). Ce biais traduit la tendance des individus à accorder plus de poids à une perte qu'à un gain de même ampleur. Dans un contexte de choix entre produits, cela signifie par exemple qu'une légère dégradation perçue sur une caractéristique (prix plus élevé, qualité plus faible) peut peser davantage dans la décision qu'un avantage équivalent sur une autre dimension. L'intégration de ce principe permet de rendre compte de comportements de choix asymétriques, souvent observés empiriquement mais difficiles à expliquer dans le cadre des modèles classiques.*

Nous proposons donc une modélisation du comportement des consommateurs fondée sur ce biais cognitif. Notre objectif n'est pas seulement de reproduire des préférences individuelles, mais aussi de montrer que, lorsqu'on agrège ces comportements, des phénomènes collectifs émergent, notamment les effets de contexte. Ces effets, tels que l'effet de leurre, l'effet de similarité ou l'effet de compromis, montrent que la présence ou l'absence d'alternatives, même non choisies, peut modifier la répartition des choix entre les produits concurrents. Leur existence constitue une violation des hypothèses fondamentales de la théorie microéconomique standard (régularité et similarité), ce qui justifie l'importance d'en proposer une modélisation adaptée.

L'enjeu de ce chapitre est donc double. D'une part, nous proposons une formalisation du choix des consommateurs qui intègre explicitement l'aversion à la perte dans le processus décisionnel. D'autre part, nous évaluons la capacité de ce modèle à reproduire des comportements observés empiriquement, en particulier l'apparition d'effets de contexte. En ce sens, notre démarche vise à articuler rigueur théorique et validation empirique : à partir d'un principe psychologique bien documenté,

nous construisons une modélisation qui permet de rendre compte de régularités comportementales observées dans les expériences de choix réelles.

3.1 Définition d'une catégorie de produit

Pour notre modélisation, nous avons besoin de définir des catégories de produits. Les catégories de produits jouent un rôle fondamental dans la compréhension des comportements d'achat. Il existe plusieurs manières de modéliser une catégorie de produits, chacune répondant à des objectifs différents. L'approche la plus classique, issue du category management (Dussart 1998), définit une catégorie comme un ensemble de produits qui satisfont un même besoin ou qui sont perçus comme substituables. Dans cette optique, la catégorie est construite à partir des comportements de consommation observés, notamment dans les points de vente. Le category management est une approche stratégique qui consiste à gérer les catégories de produits comme des unités commerciales distinctes (Dhar et al. 2001), en se concentrant sur la satisfaction des besoins des consommateurs et sur l'optimisation des performances de chaque catégorie. Cette méthode implique une collaboration étroite entre les détaillants et les fournisseurs pour maximiser la valeur offerte aux clients.

Une autre manière de modéliser une catégorie repose sur les caractéristiques techniques des produits. Cette approche, plus normative, suppose qu'un produit est défini par un vecteur de caractéristiques, et que les catégories sont des sous-ensembles de l'espace des produits partageant certaines propriétés. L'auteur de (Lancaster 1966) développe l'idée que ce sont ces caractéristiques qui guident le choix des consommateurs. Dans cette approche, un smartphone pourrait être représenté par des attributs tels que la marque, le système d'exploitation, la capacité de stockage, etc. Cette modélisation permet une représentation fine de l'espace produit, utile notamment pour les algorithmes de recommandation ou pour l'analyse de marché.

Une troisième approche repose sur la substituabilité perçue. En économie, deux biens sont dits substituables si une augmentation du prix de l'un entraîne une augmentation de la demande pour l'autre. Les auteurs de (Pillai et al. 2014) analysent comment cet effet influence l'attitude des consommateurs envers les nouveaux produits introduits par les marques déjà existantes. Ce critère peut être utilisé pour regrouper les produits dans une même catégorie : on parle alors de substituts parfaits ou imparfaits, selon que les produits sont perçus comme totalement ou partiellement interchangeables. Ce concept a été largement repris en marketing, notamment à travers la notion de consideration set, qui représente l'ensemble des produits qu'un consommateur est réellement susceptible de choisir. Les auteurs de (Hauser 2014) discutent de la manière dont la présence ou non d'une marque dans les ensembles de considération des consommateurs peut altérer les ventes.

Enfin, certaines recherches empiriques proposent de modéliser les catégories de manière endogène, à partir des données de consommation. Par exemple, des méthodes de clustering ou de co-achat permettent d'identifier des groupes de produits souvent achetés ensemble ou perçus comme similaires dans leur usage (Chen et al., 2015). Ces méthodes,

fondées sur des données empiriques, révèlent parfois des regroupements contre-intuitifs mais pertinents du point de vue du comportement réel des consommateurs. Par exemple, les auteurs de (Holy et al. 2017) réalisent une catégorisation de leurs produits à l'aide d'une méthode de clustering.

Definition 3.1.1 (Définition d'une catégorie de produits)**.** Dans ce chapitre, nous adoptons une approche intermédiaire. Nous définissons une catégorie de produits comme un ensemble de produits interchangeables dans leur usage. Cela signifie que, pour un consommateur ayant défini un besoin (par exemple, acheter un smartphone), tous les produits de la catégorie peuvent être substitués sans perte significative d'utilité. Cette approche s'aligne sur les modèles classiques de substitution en économie, tout en conservant une souplesse permettant d'ajuster les frontières des catégories selon les contextes de consommation. Nous faisons ici une hypothèse a priori sur la structure des catégories, en la fondant sur des attributs jugés pertinents du point de vue du consommateur. Cette hypothèse pourrait être révisée ou validée à l'aide d'une analyse de données réelles, notamment par le recours à des méthodes telles que le clustering, afin de catégoriser les produits étudiés.

3.1.1 Un exemple de catégorisation des produits dans un magasin de distribution de produits culturels

Pour illustrer les différentes approches de définition d'une catégorie de produits, prenons l'exemple concret d'un distributeur de produits culturels.

Approche du category management. Dans cette approche, une catégorie est pensée comme une unité de gestion commerciale. Par exemple les différentes catégories sont : *Smartphones*, *Ordinateurs portables*, *Consoles de jeux vidéo*, *Appareils photo reflex* etc. Chaque catégorie est gérée comme un rayon distinct, avec ses propres objectifs de ventes, promotions, partenariats avec les fournisseurs (Apple, Samsung, Sony, Canon, etc.), et son rôle stratégique dans l'attractivité du magasin.

Approche technique (vecteur de caractéristiques). Ici, les produits sont définis par leurs attributs objectifs. Un smartphone peut être décrit par des caractéristiques telles que le système d'exploitation (iOS, Android), la taille de l'écran, la capacité de stockage, la résolution de l'appareil photo, ou encore la compatibilité 5G. Cette logique permet aux revendeurs de proposer, par exemple sur leurs sites web, une navigation par filtres ou par comparateurs techniques.

Approche économique par substituabilité perçue. Dans cette approche, on regroupe les produits qui peuvent se remplacer dans l'usage du consommateur.

- Exemple : *Smartphones Apple vs. Smartphones Samsung*. Un consommateur à la recherche d'un smartphone haut de gamme peut hésiter entre un iPhone et un Galaxy, considérés comme substituts imparfaits (mêmes usages principaux, mais différences d'écosystèmes).
- Autre exemple : *Abonnement Deezer vs. Spotify Premium*, qui relèvent de substituts presque parfaits du point de vue de l'usage.

Approche empirique (clustering et co-achat). Ici, les catégories sont construites à partir des données de consommation réelles.

— Les clients qui achètent une console *PlayStation* achètent aussi régulièrement des casques gaming et des manettes supplémentaires. Une catégorie empirique de type *Écosystème PlayStation* peut ainsi émerger.
— Les acheteurs de smartphones achètent souvent en même temps des coques de protection et des écouteurs sans fil, ce qui permet de regrouper ces produits du point de vue des usages.

Ces regroupements ne correspondent pas nécessairement aux rayons classiques, mais reflètent les comportements d'achat effectifs.

Approche intermédiaire (adoptée dans ce chapitre). Enfin, selon l'approche retenue dans ce travail, une catégorie est définie comme un ensemble de produits interchangeables dans leur usage.

— Pour le besoin *acheter un smartphone*, la catégorie regroupe tous les modèles proposés (Apple, Samsung, Xiaomi, Oppo, etc.), interchangeables du point de vue de l'usage principal (téléphoner, naviguer sur internet, prendre des photos).
— En revanche, une tablette ou un ordinateur portable n'entrent pas dans cette catégorie, même si certains usages se chevauchent.

Nous définissons pour le reste de ce chapitre qu'une catégorie de produit est un ensemble de produits similaires ayant les mêmes caractéristiques. Reprenons notre exemple avec les smartphones, ils pourraient être représentés par un ensemble de caractéristiques tel que :

— Qualité appareil photo
— Type d'OS
— Marque
— Taille stockage
— Taille mémoire
— Type de puce

On définit alors la catégorie « smartphone » comme l'ensemble les produits ayant ces caractéristiques. Il est possible d'enlever une ou plusieurs caractéristiques afin d'agrandir la catégorie comme il est possible d'ajouter une caractéristique différenciante afin réduire le nombre de produits d'une catégorie, ou dit autrement de la séparer en deux catégories distinctes. Dans ce chapitre, nous adressons le problème en définissant nos catégories comme l'ensemble des produits interchangeables dans la consommation, les produits substituables. Par exemple si je souhaite acheter de l'eau plate, peu importe le groupement qui est fait (6x1.5L par exemple) et la marque, les produits sont interchangeables.

3.2 Modélisation du choix des agents

Les consommateurs sont quotidiennement soumis à des biais cognitifs qui influencent leurs décisions d'achat, souvent de manière inconsciente. Ces biais, largement étudiés en économie comportementale et en psychologie, peuvent être amplifiés par le contexte dans lequel les choix sont effectués. Comprendre et modéliser ces effets dans un cadre

de simulation est essentiel pour analyser les mécanismes sous-jacents aux décisions des consommateurs. Parmi ces biais, les « effets de contexte » occupent une place centrale, car ils mettent en évidence la manière dont la structure de l'offre et la présence de certaines alternatives peuvent modifier les préférences des individus. Cette section explore ces effets en définissant leurs principes fondamentaux et en illustrant leur impact sur le comportement des consommateurs.

On appelle « effets de contexte » les effets psychologiques qui affectent les consommateurs dans leur processus de décision. Les choix des clients sont modifiés par l'apparition ou la disparition de produits qu'ils n'ont pas l'intention d'acheter. Les effets de contexte ne s'observent que sur des produits similaires, qui font partie d'une même catégorie. Ces produits ont des différences de caractéristiques, et sont en concurrence. Les consommateurs choisissent alors, selon leurs préférences, le produit qui leur convient le mieux.

Le choix des consommateurs est déterminé par de nombreux facteurs. Dans (Huber, Payne et C. Puto 1982) les auteurs sont les premiers à montrer que le contexte peut influencer de manière inattendue les processus de décision. Ils montrent l'existence d'un effet de leurre aussi appelé effet d'attraction.

Cet effet apparaît lors de l'introduction d'un nouveau produit, un leurre, qui améliore les ventes d'un produit déjà sur le marché alors même qu'il est en concurrence directe avec celui-ci. Le leurre est un produit moins attirant à cause d'un mauvais rapport qualité/prix. Ainsi, ce produit n'est pas acheté par les consommateurs, mais fournit un effet de levier sur les ventes des autres produits. L'existence de cet effet, comme tous les autres effets de contexte, montre qu'un produit qui n'est pas intéressant pour les consommateurs influence néanmoins leurs choix.

3.2.1 L'effet de leurre

On considère que A domine (respectivement domine partiellement) le produit B, noté $A \succ B$ (respectivement $A \succeq B$) si toutes les (respectivement au moins une) caractéristique(s) de A sont égales ou supérieures à B. Un produit asymétriquement dominé est un produit à la fois dominé par un produit et partiellement dominé par un autre. Par exemple, soit trois produits A, B et C tel que $A \succeq C$ et $B \succ C$, alors C est asymétriquement dominé. On peut tout à fait inverser A et B dans ce dernier exemple.

Historiquement, les modèles économiques posent l'hypothèse que l'introduction d'un nouveau produit dans une catégorie fait systématiquement diminuer les ventes de tous les autres produits. De plus, les produits initiaux les plus similaires à ce nouveau produit sont les produits pour lesquels les ventes diminuent le plus. Ces deux hypothèses lorsqu'elles sont vérifiées sont appelées « contraintes de régularité et de similarité ». L'effet de leurre par son existence remet en cause ces deux hypothèses historiques. Il est décrit par (Huber, Payne et C. Puto 1982) comme *« l'ajout d'une solution dominée asymétriquement viole la contrainte de régularité et l'hypothèse de similarité. »*

Prenons l'hypothèse que chaque produit P_i dispose de deux caractéristiques : son prix $P_{i,p}$

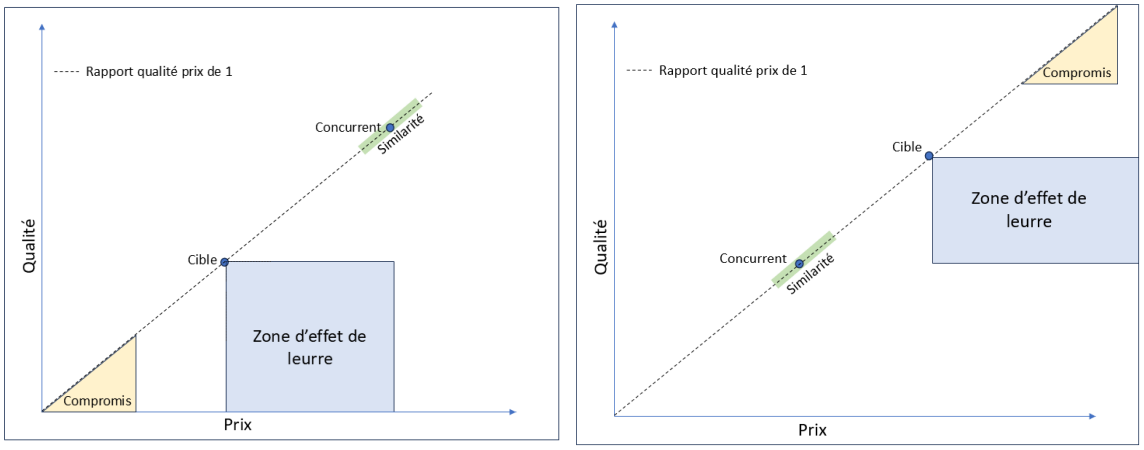


FIGURE 3.1 – *Ces deux figures présentent les zones d'action des différents effets de contexte. Chaque zone représente un ensemble de prix/qualité qu'un nouveau produit doit avoir pour activer l'effet. Ces effets ont comme impact d'augmenter les ventes du produit cible et de diminuer les ventes du produit concurrent.*

et sa qualité $P_{i,q}$. Considérons par exemple, un ensemble initial de deux produits P_a et P_b.

$$P_{a,p} = 10, P_{a,q} = 10 \tag{3.1}$$

$$P_{b,p} = 5, P_{b,q} = 5 \tag{3.2}$$

On imagine que les ventes se répartissent équitablement. Maintenant introduisons un nouveau produit P_c avec les caractéristiques suivantes :

$$P_{c,p} = 12, P_{c,q} = 8 \tag{3.3}$$

On note que $P_{b,p} < P_{a,p} < P_{c,p}$ ainsi que $P_{b,q} < P_{c,q} < P_{a,q}$. Donc $P_a \succ P_c$ et $P_b \succeq P_c$. On remarque que P_c est dominé par P_a car il est plus cher et de moins bonne qualité, mais qu'il est partiellement dominé par P_b parce qu'il est aussi plus cher, mais cette fois de meilleure qualité.

En conséquence, lorsqu'on introduit le produit P_c, on observe que les ventes de P_a augmentent.

Si les paramètres du produit leurre P_c sont $P_{c,p} = 8, P_{c,q} = 3$ on note alors $P_{b,p} < P_{c,p} < P_{a,p}$ et $P_{c,q} < P_{b,q} < P_{a,q}$. Donc $P_a \succeq P_c$ et $P_b \succ P_c$. Avec ces paramètres, on observe une augmentation des ventes du produit P_b.

On définit donc les effets de contexte comme l'ensemble des influences liées à l'environnement dans lequel se déroule l'acte d'achat, qui modifient la perception, l'évaluation et le comportement du consommateur, et par conséquent les ventes.

Dans la figure 3.1 on observe différentes zones de placement d'un éventuel nouveau produit par rapport à un produit cible et un produit concurrent. Dans cet exemple les différentes zones d'effets vont augmenter les ventes de la cible par rapport au concurrent.

En fonction de la position relative de ces deux produits, les zones d'effets de contexte changent. Ce cas simple à deux concurrents permet de facilement identifier ces zones, contrairement au scénario général dans lequel les effets peuvent s'additionner ou se compenser, rendant la position de telles zones plus difficiles.

Un point intéressant est que la zone de similarité de la cible et la zone de leurre du concurrent sont très proches. En effet, lorsqu'un nouveau produit est placé proche d'un autre, il tend à détourner une partie des préférences qui lui étaient initialement favorables : c'est l'effet de similarité. Mais en modifiant légèrement les caractéristiques de ce nouveau produit, il peut aussi jouer le rôle de leurre, rendant le produit qui lui est similaire plus attractive par comparaison. Ainsi, une configuration donnée peut avec un très léger changement soit affaiblir la cible par similarité soit la renforcer par un effet de leurre, ce qui montre la complexité et l'interdépendance des différentes zones d'effet de contexte.

3.2.2 Effet de similarité et effet de compromis

Il existe d'autres effets de contexte, les auteurs de (Noguchi et al. 2014) décrivent l'effet de leurre, l'effet de similarité et l'effet de compromis. Les paramètres des expériences sont toujours les mêmes que précédemment : deux produits sont en concurrence, le produit A cher et de bonne qualité, le produit B peu cher, mais de moins bonne qualité. Ils disposent du même rapport qualité/prix. L'ajout d'un produit C vient modifier les ventes de manière inattendue.

Sur ces effets, le produit C dispose du même rapport qualité/prix que les deux produits initiaux.

Lorsque le produit C est très similaire à A ou B, les ventes évoluent de manière asymétrique. Les deux produits voient leurs ventes baisser, mais l'effet est encore plus fort sur le produit similaire à C. Pour résumer, les consommateurs ont tendance à préférer le produit avec des caractéristiques qui se différencient des autres.

En revanche, lorsque le produit C est très différent, il crée un effet de compromis. Par exemple, si on introduit un produit C très cher et de très bonne qualité, le produit A, cher, devient alors un compromis entre les deux options (C très cher et B peu cher) et voit ses ventes augmenter. Les consommateurs préfèrent les produits qui forment un compromis entre deux autres produits.

Limites de l'effet de leurre

(Frederick et al. 2014 ; Huber, Payne et C. P. Puto 2014) montrent que l'effet de leurre est beaucoup moins fort lorsque les consommateurs n'ont pas de moyen clair d'évaluer la qualité. (Frederick et al. 2014) ont des résultats similaires à (Huber, Payne et C. Puto 1982) uniquement lorsque la qualité est définie numériquement.

Expériences réelles sur l'effet de leurre

Les effets de contexte se manifestent également en dehors de l'écosystème du marketing et sont fondamentaux dans la prise de décision. Par exemple, (Trueblood et al. 2013)

réalisent une expérience ou les sujets doivent trouver le rectangle avec la plus grande aire. Dans cette expérience d'évaluation intuitive, il n'y a pas de qualité définie numériquement pour aider les sujets. L'expérience est la même : on ajoute un leurre asymétriquement dominé par une des deux solutions en longueur et en largeur. Les auteurs montrent que même dans cette situation plus abstraite, on observe quand même l'effet de leurre.

(Pettibone et al. 2000) suggèrent que l'aversion à la perte, ce biais cognitif qui amène les individus à accorder plus d'importance à une perte qu'à un gain équivalent, serait l'une des causes possibles de l'effet de leurre. (Tversky et al. 1991) se focalisent sur l'aversion à la perte, ses effets et comment ce biais modifie les comportements. Ces mêmes auteurs proposent l'utilisation de la théorie des perspectives afin de décrire le comportement humain lors de choix risqués (Kahneman et al. 1979). Cette théorie est utilisée dans un cadre où les prises de décision ont des impacts incertains mais connus par les sujets de l'expérience. Ainsi chaque sujet connaît pour chaque choix la probabilité de chaque évènement. Ces évènements peuvent être positifs ou négatifs. L'évaluation des gains et des pertes par les sujets est difficile à cause de l'aspect aléatoire des résultats. De ce fait, des biais liés aux risques et à la perception des probabilités par les sujets interviennent. Pour notre modèle, on imagine utiliser ce cadre conceptuel sans l'aspect aléatoire afin de modéliser le processus de décision des clients.

Les auteurs de (Stanley et al. 2024) étudient l'impact de l'ensemble des choix disponibles sur les effets de contexte. Trois expériences sont réalisées. Dans une première expérience, les auteurs montrent que l'effet de leurre est de plus en plus faible à mesure que le nombre de choix initiaux disponibles augmente (allant de 3 à 9). Dans une seconde expérience, les auteurs montrent que l'ordre dans lequel les produits sont affichés à l'écran, dans le cadre d'achats en ligne, a une importance. Si les produits sont triés par prix ou qualité, l'effet de leurre a plus d'impact sur les choix des personnes testées. Enfin, dans une troisième expérience, grâce à l'utilisation d'un « eye tracker », les auteurs valident leur approche et répondent aux questions laissées en suspens. Ils montrent grâce à un index que les personnes ont tendance à créer un processus lexicographique dans leur manière d'analyser les différents produits. Lorsque le nombre d'articles disponibles augmente, les personnes sont encore plus impactées par cet effet. Les auteurs émettent l'hypothèse que s'il y a trop d'options, les choix ne se basent que sur une partie des produits disponibles. Si les produits sont triés, le processus d'évaluation des produits est facilité, les sujets peuvent évaluer un plus grand nombre de produits et sont alors davantage impactés par les effets de contexte.

Les auteurs de (Noguchi et al. 2018) parlent de « big three contextual effect » en parlant des effets de leurre, de similarité et de compromis. (Spektor et al. 2022) déclarent que toute étude sérieuse de modélisation, de prise de décision doit être capable de reproduire une série d'effets réellement observés tels que l'effet de leurre, de compromis et de similarité.

L'ensemble de ces travaux met en lumière la complexité et parfois la contradiction des résultats obtenus autour des effets de contexte, et en particulier de l'effet de leurre. D'un côté, certaines études (Frederick et al. 2014 ; Huber, Payne et C. P. Puto 2014) montrent

que cet effet tend à s'atténuer, voire à disparaître, lorsque les consommateurs n'ont pas de repère clair pour évaluer la qualité des options disponibles. D'un autre côté, d'autres expériences comme (Trueblood et al. 2013) suggèrent que même en l'absence de définition numérique explicite de la qualité, un effet de leurre persiste dans certaines tâches perceptives ou intuitives. Ces divergences indiquent que la nature et l'ampleur des effets de contexte dépendent fortement de la capacité des individus à comparer les alternatives de manière fiable. Lorsque la qualité est objectivable et mesurable, le processus décisionnel semble plus structuré et potentiellement moins sensible aux biais de présentation. À l'inverse, dans les situations où la qualité est difficile à évaluer, les consommateurs peuvent se reposer davantage sur des heuristiques, laissant plus de place aux influences contextuelles comme l'effet de leurre, de compromis ou de similarité. Ces différences soulignent que la manière dont la qualité est présentée ou perçue peut jouer un rôle dans la manifestation des effets de contexte, mais les travaux existants ne permettent pas d'en tirer une conclusion univoque. Ce constat justifie l'intérêt de considérer plusieurs configurations expérimentales et de rester prudent dans la généralisation des résultats.

3.2.3 Critiques de l'existant, les modélisations des effets de contexte

(Chica et al. 2017 ; Ulbinaitė et al. 2011 ; Zhang et al. 2007) proposent différents modèles à base d'agent, reproduisant l'effet de leurre. Les travaux des auteurs de (Zhang et al. 2007) sont basés sur la théorie de la motivation. Les auteurs définissent une fonction de motivation qui guide les agents dans leurs achats. Cette fonction est une fonction d'utilité que les agents utilisent pour évaluer les produits et déterminent quel produit acheter. Les agents choisissent systématiquement le produit le mieux évalué par la fonction d'utilité ou « fonction de motivation ». Dans ce modèle, les auteurs simulent l'achat d'une catégorie de produit à la fois avec au maximum trois produits par catégories.

La fonction d'utilité proposée est composée de quatre facteurs :

- Le facteur de bouche-à-oreille. Les agents influencent leurs voisins selon le choix qu'ils ont fait. Bien que dans cette étude la distance d'influence sur le voisinage puisse varier, dans leurs expériences les auteurs sont restés avec des distances d'influence du voisinage très petites. L'effet est assez faible à cause de ce paramètre.
- L'impact marketing. Ce paramètre est fixé sur l'environnement. Tiré aléatoirement en début de simulation, il est différent pour chacun des agents. Il est intéressant de modéliser l'impact de ce paramètre qui n'a pas été exploré par les auteurs.
- Le prix. Chaque agent dispose d'une sensibilité au prix qui lui vient de son revenu, plus l'agent a un revenu élevé, moins il est sensible au prix. L'impact du prix est calculé grâce à une fonction $-\alpha^{Prix_{article}-Prix_{moyen}}$. L'exponentiation sert à avoir une fonction qui reproduit l'aversion à la perte. La fonction croît toujours moins vite lorsque les agents sont gagnants que lorsqu'ils sont perdants.
- La qualité. Définie de la même manière que le prix. Chaque agent dispose d'une sensibilité à la qualité qui dépend de ses revenus. Plus l'agent a des revenus élevés, plus sa sensibilité à la qualité sera forte.

Cette fonction d'utilité a selon nous plusieurs défauts.

Pour l'évaluation du prix et de la qualité, les auteurs utilisent l'écart des caractéristiques

du produit évalué (Prix et Qualité), à la moyenne des caractéristiques de la catégorie. Cet écart est mis à la puissance des paramètres α et β. Ce sont les équations 3.4 et 3.5 où P_i est le prix du produit évalué, P_{avg} le prix moyen des produits, Q_i la qualité du produit évalué, Q_{avg} la qualité moyenne des produits de la catégorie. Finalement, k et L sont des constantes propres à chaque agent calculées grâce aux revenus des agents.

$$PS_i = -\alpha^{P_i - P_{avg}} + k \tag{3.4}$$

$$QS_i = \beta^{|Q_i - Q_{avg}|} + L \tag{3.5}$$

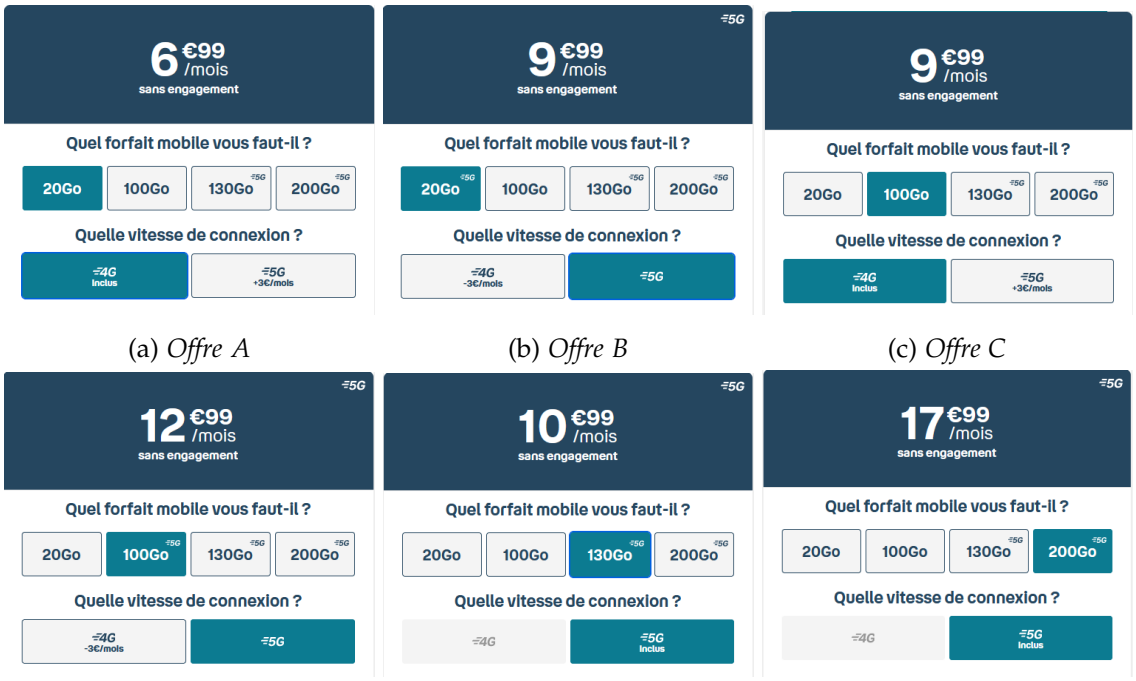
Les auteurs ne donnent pas précisément les valeurs de α et β cependant leur modèle est très sensible à ces deux paramètres. Ainsi un léger changement (de l'ordre de 1,1 vers 1,2) de l'un de ces deux paramètres et tous les agents achètent le même produit. De plus, lors de l'évaluation de la qualité, seul l'écart à la qualité moyenne est évalué. Par exemple, si le produit est de meilleure qualité, l'effet est le même que si le produit est de moins bonne qualité. De plus, les agents ne peuvent pas choisir le produit leurre. De ce fait, il est impossible d'étudier les possibilités de leurre ayant un rapport qualité/prix meilleur ou équivalent à celui des produits existants. En effet, dans cette situation, l'introduction d'un produit qui sur-performe les autres a comme impact de cannibaliser, de détourner les ventes des produits existants. Finalement, ce modèle ne reproduit pas les autres effets de contexte tels que l'effet de similarité ou l'effet de compromis.

Une autre modélisation est proposée par (Lee et al. 2014). Ce système repose sur une fonction d'utilité linéaire avec de l'aversion à la perte. Pour calculer l'utilité des produits, les agents utilisent un produit de référence et mesure l'écart du produit évalué à ce produit de référence. Ils intègrent aussi un système d'influence entre agents en réseau, introduisant ainsi des facteurs qui se confondent et qui donc compliquent l'analyse des résultats. De ce fait, les auteurs ne font pas le lien entre leurs résultats et les effets de contexte.

3.3 Quelques exemples réels

Il est indispensable aujourd'hui pour les entreprises d'utiliser tous les leviers qui sont à leur disposition afin d'améliorer leur stratégie marketing. L'effet de leurre, de similarité et de compromis doivent faire partie de ces leviers. Cependant, ces effets peuvent être décrits lorsqu'il n'y a que trois produits mis en œuvre (deux produits initiaux et un produit introduit), mais deviennent bien plus complexes dans le cas général. Les différents effets de contexte s'accumulent et il devient difficile de bien comprendre quel est l'effet final de l'introduction d'un nouveau produit. L'industrie de la téléphonie mobile est un domaine dans lequel les effets de contexte sont utilisés de manière évidente.

La figure 3.2 issue d'un exemple réel d'une entreprise de téléphonie mobile nous permet d'observer plusieurs effets de contexte. Cet exemple dispose d'un effet de compromis sur les offres C, D et E. En effet, les offres A et F sont des offres extrêmes que les clients vont éviter afin de se concentrer sur les offres dites « de compromis ». De plus, l'offre E bénéficie d'un effet de leurre évident, c'est une offre à 130 Go, avec la 5G pour 10.99



(a) *Offre A* (b) *Offre B* (c) *Offre C*

(d) *Offre D* (e) *Offre E* (f) *Offre F*

FIGURE 3.2 – *Un exemple d'offres mobiles datant du 19/03/2024. Ces six offres sont toutes sur une même page, la navigation entre les offres est fluide et la comparaison facile. Autre exemple.*

€ tandis que l'offre *B* est plus chère et dispose de 100 Go avec la 5G. Dans cet exemple, les offres 4G sont aussi disponibles en 5G pour un coût de 3 € supplémentaire, mais ne sont pas mises en avant dans l'interface graphique, il faut cliquer sur une offre 4G pour lui ajouter la 5G. Cela peut modifier la perception ou l'impact qu'a la 5G sur le consommateur. Enfin, si la présence de la 5G n'est pas importante pour le consommateur, l'offre *B* peut agir comme un leurre et améliorer la perception de l'offre *C*. À l'inverse, un consommateur qui accorde de l'importance à la 5G verra plutôt l'offre *C* comme un second leurre, améliorant les ventes de l'offre *E*. Avec ces deux derniers exemples, on saisit toute l'importance de la perception de la qualité.

Afin de résumer le positionnement des six offres, on représente chaque offre par un point sur un graphique qualité/prix sur la FIGURE 3.3. Cette fois, il est clair que l'offre *E*, entourée sur la FIGURE 3.3, est l'offre la plus mise en avant par l'entreprise. En effet, cette offre bénéficie de l'effet de leurre et de compromis. Pour que les entreprises puissent tirer le meilleur parti possible de ces effets, il leur est donc préférable d'avoir un outil de simulation afin de tester au préalable l'impact de l'introduction d'un nouveau produit ou d'une nouvelle offre. Nous proposons ici un modèle d'aide à la décision, étendant l'existant (MATHIEU, PANZOLI et al. 2011 ; SAID et al. 2002 ; ULBINAITĖ et al. 2011 ; ZHANG et al. 2007). Ce modèle multi-agents, adaptatif aux caractéristiques des produits, mais aussi aux caractéristiques spécifiques des clients, a pour but d'observer l'impact des différents effets de contexte.

