



**HAL**  
open science

## Purchase intention-based agent for customer behaviours

Arnaud Doniec, Stéphane Lecoecueche, René Mandiau, Antoine Sylvain

► **To cite this version:**

Arnaud Doniec, Stéphane Lecoecueche, René Mandiau, Antoine Sylvain. Purchase intention-based agent for customer behaviours. *Information Sciences*, 2020, 521, pp.380-397. 10.1016/j.ins.2020.02.054 . hal-02496810

**HAL Id: hal-02496810**

**<https://hal.science/hal-02496810v1>**

Submitted on 7 Mar 2022

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Purchase [intention-based](#) agent for customer behaviours

Arnaud Doniec<sup>a</sup>, Stéphane Lecoecuche<sup>a</sup>, René Mandiau<sup>b</sup>, Antoine Sylvain<sup>a</sup>

<sup>a</sup>*IMT Lille Douai, Univ. Lille, URIA, F-59000 Lille, France*  
*e-mail: {prenom.nom}@imt-lille-douai.fr*

<sup>b</sup>*LAMIH UMR CNRS 8201, Université Polytechnique Hauts-de-France (UPHF)*  
*F-59313 Valenciennes Cedex, France*  
*e-mail: rene.mandiau@uphf.fr*

---

### Abstract

Simulating human activities remains a challenging problem because the decision-making mechanisms underlying these activities are difficult to reproduce and mimic. In this article, we are interested in the simulation of in-store shoppers whose activities are generally divided into two parts: a walking activity and a purchase activity. Since the act of buying is more complex than simply following a shopping list, we propose here to model the attraction relationships that can exist between a product and a customer. This attraction model is used to build a multi-agent simulation whose realism is evaluated through various experiments.

*Keywords:* [Agent-based](#) modelling, Customer behaviours, Purchase intention, Stores' simulation

---

### 1. Introduction

2 The retail activity is a crucial challenge in terms of [employment](#) and  
3 [growth](#) of a country. In 2016, worldwide retail sales reached \$22.049 tril-  
4 lion [27]. The share of e-commerce has greatly increased, representing \$1.915  
5 trillion. In fact, many analysts predict stronger increases in the coming years  
6 and agree that e-commerce is cannibalizing traditional retailing. Thus, this  
7 industry needs to evolve, to reinvent itself, to offer customers new shopping  
8 experiences in physical stores. Therefore, traditional retailers need tools to  
9 analyse, understand and simulate customer behaviours inside their physical  
10 stores. [In our opinion](#), a store simulation should be useful:

- 11 • to give an interactive, dynamic and global view of the [store](#),

- 12 • to study the different customers (or shoppers) inside the store: en-  
13 trance flow, consistency and flow traffic in the store departments,  
14 area/product of interest, checkout management,
- 15 • to identify critical areas of congestion, and
- 16 • to predict the impact of changes in the organization for the trading  
17 area: changes of department location, changes of shelves sizes.

18 **A number** of commercial *decision support* tools related to retail **already**  
19 **exist** and can be classified in different categories:

- 20 • planogram designs and 3D visualization of products inside the store  
21 (e.g. Retail Smart<sup>1</sup>, Buzz 3D Retail<sup>2</sup>),
- 22 • queue management at the cash desk (e.g., QNomy<sup>3</sup>),
- 23 • supply chain management of the store (e.g., Llamasoft<sup>4</sup>), and
- 24 • marketing policies simulation (e.g., Markstrat<sup>5</sup>).

25 **To the best of our** knowledge, there are **still** no commercial tools to  
26 describe the customer behaviours within a store, even if such tools are useful  
27 from a manager's point of view. **We believe** that the construction of these  
28 behaviours is a convincing motivation for analysing the customer purchase  
29 intentions.

30 The customer behaviour inside a physical store depends on many fac-  
31 tors: the type of the store (supermarkets, clothes shops, DIY store), the  
32 motivations for visiting the store (weekly food provision, shopping trip with  
33 friends), their habits, etc. The behaviours **can thus vary greatly, so** it is  
34 **rather** illusory to propose a unique customer model.