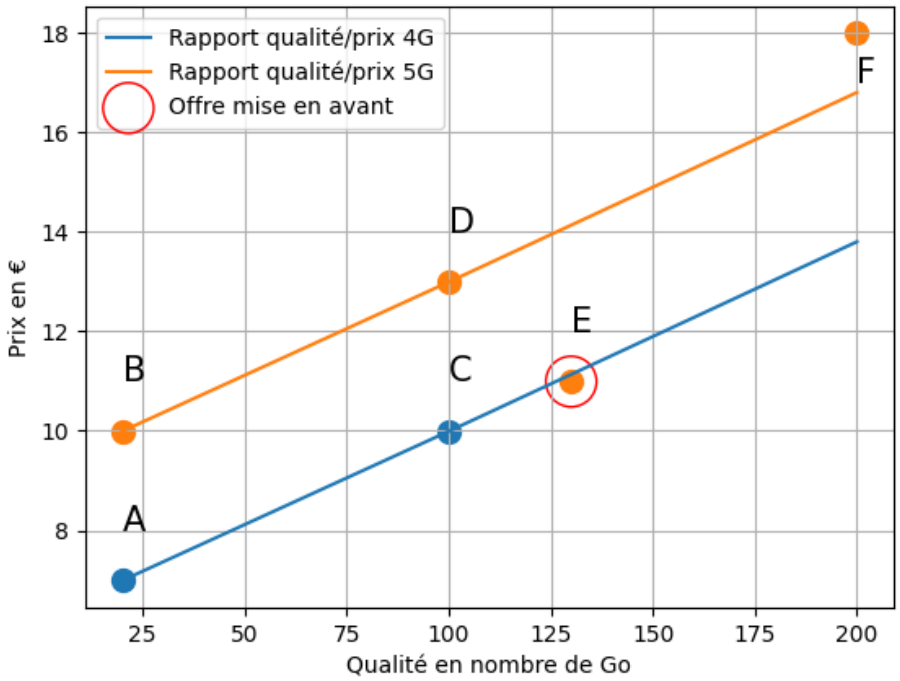


FIGURE 3.3 – *Les six points de cette figure correspondent aux offres de la* FIGURE 3.2 *dont nous avons tracé la tendance. Les points orange correspondent aux offres 5G, les bleus aux offres 4G. Le point E se trouve proche de la droite 4G par coïncidence.*

3.4 Proposition de modélisation

Dans cette section, nous allons proposer une modélisation dans la lignée directe des travaux de (LEE et al. 2014 ; ZHANG et al. 2007). Le but est de modéliser l'aversion à la perte et de montrer qu'ensuite le modèle est capable de reproduire les différents effets de contexte. Enfin, nous montrerons que nous pouvons calibrer ce modèle avec des données réelles. Nous ne considérons qu'une catégorie de produit à la fois, mais le modèle pourrait être généralisé. Enfin, l'objectif de ce modèle n'est pas de trouver un nombre de clients potentiels, mais plutôt de déterminer le choix de ces clients potentiels.

3.4.1 Les produits

Étant donné notre définition d'une catégorie (cf. définition 3.1.1), les produits sont définis par leur prix et un ensemble de caractéristiques $C_1, ..., C_n$ où chaque C_i est soit nombre, soit un booléen, soit une variable catégorielle.

Si on souhaite généraliser l'approche à d'autres types de catégories, (Par exemple si on souhaite mettre en concurrence dans une même catégorie différents types de boites de conserves), il faut alors envisager un ensemble de caractéristiques plus large où certaines caractéristiques de certains produits seront « None » pour exprimer l'absence de la caractéristique.

Dans notre cas, on note Q la qualité d'un produit définit comme une somme pondérée de ses caractéristiques :

$$Q = x_1 \cdot C_1 + ... + x_n \cdot C_n \tag{3.6}$$

L'objectif est alors de trouver l'ensemble des poids $\{x_1, ..., x_n\}$ qui permettent de modéliser la qualité Q d'un produit. Les produits sont alors représentés par le couple

$$(P, Q)$$

, on notera le produit $Product(P, Q)$. Si on reprend notre exemple d'introduction, un smartphone pourrait être représenté par les caractéristiques suivantes :

— Qualité appareil photo
— Type d'OS
— Marque
— Taille stockage
— Taille mémoire
— Type de puce

3.4.2 Les agents

Les agents représentent les clients. Afin de rester simple, ils disposent d'un revenu et d'une sensibilité à la qualité. De ce revenu, on déduit une sensibilité au prix notée S_P [0,1]. La sensibilité à la qualité notée S_Q est dépendante de la catégorie analysée. Nous proposons d'utiliser un tirage aléatoire suivant une distribution obtenue par sondages de client et d'adapter ce tirage selon la catégorie de produits analysés.

On considère ici que lorsqu'un agent consulte un produit, il est exposé à tous les produits de la même catégorie. Ce n'est pas forcément le cas dans la littérature, le nombre produits évalués sur une catégorie est estimé entre 2 et 5 (Choudhary et al. 2017) et varie selon le contexte : type de catégorie, en ligne ou non. Dans nos exemples, nous considérons une taille de catégorie assez petite (entre 4 et 8 produits) pour que cet effet ne soit pas impactant.

Les agents disposent aussi d'une fonction d'évaluation ou fonction d'utilité, qu'ils utilisent pour évaluer chacun des produits. Cette fonction donne une valeur à chaque produit qui correspond à ce que l'agent pense du produit. Plus cette valeur est élevée, plus l'agent aime le produit.

$$U_p = S_P \times (\beta \times (p - p_{avg})_+ + (p_{avg} - p)_+) \tag{3.7}$$

$$U_q = S_Q \times ((q - q_{avg})_+ + \beta \times (q_{avg} - q)_+) \tag{3.8}$$

$$UF = U_p + U_q \tag{3.9}$$

Dans cette équation, on accorde plus d'importance aux pertes qu'aux gains, pour cela, on introduit un paramètre $\beta < -1$ qui code l'aversion à la perte. La notation $(x)_+$ représente

le maximum entre 0 et une valeur x. L'utilité prix, respectivement qualité, évaluée par l'agent est notée U_p, respectivement U_q. La fonction d'utilité est notée UF.

Cette fonction d'utilité UF s'inspire des travaux de (Lee et al. 2014 ; Tversky et al. 1991 ; Zhang et al. 2007). C'est une fonction linéaire modifiée par l'aversion à la perte comme dans les travaux de (Lee et al. 2014). Par ailleurs, elle utilise un produit de référence représenté par le produit moyen (Prix et qualité moyenne de tous les produits disponibles (Lee et al. 2014 ; Zhang et al. 2007)) pour calculer l'utilité d'un autre produit. Le terme $(p - p_{avg})_+ + (p_{avg} - p)_+)$ et le terme $(q - q_{avg})_+ + (q_{avg} - q)_+)$ représentent ce calcul d'écart à la moyenne. Finalement, U_p (respectivement U_q) représente ce que l'agent perçoit en plus ou en moins par rapport à la moyenne du prix (respectivement de la qualité) en choisissant ce produit. Ces deux fonctions sont ensuite multipliées par les sensibilités internes de l'agent avant d'être additionnées.

Afin de permettre une exécution rapide, nous avons programmé le modèle en utilisant un calcul matriciel. L'ensemble du code de cette modélisation est disponible à l'adresse : https://github.com/cristal-smac/retail

3.4.3 La simulation

Le modèle est conçu pour mesurer l'impact de l'arrivée d'un nouveau produit. Nous ne simulons pas de ventes, mais plutôt des envies d'achat. Il n'y a pas de stocks. Le choix d'un produit n'influence ni les choix futurs de l'agent ni ceux des autres agents. À chaque pas de temps, tous les agents évaluent chaque produit disponible et se positionnent sur le produit qu'ils achètent. Ils choisissent le produit avec la plus grande valeur d'utilité. Par la suite, lorsque nous parlerons « d'acheteurs du produit A », en réalité, nous parlons plutôt « d'agent préférant le produit A ». Il en va de même pour les notions de parts de marché.

Pour résumer, à chaque pas de temps, chaque agent évalue tous les produits disponibles dans le modèle grâce à une fonction d'utilité et choisit le produit avec la plus grande valeur d'utilité. Nous introduisons ensuite un produit qui vient modifier les comportements et les choix des agents.

3.4.4 Calcul matriciel

Nous avons d'abord implémenté ce modèle en utilisant NetLogo. Le code source est disponible sur le GitHub de l'équipe SMAC [1]. NetLogo est particulièrement adapté au développement de systèmes multi-agents, car il offre un environnement intuitif qui permet de se concentrer sur la modélisation plutôt que sur la complexité du code. Sa syntaxe concise facilite la prise en main, même pour des utilisateurs ayant peu d'expérience en programmation. NetLogo intègre nativement la gestion d'agents indépendants et de leur environnement, ce qui évite de devoir réimplémenter des structures complexes. De plus, il propose une interface graphique intégrée qui permet de visualiser en temps réel les interactions entre agents et l'évolution du système, un atout précieux pour analyser, tester et communiquer les résultats d'un modèle. Enfin, grâce à sa communauté active et à sa

1. Lien : https://github.com/cristal-smac/retail

riche bibliothèque d'exemples, il constitue un excellent outil pour prototyper rapidement et expérimenter différents scénarios dans le domaine des SMA.

Pour valider notre travail, nous avons ensuite choisi de réimplémenter le modèle en Python. Reproduire le modèle dans un autre langage oblige à repenser certains aspects de la modélisation et offre une seconde vérification de la justesse de l'implémentation.

Python présente plusieurs atouts pour le développement de systèmes multi-agents. Sa grande richesse en bibliothèques spécialisées (comme NumPy, Pandas ou Matplotlib) permet de gérer efficacement des calculs complexes, de manipuler des données à grande échelle et de produire rapidement des analyses visuelles des résultats. Contrairement à NetLogo, Python offre un langage généraliste, ce qui facilite l'intégration d'un modèle multi-agents dans des pipelines plus larges, par exemple pour coupler la simulation avec de l'apprentissage automatique, de l'optimisation ou du traitement de données. Sa flexibilité permet également d'implémenter des algorithmes plus personnalisés ou optimisés pour des cas spécifiques, en profitant de paradigmes de programmation variés (orientée objet, fonctionnelle, impérative). Enfin, grâce à sa popularité et à sa vaste communauté, Python bénéficie d'un écosystème en constante évolution, offrant un accès rapide à des outils de pointe et à des ressources pédagogiques abondantes. Dans un premier temps, nous avons volontairement utilisé Python sans bibliothèques externes, afin de garder un contrôle total sur chaque étape du calcul.

Comme décrit dans notre algorithme, notre modélisation repose sur une boucle principale qui donne successivement la « parole » aux agents. L'ordre dans lequel les agents interviennent n'a aucun impact sur le comportement global du modèle ni sur les résultats obtenus.

De la même manière que ce qui est proposé dans la thèse de Hermellin 2016 et l'architecture des GPGPU Hermellin et al. 2014 cette propriété ouvre une nouvelle perspective : remplacer la boucle par du calcul matriciel. Bien que plus coûteux en mémoire, ce type de calcul est généralement beaucoup plus rapide, surtout en Python, qui est lent nativement mais dispose de bibliothèques optimisées comme NumPy permettant d'effectuer des opérations matricielles à très grande vitesse. Cela a mené à une question clé : Dans quelles conditions peut-on remplacer un code basé sur des boucles par du calcul matriciel ?

En théorie, n'importe quelle implémentation peut être traduite en calcul matriciel, si l'on ignore la complexité d'implémentation, ainsi que les contraintes de temps de calcul et de mémoire. En d'autres termes, le calcul matriciel est toujours possible en théorie, mais son utilisation pratique dépend de plusieurs compromis techniques. Les travaux sur l'application de SMA sur GPU vont en ce sens Hermellin 2016. Finalement, nous développons une approche similaire dans un cadre simplifié et plus accessible permis par Python.

Dans un cadre plus général, notre expérimentation nous a montré que le calcul matriciel devient particulièrement pertinent lorsque le modèle comporte un grand nombre d'agents dont les interactions ne sont pas interconnectées. De fait, lorsque l'ordre de prise de décision des agents n'influence pas le résultat final. Dans ce cas, il est possible de traiter plusieurs agents en parallèle via des opérations vectorisées. À l'inverse, lorsque les dé-

cisions des agents dépendent fortement les unes des autres (ordre d'exécution critique) ou que les agents se déplacent dans de grands espaces continus, la structure matricielle devient moins efficace, car elle exige plus de mémoire et une logique plus complexe pour gérer les interactions locales. Le choix du calcul matriciel repose donc sur un équilibre : un nombre élevé d'agents, des interactions simples et parallélisables et un environnement de taille raisonnable permettent d'exploiter pleinement ses gains de performance.

Nous présentons en annexe A la transformation d'un code classique en une version matricielle. Nous montrons qu'une telle transformation est possible même pour des modèles SMA bien établis, comme le modèle de ségrégation de Schelling (Schelling 1971). Ce modèle s'inscrivant dans un espace bidimensionnel, il peut sembler a priori peu intuitif d'y recourir au calcul matriciel. Néanmoins, l'exemple présenté illustre que cette approche est non seulement applicable dans notre contexte particulier, mais qu'elle ouvre également la voie à une utilisation plus générale du calcul matriciel dans d'autres modèles multi-agents.

Description de notre algorithme

Algorithm 2 Choix optimal de produit par agent selon sensibilité

1: $sensitivityPrice \leftarrow$ generateFromData
2: $sensitivityQuality \leftarrow$ generateFromData
3: $products \leftarrow$ generateFromData
4: $sensitivity \leftarrow \begin{bmatrix} sensitivityPrice \\ sensitivityQuality \end{bmatrix}^\top$
5: $referencePrice \leftarrow \text{mean}(products[0])$
6: $referenceQuality \leftarrow \text{mean}(products[1])$
7: $productDiffs \leftarrow products$
8: $productDiffs[0] \leftarrow -1 \cdot (products[0] - referencePrice)$
9: $productDiffs[1] \leftarrow products[1] - referenceQuality$
10: $diffsAdjusted \leftarrow$ if $productDiffs < 0$ then $productDiffs \cdot lossAversion$ else $productDiffs$
11: $diffsExpanded \leftarrow \text{transpose}(diffsAdjusted)[:, \text{newAxis}, :]$
12: $sensitivityExpanded \leftarrow sensitivityMatrix[\text{newAxis}, :, :]$
13: $utilities \leftarrow \text{sum}(diffsExpanded \cdot sensitivityExpanded, \text{axis} = -1)$
14: $bestProductIndices \leftarrow \text{argmax}(utilities, \text{axis} = 0)$
15: $selectionResults \leftarrow \text{unique}(bestProductIndices, \text{returnCounts} = \text{True})$
16: $sortedResults \leftarrow \text{zeros}(\text{len}(trueSalesSorted))$

Nous avons décidé d'utiliser pour notre seconde implémentation en python, du calcul matriciel, permis par numpy, de la même manière que précédemment avec le modèle de ségrégation. Au prix de plus de mémoire, cette implémentation va nous permettre d'effectuer plus de simulations plus rapidement.

Les quatre premières étapes de l'algorithme correspondent à l'initialisation des sensibilités des agents et des caractéristiques des produits. Cela aboutit à deux matrices :

— `products` de taille $(N_{\text{produits}}, 2)$ contenant pour chaque produit son prix et une mesure composite de qualité,
— `sensitivity` de taille $(N_{\text{agents}}, 2)$ représentant, pour chaque agent, sa sensibilité au prix et à la qualité.

Ensuite, des valeurs de référence (moyennes) pour le prix et la qualité sont calculées pour la catégorie à partir de l'ensemble des produits :

$$\texttt{referencePrice} = \text{moyenne des prix} \tag{3.10}$$

$$\texttt{referenceQuality} = \text{moyenne des qualités.} \tag{3.11}$$

Les écarts entre les caractéristiques des produits et les valeurs de référence sont alors déterminés. Afin de tenir compte de l'aversion à la perte, les écarts négatifs sont multipliés par le paramètre $\lambda > 1$:

$$\texttt{productDiffs} = \begin{cases} \texttt{diff} \times \lambda & \text{si } \texttt{diff} < 0 \\ \texttt{diff} & \text{sinon} \end{cases} \tag{3.12}$$

Cela donne une matrice `productDiffs` de dimension $(N_{\text{produits}}, 2)$.

La phase clé consiste à évaluer l'utilité perçue de chaque produit pour chaque agent. Pour ce faire, une multiplication tensorielle est réalisée entre les sensibilités des agents et les différences ajustées des produits. Cela nécessite une expansion des dimensions pour permettre une diffusion correcte lors de la multiplication :

$$\texttt{utilities}_{ij} = \sum_{k=1}^{2} \texttt{productDiffs}_{jk} \cdot \texttt{sensitivity}_{ik} \tag{3.13}$$

Ce calcul produit une matrice `utilities` de taille $(N_{\text{agents}}, N_{\text{produits}})$.

Pour chaque agent, le produit maximisant l'utilité est identifié, ce qui permet de construire un vecteur d'indices des produits préférés, et d'agréger les résultats pour obtenir une estimation des parts de marché.

Exemple illustratif avec des matrices réduites

Nous allons développer l'algorithme sur un exemple purement illustratif à trois produits et deux agents.

Étape 0 : Données initiales

$$\texttt{products} = \begin{bmatrix} 10 & 7 \\ 12 & 9 \\ 14 & 8 \\ \vdots & \vdots \end{bmatrix}, \quad \texttt{sensitivity} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \\ \vdots & \vdots \end{bmatrix}$$

Étape 1 : Références

$$\texttt{referencePrice} = \frac{10+12+14}{3} = 12, \quad \texttt{referenceQuality} = \frac{7+9+8}{3} = 8$$

Étape 2 : Écarts par rapport aux références

$$\texttt{productDiffs} = \begin{bmatrix} -(10-12) & (7-8) \\ -(12-12) & (9-8) \\ -(14-12) & (8-8) \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} +2 & -1 \\ 0 & +1 \\ -2 & 0 \\ \vdots & \vdots \end{bmatrix}$$

Étape 3 : Ajustement avec aversion à la perte ($\lambda = 2$)

$$\texttt{diffsAdjusted} = \begin{bmatrix} 2 & (-1) \times 2 \\ 0 & 1 \\ (-2) \times 2 & 0 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 0 & 1 \\ -4 & 0 \\ \vdots & \vdots \end{bmatrix}$$

Étape 4 : Expansion des dimensions

$$\texttt{diffsExpanded} \in \mathbb{R}^{(N_{\text{produits}},1,2)} = \begin{bmatrix} [2 & -2] \\ [0 & 1] \\ [-4 & 0] \\ \vdots & \vdots \end{bmatrix}$$

$$\texttt{sensitivityExpanded} \in \mathbb{R}^{(1,N_{\text{agents}},2)} = \begin{bmatrix} [0.8 & 0.2] \\ [0.3 & 0.7] \\ \vdots & \vdots \end{bmatrix}$$

Étape 5 : Utilités

$$\texttt{utilities}_{ij} = \sum_{k=1}^{2} \texttt{productDiffs}_{jk} \cdot \texttt{sensitivity}_{ik}$$

$$\texttt{utilities} = \begin{bmatrix} 2 \cdot 0.8 + (-2) \cdot 0.2 & 0 \cdot 0.8 + 1 \cdot 0.2 & (-4) \cdot 0.8 + 0 \cdot 0.2 & \cdots \\ 2 \cdot 0.3 + (-2) \cdot 0.7 & 0 \cdot 0.3 + 1 \cdot 0.7 & (-4) \cdot 0.3 + 0 \cdot 0.7 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} 1.2 & 0.2 & -3.2 & \cdots \\ -0.8 & 0.7 & -1.2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Étape 6 : Choix optimal par agent

$$\texttt{bestProductIndices} = \begin{bmatrix} 1 \\ 2 \\ \vdots \end{bmatrix}$$

3.5 Résultats

Dans cette partie, nous allons montrer que notre modèle permet de mettre en évidence les trois effets de contexte majeurs que sont l'effet de leurre, de compromis et de similarité. Ces résultats ont été présentés lors de conférences (Vanderlynden et al. 2024). Pour cela, nous réalisons 3 catégories d'expériences,[2] chacune partant du même état initial avec un produit A(80,80) et un produit B(20,20). Nous introduirons un produit C(x,y) qui jouera tour à tour le rôle de leurre, compromis et similarité et nous montrerons que l'arrivée de ce produit modifie le comportement des agents à la manière de ce qui est observé dans la réalité.

3.5.1 Reproduction des effets de contexte

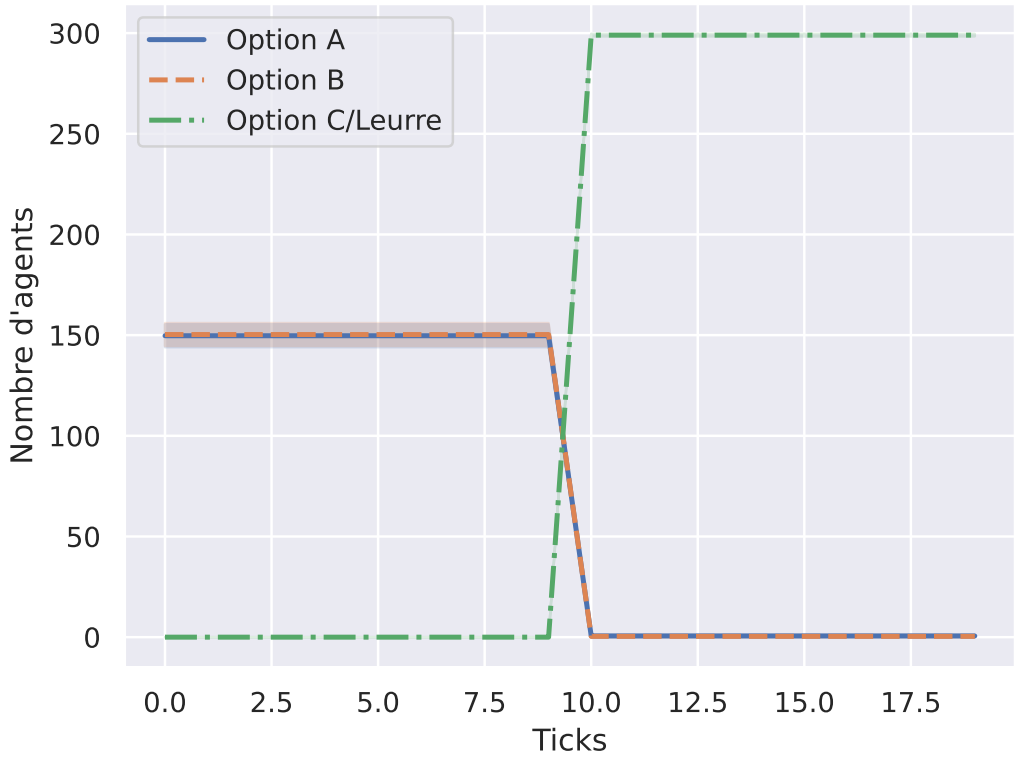


Figure 3.4 – *Introduction d'un produit C(40,60).*

Dans toutes les figures de cette section, le produit A est représenté sur les courbes en bleu, B en jaune. Ces produits ont un rapport qualité/prix de 1. Le produit C, introduit durant la simulation, est représenté en vert. Dans cette partie, la distribution des paramètres des agents ne varie pas. Le revenu des agents est tiré aléatoirement selon une loi normale. Dans nos exemples pour cette partie, $revenu \sim \mathcal{N}(3000, 500)$ avec $1000 < revenu < 6000$.

$$t(revenu) = revenu/6000 \tag{3.14}$$

Un individu avec un revenu de 6000€ par mois aura $S_P = 0$ et $S_Q = 1$. Un individu avec

2. Une feuille Jupyter permet de reproduire ces expériences : https://github.com/cristal-smac/retail

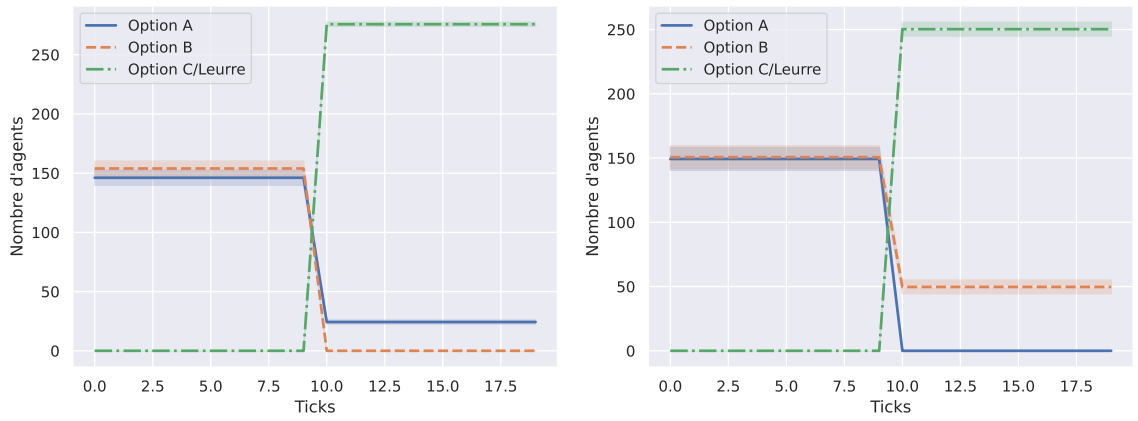


FIGURE 3.5 – *À gauche, l'introduction d'un produit C(20,35). À droite, C(80,95). Ces deux produits sur-performent en termes de rapport qualité/prix les produits initiaux.*

un revenu de 3000€ par mois est indécis, il aura $S_P = S_Q = 0.5$. Différentes fonctions t peuvent faire varier l'intensité des effets en changeant les caractéristiques internes des agents mais ne modifient pas l'apparition des effets de contexte. Les simulations réalisées sont entièrement déterministes. Il n'y a pas de changement de consommation chez les agents durant la simulation, hormis lors de l'introduction d'un troisième produit.

Sur la FIGURE 3.4 nous pouvons observer les fortes ventes du produit C (en vert). Ce produit n'est pas un leurre, mais une troisième option optimale, car son rapport qualité/prix est meilleur que les deux autres. On observe bien que ce produit cannibalise les ventes des deux autres de manière symétrique. Sur la FIGURE 3.5 cette fois-ci, le nouveau produit est toujours optimal, mais proche du produit A ou B. On observe que dans cette situation, l'introduction de ce nouveau produit optimal cannibalise entièrement les ventes du produit A ou B duquel il est proche.

L'effet de leurre

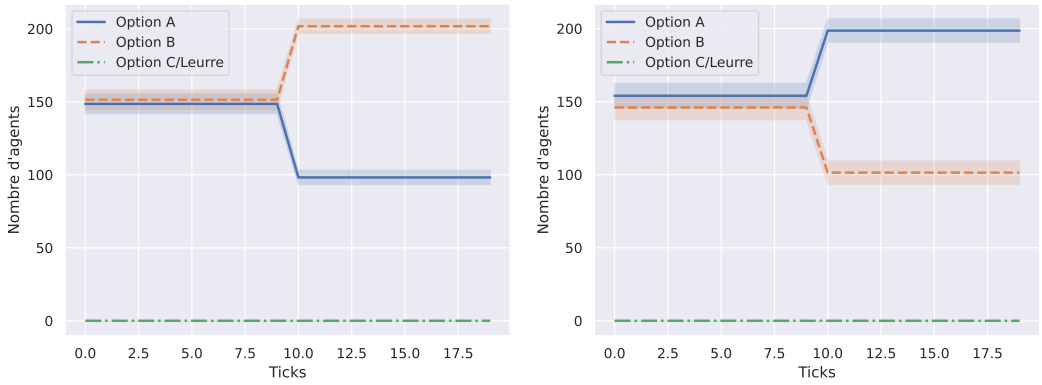


FIGURE 3.6 – *Reproduction de l'effet de leurre. Sur la figure de gauche un leurre C(30,10), sur la figure de droite un leurre C(90,70). Ces deux produits sont asymétriquement dominés par les solutions initiales. Le premier est dominé par le produit peu cher, mais pas par le produit cher et inversement pour le second leurre.*

On voit dans la FIGURE 3.8 que le leurre bien que sous optimal, est quand même acheté. Cela provient de deux effets : d'une part, le produit n'est pas complètement inintéressant,

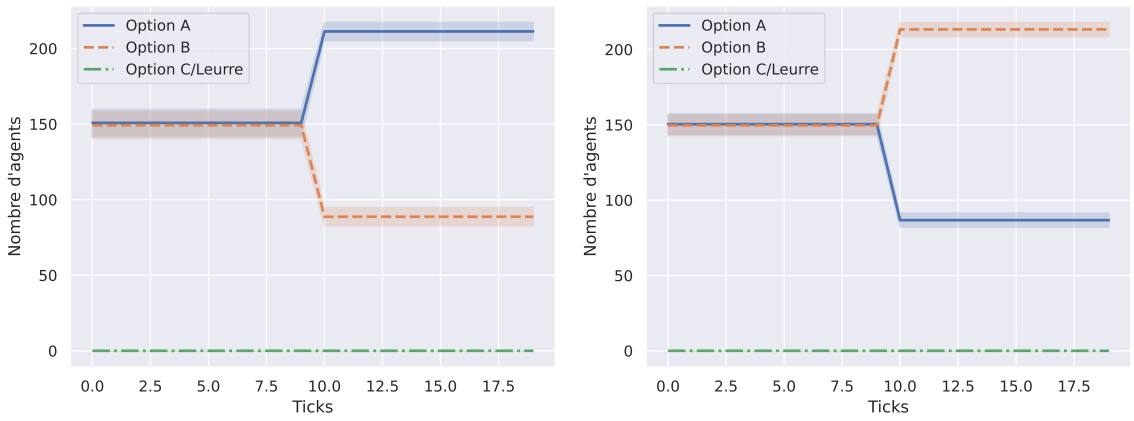


FIGURE 3.7 – *Reproduction de l'effet de leurre. Sur la figure de gauche un leurre C(100,80) sur celle de droite C(15,0). Les leurres ne sont dominés par aucune autre alternative, mais restent moins bons en rapport qualité/prix par rapport aux produits initialement présents.*

il n'est dominé par aucune autre solution, d'autre part, il est assez distant des deux autres produits. De plus, l'effet de compromis exerce une influence.

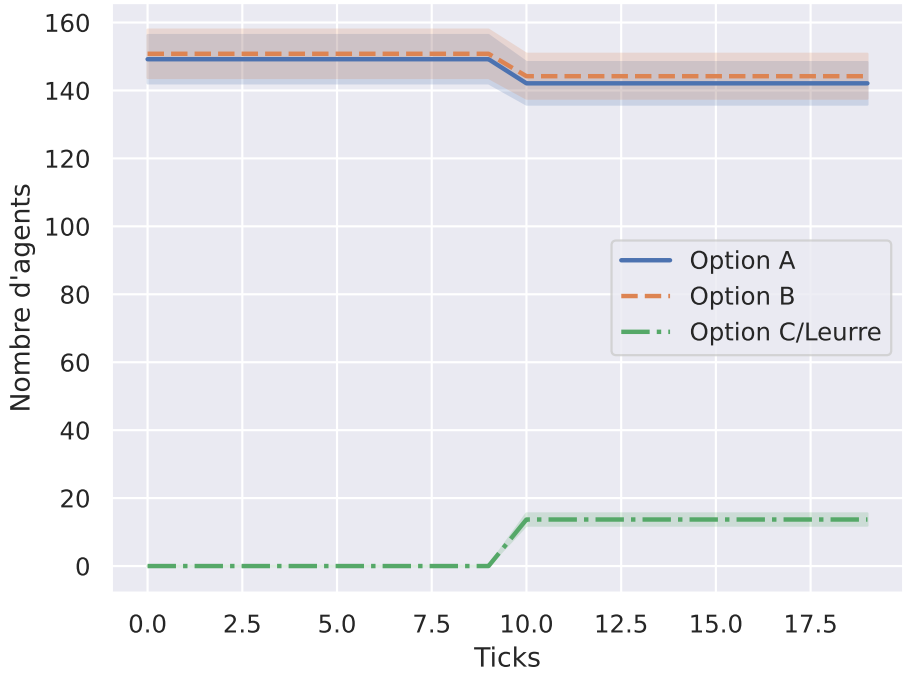


FIGURE 3.8 – *Introduction d'un produit C(55,45)*

L'effet de compromis

Sur la FIGURE 3.10 nous introduisons un produit C dont les caractéristiques sont extrêmes. De ce fait, les produits B puis A bénéficient d'un effet de compromis et voient leurs ventes augmenter. De plus, le produit C, est une alternative compétitive et bénéficie donc de ventes similaires à celles de A ou B. Sur la FIGURE 3.11 nous avons changé les paramètres de A et de B afin de faire en sorte que leurs caractéristiques, se trouvent tour à tour réellement à la moyenne des caractéristiques des produits. On peut observer que plus les caractéristiques du produit sont proches de la moyenne, meilleur est l'effet de compromis.

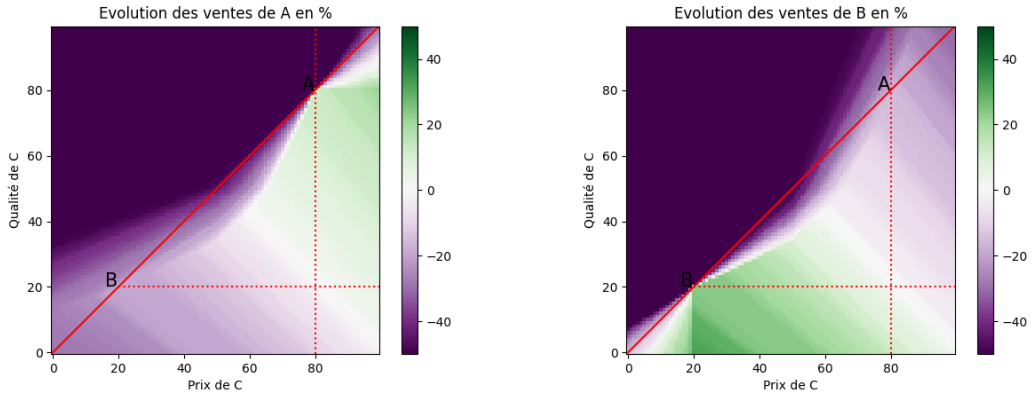


FIGURE 3.9 – *Ces deux matrices représentent l'augmentation ou la diminution des ventes en pourcentage selon les caractéristiques du leurre venant d'être introduit. En haut A, en bas B.*

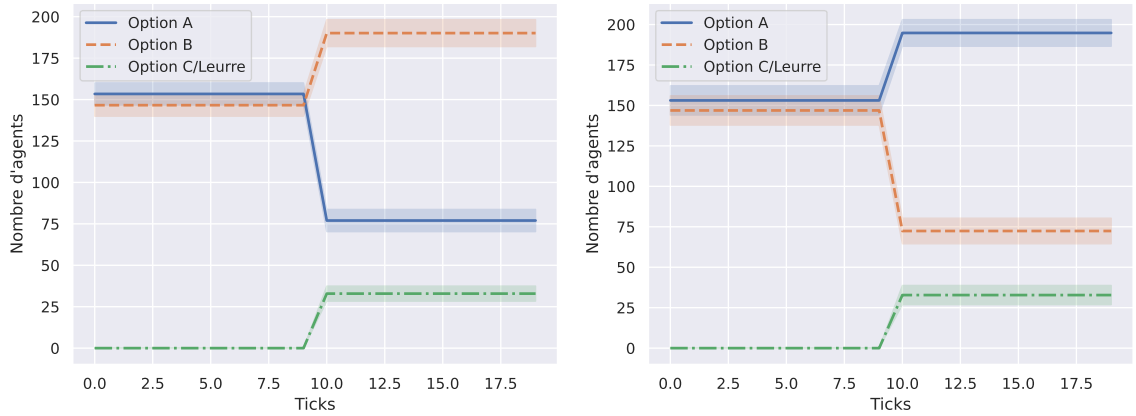


FIGURE 3.10 – *À gauche, l'introduction d'un produit C(1,1). À droite, C(100,100).*

De plus, les produits aux extrêmes étant équidistants de la moyenne, leur nombre de ventes sont quasiment les mêmes. Étant donné que la distribution des caractéristiques des agents suit une loi normale, les agents sont en nombre à peu près égal de part et d'autre des caractéristiques, ce qui explique aussi pourquoi les ventes des produits aux extrêmes sont similaires. Avec ces différentes simulations, nous montrons que notre modèle est capable de reproduire l'effet de compromis assez fidèlement.