35 From a research point of view, the study of customer behaviour clearly  
36 falls within the scope of social simulations. In the last decade, they have been

---

<sup>1</sup><http://www.retailsmart.com/3d-planograms>, Accessed: 2020-01-28

<sup>2</sup>[http://www.buzz3d.com/3d\\_planogram\\_virtual\\_store\\_index.html](http://www.buzz3d.com/3d_planogram_virtual_store_index.html), Accessed:  
2020-01-28

<sup>3</sup><https://www.qnomy.com/queue-management>, Accessed: 2020-01-28

<sup>4</sup><https://www.llamasoft.com/solutions/supply-chain-simulation/>, Accessed:  
2020-01-28

<sup>5</sup><https://web.stratxsimulations.com/simulation/strategic-marketing-simulation/>,  
Accessed: 2020-01-28

37 extensively studied from the perspective of agent-based modelling, since hu-  
38 man activities are often too difficult to express using classical mathemati-  
39 cal equation-based models [3, 33, 47]. Contrary to an analytical approach,  
40 a Multi-Agent Based Simulation (*MABS*) does not try to model the sys-  
41 tem in its entirety but splits it up in interacting entities (called agents),  
42 whose behaviours are sometimes simpler to implement rather than finding  
43 equations describing the global system. The core idea of *MABS* relies on  
44 the concept of emergence: the overall dynamic of the system results from  
45 their individual behaviours and their interactions. In addition, we can note  
46 that agent-oriented approaches have been used, for a long time, for trad-  
47 ing in virtual markets where agents could autonomously purchase or sell  
48 products [13]. The main contributions of these studies are often based on  
49 web technology and multi-agent negotiation models concerning the different  
50 transactions (e.g. variable price, bounded time for the transaction). For all  
51 these reasons, we think that a multi-agent approach is well suited to describe  
52 realistic simulations of stores.

53 In this article, the focus is on situated agent-based simulation within a  
54 physical store which should reproduce the trips and purchases of customers.  
55 These simulated customers (modelled as agents) can thus move in the shop,  
56 purchase different products, etc. To obtain these behaviours, we propose  
57 to use some existing psychological studies to build a configurable model  
58 of customers inside a store and then, to fit this model to a specific store  
59 through real-data assimilation techniques (features are learned from real  
60 data of existing supermarkets, mainly trajectories and receipts). Our con-  
61 tribution models essentially the purchase intentions, and we propose a new  
62 approach accounting for the way to characterize the relationships between  
63 the customers and the products.

64 The paper is organized as follows. Section 2 deals with the state of the  
65 art related to psychological studies of customer behaviour and *MABS* for  
66 retail. Section 3 presents our agent model based on different behaviours  
67 (in a shop), the walking and the purchase intentions. In particular, we  
68 focus on a model of product-customer attraction. Our model is evaluated  
69 in Section 4, through various experiments involving the use of real collected  
70 data. Section 5 discusses our approach in a more general context. Finally,  
71 the last section concludes and gives a few perspectives.

## 72 2. Related work

73 We briefly present the state of the art about psychological and marketing  
74 studies for customer behaviour (Sect. 2.1) and existing agent-based models

75 for marketing and retail (Sect. 2.2).

### 76 2.1. Psychological and marketing studies of customer behaviour

77 There is an extensive literature [18, 42] related to customer behaviours in  
78 stores. These studies claim that these various behaviours are closely related  
79 to the concerned type of store. In particular, they distinguish two main  
80 groups, called *shoppers* (*i.e.*, customers coming with a real purchase inten-  
81 tion) and *gatherers* (*i.e.*, customers looking for some information and enter-  
82 tainment, but without necessarily purchasing products). These two types  
83 of behaviours correspond to two views of the sale space: an *economic space*  
84 used for supplying, or an *entertainment space* used for leisure activities [10].  
85 For example, in the case of supermarkets, there are few gatherers (*i.e.*, the  
86 majority of customers are shoppers, whereas these gatherers represent a  
87 third of beauty shops' visitors [39]). Note **however** that customers inside  
88 **the same group** still present distinct behaviours. Indeed, **not all shoppers**  
89 **purchase** exactly the same products, **so they could be classified** depending  
90 on their displacements and purchases.