L'effet de similarité

La FIGURE 3.12 montre l'effet de similarité. On introduit un produit C similaire au produit A. On observe que les ventes du produit B diminuent plus faiblement en proportion que celles du produit A.

3.5.2 L'impact des paramètres

Dans les simulations réalisées, les agents et leurs caractéristiques jouent un rôle important. L'effet de leurre ainsi que les différents effets de contexte affectent principalement les agents équilibrés, donc indécis. Dans cette section, deux paramètres sont étudiés, les revenus des agents et le paramètre d'aversion à la perte. Nos résultats sont synthétisés dans les tables 3.2 et 3.1. Les revenus des agents sont tirés aléatoirement en début de simulation.

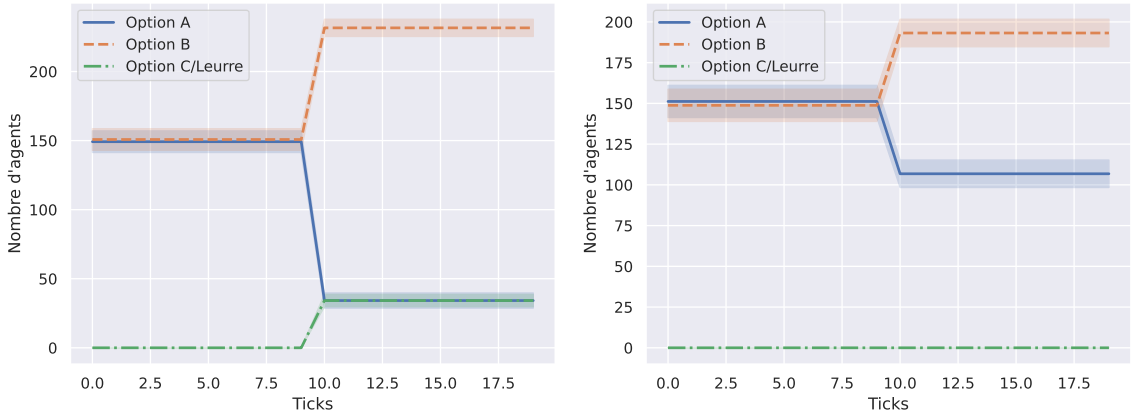


FIGURE 3.11 – *À gauche, l'introduction d'un produit $C(20,20)$, de plus les caractéristiques du produit B ont été changés pour cette simulation à $B(50,50)$. À droite, l'introduction d'un produit $C(100,100)$ de plus les caractéristiques du produit A ont été changés, pour cette simulation à $A(60,60)$.*

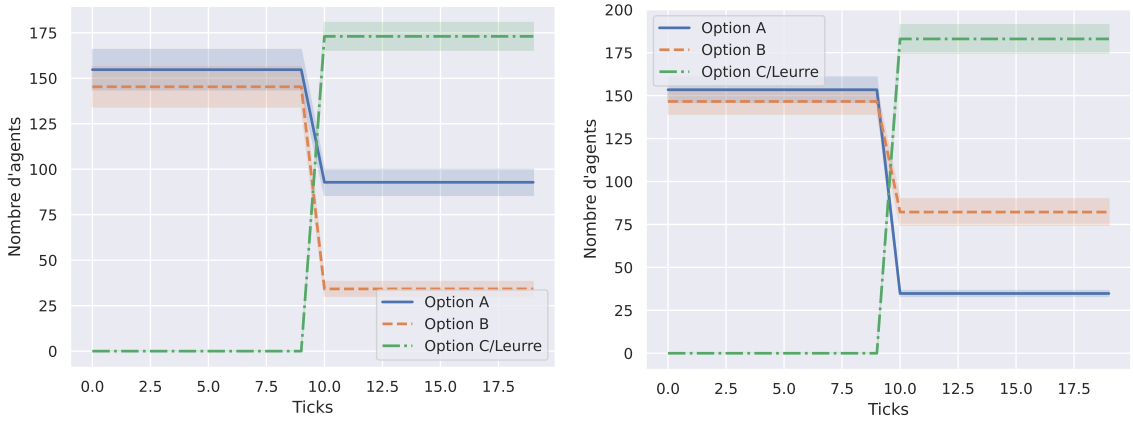


FIGURE 3.12 – *À gauche, l'introduction d'un produit C(30,30). À droite, C(70,70).*

Nous testons différentes distributions afin d'en déterminer l'impact. L'aversion à la perte est un paramètre global du modèle. Il est important, car sans lui, nous n'observons pas d'effet de contexte. De plus, une aversion à la perte plus ou moins forte peut s'observer en réalité selon les produits analysés.

Nous observons grâce à la table 3.3 que plus les agents se concentrent sur le rapport qualité/prix (σ faible), plus ils sont affectés par les effets de contexte. La table 3.2 montre que plus l'aversion à la perte est forte, (plus β est petit) plus nos agents sont sensibles aux effets de contextes.

L'introduction de plusieurs produits rend les effets de contexte bien plus complexes à observer. En effet, comme ces effets s'additionnent, il devient difficile de comprendre quels effets sont en place. De plus, la disposition des produits impacte les effets de contexte. Placer les produits uniformément sur l'espace qualité/prix ne donne pas le même environnement que si on place trois produits bon marché face à trois produits de qualité, cette différence va produire des résultats différents. Nous pouvons observer cet effet sur la FIGURE 3.13.

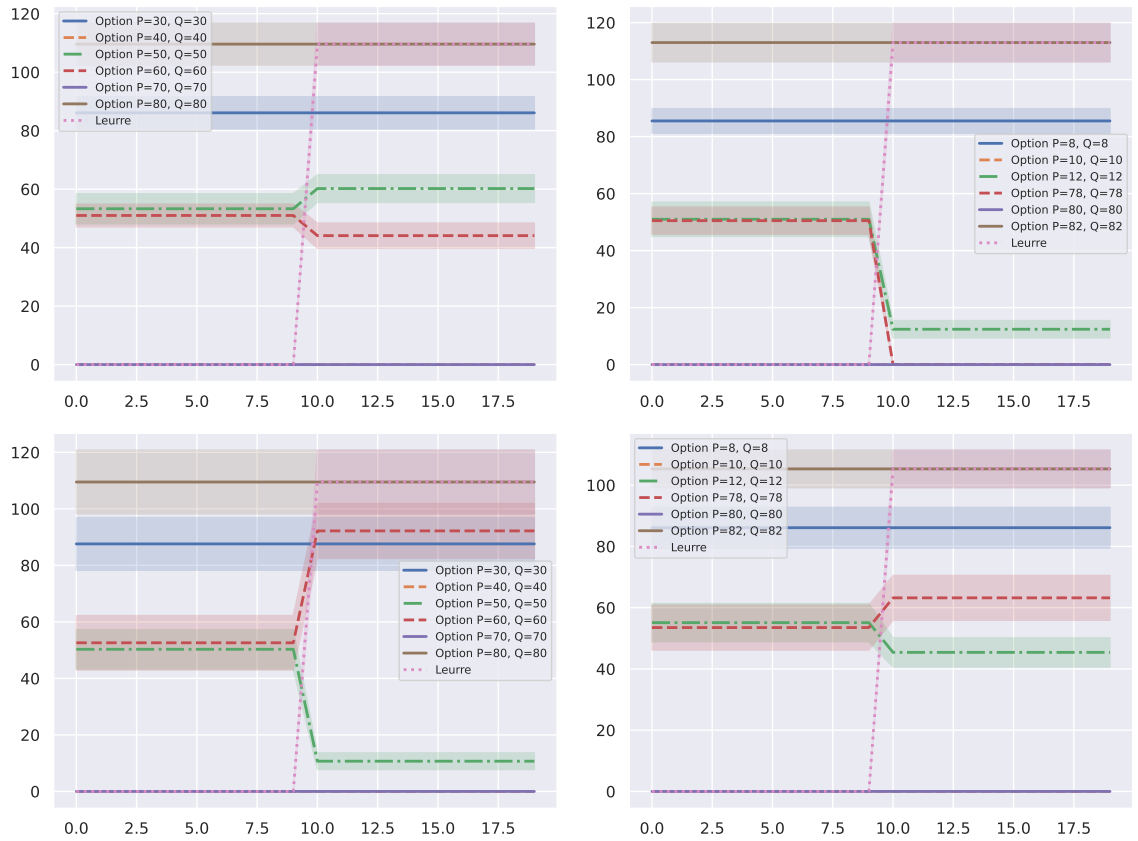


FIGURE 3.13 – *L'effet de compromis illustré sur les deux figures du haut avec un produit C(50,50). L'effet de leurre sur les deux figures du bas avec C(90,70). Sur les deux figures de gauche, les produits sont répartis uniformément entre 0 et 100 en termes de qualité et de prix, sur celles de droite, les produits sont séparés en deux groupes, les produit bon marché et les produits de qualité.*

3.5.3 Reproduction de notre exemple réel

La FIGURE 3.14 montre une simulation dans laquelle les six offres de la FIGURE 3.2 ont été reproduites. Nous pouvons observer que dans tous les cas, l'offre en bleu (l'offre mise en avant par les effets de contexte) obtient une part significative des ventes. Sur la simulation de gauche, la présence ou non de la 5G n'est pas modélisée. Une offre de 100 Go correspond alors à un produit de qualité 10, une offre à 20 Go une qualité de 2 etc. Dans cette simulation, l'offre de 100 Go à 9,99 € dont nous avions parlé précédemment ne reçoit aucune vente. Dans la simulation de droite, nous mettons en place un processus de décision dans lequel les agents ont une sensibilité à la 5G. Ainsi, un agent peut aimer la 5G et améliorer les scores de qualité des différents produits disposant de la 5G, ou au contraire, ne pas aimer la 5G et retrancher de la qualité à ces offres. Cette sensibilité est représentée comme un réel défini entre 0 et 2. Une valeur de 1 représente un agent qui n'a pas d'avis sur la question et ne prend pas en compte la 5G pour réaliser son choix. Dans cette simulation, nous observons l'émergence de ventes de l'offre 4G 100 Go à 9,99 €. Cela correspond aux exemples donnés dans la section 3.3 et illustre l'impact de la perception de la qualité par les consommateurs. Avec cet exemple, nous montrons qu'avec un travail de modélisation de perception (ici la 5G), notre modèle est capable de s'adapter à un scénario réel.

Distribution du revenu	Produit C		$A(80,80)$		$B(20,20)$	
	$C1$	$C2$	$C1$	$C2$	$C1$	$C2$
$\mathcal{N}(3000,500)$	0%	99%	75%	0.5%	25%	0.5%
$\mathcal{N}(3000,1500)$	0%	50%	60%	25%	40%	25%
$\mathcal{N}(3000,3000)$	0%	55%	28%	37%	45%	35%
$\mathcal{U}(0,5000)$	0%	18%	55%	43%	45%	39%

TABLE 3.1 – *Ventes en pourcentage des différents produits après introduction du produit C avec C1 = C(90,70) et C2 = C(50,50).*

β	Produit C		A(80,80)		B(20,20)	
	$C1$	$C2$	$C1$	$C2$	$C1$	$C2$
-1	0%	0%	50%	50%	50%	50%
-1.5	0%	47%	59%	29%	41%	24%
-2	0%	74%	68%	13%	32%	13%
-3	0%	90%	75%	5%	25%	5%

TABLE 3.2 – *Ventes en pourcentage des différents produits selon β après introduction du produit C avec $C1 = C(90,70)$ et $C2 = C(50,50)$.*

Finalement sur l'ensemble de ces résultats on peut noter que de petites variations dans la configuration des caractéristiques des produits, changent considérablement les effets observés ainsi que leurs impacts. C'est une caractéristique d'un système complexe.

3.6 Calibration

La calibration et les résultats associés de cette section ont été présentés par (Jarod VANDERLYNDEN et al. 2025a ; Jarod VANDERLYNDEN et al. 2025b).

L'objectif de cette calibration est de reproduire une distribution réelle des ventes. Par exemple, considérons un ensemble de quatre produits avec les pourcentages de ventes suivants : [15%, 25%, 40%, 20%]. Nous ajustons les paramètres de notre modèle afin que les ventes en pourcentage qu'il prédit correspondent à cette distribution.

Nous disposons d'un jeu de données provenant de tickets de caisse, à partir duquel nous avons extrait les ventes de quatre produits concurrents, de la même catégorie, sur une période de deux ans. Ces ventes demeurent globalement stables dans le temps, bien que certaines promotions ponctuelles puissent induire des variations temporaires limitées. Cette stabilité permet d'éviter que les effets exogènes altèrent les résultats.

Dans certains contextes, il est en effet essentiel de disposer de données représentatives du cadre d'analyse étudié. Par exemple, si des facteurs externes influencent de manière significative les prix ou la consommation, tels qu'un nouveau financement public, une campagne marketing d'envergure, une innovation technologique ou une évolution des comportements de consommation, leur intégration au modèle devient impérative. La prise en compte de ces facteurs externes est essentielle pour garantir l'existence d'un ensemble de

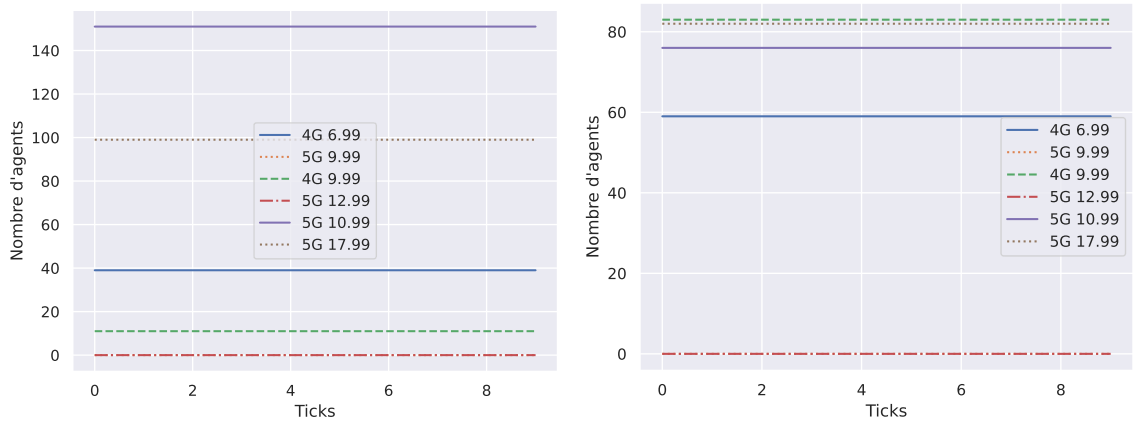


FIGURE 3.14 – *Deux simulations réalisées avec l'exemple en section 3.3. Sur la figure de gauche la présence de la 5G est ignorée, sur celle de droite la 5G est prise en compte selon un paramètre interne à l'agent.*

paramètres permettant une calibration robuste. Toutefois, cette intégration peut s'avérer complexe et requérir des données précises et détaillées.

Le prix de chaque produit est directement issu de nos données. La qualité des produits est définie comme une combinaison de deux paramètres :

— C_1 : un entier représentant le nombre de pièces (plus il est élevé, meilleure est la qualité).

— C_2 : un entier représentant une classification de la qualité de l'emballage (plus il est élevé, plus la classe est qualitative).

Nous disposons de quatre classes d'emballage, une par produit. Si un nouvel article était ajouté, il appartiendrait nécessairement à l'une de ces classes. L'équation 3.15 se ramène donc à deux terme.

$$Q = x_1 \cdot C_1 + x_2 \cdot C_2 \tag{3.15}$$

Les paramètres x_1 et x_2 sont à calibrer.

La sensibilité au prix des agents est tirée selon la distribution des revenus français de L'INSEE (2022)[3] en équivalent temps plein dans le secteur privé. Nous ajoutons un paramètre $0 < x_3 < 9000$ qui correspond au seuil à partir duquel un agent est doté d'une sensibilité minimale au prix. Cela correspond au seuil à partir duquel un agent ne regarde plus le prix pour choisir son produit. Ce seuil est dépendant du produit analysé. La borne supérieure de x_3 est simplement liée aux données utilisées qui ne référencent pas de revenus supérieurs à 9000€ par mois.

La sensibilité à la qualité S_Q de nos agents est représentée par un réel compris entre 0 et 1. 0 étant une sensibilité nulle à la qualité et 1 une sensibilité maximale. Pour obtenir cette sensibilité, nous utilisons un tirage suivant une loi normale de moyenne μ et d'écart-type σ.

Les agents sont répartis en trois catégories ayant chacune une valeur de μ et σ différente. La répartition des agents dans ces catégories est faite en s'appuyant sur un questionnaire

3. https ://www.insee.fr/fr/statistiques/7707884#onglet-1, données, figure 2

réalisé auprès des clients de cette entreprise. Les clients avaient droit à plusieurs réponses face à la question « Quels sont vos critères de choix ». Trois réponses étaient possibles « Le prix le plus bas » 47% des réponses, « L'esthétique » 7.5% des réponses et « Le respect des normes de qualité » 45.5% des réponses.

Les valeurs utilisées pour μ selon les catégories sont les suivantes :

— « Le prix le plus bas » $\mu = 0.1$
— « L'esthétique » $\mu = x_4$
— « Le respect des normes de qualité » $\mu = 0.9$

La valeur de σ est la même pour toutes les catégories et est laissée libre pour le processus de calibration : $\sigma = x_5$. Les valeurs inférieures à zéro et supérieures à un sont ramenées à ces bornes. La valeur de sensibilité à la qualité d'un agent suit donc une loi normale de moyenne et d'écart-type dépendant de sa catégorie. Les agents sont répartis entre les catégories selon une distribution issues du questionnaire précédent. Pour la calibration de nos agents et de nos produits, nous avons au total 5 paramètres. x_1 et x_2 qui interviennent dans la représentation de la qualité des produits, x_3 qui intervient dans le tirage de la sensibilité au prix et x_4 et x_5 qui interviennent dans le tirage de la sensibilité à la qualité. Il ne reste plus que le paramètre d'aversion à la perte β à calibrer, pour un total de six paramètres. La table 3.3 référence l'ensemble de ces paramètres

x_1	Poids du paramètre 1 sur la qualité des produits	$0 < x_1 < 1$
x_2	Poids du paramètre 2 sur la qualité des produits	$0 < x_2 < 1$
x_3	Revenu à partir duquel la sensibilité au prix est minimale	$0 < x_3 < 9000$
x_4	Moyenne de la catégorie « esthétique » pour le tirage de la sensibilité à la qualité	$0 < x_4 < 1$
x_5	Écart-type utilisé pour le tirage de la sensibilité à la qualité	$0 < x_5 < 1$
β	Paramètre d'aversion à la perte	$1<\beta$

TABLE 3.3 – *Résumé des paramètres à calibrer*

Pour la calibration, nous utilisons un Grid Search et discrétisons chacun des paramètres en 10 valeurs, ce qui représente un million de simulations (10^6). La fonction de perte que l'on cherche à minimiser est la différence des proportions de ventes en pourcentage de chacun des produits entre les ventes réelles et les ventes simulées :

$loss = \sum |VentesRelles - VentesSim|$.

x1	x2	x3	x4	x5	β
0.14	0.10	2624	0.05	0.87	2.72

TABLE 3.4 – *Liste des paramètres calibrés*

	P_1	P_2	P_3	P_4
Prix	0.03	0.23	1	0
Qualité	0.01	0.20	0.25	0.04

TABLE 3.5 – *Les caractéristiques des produits après normalisation et application des paramètres des différents produits.*

La table 3.4 montre les paramètres calibrés.

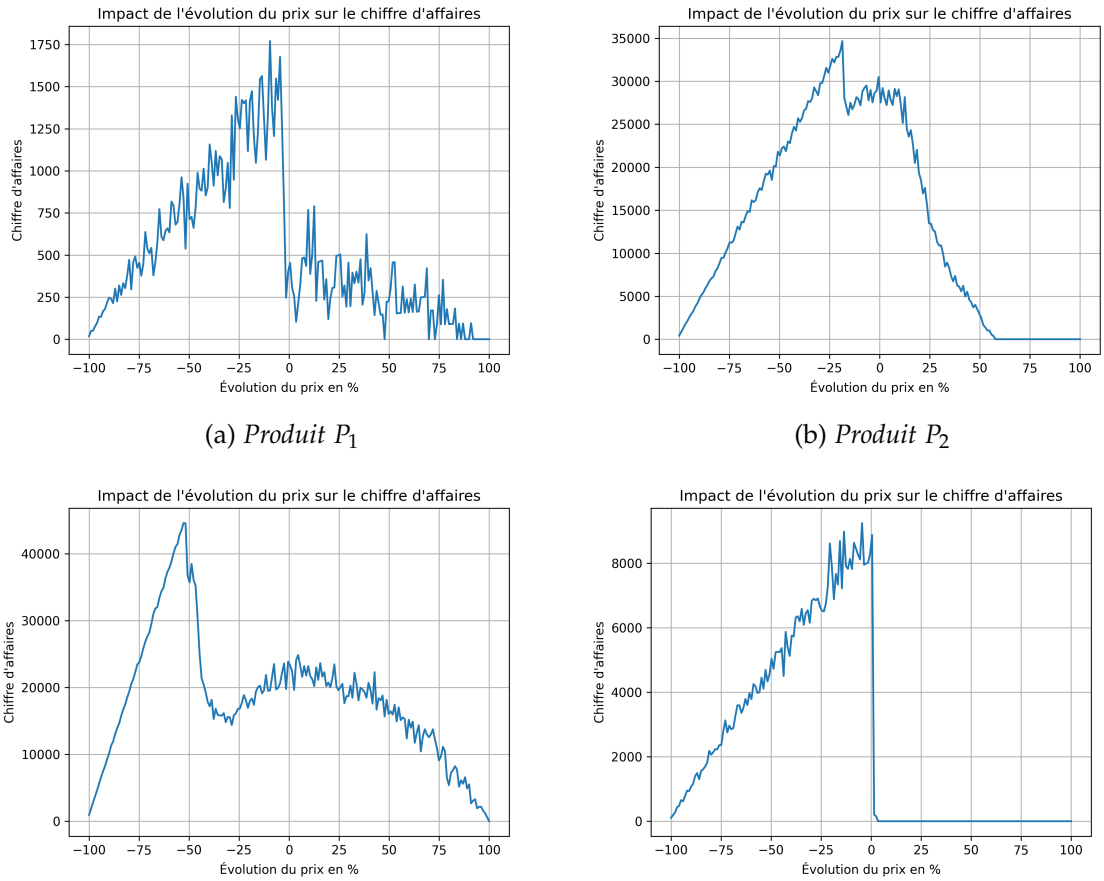
3.6.1 Résultats après calibration

Notre modèle parvient à reproduire fidèlement la distribution réelle des ventes. Il peut être considéré comme un outil fiable d'aide à la décision. En effet, cette capacité de calibration valide la pertinence des hypothèses et des mécanismes intégrés au modèle, ce qui ouvre la possibilité de l'utiliser pour simuler différents scénarios et évaluer l'impact de modifications sur les ventes. Ainsi, en ajustant certains paramètres comme le prix, la qualité perçue ou encore les préférences des consommateurs, il devient envisageable d'anticiper l'effet de ces changements sur la répartition des ventes, offrant ainsi un levier stratégique pour optimiser l'offre et maximiser la performance commerciale.

Le modèle peut apporter des éléments de réponse à une grande variété de questions stratégiques, qui varient selon le point de vue adopté. Par exemple, un magasin qui vend différents produits n'aura pas les mêmes objectifs qu'une marque cherchant à imposer son produit phare sur un marché concurrentiel. Dans le premier cas, le magasin peut choisir d'adapter les prix des différents produits en fonction de ses marges, et il est même possible que l'article le moins cher soit celui qui lui rapporte le plus. À l'inverse, dans le second cas, la marque cherchera avant tout à positionner son produit de manière optimale pour maximiser ses bénéfices, en ajustant son prix et, si possible, sa qualité. Elle devra aussi s'adapter aux variations de prix des concurrents et réévaluer son positionnement après l'introduction du produit sur le marché. Dans un premier temps, nous allons nous concentrer sur le scénario d'un seul magasin, ce qui correspond à notre situation réelle, afin d'analyser les enseignements fournis par le modèle. Ensuite, nous explorerons des scénarios fictifs pour voir comment le modèle peut répondre à des problématiques similaires dans d'autres contextes.

Optimisation des prix d'un magasin

On se place dans le scénario de notre calibration, il y a quatre produits en concurrence, leurs caractéristiques sont tirées de nos données. La table 3.5 référence le prix et la qualité des produits après l'application des paramètres. On observe que les produits P_1 et P_4 sont les alternatives peu chères. Le produit P_3 est le produit le plus cher et de meilleure qualité.



(a) *Produit* P_1 (b) *Produit* P_2

(c) *Produit* P_3 (d) *Produit* P_4

FIGURE 3.15 – *Le chiffre d'affaires de chaque produit,* P_1, P_2, P_3, P_4*, selon l'évolution de son prix. Chaque point représente une simulation. Sur chaque figure, le prix des autres produits ne bouge pas.*

Le produit P_2 est un compromis entre les différentes alternatives. On teste différents prix pour les produits avec un grid-search afin de déterminer quel ensemble de prix amène les bénéfices maximums pour la catégorie. On ne modélise pas la concurrence d'autres magasins pour rester dans notre cadre réel, car nous n'avons pas accès aux données des concurrents ou aux places de marché.

Le modèle ne permet pas de prédire le nombre de clients potentiels. Or dans ce scénario, on comprend bien que jouer sur les prix affecte la compétitivité du magasin que l'on étudie et par conséquent le nombre de clients. Afin d'étendre le modèle, nous proposons d'ajouter un paramètre de régulation du nombre d'agents. Ce paramètre est nécessaire dans ce scénario afin d'éviter que l'augmentation des prix soit toujours la meilleure solution (si le nombre d'agents est fixe). Le calcul de ce paramètre nécessiterait des informations que nous ne possédons pas tel que l'impact du prix sur la consommation du produit étudié, nous le fixons donc arbitrairement.

Le nombre d'agents est ajusté en fonction du pourcentage d'augmentation du prix moyen. Une hausse du prix moyen entraîne une diminution du nombre d'agents, traduisant une réduction de la demande ou de l'accessibilité du produit. À l'inverse, une baisse du prix moyen induit une croissance du nombre d'agents, traduisant une amélioration de l'acces-

P_1	P_2	P_3	P_4
-1.15%	+20%	-1.5%	+0.05%

TABLE 3.6 – *Proposition de modification de prix réalisé sur des données réelles.*

sibilité et de la désirabilité du produit. Ce mécanisme permet de modéliser l'adaptabilité du système en réponse aux fluctuations de prix.

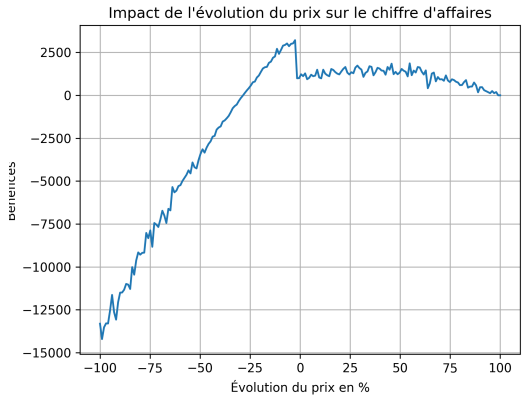
— N_a : Le nombre d'agents.
— A_{prix} : L'ancien prix moyen.
— N_{prix} : Le nouveau prix moyen.
— M Le paramètre de modulation de la population. Fixé à 1 dans nos simulations, celui-ci est dépendant du contexte.

$$N_a = N_a \times [1 + M(\frac{A_{prix}}{N_{prix}} - 1)] \tag{3.16}$$

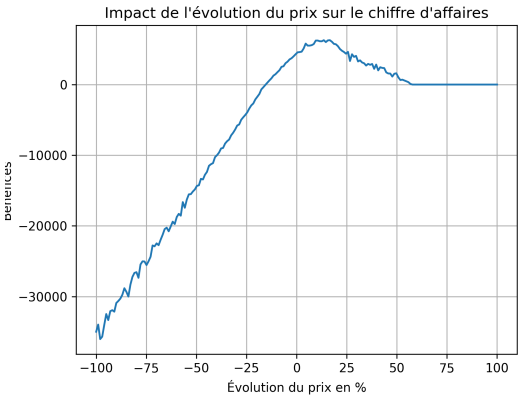
Les courbes de la figure 3.15 montrent l'évolution du chiffre d'affaires de chaque produit lorsque l'on modifie son prix individuellement. On garde le même contexte et on simule à nouveau toutes les ventes avec le nouvel ensemble de prix/produits où seul un prix a changé. On observe des résultats allant dans le sens des effets de contexte, notamment de l'effet de leurre : Lorsqu'un produit diminue suffisamment son prix, pour passer sous le prix d'un produit de moindre qualité, ce dernier agit comme un leurre et l'attractivité du premier produit augmente considérablement. L'aléatoire de ces résultats provient de la population d'agents dont les caractéristiques sont tirées aléatoirement.

Les courbes de la figure 3.16 sont similaires à la figure précédente, mais montrent l'évolution des bénéfices du produit plutôt que le chiffre d'affaires. On observe que la marge de manœuvre des différents produits est en réalité plus faible. Leurs bénéfices baissent rapidement en cas de changement trop brutal du prix. L'effet de leurre est toujours observable pour le produit P_1

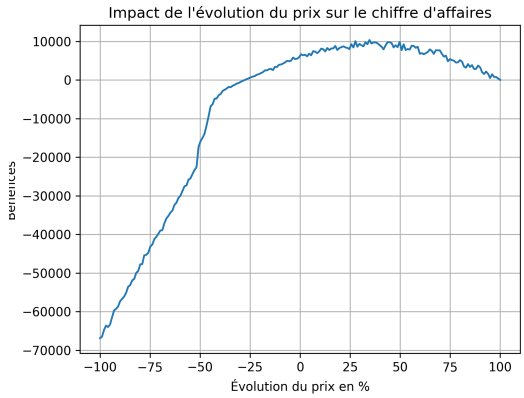
L'analyse des résultats disponible a la table 3.6 suggère une augmentation du prix du produit le plus populaire, caractérisé par le meilleur rapport qualité-prix. Ce résultat traduit l'ajustement du produit $P2$ afin qu'il joue un rôle de compromis vis-à-vis des autres produits. Du fait que sa popularité demeure élevée, l'augmentation modérée de son prix entraîne une amélioration des bénéfices du magasin dans notre simulation. Ces observations s'inscrivent dans la continuité des effets de contexte précédemment mis en évidence par le modèle. Toutefois, le produit le plus populaire peut également servir de produit d'appel ou constituer un moteur de vente à lui seul. Dans ce cas, une augmentation de son prix pourrait induire des effets plus marqués que ceux simulés, en raison des limites du paramètre de modulation du nombre d'agents que nous venons d'introduire. Une campagne de test empirique serait nécessaire afin d'évaluer la robustesse et la pertinence de ce paramètre.



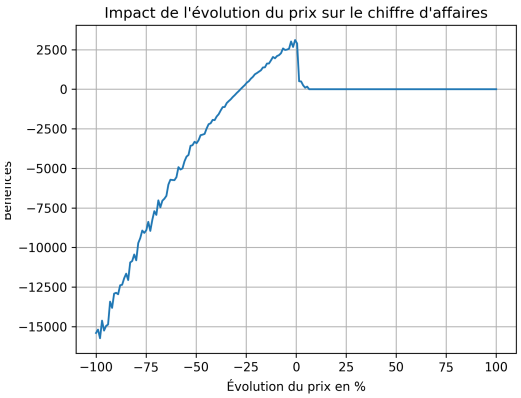
(a) *Produit* P_1



(b) *Produit* P_2



(c) *Produit* P_3



(d) *Produit* P_4

FIGURE 3.16 – *Le bénéfice de chaque produit,* P_1, P_2, P_3, P_4*, selon l'évolution de son prix. Chaque point représente une simulation. Sur chaque figure, le prix des autres produits ne bouge pas. On observe que les modifications de prix ont des conséquences rapides sur les bénéfices.*

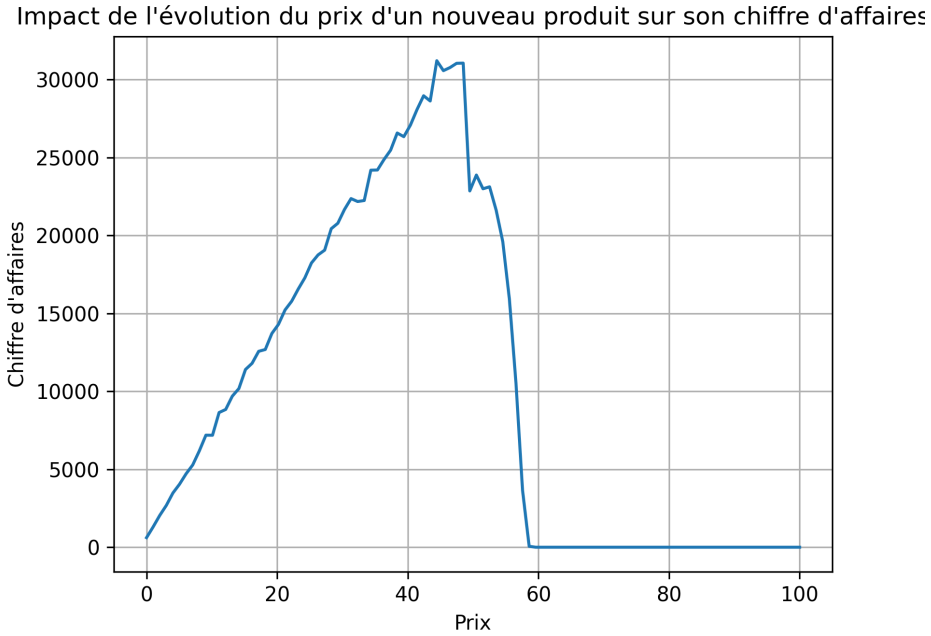


FIGURE 3.17 – *Impact du prix du nouveau produit sur son chiffre d'affaires. On observe un prix optimal pour le chiffre d'affaires entre 45 et 50.*

Introduction d'un nouveau produit

Étant donné que les paramètres qui modifient la qualité sont connus (C_1, C_2), on peut estimer la qualité d'un nouveau produit que l'on souhaite introduire sur le marché. En considérant que le marché ne change pas, on peut évaluer quel est le prix optimal pour notre nouveau produit.