91 The modelling of these behaviours has been the subject of **extensive**  
92 **research**. As suggested by different authors [5, 19], a customer behaviour in-  
93 side a supermarket results essentially from two main activities: (i) a walking  
94 activity and (ii) a purchase activity.

95 The walking activity is defined by physical parameters: for example, a  
96 *rhythm* (*i.e.*, **the way they move with slow speed or fast**), an *extent* (*i.e.*,  
97 **the number of crossed departments**), an *orientation* (*i.e.*, the field of vision)  
98 allowing customers to observe their own environment.

99 The purchase activity may be considered as a more cognitive task than  
100 the previous one. Many works have studied the purchase intention (*i.e.*,  
101 the reasons that led a customer to buy a particular product). For example,  
102 Sakas et al. [38] describe a dynamic simulation model based on four graphical  
103 items: (i) a tank representing the quantity of resources, (ii) a flow for an  
104 activity filling or draining a tank, (iii) a convector keeping certain values,  
105 and (iv) connectors providing links between the items. In their analysis,  
106 the authors express their model based on the connections between different  
107 tanks: the characteristics (for example, incentives, roles and personality)  
108 affecting the decision-making for the purchases by the customer, as well as  
109 psychological factors (for example, perception and beliefs/preferences).

110 Several features of a product influence its attractiveness and therefore  
111 the purchase intention of customers: the *brand*, the *price* [12], the *special*  
112 *offers and promotions* [24], the *position* of the product (both inside the  
113 supermarket and on the shelf) [7], the *products already put in the basket* [1].

114 Some researchers [8] explain that the purchase intentions can be estimated  
115 by these different elements (called *perceived risks*).

116 In addition, the purchase activity can be influenced by two main envi-  
117 ronmental factors [39]: a *crowd density* and an *available time*. The *crowd*  
118 *density* corresponds to the number of persons present in a given area that  
119 a customer perceives. A higher density reduces the time that a customer  
120 accepts to spend in the store: he/she can play on his/her exploration and  
121 deliberation times. This constraint negatively influences the purchases and  
122 the interactions with other agents. The *available time* is a constraint (dur-  
123 ing the cognitive step) which can imply a fast decision to purchase (or not)  
124 products and also a more limited exploration of the store. In our context,  
125 this available time produces a pressure (if this time is short) similar to the  
126 pressure caused by a higher crowd density.

127 We share the idea, initially proposed by Ladwein [23], claiming that  
128 the customer behaviour is split in four main *actoms* (*i.e.*, atomic actions),  
129 namely *mobility*, *stop*, *prehension* and *purchase*. *The mobility* actom corre-  
130 sponds to the displacement of the customer inside the store, and the obser-  
131 vation of the departments during the walking [43]. *The stop* actom considers  
132 the [situation in which the](#) customer is stopping in front of a shelve. A stop  
133 usually follows the identification of a product (*i.e.*, a visual perception of  
134 the department) during the mobility phase. *The prehension* actom means a  
135 hand gripping of a product. *The purchase* actom corresponds to an action  
136 to purchase or not (also to put or not the product in the cart).

## 137 2.2. Agent-based models for marketing and retail

138 We find a lot of contributions concerning the use of simulations for mar-  
139 keting and retail. Some [of these contributions](#) have investigated the use of  
140 macroscopic simulations to understand customers' interactions in a specific  
141 market [37], to study new sellers strategies like dynamic pricing [9] [or to](#)  
142 [train](#) sellers (based on serious games) [28].

143 Regarding the simulation of stores in terms of flows, [some solutions using](#)  
144 [MABS have already been proposed](#). For example, *MassMotion* <sup>6</sup> simulates  
145 the displacements in different types of buildings, [including, for example,](#)  
146 shopping malls. Moulin et al. [29] propose a pedestrian simulation inside a  
147 mall. Each agent (the customer) enters the mall with a list of predefined  
148 stores or [kiosks](#) to visit, often with specific motivations for services: banking,

---

<sup>6</sup><http://www.oasys-software.com/products/engineering/massmotion.html>, Ac-  
cessed: 2020-01-28

149 café and restaurant, entertainment (*e.g.*, theatres, cinema), etc.

150 *SimMarket* [40] focuses on the purchase intention. The authors propose  
151 a behavioural model of customers – agents – involving different factors like  
152 the price, the promotion [awareness](#) or the brand loyalty. The behaviour is  
153 expressed in this study by using [Bayesian networks](#). For each category of  
154 products, a Bayesian network is extracted through the search of data mining  
155 patterns from individual customer data.

156 *Format-Store* [28] proposes a serious game dedicated to employees. Based  
157 on a methodology named IODA (acronym for *Interaction Oriented Design*  
158 *of Agent simulation*), the authors describe all possible interactions occurring  
159 in the store: for example, the interactions between a customer and a prod-  
160 uct, or between an employee and a customer. All the entities defined by an  
161 interaction (customers, products, employees) are thus considered as agents.  
162 This methodology may have the benefit of describing a very homogeneous  
163 multi-agent model, [but it has](#) the drawback of a very high number of agents  
164 (due to [the large number](#) of products and customers) and of making the  
165 concept of purchase intentions less clear (furthermore, it is important to  
166 note that it was not the initial motivation of their studies).

167 Bell and Mgbemena [4] describe an agent-based modelling (in the mo-  
168 bile services sector) using the decision-tree analysis to create agents (in near  
169 *real-time*) from data analysis and social interactions. Their approach called  
170 CADET (*Customer Agent DEcision Tree*), is based on decision trees (tree-  
171 like graphs with possible paths to arrive at a decision) [employed to](#) study  
172 the interactions between customers. These authors claim that most [research](#)  
173 that [addresses](#) customer churn using data-mining techniques (such as cluster-  
174 ing and classification) do not typically account for social effects on customer  
175 retention. The influence of social networks has also been found to be a key  
176 factor to customer [retention](#) [35]. The advantage of the CADET approach is  
177 a very quick adaptability of the agent-based modelling. However, we think  
178 that the data-mining techniques remain sufficient for a validation of a model  
179 based on the purchase intentions of customers for a physical shop.

180 ABISS (Agent-Based In-Store Simulator) [44] focuses on "virtually" cus-  
181 tomers (by walking and purchase activities) within a supermarket. This tool  
182 is able to consider planned purchases (shopping list) and unplanned pur-  
183 chases (promotion items) through different purchasing probabilities. These  
184 [probabilities](#) are determined [from](#) the analysis of real past transactions. Ac-  
185 cording to us, ABISS is probably the most advanced, but few of the influ-  
186 encing factors previously presented (loyalty to a brand, products' pricing,  
187 time pressure) are taken into account.

188 The model SHOPPER [36] is not an agent-based model, but it aims to

Table 1: Synthesis of the main elements for customer activities and comparison between state-of-the-art and our proposed approach

	Walking Activities		Purchase Activities										
	Path Search	Obstacle Avoiding	Shopping List	Money Limit	Staying Time	Seasonal Effects	Price	Interaction between products	Loyalty Card	User Preferences	Promotion	Crowd Density	Checkout Management
[4]	•		•	•			•			•	•		
[20]	•		•	•	•		•			•	•		
[28]	•	•	•		•		•			•			
[29]	•		•			•	•	•	•	•			•
[36]				•			•		•	•			
[40]				•			•		•	•			
[44]	•		•		•		•	•		•	•		
Our Study	•	•	•		•		•	•		•	•	•	•

189 reproduce the customer behaviour inside a supermarket. This approach uses  
 190 learning techniques based on previous shopping trips (receipts) and proposes  
 191 to investigate seasonal effects and interactions between products.

192 We investigate an approach where the agents represent the human cus-  
 193 tomers in the store (the products are then seen as resources of this environ-  
 194 ment). Adapted from Kitazawa et al. [20] and Ruiz et al. [36], Table 1 gives  
 195 a synthesis of different elements for the two main activities, walking and  
 196 purchasing. Considering all these existing models, we can observe that the  
 197 purchase activity is either absent or very simplified and that the walking  
 198 activity is not always directly related to the purchase intention. For the  
 199 walking activity, we show that the avoidance actions are not always really  
 200 implemented, although they are well-known in the literature. Concerning  
 201 the second activity, it seems that the shopping list is usually pre-defined  
 202 (we agree with Ruiz et al. [36] that this list should be adapted dynamically  
 203 based on the purchased products or promotions, for example). Moreover,  
 204 the checkout management is not always considered, although it has an im-  
 205 pact (essentially, the time pressure) for the real behaviour of a customer.  
 206 Finally, the staying time is often considered, but the crowd density (not de-  
 207 fined in the existing studies) may change the customer behaviour during its  
 208 purchase. The literature in the field of marketing and consumer psychology  
 209 is extremely rich and shows interest in the temporal pressure and in different  
 210 actoms. We think that these concepts should be considered for our agent  
 211 model, presented in the next section.