Sur la figure 3.17 on observe un seuil ou le prix est optimal. Enfin, il y a une discussion de positionnement à avoir, car si on se place à l'optimal, le produit sera sensible aux promotions et réduction de prix de ses concurrents. Dans cette discussion, la prise en compte du coût de production est indispensable.

Analyse de différentes stratégies de prix

Nous avons montré que le modèle est capable de reproduire fidèlement les ventes d'une catégorie de produits réels et de fournir des recommandations pertinentes en matière d'évolution des prix. Ainsi, la validité du modèle est établie, confirmant sa capacité à simuler avec précision une situation réelle.

Nous nous intéressons à présent à une généralisation du modèle. Jusqu'ici, notre objectif était de maximiser le bénéfice global de la catégorie, ce qui correspond à un scénario de gestion de magasin. Désormais, nous considérons un cadre concurrentiel dans lequel chaque produit agit individuellement avec pour objectif la maximisation de son propre bénéfice. Nous modélisons l'ensemble de la catégorie, intégrant tous les concurrents. Dans ce contexte, l'impact du prix moyen sur le nombre de clients devient moins marqué. En effet, une augmentation généralisée des prix nécessiterait une coordination entre les différents acteurs du marché. De plus, étant donné que l'ensemble des concurrents est modélisé,

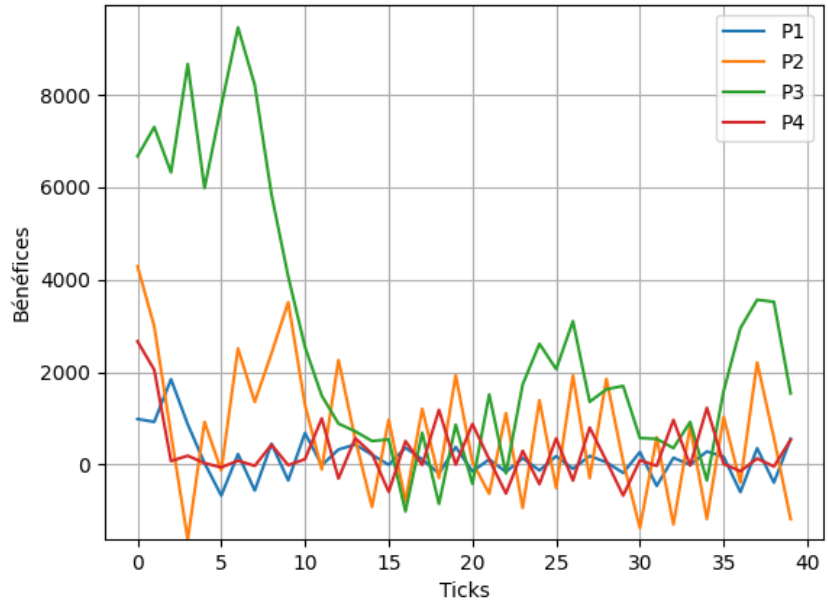


FIGURE 3.18 – *Le bénéfice de chaque produit, P_1, P_2, P_3, P_4. Chaque pas de temps représente une simulation d'où sont obtenus les bénéfices des produits. Chacun des agents contrôlant les caractéristiques des produits utilise une stratégie de maximisation de son bénéfice à court terme. On observe que les bénéfices convergent vers zéro.*

aucune alternative moins coûteuse n'existe. La seule possibilité pour les consommateurs serait alors de réduire ou d'abandonner leur consommation de cette catégorie de produits si les prix deviennent excessifs. Ce phénomène, bien que plausible, varie considérablement selon le type de produit concerné.

Dans ce nouveau cadre, chaque produit représente une « marque », c'est-à-dire un agent autonome définissant sa propre stratégie de prix et réagissant aux décisions des autres acteurs. Dans cette section, nous utiliserons le terme agent pour désigner ces entités qui contrôlent le prix des produits. Nous conservons les quatre mêmes produits que dans le scénario précédent. Chaque expérience comprend 40 cycles (tours), au cours desquels les agents ajustent leurs prix en fonction de leur stratégie. Nous observons ensuite les ventes résultant de ces décisions. Il est important de noter que chaque simulation ne représente pas une évolution temporelle réelle, mais plutôt une série d'ajustements hypothétiques des prix et leurs conséquences.

Ce modèle s'apparente à un cas classique d'oligopole (STIGLER 1964). Dans ce type de marché, la collusion constitue théoriquement la stratégie optimale pour maximiser les bénéfices des entreprises concurrentes. Toutefois, cette stratégie devient de plus en plus fragile à mesure que le nombre d'acteurs augmente. Par ailleurs, lorsque la barrière à l'entrée, définie par les investissements nécessaires au lancement d'un nouveau produit, est faible, les stratégies de collusion tendent à être moins viables (GREEN et al. 2014). En réalité, il n'existe donc pas toujours d'équilibre stable dans ce type de configuration.

Dans la figure 3.18, nous analysons un scénario dans lequel chaque agent suit une straté-

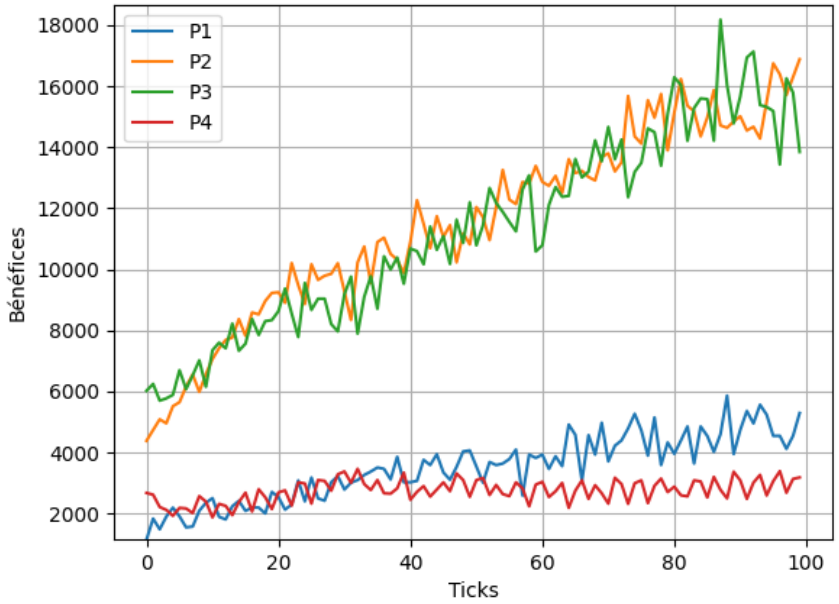


FIGURE 3.19 – *Le bénéfice de chaque produit, P_1, P_2, P_3, P_4, P_a. Chaque pas de temps représente une simulation d'où sont obtenus les bénéfices des produits. Chacun des agents contrôlant les caractéristiques des produits utilise une stratégie de tâtonnement. : « Modification du prix, observation des conséquences, adaptation ». On observe que cette stratégie avantage les produits seul sur leur segment de vente (P_2 et P_3).*

gie de maximisation immédiate de son bénéfice : à chaque tour, il identifie le prix optimal grâce au modèle et l'applique au tour suivant. Tous les agents agissent simultanément. Cette approche conduit à une guerre des prix, où les bénéfices de l'ensemble des agents finissent par atteindre zéro. Cette dynamique illustre une concurrence parfaite, où la stratégie adoptée devient collectivement inefficace pour les produits. Chaque agent base en effet sa décision sur l'hypothèse que ses concurrents ne modifieront pas leur prix, ce qui conduit à une dévaluation progressive du marché. Nous introduisons ensuite un comportement cyclique visant à éviter cet effondrement des bénéfices. Le principe est le suivant :

— L'agent tente d'abord une baisse de prix.
— Si cette réduction ne génère pas une augmentation des bénéfices, il revient à son prix initial.
— Si la baisse de prix s'avère inefficace, l'agent tente alors une augmentation de prix.
— Si cette hausse entraîne une diminution des bénéfices, il rétablit son prix initial.
— En revanche, si une variation de prix, à la hausse ou à la baisse, se traduit par une amélioration des bénéfices, l'agent adopte ce nouveau prix comme référence pour les tours suivants.

Ce comportement est illustré dans la figure 3.19, où tous les agents adoptent cette même stratégie. Nous observons que les produits P_2 et P_3 augmente significativement leurs prix. Cette stratégie peut s'avérer efficace lorsqu'un produit se distingue de ses concurrents, notamment par sa qualité, à l'instar du marché des smartphones. Toutefois, son succès

	P_1	P_2	P_3	P_4	P_a
Prix	49.9	59.9	99.9	49.4	$PrixP_a$
Qualité	0.01	0.20	0.25	0.04	0.17

TABLE 3.7 – *Caractéristiques des produits de notre scénario.*

dépend en grande partie des attentes des consommateurs et, dans notre modèle, des paramètres définissant leur positionnement. Cela souligne l'importance d'une calibration rigoureuse. Il est important de noter que, dans ces scénarios, nous conservons les mêmes paramètres que ceux calibrés précédemment. Or, ces paramètres ne correspondent pas nécessairement aux nouvelles configurations étudiées, ce qui peut influencer les résultats obtenus. Finalement, les stratégies adoptées par nos agents restent inchangées. Pourtant, les résultats révèlent une forme de coopération implicite. Si P_2 ou P_3 réduisait son prix, cela affecterait directement le second produit en raison des interactions sur la catégorie. De plus, les produits P_1 et P_4 ne peuvent pas augmenter leurs prix, car ils évoluent sur le même segment de marché et subissent ainsi une contrainte concurrentielle. On en déduit que, pour limiter les hausses de prix, il est essentiel qu'un produit dispose d'au moins un concurrent positionné sur une tranche similaire en termes de prix et de qualité.

Dans cette section, nous avons montré que ce modèle permet de tester différentes stratégies et d'analyser leurs interactions. Il serait intéressant d'identifier d'autres ensembles de stratégies conduisant systématiquement à un état stable, indépendamment des prix initiaux. C'est précisément ce que nous observons dans notre premier test (figure 3.18) : quels que soient les prix de départ, ils convergent toujours vers le coût des produits. Ce comportement illustre un système dans lequel la concurrence est parfaite et où toutes les marges finissent à zéro.

3.6.2 Un scénario fictif

Considérons un scénario dans lequel l'entreprise A introduit un nouveau produit noté P_a. La table 3.7 présente les prix avant l'application des paramètres ainsi que la qualité des produits après application des paramètres et normalisation. On observe que P_a est défini comme un produit de qualité moyenne.

La figure 3.20 illustre à travers une simulation le prix optimal maximisant le bénéfice. L'entreprise A fixe ainsi le prix de son produit à 55.10. Une fois P_a introduit sur le marché, les concurrents subissent une diminution de leur chiffre d'affaires. Deux scénarios sont alors envisageables : soit les concurrents maintiennent leurs prix inchangés, soit les plus impactés décident d'ajuster leur stratégie tarifaire.

Dans le premier cas, si aucun concurrent ne modifie son prix, la situation du marché correspond exactement à celle simulée lors de notre analyse de l'impact de l'introduction de P_a. Le modèle ne permet alors pas d'obtenir d'informations supplémentaires. Dans le second cas, les concurrents ayant subi une perte significative de bénéfices, par exemple supérieure à 10 %, ajustent leurs prix. Cette réaction initiale peut déclencher une cascade

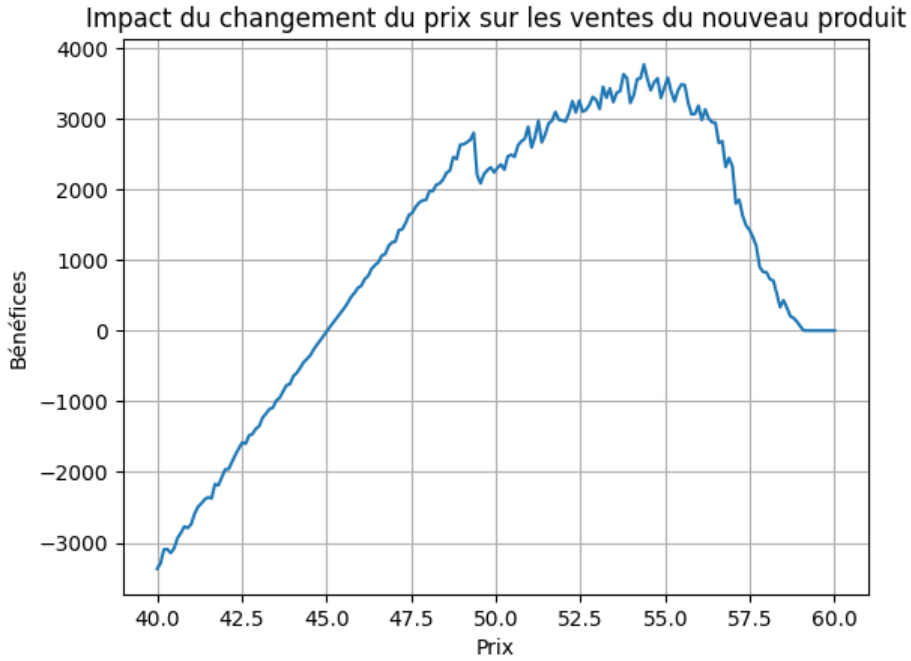


FIGURE 3.20 – *Évolution des bénéfices simulés en fonction du prix de P_a. On observe que le prix optimal est 55.1*

d'ajustements parmi les autres acteurs du marché. La figure 3.21 illustre ces dynamiques concurrentielles lorsque les réactions sont déclenchées par une perte de bénéfices de 10 % et de 20 %. Dans les deux situations, une guerre des prix s'installe, entraînant une érosion des marges pour la plupart des produits. Seuls les produits haut de gamme conservent une marge bénéficiaire significative. Ce phénomène disparaît lorsque le seuil de réaction des concurrents est suffisamment élevé, par exemple fixé à 50 % du bénéfice, empêchant aussi tout ajustement.

3.6.3 Analyse de sensibilité

Rappelons que les six paramètres sont les suivants : x_1 et x_2 influencent la représentation de la qualité du produit ; x_3 détermine la sensibilité au prix ; x_4 et x_5 régissent la distribution de la sensibilité à la qualité ; et β est le coefficient d'aversion à la perte.

Pour le reste de l'analyse, nous avons utilisé une distribution de ventes fixe obtenue à l'aide d'un ensemble défini de paramètres, et observé comment les résultats s'écartent de cette référence. Nous aurions également pu utiliser la déviation à partir de nos paramètres calibrés ; les résultats auraient été similaires.

Un paramètre à la fois

L'analyse One-Factor-At-Time (OFAT) présentée en 3.22 révèle que les paramètres ont des impacts très variables sur la distance de Wasserstein par rapport à la distribution de référence. Le paramètre x_2 présente l'effet le plus fort : lorsqu'il augmente de 0 à 1, la distance moyenne chute brusquement d'environ 30 à 2.7, indiquant une forte stabilisation. Le paramètre x_3 montre un effet non linéaire, avec un pic marqué entre 2000 et 3000, suivi d'un déclin, ce qui indique un seuil critique au-delà duquel le comportement des agents change

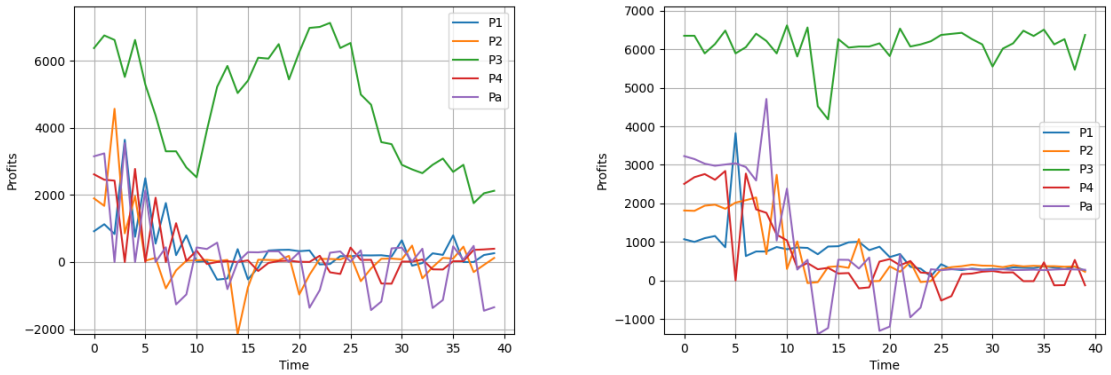


FIGURE 3.21 – *Le bénéfice de chaque produit, P_1, P_2, P_3, P_4, P_a. Chaque pas de temps représente une simulation d'où sont obtenus les bénéfices des produits. Chacun des agents contrôlant les caractéristiques des produits réagit avec une stratégie de maximisation de son bénéfice lorsque celui-ci diminue de 10% sur la figure de gauche et 20% sur la figure de droite. On observe une guerre des prix réduisant les marges sauf pour le produit cher et de haute qualité.*

significativement. Le paramètre β a un effet modéré, avec une augmentation progressive de la distance, tandis que x_1 exerce un effet plus doux, avec une baisse graduelle et plus stable. Enfin, les paramètres x_4 et x_5 ont des effets moins significatifs.

Analyse de Sobol

Paramètre	ST	$ST\ conf$	S_1	$S_1\ conf$
x_1	0.045	0.021	0.005	0.016
x_2	0.085	0.020	0.0009	0.031
x_3	0.910	0.073	0.637	0.094
β	0.090	0.023	0.033	0.028
x_4	0.018	0.002	-0.0001	0.012
x_5	0.329	0.045	0.097	0.048

TABLE 3.8 – *Résultats de l'analyse de sensibilité de Sobol*

— x_3 : L'indice de sensibilité total $ST = 0.910$ est très élevé, indiquant que ce paramètre a une influence considérable sur la sortie du modèle. L'indice de premier ordre $S_1 = 0.637$ montre également que x_3 exerce un effet direct fort, bien que la sensibilité totale inclue aussi les interactions avec les autres paramètres.

— x_5 : L'indice $ST = 0.329$ suggère une influence modérée de ce paramètre sur la sortie du modèle. L'indice de premier ordre $S_1 = 0.097$ indique que l'effet direct de x_5 est moins important que celui de x_3, mais reste néanmoins notable.

— x_2 et x_1 : Ces paramètres ont des indices de sensibilité relativement faibles, en particulier x_2 avec $ST = 0.085$ et $S_1 = 0.0009$. Cela indique que leur impact sur le modèle est minimal, même les effets d'interaction restent faibles. x_1 présente une influence légèrement plus grande avec $ST = 0.045$, mais cela reste faible en comparaison avec les paramètres plus influents comme x_3.

— β : Bien que l'indice de sensibilité soit modéré avec $ST = 0.090$ et $S_1 = 0.033$, il

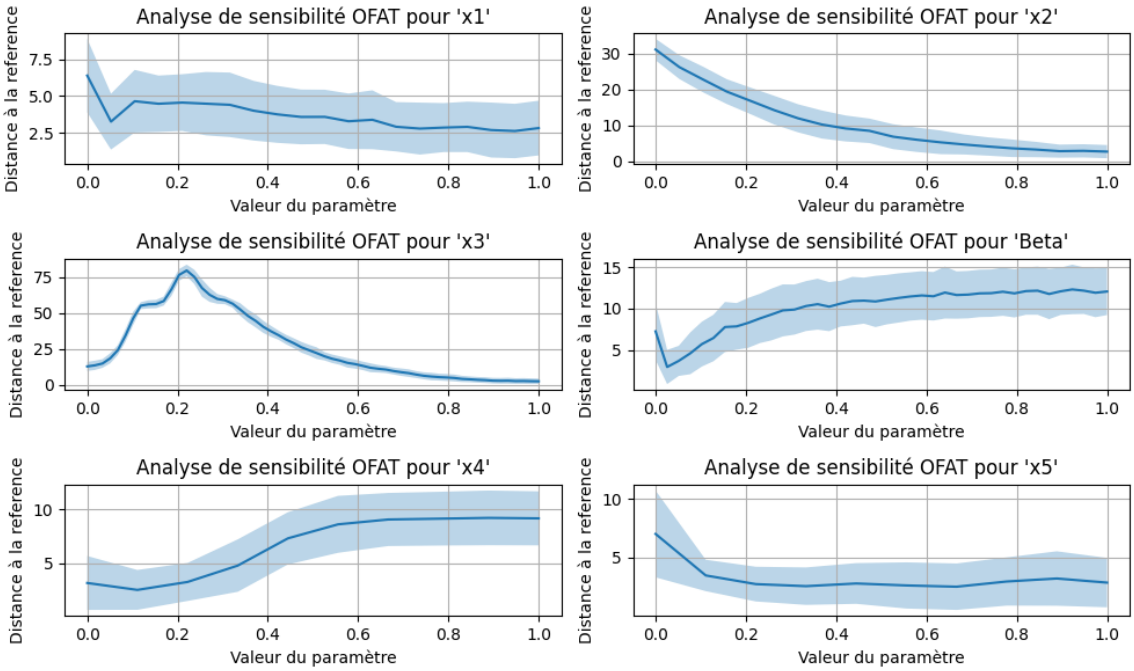


FIGURE 3.22 – *Analyse de sensibilité un-paramètre-à-la-fois (OFAT) pour chaque paramètre. La déviation est due à la stochasticité de notre modèle.*

reste inférieur à celui de x_3 et x_5, ce qui suggère que son rôle dans le modèle est moins dominant.

— x_4 : L'indice de sensibilité de premier ordre $S_1 = -0.0001$ est quasiment nul, ce qui indique que ce paramètre n'a presque aucun effet direct sur la sortie du modèle. Même l'indice de sensibilité total $ST = 0.018$ est faible, ce qui renforce l'idée que x_4 a une importance relative minimale.

— x_3 : Ce paramètre a la plus forte influence sur le modèle, avec un indice de sensibilité total $ST = 0.911$ et un effet direct important $S_1 = 0.638$.

— x_5 : Avec $ST = 0.329$ et $S_1 = 0.097$, x_5 montre un impact modéré. Bien que son effet direct soit plus faible que celui de x_3, il joue néanmoins un rôle significatif via les interactions.

— β : Le coefficient d'aversion à la perte a une influence modérée ($ST = 0.090$, $S_1 = 0.033$), contribuant principalement par les interactions.

— x_2 et x_1 : Ces deux paramètres ont des indices de sensibilité faibles, x_2 ($ST = 0.085$) devançant légèrement x_1 ($ST = 0.045$). Leurs effets sur la sortie sont minimes, ce qui suggère qu'ils peuvent être relégués au second plan dans l'étalonnage du modèle.

— x_4 : Ce paramètre a un impact négligeable, avec une sensibilité directe quasi nulle ($S_1 = 0.001$) et un indice total très faible ($ST = 0.018$), ce qui indique qu'il a peu d'effet sur les résultats du modèle.

Le paramètre x_3 est le facteur le plus influent dans le modèle, suivi de x_5. Les autres paramètres, en particulier x_1, x_2 et x_4, ont des impacts faible. Ces résultats suggèrent que le modèle est très sensible aux variations de x_3, ce qui pourrait en faire un facteur critique dans la prise de décision basée sur cette analyse. De plus, l'analyse OFAT offre

une meilleure compréhension des contributions individuelles de chaque paramètre, tandis que les indices de sensibilité fournissent une mesure précise de leurs influences directes et totales sur le comportement du modèle.

3.6.4 Un autre exemple

Notre modèle peut aussi s'appliquer à de nombreux autre type de produits. Par exemple, les télévisions avec leurs écrans LED, QLED ou OLED ; la plupart des produits vendus sur Amazon, où les notes et le nombre de votes influencent le choix ; ou encore les smartphones, dont les constructeurs et influenceurs mettent en avant de nombreuses caractéristiques. On peut aussi penser aux popcorns au cinéma, où dans notre modèle la quantité remplace la qualité.

Les résultats sont plus complexes pour les produits où la marque joue un rôle central. Dans ce cas, il faut être capable de détecter cette influence et d'ajouter un paramètre spécifique pour en tenir compte. C'est tout à fait possible, mais nous ne l'avons pas encore intégré.

Pour les produits de consommation courante, d'autres facteurs entrent en jeu. La quantité et le prix sont généralement les deux premiers critères. Mais la marque, les goûts et les préférences personnelles des consommateurs ont aussi un impact très important. Cet effet est bien plus marqué que pour les produits techniques ou pratiques, où ce sont surtout les caractéristiques objectives qui comptent.

Ainsi, notre modélisation est très pertinente pour les premiers exemples, pertinente pour les seconds si l'on intègre l'effet de la marque, mais beaucoup moins adaptée aux produits de grande consommation. En effet, les goûts et les habitudes influencent trop fortement les résultats pour que le modèle fonctionne correctement sans ajustements. Il faudrait donc l'adapter afin de prendre en compte ces aspects dans le cas des biens de consommation courante. Enfin, la concurrence dans ce type de marché est souvent inter-catégories : par exemple, les conserves de haricots sont en concurrence même partiel avec les conserves de petits pois, alors que notre définition actuelle les place dans des catégories différentes.

3.7 Une modélisation sur les forfaits de téléphonie mobile

Nous avons étudié un secteur d'activité qui utilise activement les effets de contexte pour valoriser certaines offres : les forfaits de téléphonie mobile. Les prix y varient fréquemment et sont systématiquement enregistrés par les concurrents afin d'ajuster leur positionnement sur ce marché hautement concurrentiel. Ainsi, en plus des données de ventes, les données des prix pratiqués par les concurrents sont également disponibles et régulièrement exploitées.

Plusieurs défis émergent dans l'analyse de ce marché. D'abord, il s'agit d'un secteur en constante et rapide évolution. Par exemple, en l'espace de deux ans, le prix du Go est passé de 0,40 €/Go à 0,10 €/Go, puis à 0,05 €/Go (données de juin 2025). Par ailleurs, l'introduction de la 5G a entraîné des perturbations notables sur le volume et le type des

ventes. De même, des évolutions réglementaires, telles que la généralisation de conditions avantageuses pour l'utilisation des forfaits à l'étranger, ont également modifié en profondeur les dynamiques du marché.

En outre, les campagnes de publicité et des événements exogènes influencent significativement les comportements d'achat. Ces facteurs exogènes et ces transformations technologiques génèrent une volatilité marquée dans les ventes, qui évoluent de manière erratique dans le temps. Cela complique l'application directe et généralisée de notre modèle, notamment sur le long terme, car les résultats obtenus varient fortement selon les périodes observées.

Notre modèle, en l'état, ne permet pas de capturer adéquatement ces perturbations majeures. Ce choix est délibéré : pour préserver la capacité explicative du modèle et pouvoir isoler les effets de l'aversion à la perte ainsi que les effets de contexte, nous avons fait le choix de maintenir une structure simple, avec peu de paramètres. Cette simplicité constitue sans doute l'une des forces du modèle, car elle permet de mettre en évidence de manière claire le lien entre l'aversion à la perte et les effets de contexte. Elle offre également la possibilité de calibrer le modèle sur des données réelles, démontrant ainsi sa capacité d'adaptation à différents environnements empiriques. Par ailleurs, le nombre restreint de paramètres limite le risque de surapprentissage, ce qui renforce la validité des résultats obtenus et conforte la pertinence de notre approche.

En intégrant davantage de paramètres (liées par exemple aux dépenses marketing ou au bouche-à-oreille), nous aurions risqué de diluer l'effet spécifique de l'aversion à la perte et de masquer les effets de contexte. Cette volonté de simplicité justifie que nous n'ayons pas intégré explicitement les impacts marketing, bien que leur influence soit avérée. Par conséquent, notre modèle ne vise pas une généralisation robuste à l'ensemble du marché ni à des horizons temporels étendus. Il se concentre plutôt sur la modélisation d'un aspect spécifique du comportement des consommateurs. Cette contribution demeure néanmoins significative, dans la mesure où il est rare de parvenir à isoler empiriquement un tel mécanisme dans la littérature. En effet, la complexité et la variabilité du comportement humain rendent encore aujourd'hui illusoire toute modélisation exhaustive des agents humains. Dans le cadre des SMA, cette difficulté est renforcée par le fait que des dynamiques collectives émergent de l'interaction entre individus, venant s'ajouter à la complexité déjà présente au niveau de chaque agent, ce qui justifie l'intérêt d'un modèle ciblé sur des mécanismes spécifiques.

Concernant les données exploitées, nous avons eu accès à des informations détaillées : prix des concurrents, prix des offres de la marque étudiée, et volumes de ventes correspondants. Toutefois, l'agrégation des ventes se faisait sur des périodes temporelles inconsistantes, ce qui en limite l'exploitation optimale. Sur des périodes où les niveaux d'investissement marketing et les variations de prix globaux restent relativement stables, le modèle produit des résultats cohérents.

Nous avons ainsi ciblé une période de trois mois pour notre analyse, en appliquant une calibration similaire à celle réalisée sur un jeu de données précédent, avec la même fonction de perte. Durant cette période, entre 11 et 20 offres étaient simultanément présentes

sur le marché. Ce nombre important d'options conduit à formuler l'hypothèse d'un mécanisme de « top of mind ». Concrètement, chaque agent sélectionne aléatoirement entre 1 et 4 marques (sur un total de 4 marques présentes dans les données), la probabilité de sélection étant dérivée de données d'enquête où les répondants indiquent spontanément la ou les marques qui leur viennent en tête.

Le modèle doit également tenir compte d'effets d'inertie : par exemple, si l'opérateur A a dominé le marché pendant deux mois, un changement soudain de l'offre en faveur de l'opérateur B ne se traduira pas immédiatement par un basculement massif des comportements des consommateurs. Le modèle, en revanche, capte cette évolution de manière quasi instantanée, ce qui constitue une simplification notable par rapport aux comportements réels.

Enfin, l'introduction du mécanisme de « top of mind » complexifie considérablement le code du modèle, au point de ralentir son exécution. Cela limite le nombre de simulations et d'explorations paramétriques possibles. Nous avons réalisé un grid search sur 300 simulations, chaque simulation couvrant 60 pas de temps pendant lesquels les agents choisissent une offre.

L'apparition ou la disparition de certaines offres durant ces périodes génère des discontinuités marquées, rendant la calibration délicate sur l'ensemble des données.

Les paramètres que nous cherchons à calibrer sont : le paramètre d'aversion à la perte, ainsi que les paramètres gouvernant la fonction de transformation du volume de données (Go) en qualité perçue par les consommateurs.

La figure 3.23 illustre l'évolution de la fonction de perte au fil des itérations de calibration, menée ici à l'aide d'un grid search.

Les résultats obtenus démontrent que le modèle est capable d'identifier un jeu de paramètres permettant de simuler correctement certaines périodes spécifiques. Une fois ces paramètres établis, les mêmes types d'analyses que celles menées sur les jeux de données précédents peuvent être réalisés.

Cependant, il nous manque encore un paramètre ou un mécanisme d'agrégation permettant de garantir que les paramètres obtenus puissent se généraliser à l'ensemble des périodes. De plus, nous faisons l'hypothèse que la population d'agents reste constante au fil du temps : nous ne générons pas de nouveaux agents, partant du principe que l'échantillon initial est représentatif de la population globale. Cette hypothèse constitue une limitation, dans la mesure où la composition de la population de consommateurs peut évoluer (sous l'influence des campagnes publicitaires, par exemple). Néanmoins, en l'absence d'information permettant de qualifier ces variations, nous sommes contraints de supposer que notre échantillon d'agents est suffisant.

Ces résultats illustrent à la fois la robustesse du modèle, capable de s'adapter à diverses situations, et ses limites. Des travaux complémentaires sont nécessaires pour intégrer des facteurs tels que l'impact des investissements marketing ou les dynamiques de bouche-à-oreille. On peut ainsi envisager le développement d'un modèle modulaire, combinant ces

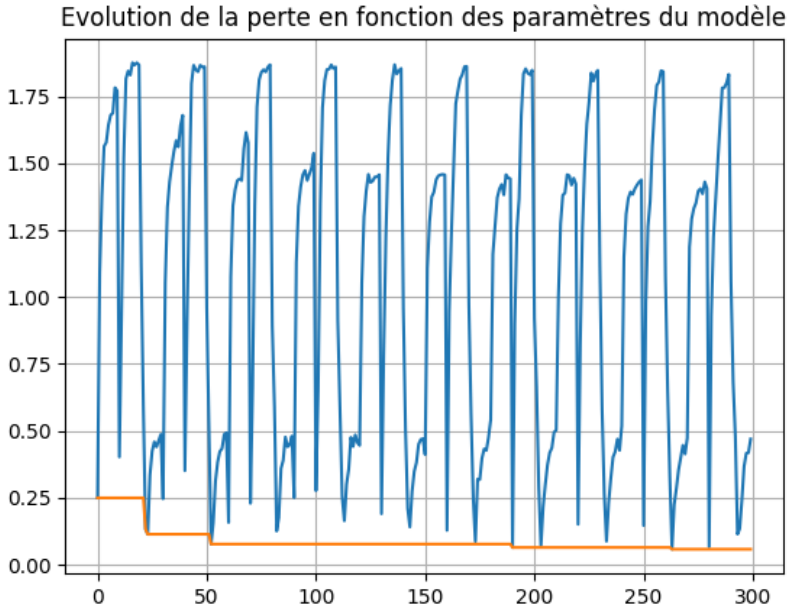


FIGURE 3.23 – *Evolution de la fonction de perte au fil des itérations de la calibration grid search du modèle sur les données de téléphonie mobile.*

différentes briques de base, et pour lequel une calibration fine permettrait de reproduire de manière satisfaisante les dynamiques observées sur les données réelles.

3.7.1 Conclusion sur la modélisation

Bien que les agents dans notre modèle soient indépendants, sans interactions sociales ou impact sur un environnement partagé, nous observons des dynamiques collectives non triviales. Cette propriété place notre approche dans une zone intermédiaire entre une modélisation centrée individus et les approches multi-agents classiques. En effet, les effets émergents que nous mettons en évidence montrent qu'une structure agrégée complexe peut émerger de décisions isolées, sans qu'il soit nécessaire d'introduire des interactions explicites. Cela renforce la capacité explicative du modèle tout en maintenant une certaine simplicité computationnelle.

3.8 Synthèse

Ce chapitre a introduit une modélisation du choix intra-catégorie à travers une fonction d'utilité, construite pour capturer les comportements individuels en situation de décision parmi des produits similaires. Nous avons fait le choix d'une modélisation minimaliste, afin de garantir à la fois l'interprétabilité, la stabilité du modèle et sa capacité de calibration.

Une contribution centrale de ce travail est de montrer qu'une modélisation simple est à la fois nécessaire et suffisante pour approcher la simulation de comportements de consommateurs. Elle est nécessaire pour assurer la validité du modèle : en limitant le nombre de paramètres, on réduit le risque d'introduire artificiellement des effets ou de masquer les

mécanismes réellement à l'œuvre. Elle est également suffisante, car cette structure épurée permet déjà de faire émerger des effets de contexte observés empiriquement, tels que l'effet de leurre. Ces dynamiques ne sont donc pas le produit d'une sur-spécification, mais résultent directement des mécanismes de décision intégrés dans le modèle. Dans cette perspective, notre approche s'inscrit dans l'esprit du rasoir d'Ockham : entre plusieurs explications possibles, la plus simple s'avère souvent la plus plausible.

Le modèle a été testé sur plusieurs jeux de données, réelles et simulées, montrant sa capacité à reproduire des comportements cohérents et à capturer l'impact de la structure de l'offre sur les décisions individuelles. L'analyse de sensibilité a permis d'identifier les paramètres déterminants et de confirmer que le modèle reste stable tout en étant expressif.

Avec ce modèle, nous proposons un système d'aide à la décision capable d'éclairer l'impact de l'introduction d'un nouveau produit sur le marché. Le processus de calibration, certes nécessaire, contribue à une meilleure compréhension des mécanismes de formation et de perception de la qualité chez les consommateurs. Ainsi, notre approche peut intéresser aussi bien les marques que les distributeurs, en leur offrant un outil pour orienter leurs choix stratégiques, qu'il s'agisse de mettre en avant certains produits, de gérer les stocks promotionnels ou encore de faciliter l'écoulement de références spécifiques.

Nous montrons également que, lorsque le modèle intègre plusieurs « méta-agents » représentant des entreprises concurrentes, il met en évidence des dynamiques de marché réalistes. En particulier, si certains acteurs ajustent leurs prix à la suite d'une baisse significative de leurs bénéfices, cela peut déclencher une cascade de réactions concurrentielles menant à une guerre des prix, au détriment des marges de la majorité des produits. Seuls les produits haut de gamme parviennent alors à maintenir une rentabilité notable. À l'inverse, lorsque le seuil de réaction des concurrents est fixé à un niveau plus élevé, aucun ajustement n'est enclenché et le marché reste stable. Ces résultats peuvent ainsi être mobilisés dans le cadre d'un système d'aide à la décision, en fournissant aux entreprises des scénarios prospectifs sur les conséquences possibles de leurs propres choix tarifaires, ainsi que de leurs réactions aux modifications tarifaires des concurrents.

Enfin, cette modélisation constitue un élément autonome intégrable dans un SMA plus large. De plus elle permet ainsi de représenter les choix des agents au sein d'une catégorie de produits, en s'appuyant sur quelques caractéristiques individuelles, et peut être utilisée comme système d'aide à la décision dans des contextes marketing variés.

Chapitre 4

Réalisation d'un système multi-agents appliqué au marketing : le cas des promotions

Dans ce chapitre, nous explorons l'utilisation d'un Système Multi-Agents (SMA) pour l'étude des promotions dans un supermarché. L'objectif est de proposer un modèle de simulation appliqué aux produits de consommation courante, c'est-à-dire des articles achetés régulièrement par les ménages (produits alimentaires, hygiène, entretien, etc.).

Le comportement des agents est fondé sur un mécanisme psychologique déjà étudié : L'aversion à la perte. Dans ce chapitre nous introduisons la notion d'inertie dans les achats. Dans le cas des biens de consommation courante, les consommateurs tendent à reproduire leurs habitudes d'achat, car ces biens répondent à des besoins récurrents et peu différenciés.

Bien que nous ne disposions pas de données réelles de supermarchés dans le cadre de cette thèse, le modèle est conçu de manière à pouvoir intégrer ultérieurement des données d'achat empiriques, par exemple issues des cartes de fidélité. Ces dernières constituent une source d'information précieuse, car elles reflètent la régularité des achats et le degré de fidélité des clients. Ce phénomène est particulièrement marqué dans le cas des supermarchés, davantage que dans des secteurs tels que le bricolage ou la téléphonie mobile sans engagement.

Ainsi, ce modèle vise à tirer parti de ces historiques d'achats afin de représenter plus finement les besoins des agents, tout en permettant d'analyser l'impact des promotions commerciales. Le cadre de simulation ouvre également la voie à l'étude des phénomènes émergents liés à ces promotions, comme les variations collectives de la demande ou l'évolution de la fidélité des clients.

4.1 Simulation d'un supermarché

Nous considérons le cas d'un supermarché de type « grande surface », caractérisé par un flux régulier de clients fidèles. Ces derniers achètent de manière répétée des produits périssables ou semi-périssables, dont la durée de vie limitée renforce la redondance des achats. Cette régularité, particulièrement observable chez les porteurs de cartes de

fidélité, peut être exploitée pour mieux modéliser le comportement des agents dans la simulation.

Dans cette version du modèle, nous faisons deux simplifications méthodologiques :

Il n'y a pas de modélisation spatiale. Nous ne tenons pas compte du déplacement des clients dans les rayons, ce qui exclut l'apparition de comportements d'achats impulsifs ou imprévus. La concurrence externe n'a pas d'impact, l'influence des autres enseignes de distribution n'est pas prise en compte. Ce parti pris de simplicité méthodologique n'est pas une faiblesse mais un choix assumé : il permet de mieux isoler et analyser les mécanismes liés à l'impact des promotions sur les comportements des clients fidèles. En réduisant ainsi la complexité du système, nous obtenons un modèle plus lisible et plus explicable, offrant un cadre pertinent pour l'étude des phénomènes émergents associés aux politiques promotionnelles.

L'attention est donc concentrée sur l'impact des promotions à l'intérieur du supermarché. Notre objectif est d'analyser comment les promotions influencent la consommation des clients existants et comment ces modifications de comportements individuels produisent des dynamiques collectives au sein du magasin.

4.1.1 Faits stylisés à reproduire

Pour évaluer la pertinence d'un modèle multi-agents appliqué au contexte des promotions, il est indispensable qu'il soit capable de reproduire certains faits stylisés bien établis dans la littérature marketing, comme ceux décrits par (Blattberg et al. 1995). Parmi ces dynamiques classiques, on retrouve :

- l'augmentation du volume des ventes sous l'effet d'une promotion, qui doit s'accompagner de l'estimation d'une élasticité prix crédible. Cette dernière traduit la sensibilité de la demande aux variations de prix et permet de vérifier que le modèle reflète correctement les réactions attendues des consommateurs ;
- la cannibalisation, phénomène par lequel la promotion d'un produit détourne les achats d'autres produits de la même catégorie, sans nécessairement accroître la consommation globale ;
- l'impact négatif des promotions successives, lié à l'effet d'accoutumance ou de saturation des clients, qui tend à réduire l'efficacité marginale de nouvelles opérations commerciales lorsqu'elles sont trop rapprochées dans le temps.

Au-delà de ces régularités globales, notre approche permet également d'intégrer des phénomènes observables au niveau individuel, qui échappent aux modèles plus agrégés. Il s'agit par exemple de l'impact des guerre des prix, ou encore du rôle des promotions dans l'acquisition de nouveaux clients et dans leur fidélisation à long terme.

En revanche, certains facteurs restent volontairement exclus de notre modélisation, tels que le placement optimal des produits dans le magasin, l'image perçue du point de vente, ou encore celle de l'enseigne dans son ensemble. Compte tenu de l'état actuel des connaissances en modélisation multi-agents appliquée au marketing, il n'est pas encore possible de construire un modèle à la fois explicable et intégrant simultanément l'ensemble de ces dimensions. Nous faisons donc le choix d'une approche plus épurée, mais enrichie par

rapport au modèle précédent. Cette montée en complexité contrôlée est rendue possible grâce à la compréhension fine du modèle présenté dans le chapitre précédent, tout en maintenant un modèle interprétable et scientifiquement robuste. Nous considérons ainsi qu'il est nécessaire de valider progressivement chacune des « briques » constitutives de notre approche, afin de pouvoir les combiner ultérieurement dans un modèle plus complet, capable d'intégrer de manière cohérente l'ensemble de ces effets. Dans cette perspective, nous proposons avec ce chapitre une première étape de complexification dans une direction précise : celle des grandes surfaces et des produits de grande consommation.

4.1.2 Le contexte de ce modèle

Dans la suite de ce chapitre, nous nous plaçons dans le contexte d'un magasin de grande surface proposant des biens de consommation courante. Nous posons l'hypothèse que les clients réalisent leurs achats de manière récurrente, que ce soit dans ce magasin ou auprès d'enseignes concurrentes. Cette régularité, propre aux produits du quotidien (alimentaire, hygiène, entretien, etc.), permet de structurer le comportement des agents et de définir des besoins récurrents dans le temps. Elle offre ainsi un cadre favorable à la modélisation, car elle réduit l'incertitude sur les dynamiques de demande.

Le modèle est conçu comme une application générique, capable d'intégrer des données réelles issues de programmes de fidélité. Ces données, lorsqu'elles sont disponibles, constituent une source particulièrement riche pour calibrer les comportements des agents, puisqu'elles permettent de relier chaque transaction à un profil client détaillé et d'observer directement la fréquence et la régularité des achats.

Néanmoins, dans le cadre de cette thèse, nous ne disposons pas de telles données pour valider empiriquement le modèle par comparaison directe avec un cas réel, comme cela avait été fait dans les chapitres précédents. Ce manque de données ne remet cependant pas en cause la pertinence de l'approche : la validation repose ici sur la capacité du modèle à reproduire des faits stylisés bien documentés dans la littérature marketing, ce qui constitue un critère de robustesse reconnu dans le domaine de la simulation multi-agents.

4.2 Simulation d'un magasin

On utilise l'historique d'achats obtenu via les cartes de fidélité pour approximer le besoin de chaque agent. La stratégie de récupération des données via les clients possédant une carte de fidélité est une stratégie efficace mise en place par beaucoup de grandes surfaces, elle permet principalement l'analyse des comportements de consommateurs. On pose l'hypothèse que les besoins des agents ne changent pas au cours de la simulation.

Depuis de nombreuses années, les entreprises commerciales mesurent l'importance et les apports d'une campagne promotionnelle, qu'elle soit de type réduction de prix ou de type publicitaire. Leur souhait est évidemment que cette campagne soit la plus efficace possible, et cela, en cernant au mieux les moyens à utiliser et la cible concernée. Malheureusement, ces campagnes ont en général un coût très élevé et peuvent, si elles ne sont pas

exécutées correctement, avoir peu d'utilité, voire même avoir une influence négative sur l'image d'une marque ou d'un produit. Comme pour de nombreux phénomènes coûteux et complexes à mettre en œuvre *in vivo* il est préférable d'essayer de mesurer à priori, *in silico* l'impact des campagnes que l'on souhaite réaliser. Par ailleurs, le test *in silico* est depuis quelques années communément admis (Axtell et al. 2022 ; Delre et al. 2007 ; Jager 2007 ; Negahban et al. 2014 ; Said et al. 2002). Enfin, les effets des campagnes marketing (mix marketing (Borden 1964)) sont principalement modélisés dans l'industrie par des méthodes statistiques (Tellis 2006 ; Wigren et al. 2019) et d'apprentissage automatique (Hung et al. 2019 ; Tellis 2006), bien souvent pour faire de la segmentation clientèle et de l'aide à la décision. Ces approches globalisantes ne permettent ni la mise en place de comportements fins des consommateurs ni la modélisation d'une multitude de différentes campagnes. Par exemple : connaître la différence d'impact entre une promotion de 10% et de 20% ou savoir sur quels consommateurs la différence est forte ou faible. En effets, les approches globalisantes ne permettent de répondre que partiellement à ces questions. Nous soutenons que les systèmes multi-agents (SMA) avec leur approche centrée individus facilitent la conception et la calibration de comportements à un niveau de finesse permettant de mieux comprendre la multitude de facteurs auxquels est confrontée une campagne marketing. Nous montrons par ailleurs qu'une telle approche permet une adaptation plus aisée aux changements de l'environnement comme l'arrivée ou le changement d'un produit et fournit, de ce fait, des résultats robustes et exploitables.

Dans ce chapitre, nous nous plaçons dans le cadre d'une grande surface de type supermarché avec pour objectif la compréhension des dynamiques de consommateurs à travers leurs réactions, au fil du temps, face aux changements de prix ou de packaging des produits. Pour ce faire, nous proposons une modélisation centrée individus, permettant à travers une simulation de déclencher des campagnes promotionnelle à une date et une durée choisie, et permettant d'en mesurer l'impact sur des populations variées. La grande diversité des comportements clients (inertiels, dépendants du prix ou de la qualité, promophiles, équilibrés...) et des campagnes promotionnelles (baisse du prix, réduction en pourcentage, bons d'achats, lots achetés/offerts...) motive l'utilisation de ce type de modèle.

Dans notre modèle les interactions sont basées sur des critères issus de différentes disciplines comme les sciences sociales pour l'aversion à la perte (Prospect theory) (Hardie et al. 1993), le marketing pour l'inertie influant le choix des marques (Bawa 1990 ; Cantillo et al. 2007), ou les sciences économiques pour l'évaluation de la qualité (Khandelwal 2010). Dans ce chapitre nous proposons un modèle capable de prendre en compte ces différents aspects en nous inspirant par ailleurs de (Delre et al. 2007 ; Negahban et al. 2014 ; Said et al. 2002) pour la mise en place du choix des produits. Cependant, contrairement à ces articles qui se focalisent sur le comportement des consommateurs, la dynamique des prix est ici le sujet d'étude principal.

Nous montrons dans ce chapitre que notre approche centrée individus, focalisée sur les prix des promotions, permet une meilleure compréhension des comportements que l'on modélise et engendre les faits stylisés connus du domaine comme l'augmentation du volume des ventes et la baisse des ventes concurrentes lors d'une réduction de prix. Ces

deux phénomènes sont impératifs à exhiber puisqu'ils sont communs à toute campagne de réduction de prix. Notre approche les met en évidence, mais elle permet de plus d'observer l'impact des réductions de prix sur les phénomènes d'acquisition et de fidélisation selon les profils de consommateurs, ou encore les effets des guerres de prix sur les différents comportements. Finalement, notre modèle multi-agents s'adapte aux changements de l'environnement, notamment à l'apparition de nouveaux produits, aux changements de prix, ou à la présence d'une nouvelle campagne marketing et ses différents paramètres. Il permet donc de mesurer efficacement les conséquences de différents scenarios « *what if* » de test d'hypothèses sur la campagne à réaliser. Sans avoir à réécrire les différents comportements il est donc possible de changer les différents paramètres et d'en calculer les conséquences.

Dans la première partie, nous présentons l'état de l'art du domaine à travers des travaux réalisés dans un cadre similaire au nôtre et nous critiquons le manque de lien entre l'évolution des prix et comportement des consommateurs. Dans la seconde partie, nous présentons un modèle multi-agents de test de campagnes de réductions qui s'appuie sur une fonction d'utilité individuelle que possède chaque agent pour évaluer les produits. La troisième partie décrit plusieurs expériences réalisées avec ce modèle ainsi que les résultats qu'il permet d'obtenir, comme la fidélisation ou les volumes de vente. Enfin, la dernière partie aborde les avantages et les extensions potentielles de ce modèle, notamment la prise en compte de l'influence sociale.

4.3 État de l'art

4.3.1 L'étude des promotions avec des SMA

L'étude des stratégies marketing par la simulation n'est pas nouvelle. D'autres travaux s'y sont déjà essayés, souvent à l'aide d'approches équationnelles ou statistiques (Bass 1969). Peu d'entre eux se sont appuyés sur une approche centrée individus pourtant nécessaires pour assurer de la différentiation comportementale. On peut néanmoins citer (Axtell et al. 2022 ; Delre et al. 2007 ; Jager 2007 ; Negahban et al. 2014 ; Said et al. 2002) qui utilisent tous une approche à base d'agents.

Les auteurs de (Delre et al. 2007) modélisent le lancement d'un produit dans une population soumis au phénomène de bouche-à-oreille (WOM pour « Word Of Mouth »). Les possibilités d'interactions sont modélisées via un graphe de type « small world » (Watts–Strogatz) et les interactions sont élémentaires, si un agent adopte le nouveau produit, il tente de convaincre ses voisins de faire de même. Le bouche-à-oreille est un phénomène complexe et les modèles qui l'étudient intègrent des graphes afin de représenter des notions de contacts sociaux, contacts que les agents peuvent influencer. Nous supposons que ces modèles de bouche-à-oreille, et plus globalement d'influence sociale, sont un possible développement futur, mais s'intègrent mal avec ce que nous essayons de montrer. En effet, la superposition des effets de dynamiques des prix et de bouche-à-oreille réduirait l'interprétablilité des résultats, de plus nous nous intéressons avant tout à comprendre l'impact des prix sur les différents comportements. Les auteurs de (Negahban et al. 2014) proposent des lignes directrices afin de modéliser correctement des aspects

marketing dans des SMA. Par ailleurs, les auteurs proposent une approche basée sur l'évaluation des produits suivant une fonction d'utilité. Cela permet de modéliser des agents qui, selon leurs caractéristiques personnelles, évalueront différemment les articles. Ainsi, il devient possible de créer et moduler les caractéristiques des agents, de manière à reproduire des comportements classiquement observés en marketing. Les auteurs de (Said et al. 2002) présentent un SMA utilisant des paramètres comportementaux servants à réguler la diffusion ou le reçu de stimuli externe à eux-mêmes, ces paramètres sont similaires à ce que nous proposons. Ils montrent que leur modèle est capable de reproduire des faits stylisés sur le choix des marques des consommateurs, notamment l'émergence d'un équilibre entre les parts de marchés et un effet de verrouillage des parts de marché d'une marque dominante ou encore une compétition cyclique entre les marques dominantes, argumentant que leur outil de simulation permet le reproduire une évolution réaliste d'un marché. L'auteur de (Jager 2007) applique les quatre éléments marketing ((Borden 1964)), produit, prix, place de distribution et promotion, dans un modèle centré individus comportemental tiré des simulations sociales. Les agents ont des préférences individuelles et des préférences sociales (qui consomme les produits). Les préférences individuelles sont définies par les caractéristiques de chaque agent et les préférences sociales sont déterminés en regardant la consommation des individus connectés socialement dans un graphe.

On notera cependant que les modèles présentés par (Delre et al. 2007 ; Said et al. 2002) n'intègrent pas explicitement la composante prix, et que (Axtell et al. 2022 ; Jager 2007 ; Negahban et al. 2014) intègrent bien le prix, mais s'intéressent avant tout à l'influence sociale sans étudier le lien prix - comportement, point central de notre travail.

D'une manière générale, certaines propriétés sont plus faciles à modéliser à l'aide de SMA et de modèles centrés individus qu'avec des modèles équationnels. C'est le cas notamment de l'influence sociale qui nécessite d'expliciter les liens entre les différents individus, par opposition à la notion de publicité qui peut plus facilement être étudiée avec un modèle équationnel. Cela se reflète dans la littérature.

4.3.2 Qu'est-ce qui fait un comportement ?

La modélisation d'un processus de comportement d'achat s'appuie sur deux aspects fondamentaux : les influences internes (caractéristiques propres à chaque individu qui influence l'envie d'acheter tel ou tel produit) et les influences externes (la publicité, la promotion, le bouche à oreilles). Dans ce travail on se focalise uniquement sur les influences internes et la promotion qui est la seule influence externe. Ce choix repose sur le fait qu'il existe déjà des articles se focalisant sur les influences externes (Delre et al. 2007 ; Said et al. 2002), de plus nous montrons qu'il est possible de reproduire des faits marketings connus uniquement avec les influences internes.

Il semble naturel pour tous les auteurs d'utiliser le prix et la qualité de chaque produit comme influence interne. (Bawa 1990 ; Cohen et al. 2020 ; Hardie et al. 1993 ; Seetharaman et al. 1998) proposent d'y ajouter deux critères supplémentaires permettant de rendre les modèles plus réalistes : l'aversion à la perte et l'inertie.

— l'aversion à la perte (prospect theory), s'appuie sur l'idée que perdre 1€ a plus d'impact sur un consommateur que de gagner 1€.
— l'inertie ou la loyauté envers les marques s'appuie sur l'idée qu'un consommateur ne prendra pas forcément le « meilleur » produit disponible, car il est aussi influencé par ses habitudes et sa loyauté envers les marques.

(Cohen et al. 2020 ; Hardie et al. 1993) suggèrent d'utiliser un produit de référence pour prendre en compte l'aversion à la perte. Celui-ci peut être propre à chacun grâce à l'utilisation de SMA. On imagine que chaque client dispose d'une référence mentale sur chacune des catégories de produit représentant un produit désirable pour lui. Cette référence mentale peut être réelle ou imaginaire. En comparant ce produit de référence avec les produits de la catégorie il est possible d'obtenir ce que le client perd ou gagne sur les différentes caractéristiques des produits. L'inertie peut être simplement prise en compte par un processus de renforcement (plus j'en ai acheté, plus j'en achète) (Bawa 1990 ; Seetharaman et al. 1998). Ces différentes notions peuvent être combinées à l'aide d'une fonction d'utilité qui sera utilisée lors de l'évaluation d'un produit ((Negahban et al. 2014)).

4.3.3 L'impact des promotions

Les promotions ont des effets indiscutables sur les ventes. (Blattberg et al. 1995) décrit les effets connus de ces promotions telles que l'augmentation du volume des ventes, l'impact asymétrique de ces promotions sur les autres produits en fonction de leur qualité (la qualité permet de limiter la perte de vente), ou encore l'impact dégressif d'une promotion au fil des répétitions. L'apparition de ces différents effets nous semblent indispensables pour l'évaluation d'un modèle de campagnes promotionnelles.

En conclusion, notre modèle utilise une approche centrée individus comme (Axtell et al. 2022 ; Delre et al. 2007 ; Jager 2007 ; Negahban et al. 2014 ; Said et al. 2002), implémente l'aversion à la perte et l'inertie comme (Bawa 1990 ; Cohen et al. 2020 ; Hardie et al. 1993 ; Seetharaman et al. 1998) et il permet d'exhiber les propriétés classiques d'impact des promotions décrites par (Blattberg et al. 1995 ; Kienzler et al. 2017).

4.4 Description du modèle

Le modèle décrit à été présenté par (Jarod Vanderlynden et al. 2023) et (Jarod Vanderlynden et al. 2024b).

Dans cette section nous nous focalisons sur un modèle de magasin dans lequel il est possible de simuler différentes promotions sur différents packs de produits. Par exemple une promotion en pourcentage de prix ou une promotion de type *y offert pour l'achat de x*. Le modèle est aussi capable de réagir à des modifications de prix hors du cadre d'une promotion temporaire, ou encore à l'arrivée de nouveaux packs dans le magasin. Dans ce modèle, il n'y a pas de représentation spatiale, les agents sont ici omniscients et connaissent tous les packs et leurs caractéristiques. Cependant, ils n'ont pas conscience des manipulations marketing, ils voient simplement tous les produits ainsi que leurs changements. Nous ne prenons pas en compte le positionnement géographique du magasin et des packs ni l'influence sociale pour se concentrer sur l'influence du prix et des promotions. La notion

de pack est à prendre au sens large, un pack est un lot de produit, il dispose donc d'une quantité d'un produit. Pour chaque produit, il existe une quantité unitaire (en gramme, en ml, etc.) pour des raisons de simplicité, nous ne modélisons pas cette quantité unitaire, par contre, nous nous intéressons aux promotions des lots de produits que nous appellerons pack.

Nous présentons dans un premier temps les packs et les agents qui constituent les clients du magasin, dans un second temps l'environnement qui caractérise le magasin et ses packs, dans une troisième partie la dynamique du modèle et finalement, nous présentons le modèle comportemental, en d'autres termes, la manière dont les agents raisonnent pour choisir les packs.

4.4.1 Les paramètres des packs

Un pack représente un article quelconque dans un supermarché. Cet article peut être vendu en quantité unique, il correspond donc à un article seul, ou en lot. Cette information est représentée par la caractéristique quantité. Par exemple, un pack d'eau classique contient 1.5L six fois. Mais ce pack peut être dérivé si le nombre de bouteilles individuelles ou la taille de chacune de ces bouteilles est modifiée. Soit p le prix, Qte la quantité, Qa la qualité et D un booléen indiquant la promotion. Un pack est représenté par un quadruplet noté $P(p, Qte, Qa, D)$. Chaque pack appartient à une catégorie $C_i \in C = C_1, C_2, ...$ Si on reprend l'exemple d'un pack d'eau de six bouteilles de 1.5L cela donne $P_eau(1.5, 9, 0.5, 0)$ un pack d'eau coûtant 1.5€, de 9 litres, ayant une qualité de 0.5 et n'étant pas en promotion.

$$\forall P_j \exists C_i | P_j \in C_i, C_i \in C \tag{4.1}$$

4.4.2 Les paramètres des clients/agents

Un agent, noté $a(H_i, \lambda_i, P_{ref_i}, (\gamma_p, \gamma_q, \gamma_i, \gamma_l))$, représente une entité (une personne, une famille ou autre) qui vient faire ses courses régulièrement dans le magasin. Son comportement doit être basé sur ses habitudes. Un agent est caractérisé par des paramètres internes qui le différencient de ses semblables ainsi que d'un historique individuel et de références mentales pour chacune des catégories. Soit :

- $H_{i,a}$ une liste de produits pour chaque catégorie correspondant à une fenêtre glissante sur l'historique d'achat de l'agent. L'agent ne garde pas tous ses achats en mémoire il finit par « oublier ». À l'initialisation l'agent dispose déjà d'un historique plein. Cet historique peut être généré ou provenir de données réelles pour plus de réalisme. Dans cette étude nos historiques initiaux seront générés aléatoirement par manque d'accès à des données clients.
- $P_{i,a,t}$ Le produit acheté par l'agent a dans la catégorie C_i au temps t.
- l_h la longueur des historiques d'achats. Cela correspond à la capacité de mémoire qu'on souhaite donner aux agents.
- $\lambda_{i,a}$ le besoin pour chaque catégorie C_i. Cette caractéristique reflète la quantité de produit dont un agent a besoin à chaque pas de temps. Par exemple un besoin de 1kg de pâtes représente un agent souhaitant acheter 1kg de pâtes à chaque pas de temps. Ce souhait influe sur son processus de décision et n'est qu'indicatif. Un

agent pourrait suite à son processus de décision acheter 2kg de pâtes comme 500g de pâtes.

— $P_{ref,i} \in C_i$ un pack de référence, toujours unitaire $Qte = 1$, à l'instar de (Hardie et al. 1993). Dans notre modèle ce pack de référence est toujours imaginaire. Cela n'empêche pas que ses caractéristiques peuvent cependant être identiques à un pack existant dans le magasin.

— $\gamma_p, \gamma_q, \gamma_i, \gamma_l$ respectivement la sensibilité au prix, à la qualité, à l'inertie et aux promotions.

Comme le calcul du besoin dépend d'un historique, il devient impossible d'intégrer de nouveaux agents dépourvus de données passées.

4.4.3 L'environnement

L'environnement représente le magasin. Il contient les agents et les packs. Les agents interagissent avec l'environnement en achetant des packs. On souhaite, avec les paramètres de l'environnement, pouvoir moduler le fonctionnement global du modèle. Par exemple donner plus d'importance à la promotion ou augmenter la puissance de l'aversion à la perte. L'environnement est donc caractérisé par les propriétés suivantes :

— β paramètre de l'aversion à la perte (identique pour le prix et la qualité), $\beta > 1$. Il indique combien de fois il est plus impactant de perdre que de gagner. Par exemple $\beta = 1.5$ signifie qu'un prix supérieur de 1€ au prix de référence (l'agent perd 1€) aura un impact 1.5 fois supérieur à un prix inférieur de 1€ (l'agent gagne 1€). $\beta > 1$, ce qui signifie que les agents lorsqu'ils évaluent les produits, ont toujours plus peur de perdre qu'envie de gagner.

— Up limite de quantité d'achat, $Up >= 1$ est la limite de quantité qu'un agent peut acheter en plus de ce dont il a besoin. Un tel achat peut se produire si le produit est en promotion ou qu'il est moins cher que d'habitude par exemple.

— α_{sat} paramètre de saturation. Ce paramètre gère la « pente » de la fonction de saturation

— G_p, G_q, G_i, G_d les paramètres de régulation de l'impact respectivement du prix, de la qualité, de l'inertie et de la promotion (discount). Si on augmente le paramètre G_p tous les agents auront une plus grande sensibilité au prix. Un paramètre G_p plus élevé est utilisé si le magasin étudié est réputé pour ses prix bas.

L'aversion à la perte et la longueure de l'historique sont des paramètres globaux. Ce choix permet de considérer l'historique comme une fenêtre glissante, on ne prend en compte que les H_l derniers pas de temps.

La figure 4.1 représente un diagramme UML de notre modèle avec les caractéristiques internes aux agents, aux produits et à l'environnement.

4.4.4 Hypothèses

Nous considérons les packs comme des biens de consommation courante, ce qui nous permet d'émettre l'hypothèse que les achats sont fréquents et donc qu'à chaque pas de temps chaque agent s'interroge sur l'achat. On pourrait considérer qu'un pas de temps représente une semaine, et que chaque client vient chaque semaine au magasin faire ses

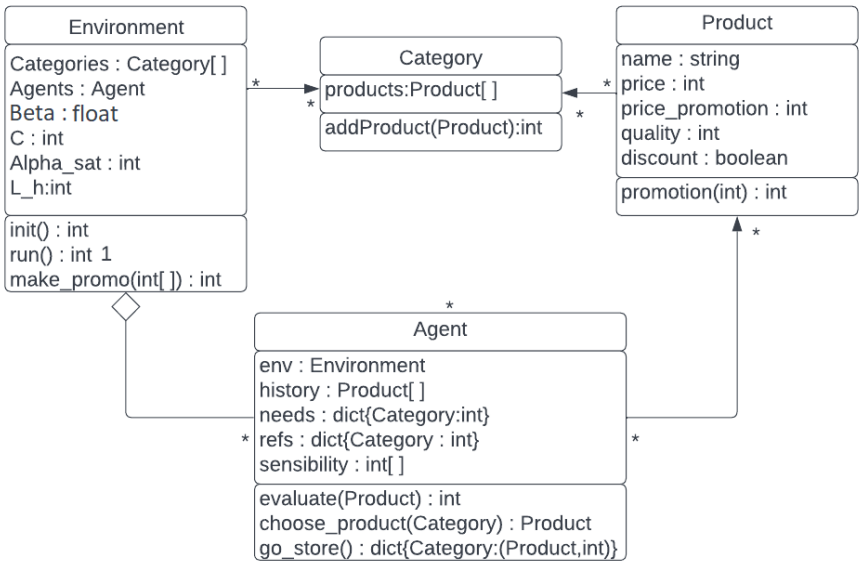


FIGURE 4.1 – *Diagramme UML du modèle.*

courses. On suppose que le besoin ($\lambda_{i,a}$) peut être calculé grâce à la moyenne de l'historique des quantités achetées, quel que soit le pack de la catégorie $\lambda_{i,a} = mean(H_{i,a}(qte))$. On exclut les achats de différents packs d'une même catégorie afin de simplifier le choix des agents. Cela revient à exclure l'achat de deux packs similaires, mais de marques différentes. Par exemple, un agent n'achètera pas deux packs d'eau de six bouteilles de 1.5L de marque différente, il achètera plutôt deux fois le même pack. L'objectif des différents agents sera donc de choisir, à chaque pas de temps, au plus un pack, dans chacune des catégories sur lesquels l'achat a été considéré. Cette hypothèse n'interdit pas aux agents d'acheter plusieurs fois le même pack ni de ne rien acheter. On suppose que les agents sont des entités représentant un foyer, une famille... Cette hypothèse permet à un magasin de relier un agent à une carte de fidélité.

4.4.5 Le choix des packs

L'intérêt de l'agent sur une catégorie

Chaque agent, à chaque pas de temps, pour chaque catégorie, choisit s'il a besoin d'un pack de cette catégorie. L'agent a évalue C_i et calcule $B(C_i, t+1)$ qui est la probabilité d'avoir besoin d'un pack de cette catégorie au temps $t+1$. Enfin, soit $N(C_i, t, t-n)$ la quantité de pack acheté par l'agent lors des n dernières étapes.

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{4.2}$$

$$B(C_i, t+1) = sigmoid(\frac{\lambda_{i,a}}{N(C_i, t, t-n)}) \tag{4.3}$$

Intuitivement, cette formule permet à l'agent a d'augmenter la probabilité de s'intéresser à la catégorie C_i s'il a acheté peu de cette dernière par rapport à son besoin sur les

derniers pas de temps et vice et versa. Étant donné que nous considérons des biens de consommation courante, nous justifions ce comportement en considérant que le besoin sur chaque catégorie doit être rempli. S'il ne l'est pas sur un pas de simulation, il aura un très forte probabilité de l'être sur le pas de temps suivant. Nous n'introduisons pas de catégories nécessaires ou prioritaires, mais cela pourrait être introduit afin de modéliser d'autres catégories de produits.

L'évaluation des packs

Lorsqu'un agent s'intéresse à une catégorie, il en évalue tous les packs, leur donne un score et en choisit un seul. Un score élevé signifie que le pack correspond aux attentes de l'agent et a plus de chances d'être choisi. Pour donner un score aux packs, l'agent utilise ses sensibilités personnelles ainsi que les paramètres des packs.