### 212 3. Proposition of a multi-agent model for customers in a store

213 We **start with** the representation of a store, and **we list** the specific char-  
214 acteristics of products and features of agents (Sect. 3.1). Then, we describe  
215 the model based on states for the different agents (Sect. 3.2). We **also focus**  
216 on a crucial **aspect** concerning our modelling, **namely** the attraction between  
217 products and agents (Sect. 3.3).

#### 218 3.1. Presentation of our multi-agent model

219 Our approach consists of a set of agents (*i.e.*, customers) evolving in  
220 the same environment (*i.e.*, a virtual representation of a store). The agents  
221 may move in the store **where** they may be **located in** different places (*e.g.*,  
222 shelves, checkouts). The products are viewed as resources situated in the  
223 same environment. The agent can purchase a product, if there exists an  
224 *attraction* between the product and the agent.

##### 225 3.1.1. Environment: the store

226 To plan their **movements** inside a store, the agents (simulated customers)  
227 need a computational representation of their environment. Since a super-  
228 market is (most of the time) well structured with aisles and shelves, a graph  
229 models the environment of our simulation: the nodes of the graph cover  
230 all intersections between aisles and shelves while edges represent the lines  
231 along which customers can move. This graph needs to be enhanced with  
232 additional information like, for example, the department name, the list of  
233 items **located** around a graph node.

##### 234 3.1.2. Resources: the products

235 Each product  $p$  situated in this environment is defined by  $\langle Cat_p, Rpr_p, A_p \rangle$ :

- 236 •  $Cat_p$  refers to the *category of product* (or a hierarchy of categories).  
237 For example, a packet of tortillas can be classified in the following hi-  
238 erarchy: *Snacks*  $\rightarrow$  *Salty Snacks*  $\rightarrow$  *Chips* (this classification may also  
239 facilitate the comparison between two products of the same category).
- 240 •  $Rpr_p$  represents the *real price* of product  $p$ .
- 241 •  $A_p$  is a vector of parameters describing the attractiveness of this prod-  
242 uct (in this study, we consider only four parameters for this vector,  
243 and will show that they are sufficient for our validation):

- 244 – *Attractiveness to height* ( $A_p^1 \in [0, 1]$ ): Many marketing studies  
245 show that a product situated on a shelf at eye level has a higher  
246 probability to be bought than a product located at ground level.
- 247 – *Attractiveness to price positioning* ( $A_p^2 \in [0, 1]$ ) in a same cate-  
248 gory: This parameter is directly related to the purchase power of  
249 customers.
- 250 – *Attractiveness to brand* ( $A_p^3 \in \{0, 0.5, 1\}$ ): This parameter refers  
251 to its location in a product range, the reputation and the prestige  
252 brand of a product. For example, food retail can be distinguished  
253 by: low cost brands ( $A_p^3 = 0$ ), store brands ( $A_p^3 = 0.5$ ) or big  
254 brands ( $A_p^3 = 1$ ).
- 255 – *Attractiveness to promotion* ( $A_p^4 \in \{0, 0.5, 1\}$ ): A set of special  
256 offers may be applied to products based on three values corre-  
257 sponding to “no-reduction”, “discount” or “de-stocking”, for ex-  
258 ample.

### 259 3.1.3. Features for the agents

260 Each simulated customer  $a$  is also located in the environment, and is  
261 defined by a tuple  $\langle Ln_a, Lp_a, Sp_a, \eta_a(t), R_a \rangle$ :

- 262 •  $Ln_a$  is a *list of places* (a predefined path inside the store) that agent  
263  $a$  has to visit.
- 264 •  $Lp_a$  defines a *hash map containing all products of the store*. To each  
265 key (product)  $p \in Lp_a$  is associated a purchase intention at time  $t$ ,  
266 denoted  $I_a^p(t)$ . This intention influences the *agent attractiveness* for  
267 products w.r.t. a list of possible purchases.
- 268 •  $Sp_a$  is the *locomotion speed* for walking in the store. Recent studies  
269 concerning the speed for human beings show that this value is very  
270 hard to determine [6]. In the following, we consider that the speed for  
271 each agent is between 1 and 1.9  $\text{ms}^{-1}$ .
- 272 •  $\eta_a(t)$  represents a *temporal pressure* expressing what a customer is  
273 able to spend in “energy” (normalized between 0 and 1). From an  
274 initial temporal pressure  $\eta_a(0)$  corresponding to the temporal pressure  
275 felt before entering the shop, the temporal pressure evolves during the  
276 simulation *as function of time*  $t$ : the more the pressure rises, the less  
277 agent  $a$  has “energy” to continue its shopping; and consequently, this  
278 temporal pressure tends to zero.

- 279 •  $R_a$  represents a vector of parameters describing four receptiveness pa-  
 280 rameters:
- 281 – *Receptiveness to height* ( $R_a^1 \in [0, 1]$ ) which influences its eye-  
 282 catching and hand-catching capacities.
  - 283 – *Receptiveness to budget* ( $R_a^2 \in [0, 1]$ ) which expresses a price level  
 284 that it is willing to support financially, more than a spending  
 285 limit.
  - 286 – *Receptiveness to brand* ( $R_a^3 \in \{0, 0.5, 1\}$ ) which represents the in-  
 287 clination for low costs, prestige brands or greater range of prod-  
 288 ucts.
  - 289 – *Receptiveness to special offers* ( $R_a^4 \in \{0, 0.5, 1\}$ ) which character-  
 290 izes the sensitivity to marketing operations.

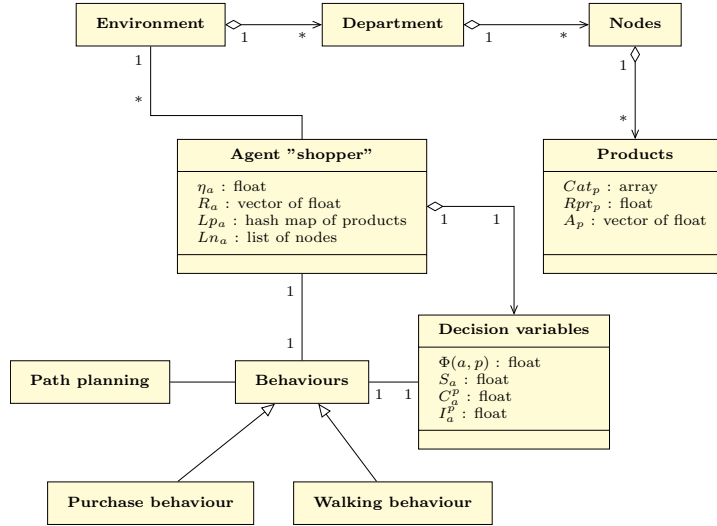


Figure 1: UML diagram of our simulation model

291 In short, Figure 1 gives an overview of our multi-agent model: an envi-  
 292 ronment divided into departments, each of them composed of several nodes;  
 293 products located on these nodes and finally agents playing the role of hu-  
 294 man customers. Each agent owns a behaviour subdivided in two activi-  
 295 ties: a walking activity and a purchase activity. They use different decision  
 296 variables and the computation based on agents and products features is  
 297 explained in detail in the next subsection.

298 3.2. Description of agents' behaviour

299 Within the simulation, the agents act according to a **finite-state** automaton, where all the transitions between states are deterministic (Figure 2).  
 300 The agents' behaviour is thus expressed through five elementary states (related to the different *actoms* introduced previously). In State 1, agent  $a$   
 301 has two possibilities: *to move towards its next location* if its list  $Ln_a$  is not empty (State 2), otherwise *to pay at checkout and exit the shop* (State 5).  
 302 From State 2, the agent can move again to the next location if it is far from the next location for a product (it remains in State 2). If this condition is  
 303 not true,  $a$  chooses the two following steps: *to evaluate products situated close to the current location* (State 3), and then *to deliberate* (*i.e.*, to decide  
 304 to purchase or not a product) (State 4). After its decision to buy or not a product, agent  $a$  is again in State 1. These different states are detailed  
 305 below.