Nous définissons quatre fonctions d'utilité U_1, U_2, U_3, U_4, une pour chaque caractéristique évaluée, respectivement : prix, qualité, inertie et promotion. On notera le prix et la qualité du produit évalué p et q ainsi que le prix et la qualité du produit de référence p_{ref} et q_{ref}. Les équations 4.4 et 4.5 comparent le pack en cours d'évaluation, au pack de référence sur le prix et la qualité. L'idée de ces équations ainsi que l'utilisation d'un pack de référence et de l'aversion à la perte provient de (Cohen et al. 2020 ; Hardie et al. 1993 ; Kahneman et al. 1979 ; Tversky et al. 1991). L'équation 4.6 prend en compte l'inertie, on la modélisera de la même manière que (Bawa 1990). Pour représenter le phénomène d'inertie, nous considérons le nombre d'achats passés, noté nb_{achats}. L'évolution de cette variable influence la fonction d'achats, modélisée sous la forme d'une parabole orientée vers le bas. Dans un premier temps, la valeur de la fonction croît avec les achats successifs, traduisant l'effet d'inertie lié à l'habitude d'achat. Toutefois, au-delà d'un certain seuil, cette valeur décroît progressivement, reflétant un effet de lassitude associé à la répétition de l'achat du même produit. Finalement, la dernière fonction calcule l'impact d'une promotion. On pondère ensuite les résultats des quatre fonctions précédentes par les sensibilités de l'agent. Cette idée reprend les travaux de (Said et al. 2002). La somme obtenue nous indique le score du pack pour cet agent. Plus le score est élevé, plus le pack intéresse l'agent. [1]

$$U_1 = G_p(\beta(p - p_{ref})_+ + (p_{ref} - p)_+) \tag{4.4}$$

$$U_2 = G_q(\beta(q - q_{ref})_+ + (q_{ref} - q)_+) \tag{4.5}$$

$$U_3 = G_i(10 \times nb_{achat} - nb^2_{achat}) \tag{4.6}$$

$$U_4 = G_d \times D \tag{4.7}$$

$$U = \sum_{k=1}^{4} U_k * \gamma_{a,k} \tag{4.8}$$

On note que l'impact de la baisse du prix est calculée dans U_1 et non U_4. On modélise avec U_4 uniquement l'impact de la présence ou non d'une promotion.

1. $(x)_+$ représente le maximum entre 0 et x. nb_{achat} représente le nombre d'achats passé de ce pack par l'agent.

Après calcul, l'utilité associée à chaque produit par l'agent est convertie en une probabilité d'achat. Dans un premier temps, seules les utilités strictement positives sont retenues pour le tirage aléatoire, à l'exception du cas où toutes les utilités sont négatives, auquel cas seul la meilleure est considérée. Cette transformation correspond à une normalisation proportionnelle, où chaque utilité est rapportée à la somme des utilités positives.

Afin de capturer différents degrés de rationalité dans le processus de décision, il est possible d'introduire un paramètre de rationalité $R > 0$. Celui-ci agit comme un exposant sur les utilités, de telle sorte que la probabilité de sélection s'écrit :

$$P(i) = \frac{U_i^R}{\sum_j U_j^R} \tag{4.9}$$

Cette transformation est analogue à une loi logit généralisée (ou « softmax » en apprentissage automatique), couramment utilisée en économie comportementale et en modélisation multi-agents pour relier l'utilité perçue à un choix probabiliste (McFadden 1972).

— Lorsque $R = 1$, la probabilité $P(i)$ est proportionnel à U_i. Le R pouvant se simplifier, on a $P(i) = \frac{U_i}{\sum_j U_j}$.
— Lorsque R>1, les utilités les plus élevées sont amplifiées, ce qui augmente la probabilité d'un choix quasi-déterministe en faveur des options dominantes.
— À la limite $R \to +\infty$, le comportement de l'agent tend vers un choix déterministe correspondant au choix de l'utilité maximal soit $argmax(U)$.

Ainsi, ce paramétrage permet de passer en continu d'un comportement déterministe à un comportement stochastique.

Dans le cadre des simulations réalisées, le paramètre de rationalité R a été maintenu à sa valeur par défaut, $R = 1$. Ce choix permet de s'appuyer sur la normalisation proportionnelle classique, garantissant une interprétation simple et stable des résultats. Bien que l'impact de ce paramètre sur les dynamiques observées se soit révélé limité dans nos tests préliminaires, son exploration systématique constituerait une piste intéressante pour de futurs travaux si jamais un besoin de modularité supplémentaire est nécéssaire.

La quantité achetée

Le calcul de la quantité achetée est indépendant de la quantité interne à chaque pack. Ce calcul sert à connaître le nombre de packs achetés par l'agent. Si $P(p, Qte, Qa, D)$ est le pack choisi, et que Qte est 1, l'agent va alors grâce à la formule 4.10 calculer la quantité dont il a besoin, disons 3. L'agent achète alors 3 fois le même pack. $N(C_i, t)$ la quantité achetée au temps t de C_i.

$$Qt(P, t+1) = max(1, \lambda_{i,a} + N) \times S \tag{4.10}$$

$$N = \sum_{\tau=0}^{T} \frac{\lambda_{i,a} - N(C_i, t-\tau)}{T+1} \tag{4.11}$$

$$S = Sat(U(P) - U(P_{ref}, t)) \tag{4.12}$$

$$Sat(x) = \frac{Up}{1 + e^{-\frac{x}{\alpha_{sat}} + log(C-1)}} \tag{4.13}$$

Le calcul de la quantité est séparé en 2 parties : Une première liée au besoin $\lambda_{i,a}$ correspond à la partie de gauche de l'équation 4.10 (écart de $\lambda_{i,a}$ au T derniers pas de temps). Une seconde déterminée grâce au score du pack, définissant à quel point l'agent a trouvé un meilleur pack qu'à son habitude, traduit par la fonction de saturation. Celle-ci augmente ou réduit la quantité achetée en fonction du score du pack choisi P par rapport au pack de référence P_{ref}. Si le score de P est plus élevé l'agent achète en plus grande quantité sinon il en achète moins.

Pour résumer, la quantité achetée est fonction de l'écart du besoin $\lambda_{i,a}$ aux T derniers achats multiplié par la fonction de saturation.

Algorithm 3 Raisonnement

procedure RunOneTimeStep
  $t \leftarrow$ time step t
  for all agents $a \in A$ **do**
    $go\_store(a)$
    for all category $c_i \in C$ **do**
      $H_{(i,a,t-l_h)} \leftarrow None$ (delete)
      $prob \leftarrow B(c_i, t+1)$
      $x \sim U(0,1)$ ▷ Tirage aléatoire uniforme entre 0 et 1
      if $x > prob$ **then**
        $H_{(i,a,t)} \leftarrow (None, 0)$
      else
        for every pack $p \in c_i$ **do**
          $evaluate(p) \leftarrow U(prix, a)$
          l append $(score(p), prix)$
        end for
        $P_{i,a,t} \leftarrow Choose_p(l)$
        $q \leftarrow$ Qt($P_{i,a,t}$,ref)
        $H_{(i,a,t)} \leftarrow (P_{i,a,t}, q)$
      end if
    end for
  end for
end procedure

4.4.6 La dynamique du modèle

À chaque pas de temps, tous les agents vont au magasin et déterminent pour chaque catégorie $C_i \in C$ si celle-ci les intéresse ou non suivant une probabilité $B(C_i, t+1)$. Sur les catégories choisies à ce pas de temps par l'agent, celui-ci évalue tous les packs puis il choisit quel pack acheter pour cette catégorie et en quelle quantité. Ce choix suit le processus de décision interne à chaque agent décrit par l'algorithme 3.

La méthode `Choisir` choisit un pack en utilisant une loi de probabilité proportionnelle au score du pack. La méthode `Qt` utilise les équations 4.10 à 4.13 pour calculer la quantité que l'agent a achète.

4.5 Expérimentations

Les résultats de cette section sont présentés par (Jarod Vanderlynden et al. 2023 ; Jarod Vanderlynden et al. 2024a). Toutes nos expériences [2] dans cette section sont réalisées avec des paramètres d'environnement fixés pour permettre la comparaison. Les paramètres du modèle : G_p, G_q, G_i, G_d sont égaux à 1. Ce qui signifie que les effets du prix, de la qualité, de l'inertie et de la promotion ont un impact global similaire. On se place en quelque sorte dans un magasin de grande surface classique qui ne se démarque pas par ses prix bas ou sa bonne qualité. Enfin, le paramètre β est fixé à 0.5, Up à 3 et α_{sat} à 10. Sur une même expérience, les agents et packs ont les mêmes caractéristiques. Les procédures permettant de générer les agents et les packs sont des procédures stochastiques. Les agents sont catégorisés selon leurs sensibilités : c'est ce qui définit leur appartenance à un profil de consommateur. Toutes les expériences sont réalisées 20 fois pour plus de précision. Du fait de la génération stochastique des agents, une phase de stabilisation a lieu en début de simulation. En effet, les agents disposent d'un historique initial qui ne correspond pas forcément à leurs caractéristiques internes. Par exemple, un agent ayant une forte sensibilité à la qualité et ayant un historique composé de packs bon marché, mais de faible qualité. Au fil des achats, l'historique des agents se met à jour et se met à correspondre au mix historique/caractéristiques adéquat. Enfin, nous montrons que dans une même situation (agents et packs similaires), deux promotions identiques ont quasiment le même effet. Par la suite, nous allons tester différents scénarios dans lesquels nous modifions uniquement les caractéristiques des produits. Nous étudions une unique catégorie de produits ne possédant pas de caractéristiques particulières. L'objectif est de montrer que dans ces différents scénarios, le modèle reproduit des phénomènes intéressants et liés directement aux comportements des consommateurs.

Le modèle est capable de reproduire des phénomènes promotionnels macroscopiques classiques en marketing comme :

- L'augmentation du volume des ventes sur un pack en promotion ((Blattberg et al. 1995))
- La cannibalisation : baisse des ventes des packs en concurrence avec un pack en promotion.

2. Une feuille Jupyter : https://github.com/cristal-smac/retail permet de reproduire les expériences décrites ici.

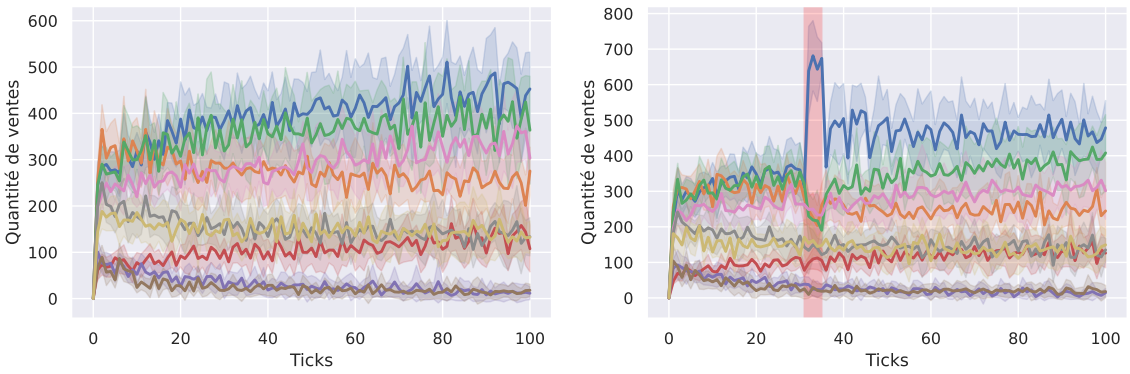


FIGURE 4.2 – *Quantité des ventes pour chacun des 9 packs en fonction du temps dans une simulation. À gauche sans promotion, à droite avec une promotion de 40% entre les pas de temps 30 et 34 représenté par le ruban rouge sur la figure.*

— La diminution de l'impact des promotions répétées sur un même pack.

Mais surtout, le modèle peut reproduire des phénomènes observables uniquement au niveau des agents, émergents, impossible à observer avec une approche qui ne serait pas centrée individu, comme :

— De multiples promotions successives changent à long terme le prix de références des agents.

— La guerre des prix est un phénomène avec des impacts macroscopique, mais aussi microscopique en changeant les perceptions que les consommateurs ont de certaines marques. Cette influence a aussi des effets sur la fidélisation des consommateurs.

— L'impact de la promotion sur l'acquisition et la fidélisation de nouveaux consommateurs, notamment selon les différents profils. Par exemple, un consommateur promophile changera régulièrement de pack si ceux-ci sont en promotions.

4.5.1 L'augmentation du volume des ventes

C'est le phénomène le plus classique, lors d'une promotion il y a une augmentation du volume des ventes du pack en promotion. Notre expérience consiste à comparer le volume des ventes d'un pack, en promotion puis sans promotion.

La figure 4.2 montre l'évolution du volume des ventes des neuf packs d'une catégorie observée. Sur une réduction de 40% du prix, nous observons en moyenne une augmentation de 50% du volume des ventes sur ce pack.

Ce phénomène d'augmentation du volume des ventes est observable peu importe la promotion ou le pack. On peut observer en figure 4.3 que les résultats diffèrent selon le niveau de promotion, mais génèrent à chaque fois un pic de ventes allant de 40 à 80% d'augmentation. La seule exception est lors d'une promotion à 5%, sur ce jeu de paramètres, le changement n'est pas assez conséquent pour influencer suffisamment les clients. Ces résultats donnent une élasticité prix du pack allant de -4 à -1.33. Cette élasticité est calculée en réalisant la moyenne des augmentations des ventes selon l'intensité du changement de prix. C'est le rapport modification des ventes/modification du prix. Une élasticité de -2 représente une augmentation du volume des ventes de 20% pour une réduction de prix

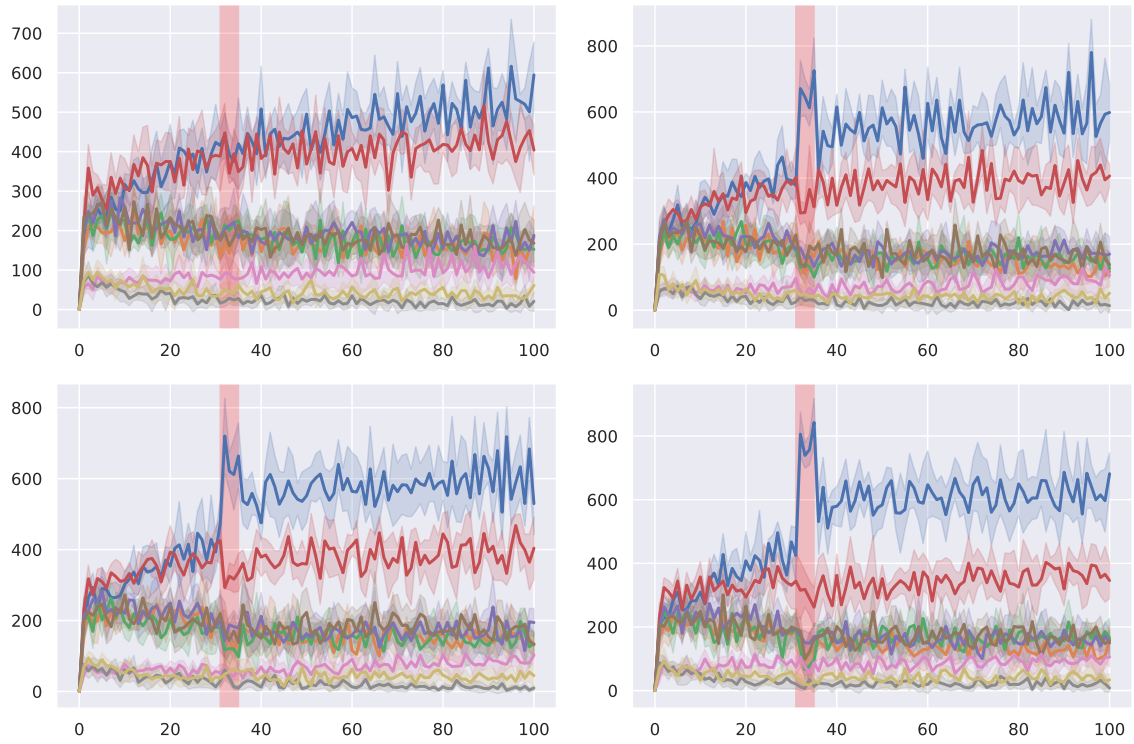


FIGURE 4.3 – *Quantité des ventes : chaque courbe sur chaque graphique représente chacun des 9 packs en fonction du temps dans une simulation. De gauche à droite, une promotion à 5%, 10%, 20% et 60% entre le pas de temps 30 et 34 représenté par le ruban rouge.*

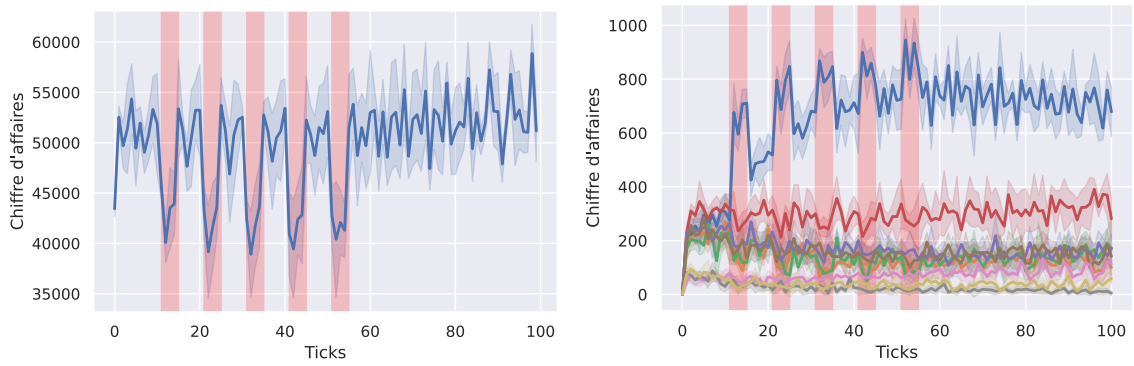


FIGURE 4.4 – *À gauche le chiffre d'affaires, à droite le nombre de ventes. Chaque courbe sur le nombre de ventes représente un pack. Les différents packs sont de la même catégorie. La même promotion sur le même pack est répétée plusieurs fois au temps 30 à 34, 40 à 44, ... jusqu'au temps 60 à 64, après, il n'y a plus de promotion.*

de 10%. L'élasticité est une mesure qui varie beaucoup selon le contexte et le produit étudié, nous avons donc choisi d'étudier la plage des élasticités que nous mesurons avec le modèle. Nos résultats sont du même ordre de grandeur que les élasticités des différents produits réellement observées par (Bijmolt et al. 2005 ; Hoch et al. 1995).

On observe que les résultats d'une promotion dans notre modèle sont très dépendants du contexte de la promotion. Ce contexte est défini par les paramètres du modèle, les agents et les packs. Par exemple :

— La proportion des différents profils d'agents.
— Les paramètres des packs et la similarité entre les packs.
— Le nombre de packs en concurrence (d'une même catégorie).

En effet, si le modèle dispose de 5 profils différents en proportion égale, on observe une augmentation de ventes de 95% lors de la promotion et de 54% après la promotion. Cependant, si le modèle contient 70% d'agents promophiles et 30% d'agents équilibrés, on observe une augmentation des ventes de 173% lors de la promotion et de 81% après la promotion.

Si on applique une promotion sur un pack d'une catégorie ne disposant que de deux packs similaires, on obtient une augmentation des ventes de 40% lors de la promotion et de 27% après la promotion. Si les deux packs sont bien différents, par exemple un pack peu cher de faible qualité et un pack très cher mais de haute qualité, on obtient une augmentation des ventes de 7.5% lors de la promotion et de 12.5% après la promotion si la promotion est sur le pack de faible qualité. Si la promotion est sur le pack de haute qualité, on obtient 31% d'augmentation des ventes lors de la promotion et 17% d'augmentation des ventes après la promotion.

Enfin, si on applique une promotion sur un pack parmi 9 on obtient une augmentation de ventes de 95% lors de la promotion et de 54% après la promotion. si on applique cette même promotion sur le même pack parmi 3 on obtient une augmentation de ventes de 31.5% lors de la promotion et de 15% après la promotion. Cet effet tout à fait naturel apparaît car un pack parmi neuf disposera d'un plus grand nombre de packs concurrents auxquels il est aisé de prendre des ventes. De plus la part de marché initiale est plus petite, nos mesures étant en pourcentage d'augmentation par rapport aux ventes initiales, il est logique d'y voir alors une augmentation de ce pourcentage malgré une augmentation du volume de vente qui peut parfois être plus faible en volume.

Nous montrons grâce à ces résultats qu'avec ce modèle il est possible de quantifier les effets des promotions et les relier à une situation particulière, un contexte, qui peut être réel. Les résultats en pourcentage sont visibles sur les tables 3.2, 3.4, **??**

4.5.2 Baisse du volume des ventes

La baisse des ventes des packs qui ne sont pas en promotion est un phénomène aussi couramment observé. On dit que le pack en promotion « cannibalise » les ventes des autres packs du même type. On peut observer cet effet sur la figure 4.2 : la baisse des ventes des packs ne bénéficiant pas de promotions va de 2 à 20 % et varie selon les packs et leur similarité avec le pack en promotion. Dans ces expériences, la baisse des ventes est

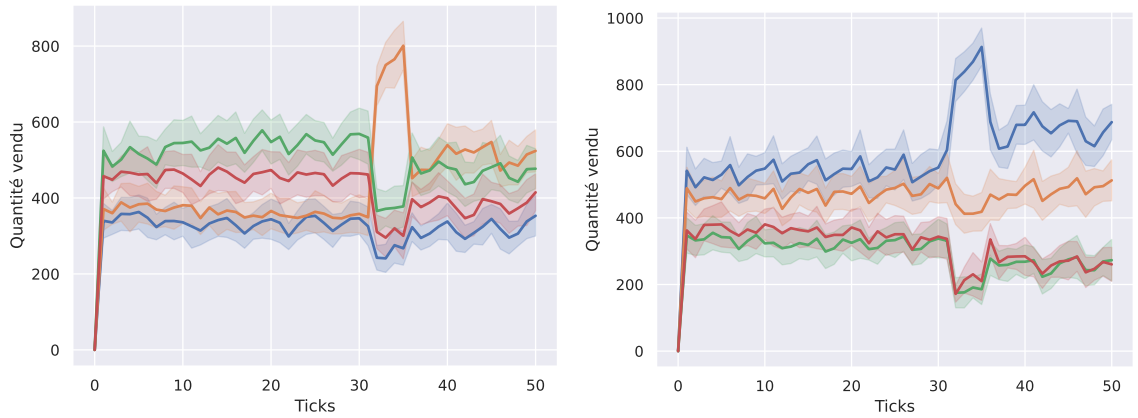


FIGURE 4.5 – *Tous les packs ont le même rapport qualité prix. En orange le pack le moins cher et de moins bonne qualité. En bleu le pack le plus cher et de meilleure qualité. En rouge et vert des packs de qualité et prix moyen. On observe que lorsque la promotion est sur le pack le moins cher (figure de gauche) les ventes du pack le plus cher diminue moins que les autres. Sur la figure de droite, lorsque le pack le plus cher est en promotion, les ventes du pack le moins cher diminuent moins que les autres.*

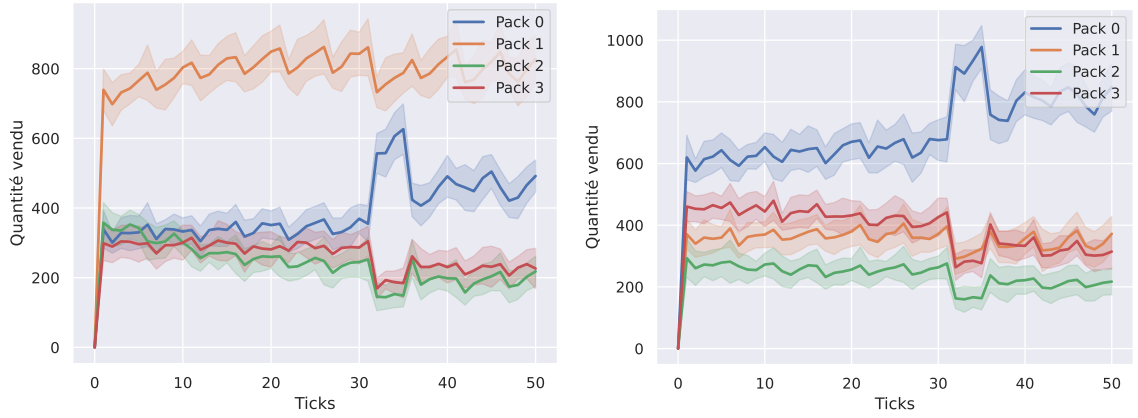


FIGURE 4.6 – *Tous les packs ont le même rapport qualité prix. En orange le pack le moins cher et de moins bonne qualité. En bleu le pack le plus cher et de meilleure qualité. En rouge et vert des packs de qualité et prix moyen. On observe que lorsque la population d'agents est plutôt orientée prix (figure de gauche) les ventes du pack le moins cher (en orange) diminuent moins que sur une simulation ou la population d'agents est plus équilibrée. Sur la figure de droite, durant une simulation lorsque la population est orientée qualité, les ventes du pack le moins cher diminuent plus que lors d'une simulation ou les profils d'agents sont plus équilibrés.*

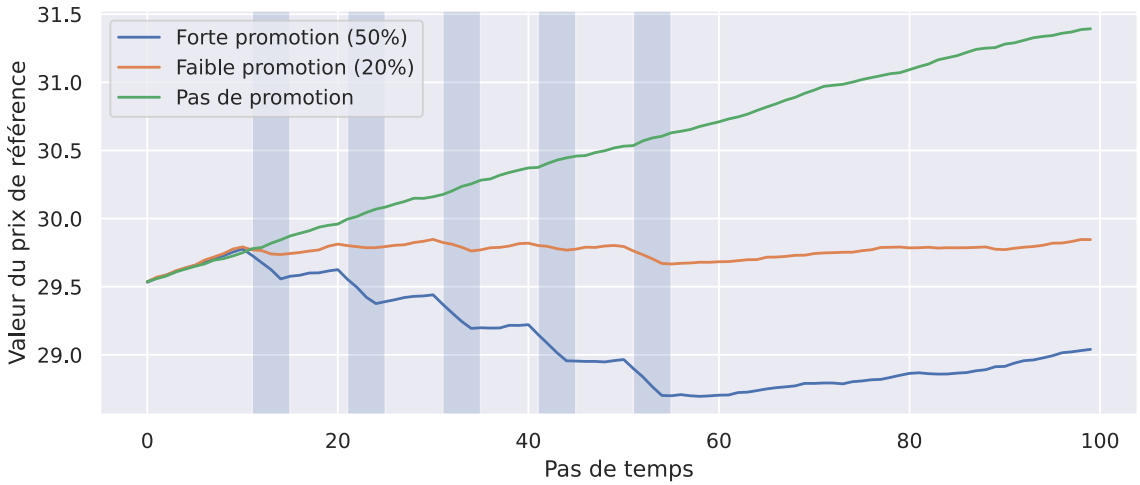


FIGURE 4.7 – *Le prix de référence moyen en fonction du temps. On observe une diminution du prix de référence moyen lors de promotions, ce qui a pour effet d'augmenter l'attractivité des packs les moins chers.*

pour au moins un pack toujours supérieur à 0%. Nous observons que les packs ne sont pas cannibalisés uniformément, c'est un phénomène couramment observé en marketing. Les ventes d'un produit de forte qualité ont tendance à mieux résister au phénomène de cannibalisation lorsque le produit en promotion est de faible qualité et vice-versa. En conclusion, on observe sur la figure 4.5 que plus un pack est similaire au pack en promotion plus ses ventes baissent. En effet, sur la figure de gauche le pack en promotion est de faible qualité et peu cher. Les ventes du pack 0, pack le plus cher mais le plus qualitatif, représenté par la courbe bleue baissent de 20% tandis que les ventes des packs 1 et 2 en rouge et vert (qualité et prix moyen) baissent de 45%. La seconde figure représente le même phénomène mais avec le pack 0, le plus cher, en promotion (courbe bleue). Le pack le moins cher, en orange, voit ses ventes être moins impactées que pour les autres packs. Enfin ces différents effets varient aussi en fonction de la population d'agent. Plus, il y a d'agents promophiles plus la cannibalisation est forte et uniforme. Plus il y a d'agent ayant un profil « prix » plus la promotion à un impact asymétrique sur les packs, de même avec la qualité. Cet impact asymétrique est observable sur la figure 4.6.

4.5.3 Impact des promotions répétées

Des promotions répétées rapidement diminuent le pic d'augmentation des ventes généré chez les clients lors d'une promotion (BLATTBERG et al. 1995). Plus les clients sont habitués aux promotions, moins ces clients changent de pack lors d'une promotion. À l'inverse, si les promotions sont exceptionnelles, alors le comportement des clients change. Ils voudront davantage profiter de la promotion, c'est ce qu'on observe en figure 4.4. Sur chaque simulation, nous obtenons une augmentation du nombre de ventes de 140% sur la première promotion pour atteindre une augmentation de 30% sur la dernière promotion. Enfin, nous avons pu observer un phénomène de diminution du prix du pack de référence des agents, ce qui a pour effet de diminuer leurs fonctions d'utilités. Cette diminution affecte la quantité que les agents achètent et avantage les packs les moins chers.

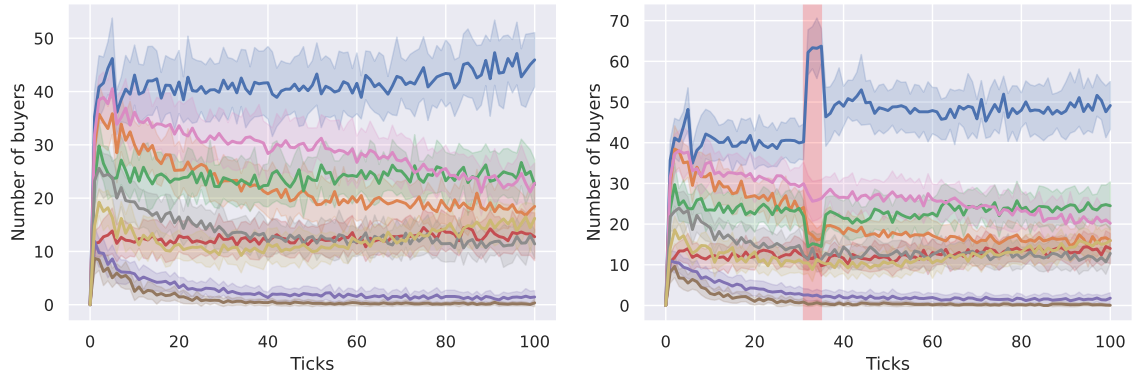


FIGURE 4.8 – *Nombre d'acheteurs pour chaque pack en fonction du temps dans une simulation. À gauche sans promotion, à droite avec une promotion de 40% entre les pas de temps 30 et 34.*

Profil d'agent	Évolution des ventes pendant la promotion	Évolution des ventes après la promotion
Prix	88,10%	-9,8%
Qualité	1,4%	-9,4%
Promophile	132,6%	79,1%
Inertiels	0,4%	-8,7%
Equilibré	57%	35,2%

TABLE 4.1 – *Compare la période après la promotion avec la période avant la promotion de 40% du pack P_0.*

4.5.4 Fidélisation des clients

Sur la figure 4.8 nous montrons qu'une promotion entraîne sur le pack en promotion une augmentation du nombre de clients suivie d'une baisse moins forte au moment de l'arrêt de la promotion, correspondant à une fidélisation. La figure 4.9 représente le nombre d'agents pour chaque profil ayant acheté le pack P_0 et le pack P_3. Le pack P_0 est le seul à recevoir une promotion. Ainsi, on observe l'impact de la promotion selon les différents profils d'agents sur P_0. Nous montrons que lors d'une promotion, il y a une augmentation du nombre d'acheteurs (50% en moyenne) sur les profils de consommateurs promophiles uniquement (sensibles à la publicité). Nous en concluons qu'une partie des agents, notamment les plus promophiles, ont acheté le pack P_0 plutôt qu'un autre, parce que le pack était en promotion. On note aussi que les agents sensibles à la qualité achètent peu pour le pack P_0 et inversement ces agents achètent beaucoup pour le pack P_3 car le pack P_3 est plus cher, mais de meilleure qualité. La promotion a très peu d'impact sur cet effet. Ce phénomène s'observe aussi sur les tables 4.1,4.2,4.3 qui montrent, l'évolution selon les profils d'agents, des ventes pendant et après la promotion, comparées aux ventes avant la promotion.