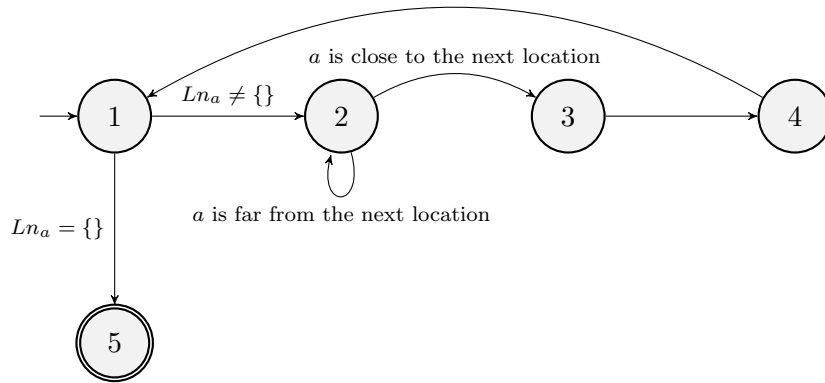


Figure 2: **Finite-state** automaton describing agents' behaviour

312 3.2.1. State 1: *Choosing the next location*

313 When agent  $a$  enters the store (or leaves its current position), it has to  
 314 choose its next destination. Based on  $Ln_a$ , this agent estimates its current  
 315 position on the graph to determine the next node to move. At its creation,  
 316 the first move corresponds by default to the first step of its *ideal path* inside  
 317 the store ( $Ln_a$  is initially computed).

318 For all connected nodes  $n$  from the current one, a score  $S_a(n)$  is computed  
 319 (Equation 1), reflecting the overall attraction of these nodes, according to  
 320 the products that  $n$  contains. To find the next location to visit, a score is  
 321 evaluated for all products  $p_i$  situated at node  $n$  using a matching function,

322 noted by  $\Phi(a, p_i)$ . The matching function models the attraction between a  
 323 product and a customer, and it could be both used to decide for the purchase  
 324 intention and to determine the next location for customer. Details on the  
 325 function  $\Phi$  will be given in Section 3.3.

$$S_a(n) = \sum_{p_i \in Lp_a(n)} \Phi(a, p_i) + \begin{cases} 1, & \text{if } n \in Ln_a \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

326 where  $S_a(n)$  represents the score of considered node  $n$ ,  $Lp_a(n)$  accounts for  
 327 the set of products attached to this node, lastly  $Ln_a$  is the pre-calculated  
 328 trajectory for  $a$ .

329 The choice of the next location is based on the following idea: rather  
 330 than visiting the next node in  $Ln_a$ , agent  $a$  could change its trajectory, if an  
 331 alternative node  $n'$  gives a higher score (*i.e.*, it owns a significant number  
 332 of attractive products). If  $n'$  is not the next element of its trajectory  $Ln_a$ ,  
 333 we add it to  $Ln_a$ , between the current and next nodes. Note that the node  
 334  $n$  draws a direct benefit, if this node has initially been specified in  $Ln_a$ .

### 335 3.2.2. State 2: *Walking towards the next location*

336 From its current node, the agent moves towards the next node of  $Ln_a$ .  
 337 This activity involves two complementary sub-tasks: (i) *walking* in the store  
 338 and *avoiding* different obstacles or customers, and (ii) *perceiving* the envi-  
 339 ronment.

340 The walking sub-task is derived from two standard algorithms. To reach  
 341 a goal, each agent may produce a *plan* based on a graph by applying a path  
 342 planning algorithm. The plan is based on a set of *transit nodes*. Then, at  
 343 each iteration, the agent can move between two nodes by avoiding the differ-  
 344 ent obstacles (shelves, other agents). More precisely, the agents / customers  
 345 plan their trip inside the store with a standard A\* algorithm [15]. Between  
 346 two nodes of the graph, the agent's displacements take place in a continu-  
 347 ous space. This walking activity is thus ensured by a well-known steering  
 348 algorithm [34].