Profil d'agent	Évolution des ventes pendant la promotion	Évolution des ventes après la promotion
Prix	-48,32%	-12.25%
Qualité	-76.92%	-23.67%
Promophile	-92.81%	-71.39%
Inertiels	-47.40%	-63.74%
Equilibré	-63.84%	-60.23%

TABLE 4.2 – *Compare la période après la promotion avec la période avant la promotion de 40% du pack P_1.*

Profil d'agent	Évolution des ventes pendant la promotion	Évolution des ventes après la promotion
Prix	-8%	-3%
Qualité	-82%	-42%
Promophile	55%	22%
Inertiels	-10%	-22%
Equilibré	18%	6%

TABLE 4.3 – *Même expérience que pour la table 4.1 mais avec 3 packs dans la catégorie au lieu de 9 packs.*

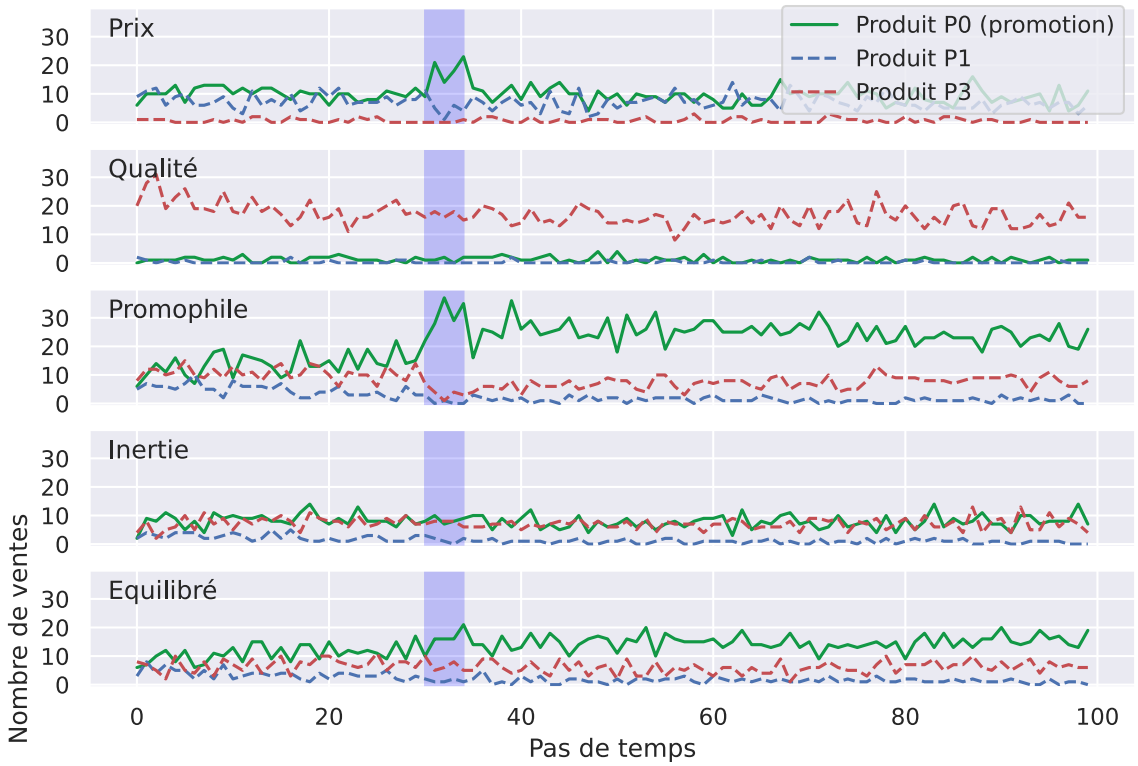


FIGURE 4.9 – *Évolution du nombre d'acheteurs lors d'une promotion de 40% sur P_0 en $30 \leq t \leq 40$. On observe une augmentation durable des ventes sur le pack P_0 chez les promophiles et une augmentation temporaire chez les agents portés sur le prix*

4.5.5 Impact de la guerre des prix

Afin d'explorer les réactions du modèle face à une guerre des prix, nous réalisons une simulation plus simple dans laquelle il n'y a que deux packs correspondant à la figure 4.10. Dans cette situation, nous notons une baisse du chiffre d'affaires dès qu'une réduction de prix a lieu, même s'il n'y a une baisse de prix que d'un seul pack. De plus, nous notons que lors d'une guerre des prix, le pack qui baisse le plus son prix récupère la plus grande part de marché, mais les deux packs perdent en chiffre d'affaires. On observe une phase de stabilisation du modèle. La mesure des "favoris", définie comme le pack le plus acheté dans l'historique d'un agent, est au départ influencée par un historique aléatoire non lié aux caractéristiques des agents. Il faut donc un certain temps, variable selon les tirages initiaux, pour que les agents qui devraient préférer le pack bleu en raison de leurs caractéristiques voient effectivement leur "favori" basculer du pack orange vers le pack bleu, et inversement.

4.6 Analyse de l'impact des paramètres

Nous avons réalisé une étude d'impact des paramètres d'environnement du modèle. Ces paramètres sont décrits en section 4.4. Ce sont les paramètres G_p, G_q, G_i, G_d, α_{sat}, C β. Bien évidemment, si nous voulons tester toutes les combinaisons possibles, il y a une explosion combinatoire. C'est pour cela que nous nous limitons à ces 7 paramètres et que pour chaque paramètre, nous ne testerons que 5 valeurs différentes afin d'essayer

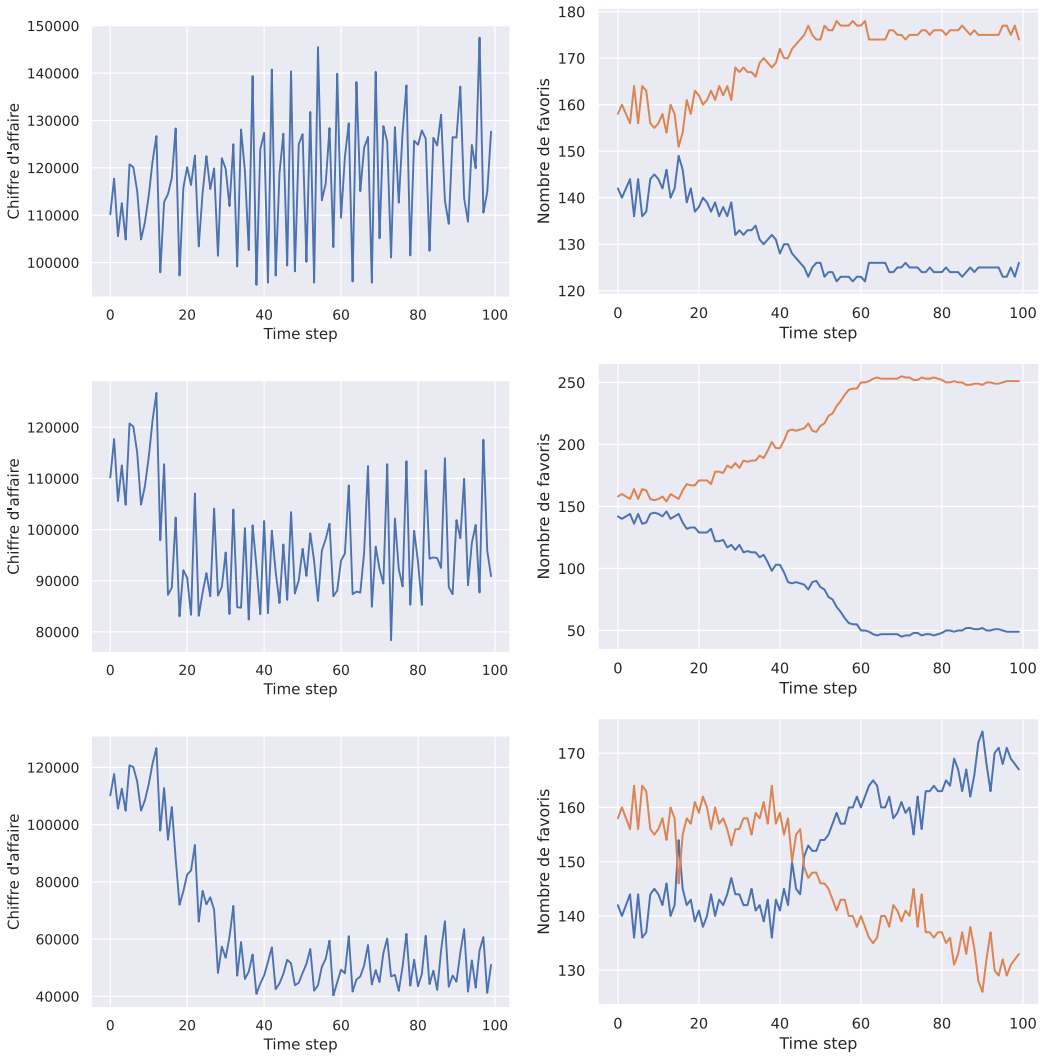


FIGURE 4.10 – *À gauche le chiffre d'affaires total de la catégorie, à droite pour chaque pack le nombre d'agents ayant ce produit en « favoris » (le plus acheté par l'agent dans l'historique). De haut en bas, une simulation sans changement de prix, une où seule une marque baisse son prix de 20% à partir de* $t = 15$ *et une où les 2 packs diminuent leurs prix à certains pas de temps. On observe une phase de stabilisation du modèle sur les premiers pas de temps. Dans la dernière série de figures, le pack bleu, dernier à réduire son prix, finit par remporter la guerre des prix : ses ventes se stabilisent à un niveau supérieur à celui de son concurrent.*

Paramètres	G_p	G_q	G_i	G_d	α_{sat}	β	Up
Effet A	0.63	0.32	1.38	0.55	0.04	0.25	0.43
Effet B	0.10	0.06	0.10	0.79	0.03	0.08	0.14
Effet C	0.07	0.19	0.53	0.13	0.03	0.01	0.05
Effet D	0.22	0.36	0.07	0.23	0.07	0.05	0.50
Effet E	0.25	0.07	0.02	0.03	0.04	0.03	0.15

TABLE 4.4 – *Score d'impact de chaque paramètre (ligne) sur les différents effets (colonnes) de la simulation. Le score calculé représente la variance expliquée par le paramètre de l'effet observée. Plus le score est élevé, plus l'impact du paramètre est important. Les scores les plus impactants sont en rouge.*

d'observer l'impact de chaque paramètre. Cela fait un total de 7^5 environ 80 000 ensembles de paramètres. De plus, étant donné que notre modèle est stochastique, nous avons réalisé pour chaque ensemble de paramètres 10 simulations, qui ont été ensuite moyennés, ce qui nous donne un total d'environ 800 000 simulations.

Nous souhaitons observer comment chaque paramètre fait évoluer nos résultats. Nous faisons évoluer chacun des paramètres et nous mesurons sur chaque simulation différents effets marketing. Les cinq effets marketing étudiés sont :

- Effet A : Augmentation des ventes du pack en promotion pendant la période de promotion.
- Effet B : Augmentation des ventes du pack en promotion après la période de promotion (fidélisation).
- Effet C : Perte de ventes d'un autre pack pendant la période de promotion (Cannibalisation).
- Effet D : Perte de ventes d'un autre pack après la période de promotion (perte de fidélisation).
- Effet E : Part de marché du pack en promotion lors du dernier pas de temps.

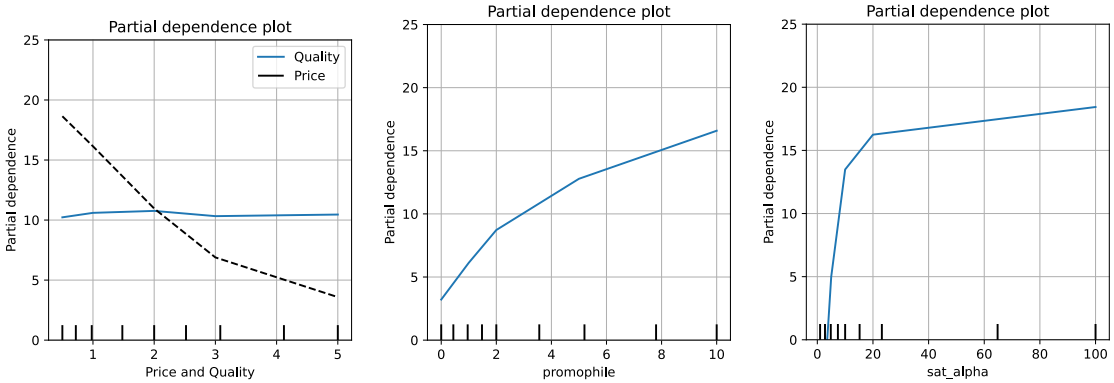
. Nous comparons l'évolution de chacun des effets avec ceux des simulations effectuées sur la valeur de paramètre précédent. Par exemple, nous réalisons toutes les simulations ou $G_p = 0.5$, cela représente 7^4 simulations sur lesquels cinq effets sont à chaque fois mesurés. La distribution ainsi obtenue est ensuite comparée à celle obtenue lorsque $G_p = 1$. Pour comparer ces distributions, nous utiliserons la mesure de divergence de Kullback-Leibler (KL divergence) :

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) log(\frac{P(x)}{Q(x)}) \tag{4.14}$$

Une autre technique de visualisation que nous avons employée est la création de courbes de dépendance partielle 4.11. Ces courbes sont obtenues en utilisant les moyennes de chaque distribution et un modèle de régression pour mesurer l'impact d'un paramètre sur un phénomène macroscopique. En analysant ces tracés, nous obtenons des informations sur la relation entre un paramètre et l'effet macroscopique observé.

Nous observons que le paramètre de prix G_p a un impact négatif sur l'augmentation des

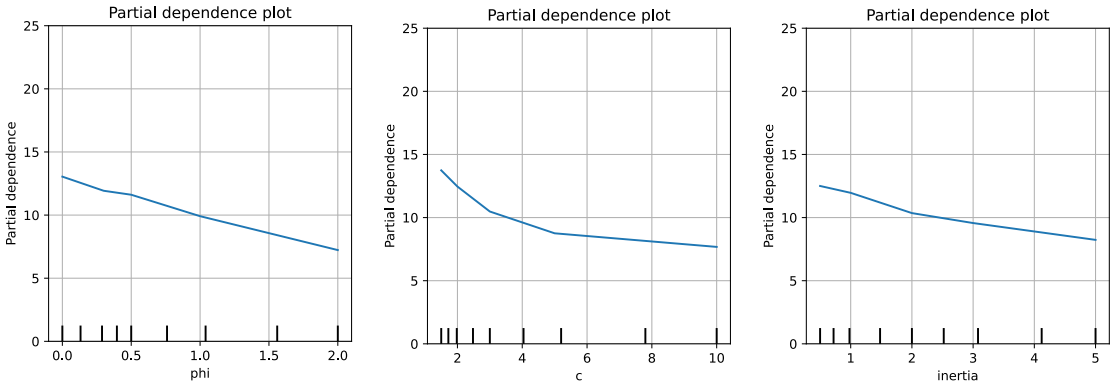
FIGURE 4.11 – *L'influence des paramètres globaux sur différents résultats de simulations.*



(a) *Dépendance entre l'effet d'augmentation des ventes et les paramètres G_p et G_q.*

(b) *Dépendance entre l'effet d'augmentation des ventes et le paramètre G_d.*

(c) *Dépendance entre l'effet d'augmentation des ventes et le paramètre α_{sat}.*



(d) *Dépendance entre l'effet d'augmentation des ventes et le paramètre β.*

(e) *Dépendance entre l'effet d'augmentation des ventes et le paramètre Up.*

(f) *Dépendance entre l'effet d'augmentation des ventes et le paramètre G_i.*

ventes de produit. Ce qui se produit, c'est qu'une augmentation de ce paramètre rend le produit en promotion initialement plus attractif, boostant ainsi sa part de marché avant la promotion. En conséquence, il y a moins de parts de marché disponibles à capturer pendant et après la promotion, ce qui entraîne une diminution de l'augmentation des ventes pendant et après la promotion. Le paramètre de qualité G_q a peu d'impact sur ce phénomène. Le paramètre G_d exerce une influence prévisible, car il régit l'intensité de la promotion. Avec son augmentation, on observe une hausse correspondante des ventes supplémentaires réalisées pendant les périodes promotionnelles. De plus, nous observons que lorsque le paramètre α_{sat} est très petit, le modèle s'écarte du comportement souhaité. En effet, dans de tels cas, les promotions entraînent une baisse des ventes. De même, le paramètre β, qui régule le degré d'aversion à la perte, produit également l'effet attendu. Plus l'aversion à la perte est forte, plus il est difficile d'attirer de nouveaux clients, malgré les efforts promotionnels. Un schéma comparable est observé avec le paramètre G_i, qui régit l'inertie. Une valeur plus élevée pour ce paramètre correspond à une probabilité moindre que les clients modifient leurs habitudes et son impact sur l'augmentation des ventes pendant la période de promotion est similaire au paramètre β.

Ces résultats sont présentés dans un contexte spécifique, car à la fois les produits et les agents, ainsi que leurs paramètres, sont fixes et ne changent pas d'une simulation à l'autre.

Cependant, cette méthode pourrait être facilement appliquée dans un contexte différent. De plus, ces résultats peuvent être exploités à des fins de calibration en évaluant l'impact de chaque paramètre sur les phénomènes macroscopiques. Par exemple, s'il est nécessaire de calibrer le modèle sur un phénomène macroscopique précis, nos résultats indiquent quels paramètres doivent être modifiés en conséquence. Ainsi, il serait possible de calibrer chaque phénomène en fonction d'un nombre réduit de paramètres et d'éviter une explosion combinatoire. Par exemple, nos résultats montrent que, dans ce contexte spécifique, lors de l'augmentation des ventes d'un produit grâce aux promotions, le paramètre de qualité a peu d'influence et peut-être négligé lors de la calibration de ce phénomène. En fin de compte, ces résultats démontrent que chaque paramètre du modèle est significatif et contribue à la modularité sans altérer les phénomènes observés du modèle.

4.7 Synthèse

Dans ce chapitre, nous montrons dans la partie 4.5 que l'approche centrée individus proposée dans la partie 4.4 est capable de reproduire des phénomènes émergents connus en marketing comme l'augmentation du volume des ventes, la « cannibalisation » lié à la concurrence ou encore les changements chez les clients que provoque la répétition rapide de promotions. De plus, l'approche centrée individus permet de montrer des phénomènes qui ne sont observables qu'au niveau des individus comme la fidélisation lors d'une promotion ou les effets des guerres de prix directement sur les consommateurs. L'importance de ces phénomènes montre qu'une utilisation des systèmes multi-agents peut apporter des informations sur des phénomènes impossibles à traiter avec des méthodes agrégés. Conformément à nos hypothèses, le modèle et nos résultats s'appliquent principalement aux biens de consommation courants. D'autres hypothèses permettraient d'adapter le modèle à d'autres types de biens.

Afin d'approfondir le modèle, il est possible d'ajouter un système d'influence sociale similaire à ceux décrits dans la partie 4.3 qui permettrait aux agents d'échanger et d'interagir entre eux afin de s'influencer. Dans un tel système, une discussion est à prévoir sur les moyens d'influence des agents (système de graphe par exemple) et de leurs effets, est-il pertinent de faire varier des sensibilités personnelles ? De plus, il est possible de ne donner aux agents qu'une connaissance partielle des packs, ainsi les agents devraient découvrir eux-mêmes les packs qu'ils ne connaissent pas ou être influencé socialement. Enfin, il serait intéressant d'étudier les effets sur les promotions de la similarité et de la concurrence entre produits.

Finalement, le modèle proposé dispose de la possibilité d'intégrer facilement des données réelles (via l'historique) ce qui permettrait d'améliorer le réalisme, mais surtout d'appliquer le modèle à des scénarios concrets. Cette application permettrait de pousser la calibration des agents et des paramètres du modèle sur des données et des phénomènes précis. Nous justifions son adaptabilité aux données grâce aux paramètres globaux intégrés au modèle. Une étude de l'impact des différents paramètres sur le modèle est en cours.

Chapitre 5

Conclusion

Cette thèse cherche à comprendre et reproduire les comportements des consommateurs à travers la modélisation multi-agents. Trois modèles complémentaires ont été proposés, chacun explorant une dimension spécifique du processus d'achat : la construction d'une population synthétique, la décision d'achat influencée par l'aversion à la perte, et la dynamique des achats récurrents. Ensemble, ils constituent une approche modulaire du comportement consommateur, à la fois explicative et prédictive.

Les figures 5.1, 5.2, 5.3, 5.4 illustrent les différentes étapes d'un parcours d'achat simplifié. Chacune d'elles est associée à une phase spécifique du processus et met en évidence les influences qui s'exercent à ce stade. Ces influences peuvent être de nature endogène (liées aux caractéristiques propres au consommateur) ou exogène (provenant de l'environnement ou du marché). Dans nos travaux, nous nous sommes principalement concentrés sur les phénomènes surlignés en vert, correspondant pour l'essentiel à des facteurs endogènes. Le comportement de nos agents se révèle ainsi particulièrement réactif face aux promotions. Ces représentations visent à clarifier notre positionnement scientifique en distinguant les dimensions étudiées, celles déjà couvertes par la littérature, et celles restées hors du périmètre de cette thèse. Par exemple, la phase de comparaison (figure 5.4) est intégralement couverte par notre modèle de choix, tandis que la phase de besoin (figure 5.2) est largement documentée par les travaux existants mobilisés dans cette thèse. En revanche, les phases de recherche et d'achat (figure 5.3) soulèvent encore des questions non résolues, en raison de contraintes de temps ou d'un manque de ressources empiriques disponibles. Certaines influences, telles que l'impact des avis clients ou des comparateurs en ligne, sont bien établies qualitativement. Toutefois, leur quantification demeure complexe et nécessiterait la mobilisation de données massives afin d'évaluer précisément la force de ces effets et leur rôle dans la formation du comportement des consommateurs.

Le premier modèle permet de reconstruire une population synthétique d'acheteurs potentiels répartis sur plusieurs catégories de produits. Cette étape est essentielle, car elle pose les bases d'un environnement artificiel crédible sur lequel des scénarios d'achat ont été simulés.

Le deuxième modèle représente le choix d'achat à travers le prisme de l'aversion à la perte.

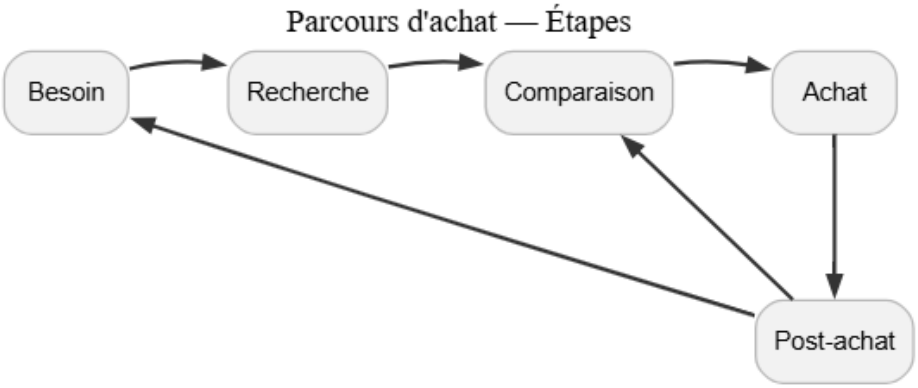


FIGURE 5.1 – *Schéma simplifié du processus d'achat d'un client. La phase de post-achat regroupe l'ensemble de l'expérience vécue pendant et après l'acquisition, laquelle influence ultérieurement le comportement de l'individu ainsi que celui de son entourage. Cette dimension n'est pas abordée dans la thèse, car elle a déjà fait l'objet de nombreuses analyses dans la littérature.*

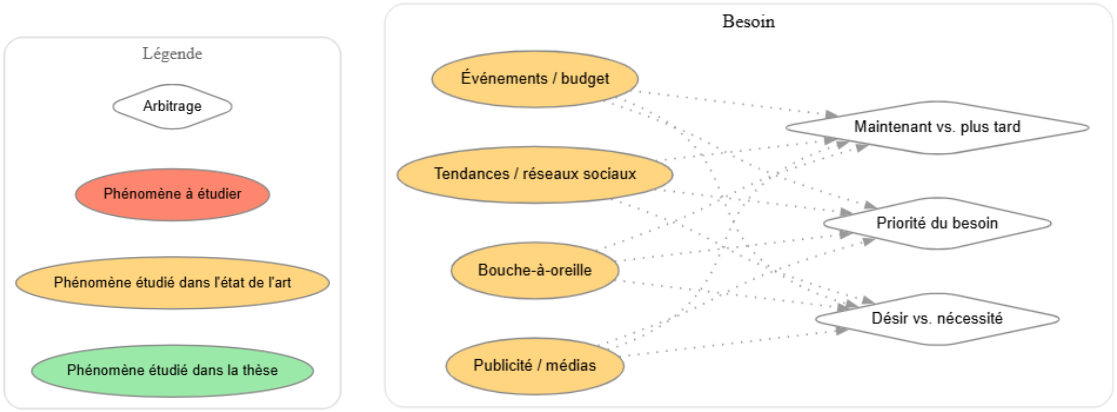


FIGURE 5.2 – *Schéma des influences du besoin.*

Nous montrons que cette approche permettait de reproduire des faits stylisés complexes observés dans le domaine du marketing, en particulier les effets de contexte. En renforçant le lien entre aversion à la perte et prise de décision, ce modèle offre une explication robuste des comportements observés dans les données réelles. Sa capacité à reproduire des comportements d'achats réels, après calibration, confirme la pertinence de cette modélisation. De plus, nous montrons que ce modèle peut être utilisé dans différents contextes de prise de décision : lors de l'introduction d'un nouveau produit, pour analyser les stratégies de modifications des prix d'entreprises concurrentes ou encore coté distributeur pour mettre en avant les produits les plus rentables.

Le troisième modèle se concentre sur les achats récurrents, influencés à la fois par les habitudes des consommateurs et par les promotions. Les résultats obtenus mettent en évidence des dynamiques cohérentes avec les régularités empiriques en marketing, confirmant que cette approche est adaptée pour étudier les comportements de fidélisation. Enfin, ce modèle complète le modèle précédent. Il montre qu'il est possible d'intégrer la modélisation du choix des consommateurs dans des contextes variés en ajoutant plus de facteurs internes et externes aux agents.

Ces trois modèles s'inscrivent dans une continuité scientifique et méthodologique : le premier modèle établit la base en générant une population synthétique représentative des acheteurs potentiels ; le second affine cette base en décrivant les mécanismes de décision individuels à travers l'aversion à la perte ; enfin, le troisième étend cette logique en intégrant la dimension temporelle, en observant comment les comportements d'achat évoluent et se stabilisent dans le temps. Cette articulation progressive permet de passer d'une représentation statique du consommateur à une modélisation dynamique de ses comportements, offrant ainsi un cadre intégré où chaque modèle prépare et enrichit le suivant. L'ensemble constitue une architecture cohérente, apte à simuler différentes facettes du processus d'achat tout en préservant la cohérence globale des mécanismes étudiés.

Ces trois contributions illustrent la complexité du comportement des clients. Par exemple, le modèle de choix présenté au chapitre 3 montre que les décisions d'achat ne dépendent pas d'un seul facteur, mais d'un ensemble d'influences contextuelles et psychologiques. Cette complexité constitue un véritable défi méthodologique : plus un modèle intègre d'interactions et de paramètres, plus sa validation devient délicate. En effet, l'espace des paramètres croît de manière exponentielle, rendant la calibration difficile et risquant d'affaiblir la validité du modèle. De plus, afin d'assurer que le modèle ne souffre pas de surapprentissage lors de la calibration, il est nécessaire de disposer d'autant de phénomènes observés que de paramètres estimés. L'enjeu est donc de concevoir des modèles parcimonieux, qui saisissent l'essentiel des mécanismes en jeu tout en demeurant opérationnels. En complément des étapes de calibration, une attention particulière a été portée à la validation empirique des modèles. Chaque modèle a été confronté, dans la mesure du possible, à des données réelles issues d'enquêtes ou de bases de transactions, afin d'évaluer sa capacité à reproduire des comportements observés. Des tests de robustesse ont également été menés pour vérifier la stabilité des résultats face à des variations de paramètres ou de contextes. Ces démarches renforcent la crédibilité et la portée opérationnelle des modèles proposés.

Au-delà de leur valeur prédictive, les SMA possèdent également une dimension explicative : ils permettent de mieux comprendre les comportements en simulant des dynamiques collectives à partir de règles individuelles simples. C'est notamment ce qui motive l'utilisation des modèles de seuil pour étudier la diffusion du bouche-à-oreille, un phénomène déjà largement exploré dans la littérature. À l'inverse, les approches épidémiologiques par contagion se révèlent moins adaptées à ce type de problématique.

Ainsi, cette thèse met en évidence l'intérêt d'une approche modulaire et progressive : chaque modèle apporte un élément spécifique à la compréhension du comportement des consommateurs. À terme, on peut imaginer que plusieurs SMA interconnectés, chacun représentant un aspect du processus d'achat (choix, récurrence, influence sociale, etc.), puissent être intégrés pour former un système global. Un tel cadre offrirait un outil puissant pour analyser, expliquer et prédire les comportements des consommateurs de manière plus complète.

Au-delà de l'intérêt scientifique et méthodologique, ces travaux offrent également des perspectives concrètes pour la stratégie marketing. En simulant le comportement des

FIGURE 5.3 – *Schéma des influences du processus d'achats dans les étapes de Recherche et d'achat. Ces diagrammes sont regroupés car interconnectés.*

consommateurs dans des environnements variés, les modèles développés permettent aux acteurs du marketing d'anticiper les réactions face à des politiques promotionnelles, de tester virtuellement différentes stratégies de prix, ou encore d'analyser les leviers de fidélisation. Ils constituent ainsi des outils d'aide à la décision, capables de relier la compréhension théorique du comportement d'achat à des applications opérationnelles dans la gestion et la planification marketing.

Ces travaux ouvrent la voie à une modélisation intégrée du comportement d'achat, capable d'articuler choix individuel, influences sociales et stratégies concurrentielles, au service d'une meilleure compréhension et anticipation des dynamiques de marché.

5.1 Perspectives

Les travaux présentés dans cette thèse ouvrent plusieurs pistes de recherche prometteuses. Bien que notre démarche porte principalement sur les dimensions endogènes du comportement des consommateurs et sur les promotions, de nombreux prolongements peuvent être envisagés afin d'enrichir et d'approfondir la compréhension des dynamiques d'achat.

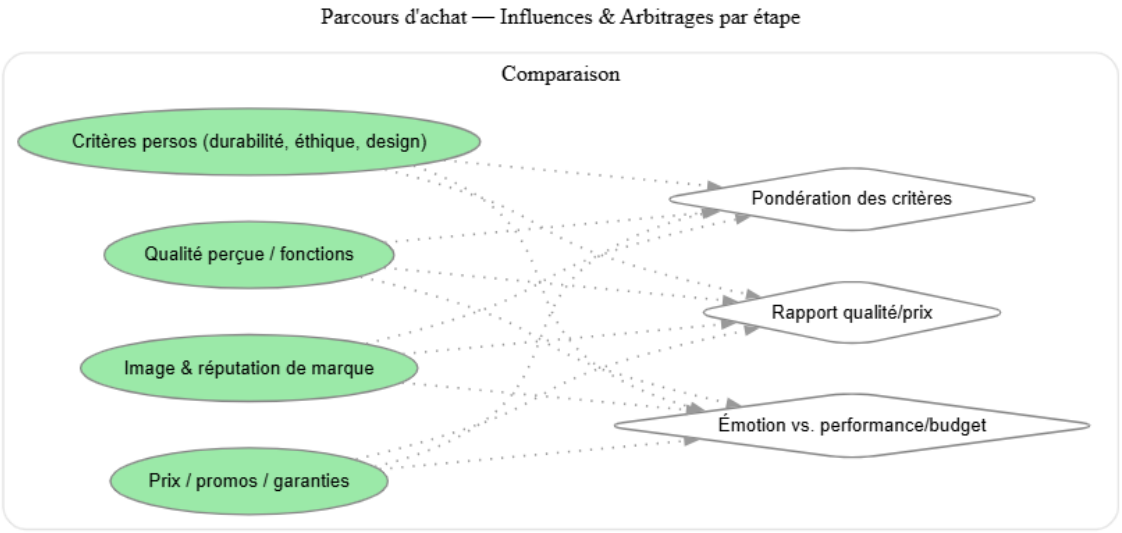


FIGURE 5.4 – *Schéma des influences du processus d'achats dans l'étape de comparaison*

5.1.1 À court terme : interactions sociales et spatialisation

Un premier axe concerne l'intégration des interactions sociales entre agents. Cet axe est envisageable à court terme car il s'appuie sur des modèles déjà développés : il s'agit principalement d'ajouter des mécanismes d'influence entre agents, sans refondre l'architecture existante. Les modèles développés jusqu'ici considèrent les individus comme largement indépendants, alors que les comportements de consommation sont fortement influencés par les réseaux sociaux, le bouche-à-oreille ou la diffusion d'opinions. L'introduction de mécanismes d'influence qu'ils soient basés sur des seuils, des structures de voisinage, ou des graphes dynamiques, permettrait de mieux représenter la propagation de l'information, des innovations ou des modes de consommation. Cette extension s'inscrirait dans la continuité de la littérature sur la diffusion et la contagion, tout en renforçant la dimension explicative des modèles développés.

Un second axe de recherche réside dans la prise en compte du mouvement et de la spatialisation des agents. La prise en compte de la dimension spatiale peut également être mise en œuvre rapidement, les outils de simulation et les données géographiques ou numériques étant déjà disponibles pour enrichir les modèles actuels. La modélisation de la mobilité des consommateurs, que ce soit dans un espace physique (déplacement entre zones commerciales) ou virtuel (navigation entre plateformes en ligne), offrirait un cadre plus réaliste pour étudier certains comportements d'achat. L'intégration de cette dimension spatiale permettrait d'examiner les effets de proximité, d'accessibilité ou de concurrence locale entre points de vente et produits.

5.1.2 À moyen terme : dynamiques concurrentielles

À moyen terme, une perspective importante réside dans la modélisation conjointe des interactions entre consommateurs et entreprises au sein d'un environnement concurrentiel évolutif. Ce développement relève plutôt du moyen terme, car il implique de faire co-évoluer plusieurs types d'agents (consommateurs et entreprises) et d'introduire des

boucles de rétroaction complexes nécessitant une calibration et une validation plus approfondies. Ce travail a déjà amorcé l'étude de l'environnement concurrentiel à travers la modélisation de méta-agents représentant les entreprises, capables d'ajuster leurs stratégies tarifaires en fonction des réactions du marché. Les simulations menées autour de l'introduction d'un nouveau produit ont mis en évidence des dynamiques caractéristiques d'un marché compétitif, telles que l'apparition de guerres de prix ou la stabilisation des marges à différents seuils de réaction. Ces premiers résultats ouvrent des perspectives vers une modélisation plus complète des interactions concurrentielles, où les entreprises et les consommateurs co-évoluent au sein d'un même système. À terme, il s'agirait de représenter explicitement la diversité des stratégies d'acteurs (positionnement, innovation, communication, différenciation) et d'étudier leurs effets sur la structure du marché. Un tel cadre permettrait d'analyser les équilibres émergents entre firmes et d'évaluer l'impact de politiques économiques, de régulations ou de chocs exogènes sur la stabilité et la performance du système concurrentiel.

5.1.3 À long terme : durabilité et écosystème intégré

Enfin, dans un contexte marqué par la transition écologique et les mutations des modes de consommation, une perspective importante consisterait à intégrer les dimensions de durabilité et de responsabilité sociale dans les modèles de comportement. Cette perspective s'inscrit dans le long terme, car elle suppose une évolution conceptuelle majeure vers une modélisation intégrée et systémique, mobilisant à la fois de nouvelles données, des dimensions sociales et environnementales, et une interconnexion complète entre modèles. Les agents pourraient être dotés de préférences évolutives reflétant des considérations environnementales, éthiques ou sociales. L'étude des interactions entre motivations économiques et valeurs sociétales offrirait une compréhension plus complète des compromis réalisés par les consommateurs et de la manière dont les politiques publiques ou les initiatives privées peuvent encourager ces transitions.

Finalement, à plus long terme, l'ambition pourrait être de construire un écosystème de modèles interconnectés, représentant les différentes composantes du processus d'achat, du besoin initial à la récurrence, en passant par l'influence sociale et la compétition entre marques. Un tel cadre intégré offrirait un outil puissant pour analyser conjointement les comportements individuels et les dynamiques de marché, ouvrant la voie à des applications opérationnelles en marketing, en économie comportementale ou en politique publique.

Ainsi, les perspectives ouvertes par ce travail ne se limitent pas à la simple amélioration des modèles existants : elles dessinent les contours d'une approche intégrée, évolutive et empirique de la modélisation du comportement des consommateurs, à la croisée des sciences sociales, des données et de la simulation multi-agents.

Annexe A

Programmation d'un SMA par calcul matriciel

Dans cette annexe, nous présentons une approche de programmation de différents modèles et processus itératifs reposant sur le calcul matriciel. L'usage de Python *et* Jupyter Notebook *constitue un atout pour le prototypage rapide, en facilitant l'expérimentation et la visualisation des résultats. Les bibliothèques telles que* NumPy *assurent des calculs matriciels à grande vitesse, tandis que des frameworks plus avancés comme* TensorFlow *ou* PyTorch *offrent la possibilité d'exploiter pleinement le calcul vectorisé et l'accélération matérielle via GPU. Cette approche est similaire à ce qui est proposé par* Hermellin 2016*, cependant l'utilisation du Python permet une plus grande flexibilité sur certains aspects de programmation.*

Enfin, la généralisation des infrastructures de calcul en ligne à faible coût, notamment les machines virtuelles équipées de GPU, rend désormais accessible une puissance de calcul autrefois réservée à des environnements spécialisés. Ces éléments motivent notre choix d'adapter et d'optimiser notre code Python afin de tirer parti de ses atouts pour la modélisation multi-agents.

A.1 Un exemple de compréhension

Pour illustrer le passage d'une boucle classique au calcul matriciel, prenons un exemple : nous avons 100 agents, chacun avec une position représentée par ses coordonnées (x, y), et à chaque étape, nous voulons augmenter leur position en fonction de leur vitesse (v_x, v_y). En Python « classique », on écrirait :

```
for i in range(n_agents):
    x[i] += vx[i]
    y[i] += vy[i]
```

Cette approche est intuitive, mais elle exécute la mise à jour agent par agent, ce qui peut être lent pour un grand nombre d'agents. En calcul matriciel, avec NumPy, on peut stocker toutes les positions dans une seule matrice positions de taille (n_agents, 2) et toutes les vitesses dans une autre matrice vitesses de même taille, puis mettre à jour toutes les positions en une seule instruction :

```
positions += vitesses
```

Ici, NumPy effectue l'opération en interne via du code optimisé en C, ce qui permet de traiter des milliers d'agents beaucoup plus rapidement qu'avec une boucle Python. Cet exemple montre bien l'avantage du calcul matriciel : moins de code, une exécution plus rapide, et une logique qui s'exprime directement sur l'ensemble des données.

A.2 Un exemple sur un problème connu

Pour démontrer qu'il est possible de transformer un algorithme typiquement multi-agents en une version matricielle, nous allons reformuler l'algorithme de ségrégation spatiale de Schelling à l'aide du calcul matriciel. L'objectif de cet exemple est de montrer que l'utilisation des opérations vectorisées ne se limite pas à notre approche intermédiaire entre les modèles multi-agents et les simulations centrées individu, mais peut également s'appliquer à des dynamiques spatiales plus classiques.

Dans le modèle original de Schelling, (Schelling 1971) chaque agent occupe une cellule d'une grille et prend des décisions en fonction de la composition locale de son voisinage (généralement les huit cellules adjacentes). Cette logique est souvent implémentée à l'aide de boucles, où chaque agent inspecte ses voisins un par un. Cependant, il est possible de reformuler cette mécanique en utilisant des opérations matricielles, ce qui permet des gains significatifs en performance et une meilleure exploitation du parallélisme matériel.

La clé de cette reformulation réside dans l'utilisation de « matrices décalées ». Pour chaque cellule de la grille, on peut construire huit versions décalées de la matrice représentant la population, chacune correspondant à un voisinage cardinal ou diagonal (haut, bas, gauche, droite, et les quatre diagonales). Il faut comparer ces huit matrices décalées, et stocker les résultats dans une dernière matrice. On obtient pour chaque cellule le nombre total de voisins correspondant au type d'agent dans la cellule. Il est ensuite possible d'évaluer la satisfaction d'un agent en comparant ce nombre avec un seuil de tolérance.

A.2.1 Un exemple restreint du modèle de ségrégation de schelling

A	$\cdot$	B	A
$\cdot$	B	A	B
A	A	$\cdot$	B
$\cdot$	B	$\cdot$	A

Table A.1 – *L'environnement initial : init*

$\cdot$	B	A	B
A	A	$\cdot$	B
$\cdot$	B	$\cdot$	A
A	$\cdot$	B	A

Table A.2 – *La table décalée vers le Nord* $\uparrow : N\hat{}$

$\cdot$	$\cdot$	$\neq$	$\neq$
$\cdot$	$\neq$	$\cdot$	$=$
$\cdot$	$\neq$	$\cdot$	$\neq$
$\cdot$	$\cdot$	$\cdot$	$=$

Table A.3 – *La table des comparaisons entre init et* $N\hat{}$

$$T^{(1)} = \begin{array}{|c|c|c|c|} \hline 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 1 \\ \hline 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 1 \\ \hline \end{array} \qquad S^{(1)} = \begin{array}{|c|c|c|c|} \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}$$

Table A.4 – *T compte le nombre de voisins et S le nombre de voisins similaires.*

$\cdot$	B	A	A
B	A	B	$\cdot$
A	$\cdot$	B	A
B	$\cdot$	A	$\cdot$

Table A.5 – *La table décalée vers l'Ouest* $\leftarrow : O_{\text{¸}}$

$\cdot$	$\cdot$	$\neq$	$=$
$\cdot$	$\neq$	$\neq$	$\cdot$
$=$	$\cdot$	$\cdot$	$\neq$
$\cdot$	$\cdot$	$\cdot$	$\cdot$

Table A.6 – *La table de comparaison entre init et* $O_{\text{¸}}$

$$T^{(2)} = \begin{array}{|c|c|c|c|}\hline 0 & 2 & 2 & 2 \\ \hline 2 & 2 & 1 & 1 \\ \hline 1 & 1 & 1 & 2 \\ \hline 2 & 0 & 2 & 1 \\ \hline \end{array} \qquad S^{(2)} = \begin{array}{|c|c|c|c|}\hline 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}$$

TABLE A.7 – *T et S après la seconde comparaison*

On répète exactement le même procédé avec les 6 autres décalages (bas, droite, et les quatre diagonales). Finalement, les matrices de résultats sont la somme des contributions de chacun des huit décalages :

— T = nombre total de voisins occupés (par cellule),
— S = nombre total de voisins similaires (par cellule).

La satisfaction des agents se calcule alors en comparant $\frac{S}{T}$ aux seuils de tolérance. On peut considérer que le déplacement des agents doit se faire avec une boucle car l'ordre de décision des agents a un impact sur le modèle : l'agent qui bouge en premier prend une place qu'aucun autre agent ne pourra prendre.

Il est aussi envisageable de n'utiliser que deux matrices :

— la matrice originale représentant l'état courant de la grille,
— et une version décalée (ou une série de versions décalées calculées à la volée).

Dans cette approche plus minimaliste, on itère sur les huit directions, et à chaque itération, on effectue un décalage de la grille, puis une comparaison cellule par cellule. Cette méthode est plus économe en mémoire, mais moins directe que l'utilisation simultanée de huit matrices décalées.

Dans la figure A.1, nous présentons l'évolution des temps d'exécution du modèle de ségrégation de Schelling selon deux scénarios distincts. Le scénario A correspond à un environnement fortement rempli (90 % d'agents) avec un seuil de satisfaction élevé (il faut 90 % de voisins similaires pour être satisfait). Dans ce cas, les agents restent durablement insatisfaits, de sorte que les simulations n'atteignent jamais un état stable et effectuent systématiquement les 100 pas de temps prévus. À l'inverse, le scénario B s'appuie sur un taux de remplissage plus faible (80 %) et un seuil de satisfaction modéré (50 %). Dans cette configuration, l'insatisfaction tend à disparaître rapidement, ce qui conduit les simulations à s'interrompre avant d'avoir parcouru l'ensemble des pas de temps, expliquant des temps d'exécution plus courts.

Dans nos expériences, nous considérons un taux de satisfaction uniforme appliqué à tous les agents. Bien qu'il soit possible de définir des seuils individuels de satisfaction, une telle hétérogénéité n'affecterait pas substantiellement les résultats de comparaison de performance présentés ici. Du point de vue de l'implémentation, l'utilisation du type INT8 pour représenter l'environnement est particulièrement adaptée, puisque la matrice ne prend que trois valeurs possibles (vide, agent de type A, agent de type B).

Enfin, notons que le nombre d'agents est proportionnel à la taille de l'environnement et reste fixé selon des paramètres classiques du modèle de Schelling. La diminution du

Algorithm 4 Modèle de ségrégation de Schelling matricialisé

$M \leftarrow \text{initializeGrid}(size, p_{empty}, p_{type1}, p_{type2})$
$T \leftarrow \text{toleranceThreshold}$
$shifts \leftarrow \{(-1,-1), (-1,0), \ldots, (1,1)\} \setminus (0,0)$
$shiftedMatrices \leftarrow [\text{shiftMatrix}(M, \Delta i, \Delta j),\ \forall (\Delta i, \Delta j) \in shifts]$
$iteration \leftarrow 0$
while iteration $< maxIter$ **do**
    $sameNeighbors \leftarrow \text{zerosLike}(M)$
    for each $shifted \in shiftedMatrices$ **do**
        $isSame \leftarrow (shifted == M) \wedge (M \neq empty)$
        $sameNeighbors \leftarrow sameNeighbors + isSame$
    end for
    $unsatisfied \leftarrow (sameNeighbors < T) \wedge (M \neq empty)$
    $unsatIndices \leftarrow \text{argWhere}(unsatisfied)$
    $emptyIndices \leftarrow \text{argWhere}(M == empty)$
    $numMoves \leftarrow \min(\text{len}(unsatIndices), \text{len}(emptyIndices))$
    if $numMoves == 0$ **then**
        break
    end if
    $emptyIndices \leftarrow \text{shuffle}(emptyIndices)$
    for $i \leftarrow 0$ **to** $numMoves - 1$ **do**
        $fromPos \leftarrow unsatIndices[i]$
        $toPos \leftarrow emptyIndices[i]$
        $M[toPos] \leftarrow M[fromPos]$
        $M[fromPos] \leftarrow empty$
    end for
    $shiftedMatrices \leftarrow [\text{shiftMatrix}(M, \Delta i, \Delta j),\ \forall (\Delta i, \Delta j) \in shifts]$
    $iteration \leftarrow iteration + 1$
end while

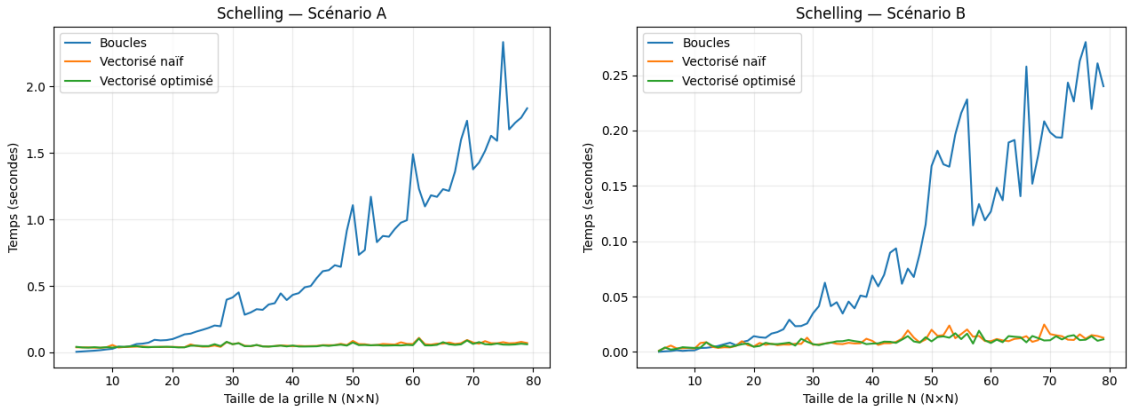


FIGURE A.1 – *Comparaison des temps d'exécution du modèle de ségrégation de Schelling en fonction de la taille de l'environnement (100 pas de temps) pour trois implémentations : une version à boucles, une version vectorisée en NumPy et une version optimisée en NumPy utilisant le type* INT8. *Le scénario A correspond à un taux de remplissage de 90 % et un seuil de satisfaction de 90 %. Le scénario B correspond à un taux de remplissage de 80 % et un seuil de satisfaction de 50 %. Dans ce second scénario, les simulations s'interrompent généralement avant d'atteindre les 100 pas de temps.*

nombre d'agents favorise légèrement la version à base de boucles, c'est ce que nous montrons dans la figure A.2 car moins de cellules doivent être évaluées. Toutefois, la légitimité de telles expériences reste discutable : dans une configuration aussi extrême, le modèle de Schelling se stabilise presque immédiatement (1 ou 2 pas de temps), rendant la comparaison peu pertinente. Par ailleurs, dans cette expérience aussi, le nombre d'agents est directement lié à la taille de l'environnement, de sorte que l'on observe la même tendance lorsque la taille de la grille augmente.

Finalement, ces résultats mettent en évidence que l'écart de temps d'exécution se creuse progressivement à mesure que le nombre d'agents (et, par conséquent, la taille de l'environnement) augmente. Ils illustrent clairement l'avantage des approches vectorisées pour ce type de simulations en Python.

L'évaluation du coût mémoire se révèle plus délicate que celle du temps d'exécution, car il est difficile d'obtenir une mesure parfaitement représentative de la charge effective. Nous avons néanmoins réalisé plusieurs expériences sur les mêmes scénarios que pour l'analyse des performances temporelles. Ces tests indiquent que la version à base de boucles présente un avantage relatif en termes de consommation mémoire. Toutefois, il convient de souligner que les optimisations offertes par NumPy, combinées à l'utilisation du type int8 pour représenter l'environnement, permettent de réduire considérablement l'écart : dans ce cas, le coût mémoire de la version NumPy optimisée ne dépasse celui de la version à boucles que d'environ 10 %. À l'inverse, la version NumPy naïve (s'appuyant sur les types par défaut int64 et float64) présente une consommation environ 2,5 fois plus élevée que celle de la version à boucles.

Ces deux techniques montrent que l'on peut, sans boucle explicite dans la partie calcul des voisins, simuler la logique de décision locale des agents du modèle de ségrégation de

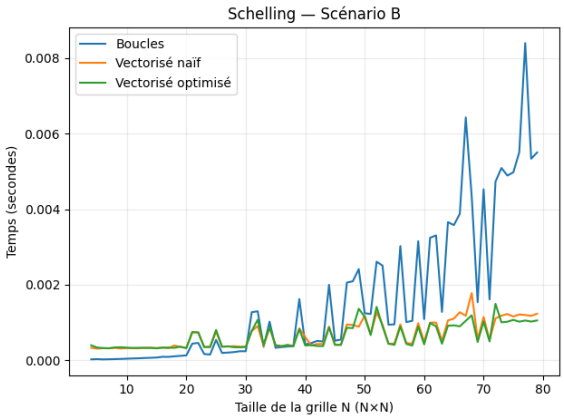


FIGURE A.2 – *Même affichage que pour les scénarios de la figure A.1 avec un scénario ou le taux de remplissage est de 1%. Il y a donc très peu d'agents et la simulation s'arrête généralement après 2 ou 3 pas de temps.*

Schelling par des opérations matricielles globales. Cette application illustre la versatilité de l'approche proposée. À notre connaissance, la principale limite inhérente à cette méthode concerne la gestion du tour de parole. En effet, lorsqu'une action effectuée par un agent influence directement le comportement des agents suivants, il devient impossible de reproduire fidèlement cette dynamique au moyen d'un calcul matriciel. En revanche, lorsque les interactions entre agents sont parallélisables, il est alors possible de mettre en œuvre une approche vectorisée reposant sur le calcul matriciel. Il convient néanmoins de souligner que cette stratégie implique un coût mémoire potentiellement élevé. Selon la nature du problème étudié, le recours au calcul matriciel peut se révéler peu adapté, entraînant une consommation mémoire plusieurs centaines de fois supérieure à celle d'une approche classique. Cet aspect doit donc être soigneusement pris en compte lors du choix méthodologique.

Ces expériences ouvrent la voie à une hybridation efficace entre simulation multi-agents et traitement vectorisé, adaptée à des environnements de calcul à grande échelle. Il faut noter que cette approche reste spécifique mais ne se limite pas à notre approche ni à la simulation centré individu.

Bibliographie

Abbasi Siar, Salman, Mohammad Ali Keramati et Mohammad Reza Motadel (2022). « Emergence of consumer impulse buying behavior with agent-based modeling approach ». In : *Journal of applied research on industrial engineering* 9.3, p. 203-230.

Albert, Réka et Albert-László Barabási (2002). « Statistical mechanics of complex networks ». In : *Reviews of modern physics* 74.1, p. 47.

Arthur, W Brian, John H Holland et Blake LeBaron (1999). « An artificial stock market ». In : *Artificial Life and Robotics* 3, p. 27-31.

Axtell, Robert L et J Doyne Farmer (2022). « Agent-based modeling in economics and finance : Past, present, and future ». In : *Journal of Economic Literature*.

Barthelemy, Johan et Philippe L Toint (2013). « Synthetic population generation without a sample ». In : *Transportation Science* 47.2, p. 266-279.

Bass, Frank M (1969). « A new product growth for model consumer durables ». In : *Management science* 15.5, p. 215-227.

Bawa, Kapil (1990). « Modeling inertia and variety seeking tendencies in brand choice behavior ». In : *Marketing science*.

Beckman, Richard J, Keith A Baggerly et Michael D McKay (1996). « Creating synthetic baseline populations ». In : *Transportation Research Part A : Policy and Practice* 30.6, p. 415-429.

Bijmolt, Tammo HA, Harald J Van Heerde et Rik GM Pieters (2005). « New empirical generalizations on the determinants of price elasticity ». In : *Journal of marketing research* 42.2, p. 141-156.

Blattberg, Robert C, Richard Briesch et Edward J Fox (1995). « How promotions work ». In : *Marketing science*.

Bonabeau, Eric (2002). « Agent-based modeling : Methods and techniques for simulating human systems ». In : *Proceedings of the national academy of sciences* 99.suppl_3, p. 7280-7287.

Borden, Neil H (1964). « The concept of the marketing mix ». In : *Journal of advertising research* 4.2, p. 2-7.

Bruderer, Erhard et Jitendra V Singh (1996). « Organizational evolution, learning, and selection : A genetic-algorithm-based model ». In : *Academy of management journal* 39.5, p. 1322-1349.

Cantillo, Víctor, Juan de Dios Ortuzar et Huw CWL Williams (2007). « Modeling discrete choices in the presence of inertia and serial correlation ». In : *Transportation Science* 41.2, p. 195-205.

Chapuis, Kevin, Patrick Taillandier et Alexis Drogoul (2022). « Generation of synthetic populations in social simulations : a review of methods and practices ». In : *Journal of Artificial Societies and Social Simulation* 25.2.

Chica, Manuel et William Rand (2017). « Building agent-based decision support systems for word-of-mouth programs : A freemium application ». In : *Journal of Marketing Research* 54.5, p. 752-767.

Choudhary, Vidyanand et al. (2017). « Evaluation set size and purchase : evidence from a product search engine ». In : *Journal of Interactive Marketing* 37.1, p. 16-31.

Cohen, Maxime et al. (2020). « An efficient algorithm for dynamic pricing using a graphical representation ». In : *Production and Operations Management* 29.10, p. 2326-2349.

Conway, John H (1970). « An enumeration of knots and links, and some of their algebraic properties ». In : *Computational problems in abstract algebra*. Elsevier, p. 329-358.

Danaye, Nasser, Ramez Kian et Nazan Colmekcioglu (2023). « An agent-based modeling approach for understanding drivers of consumer decisions on foreign versus domestic products : case study of a local refrigerator market ». In : *International Journal of Information Technology & Decision Making* 22.03, p. 1107-1134.

Delre, Sebastiano A et al. (2007). « Targeting and timing promotional activities : An agent-based model for the takeoff of new products ». In : *Journal of business research* 60.8, p. 826-835.

Dhar, Sanjay K, Stephen J Hoch et Nanda Kumar (2001). « Effective category management depends on the role of the category ». In : *Journal of Retailing* 77.2, p. 165-184.

Doshi, Ronak, Ajay Ramesh et Shrisha Rao (2022). « Modeling influencer marketing campaigns in social networks ». In : *IEEE Transactions on Computational Social Systems* 10.1, p. 322-334.

Dussart, Christian (1998). « Category management : : Strengths, limits and developments ». In : *European Management Journal* 16.1, p. 50-62.

Epstein, Joshua M et Robert Axtell (1996). *Growing artificial societies : social science from the bottom up*. Brookings Institution Press.

Frederick, Shane, Leonard Lee et Ernest Baskin (2014). « The limits of attraction ». In : *Journal of Marketing Research* 51.4, p. 487-507.

Garcia, Rosanna (2005). « Uses of agent-based modeling in innovation/new product development research ». In : *Journal of product innovation management* 22.5, p. 380-398.

Goldstein, Jeffrey (1999). « Emergence as a construct : History and issues ». In : *Emergence* 1.1, p. 49-72.

Green, Edward J, Robert C Marshall et Leslie M Marx (2014). « Tacit collusion in oligopoly ». In : *The Oxford handbook of international antitrust economics* 2, p. 464-497.

Hardie, Bruce GS, Eric J Johnson et Peter S Fader (1993). « Modeling loss aversion and reference dependence effects on brand choice ». In : *Marketing science* 12.4, p. 378-394.

Hauser, John R (2014). « Consideration-set heuristics ». In : *Journal of Business Research* 67.8, p. 1688-1699.

Hazelbag, C Marijn et al. (2020). « Calibration of individual-based models to epidemiological data : A systematic review ». In : *PLoS computational biology* 16.5, e1007893.

Hermellin, Emmanuel (2016). « Modélisation et implémentation de simulations multi-agents sur architectures massivement parallèles ». Thèse de doct. Université Montpellier.

Hermellin, Emmanuel, Fabien Michel et Jacques Ferber (2014). « Systèmes multi-agents et GPGPU : état des lieux et directions pour l'avenir. » In : *JFSMA*, p. 97-106.

Hoch, Stephen J et al. (1995). « Determinants of store-level price elasticity ». In : *Journal of marketing Research* 32.1, p. 17-29.

Holland, John H (1992). « Complex adaptive systems ». In : *Daedalus* 121.1, p. 17-30.

Holy, Vladimir, Ondrej Sokol et Michal Cerny (2017). « Clustering retail products based on customer behaviour ». In : *Applied Soft Computing* 60, p. 752-762.

Hörl, Sebastian et Milos Balac (2021). « Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data ». In : *Transportation Research Part C : Emerging Technologies* 130, p. 103291.

Huber, Joel, John W Payne et Christopher Puto (1982). « Adding asymmetrically dominated alternatives : Violations of regularity and the similarity hypothesis ». In : *Journal of consumer research* 9.1, p. 90-98.

Huber, Joel, John W Payne et Christopher P Puto (2014). « Let's be honest about the attraction effect ». In : *Journal of Marketing Research* 51.4, p. 520-525.

Hung, Phan Duy, Nguyen Duc Ngoc et Tran Duc Hanh (2019). « K-means clustering using RA case study of market segmentation ». In : *Proceedings of the 2019 5th International Conference on E-Business and Applications*, p. 100-104.

Jager, Wander (2007). « The four P's in social simulation, a perspective on how marketing could benefit from the use of social simulation ». In : *Journal of Business Research* 60.8, p. 868-875.

Kahneman, Daniel et Amos Tversky (1979). « Prospect Theory : An Analysis of Decision under Risk ». In : *Econometrica* 47.2, p. 263-292.

Khandelwal, Amit (2010). « The long and short (of) quality ladders ». In : *The Review of Economic Studies* 77.4, p. 1450-1476.

Kienzler, Mario et Christian Kowalkowski (2017). « Pricing strategy : A review of 22 years of marketing research ». In : *Journal of Business Research* 78, p. 101-110.

Lancaster, Kelvin J (1966). « A new approach to consumer theory ». In : *Journal of political economy* 74.2, p. 132-157.

Langton, Christopher G (1986). « Studying artificial life with cellular automata ». In : *Physica D : nonlinear phenomena* 22.1-3, p. 120-149.

Lee, Keeheon, Hoyeop Lee et Chang Ouk Kim (2014). « Pricing and timing strategies for new product using agent-based simulation of behavioural consumers ». In : *Journal of Artificial Societies and Social Simulation* 17.2, p. 1.

Li, Jing et XiaoWen Liu (2024). « An agent-based simulation model for analyzing and optimizing omni-channel retailing operation decisions ». In : *Journal of Retailing and Consumer Services* 79, p. 103845.

Liang, Chyi-Lyi et Bryan Collins (2025). « Applying Agent-Based Modeling to Examine Business Strategies–Tools and Examples for Researchers and Practitioners ». In : *Small Business Institute Journal* 21.1, p. 16-24.

Mathieu, Philippe, David Panzoli et Sébastien Picault (2011). « Format-store : a multi-agent based approach to experiential learning ». In : *2011 Third International Conference on Games and Virtual Worlds for Serious Applications*. IEEE, p. 120-127.

Mathieu, Philippe et Sébastien Picault (2013). « Des données aux agents : la simulation réaliste de populations diversifiées de clients ». In : *21e Journées francophones sur les systèmes multi-agents (JFSMA 2013)*. Cépaduès, p. 41-50.

McFadden, Daniel (1972). « Conditional logit analysis of qualitative choice behavior ». In : *Working paper*.

Negahban, Ashkan et Levent Yilmaz (2014). « Agent-based simulation applications in marketing research : an integrated review ». In : *Journal of Simulation* 8.2, p. 129-142.

Noguchi, Takao et Neil Stewart (2014). « In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions ». In : *Cognition* 132.1, p. 44-56.

— (2018). « Multialternative decision by sampling : A model of decision making constrained by process data. » In : *Psychological review* 125.4, p. 512.

Pettibone, Jonathan C et Douglas H Wedell (2000). « Examining models of nondominated decoy effects across judgment and choice ». In : *Organizational behavior and human decision processes* 81.2, p. 300-328.

Pillai, Rajani Ganesh et Vishal Bindroo (2014). « The moderating roles of perceived complementarity and substitutability on the perceived manufacturing difficulty–extension attitude relationship ». In : *Journal of Business Research* 67.7, p. 1353-1359.

Plata, Serge et al. (2020). « Simulating human interactions in supermarkets to measure the risk of COVID-19 contagion at scale ». In : *arXiv preprint arXiv :2006.15213*.

Rand, William et Roland T Rust (2011). « Agent-based modeling in marketing : Guidelines for rigor ». In : *International Journal of research in Marketing* 28.3, p. 181-193.

Reynolds, Craig W (1987). « Flocks, herds and schools : A distributed behavioral model ». In : *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, p. 25-34.

Rogers, Everett M, Arvind Singhal et Margaret M Quinlan (2014). « Diffusion of innovations ». In : *An integrated approach to communication theory and research*. Routledge, p. 432-448.

Romero, Elena et al. (2023). « Two decades of agent-based modeling in marketing : A bibliometric analysis ». In : *Progress in Artificial Intelligence* 12.3, p. 213-229.

Said, Lamjed Ben, Thierry Bouron et Alexis Drogoul (2002). « Agent-based interaction analysis of consumer behavior ». In : *Proceedings of the first international joint conference on AAMAS : part 1*, p. 184-190.

Sart, Mathieu (2017). « Estimating the conditional density by histogram type estimators and model selection ». In : *ESAIM : Probability and Statistics* 21, p. 34-55.

Schelling, Thomas C (1971). « Dynamic models of segregation ». In : *Journal of mathematical sociology* 1.2, p. 143-186.

Seetharaman, PB et Pradeep Chintagunta (1998). « A model of inertia and variety-seeking with marketing variables ». In : *International Journal of Research in Marketing* 15.1, p. 1-17.

Spektor, Mikhail S, David Kellen et Karl Christoph Klauer (2022). « The repulsion effect in preferential choice and its relation to perceptual choice ». In : *Cognition* 225, p. 105164.

Stanley, Jacob M et Douglas H Wedell (2024). « Impact of choice set complexity on decoy effects ». In : *Journal of Behavioral Decision Making* 37.2, e2373.

Stigler, George J (1964). « A theory of oligopoly ». In : *Journal of political Economy* 72.1, p. 44-61.

Sturley, Charlotte, Andy Newing et Alison Heppenstall (2018). « Evaluating the potential of agent-based modelling to capture consumer grocery retail store choice behaviours ». In : *The International Review of Retail, Distribution and Consumer Research* 28.1, p. 27-46.

Tellis, Gerard J (2006). « Modeling marketing mix ». In : *Handbook of marketing research*, p. 506-522.

Trueblood, Jennifer S et al. (2013). « Not just for consumers : Context effects are fundamental to decision making ». In : *Psychological science* 24.6, p. 901-908.

Tversky, Amos et Daniel Kahneman (1991). « Loss aversion in riskless choice : A reference-dependent model ». In : *The quarterly journal of economics* 106.4, p. 1039-1061.

Ulbinaitė, Aurelija, Marija Kučinskienė et Yannick Le Moullec (2011). « Integration of the decoy effect in an agentbased-model simulation of insurance consumer behaviour ». In : *2011 International Conference on Software and Computer Applications IPCSIT. Conference Proceedings*. T. 9, p. 152-157.

Vanderlynden, J, P Mathieu et R Warlop (2024). « Stratégies marketing et effets de contexte ». In : *Trente-deuxièmes journées francophones sur les systèmes multi-agents (JFSMA)*. Cépaduès, p. 161-170.

Vanderlynden, Jarod, Philippe Mathieu et Romain Warlop (2023). « Comprendre l'impact des stratégies de prix sur le comportement des consommateurs ». In : *Trente-et-unièmes journées francophones sur les systèmes multi-agents (JFSMA)*. Cépaduès, p. 55-64.

— (2024a). « Simulation of Consumers Behavior Facing Discounts and Promotions. » In : *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART)*. T. 3. SCITEPRESS-Science et Technology Publications, p. 260-267.

— (2024b). « Understanding the Impact of Promotions on Consumer Behavior. » In : *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, p. 2531-2533.

— (2025a). « Agent-Based Modeling of Context Effects in Consumer Choice ». In : *28TH European Conference on Artificial Intelligence (ECAI)*.

— (2025b). « Calibration d'un Système Multi-Agent pour l'Optimisation des Prix de Produits Concurrentiels ». In : *Journées Francophones sur les Systèmes Multi-Agents (JFSMA)*. Cépaduès, p. 175-195.

Watts, Duncan J et Steven H Strogatz (1998). « Collective dynamics of 'small-world'networks ». In : *nature* 393.6684, p. 440-442.

Wigren, Richard et Filip Cornell (2019). *Marketing Mix Modelling : A comparative study of statistical models*.

Ying, Fabian et Neave O'Clery (2021). « Modelling COVID-19 transmission in supermarkets using an agent-based model ». In : *Plos one* 16.4, e0249821.

Ying, Fabian, Alisdair OG Wallis et al. (2019). « Customer mobility and congestion in supermarkets ». In : *Physical Review E* 100.6, p. 062304.

Zhang, Tao et David Zhang (2007). « Agent-based simulation of consumer purchase decision-making and the decoy effect ». In : *Journal of business research* 60.8, p. 912-922.

Résumé

Cette thèse propose une approche modulaire et explicable de la modélisation du comportement des consommateurs à l'aide des Systèmes Multi-Agents (SMA). L'objectif est de montrer que le comportement d'achat peut être décomposé en composantes distinctes : estimation du nombre d'acheteurs, modélisation du choix et étude des achats récurrents, afin de concevoir des modèles à la fois interprétables, calibrables et empiriquement validés.

Dans un premier temps, nous développons une méthode probabiliste de reconstitution de population à partir de données de tickets de caisse partiellement anonymisés, permettant d'estimer de manière fiable le volume de clients. Ensuite, nous proposons un modèle de choix fondé sur le biais cognitif d'aversion à la perte, capable de reproduire les effets de contexte observés dans les décisions d'achat. Enfin, nous appliquons un SMA à l'étude des promotions sur les produits de consommation courante, en intégrant la notion d'inertie d'achat et de fidélité des consommateurs.

Ces trois modèles, complémentaires et validés séparément, offrent une compréhension approfondie des dynamiques individuelles et collectives du comportement d'achat. Au-delà de leur dimension prédictive, ils contribuent à une meilleure explicabilité des mécanismes de décision des consommateurs. Ce travail ouvre la voie à la construction de SMA interconnectés intégrant besoins, choix, influence sociale et concurrence, dans la perspective d'une modélisation unifiée du marché.

Mots-clés : Simulation, Multi-agents, Émergence, Marketing , Dynamique des prix, effet de leurre, Comportements, Calibration

Abstract

This dissertation presents a modular and explainable approach to consumer behavior modeling using Multi-Agent Systems (MAS). The central aim is to demonstrate that purchasing behavior can be decomposed into distinct components : estimating the number of buyers, modeling product choice, and analyzing recurrent purchases, in order to design interpretable, data-driven, and empirically validated models.

First, a probabilistic reconstruction method is developed to estimate customer populations from partially anonymized transaction data. Second, consumer choice is modeled through the psychological bias of loss aversion, reproducing empirically observed context effects. Third, a MAS framework is applied to supermarket promotions, incorporating purchase inertia and consumer loyalty to simulate realistic market responses.

Together, these complementary models enhance both the explanatory and predictive power of consumer simulations. Beyond prediction, they contribute to a deeper understanding of decision mechanisms and market dynamics. Ultimately, this research paves the way for an integrated ecosystem of interconnected MAS representing the full purchasing process, from need formation to choice and recurrence, toward a unified modeling of consumer behavior.

Keywords : Simulation, Agent based, Emergence, Marketing, Price dynamics, Decoy effect, Behaviors, Calibration.