349 In this model, a goal node may be defined by the purchase of products,  
 350 checkouts or exits of a store, and unavoidable nodes (*i.e.*, the agent must pass  
 351 by determined nodes because it belongs to a specific group of customers).  
 352 At the end of its move, the current node is equal to the next node in list  
 353  $Ln_a$ .

354 During its walking in the metric space, the simulated customer perceives  
 355 the other agents inside a cone of vision. Then, it computes a perceived  
 356 density  $\delta_a(t)$ , which is equal to the ratio between the space occupied by these

357 agents and the surface of this perceived cone. This density thus changes its  
 358 temporal pressure (Equation 2).

$$\eta_a(t) = \eta_a(0) - \delta_a(t), \quad (2)$$

359 where  $\eta_a(0) \in [0, 1]$  is the initial temporal pressure and  $\delta_a(t) \in [0, 1]$  is a  
 360 perceived density at time  $t$ .

361 A high density of agents only temporarily influences the temporal pres-  
 362 sure. It means that an agent arriving in a very busy department will have a  
 363 high temporal pressure for visiting it. Once arrived in a less crowded aisle,  
 364 the agent will get back to a normal temporal pressure. This change should  
 365 still have an impact on the purchase intention. At each change of the tem-  
 366 poral pressure, the purchase intention  $I_a^p(t)$  for each product  $p$  of list  $Lp_a$  is  
 367 also updated (Equation 3).

$$I_a^p(t) = \eta_a(t) \cdot I_a^p(t - 1), \quad (3)$$

368 where  $\eta_a(t) \in [0, 1]$  is the temporal pressure at instant  $t$  and  $I_a^p(0) \in [0, 1]$  is  
 369 the initial purchase intention of product  $p$  defined by  $Lp_a$ .

370 We suppose that the profiles of human customers should be known, which  
 371 will **make our** multi-agent simulation more **realistic**. To characterize these  
 372 profiles into different groups of agents, we propose to use clustering tech-  
 373 niques determining the typical trajectories (see Subsection 4.2.2).

### 374 3.2.3. State 3: *Evaluating products close to the current node*

375 When the agent is present at a current node  $n$ , it retrieves the list of  
 376 products and estimates the interest of different products associated with this  
 377 node. For each product  $p \in Lp_a(n)$ , the agent will evaluate its contextualized  
 378 purchase intention  $C_a^p(t)$  (Equation 4).

$$C_a^p(t) = \frac{I_a^p(t) + \Phi(a, p)}{2}, \quad (4)$$

379 where  $I_a^p(t)$  represents the purchase intention of agent  $a$  for product  $p$  (at  
 380 time  $t$ ) and  $C_a^p(t)$  is a contextualized intention for a product  $p$  (at time  $t$ ).

381 Agent  $a$  selects subsets of products having an interest called  $Lp_a^{Sel}(n)$ . It  
 382 determines products  $p_k$  for each category  $Cat_j$  (*i.e.*, it searches the *equiva-*  
 383 *lent* products belonging to the same category), and estimates the set of the  
 384 maximum values  $C_a^{p_k}(t)$  (Equation 5). At the end of this step, it obtains the  
 385 products with the highest scores available on this node.

$$Lp_a^{Sel}(n) = \left[ \max_{p_k \in Cat_1} C_a^{p_k}(t), \dots, \max_{p_k \in Cat_j} C_a^{p_k}(t), \dots, \max_{p_k \in Cat_l} C_a^{p_k}(t) \right] \quad (5)$$

386 where  $l$  is the number of available categories on node  $n$ .

#### 387 3.2.4. State 4: Deliberation

388 Agent  $a$  evaluates individually each product of [all](#) previous lists  $Lp_a^{Sel}$  and  
 389 decides to purchase the product or not by a comparison of its contextualized  
 390 purchase intention with a certain *purchase threshold*  $SdI$ . Equation 6 refers  
 391 to the set of products bought by agent  $a$  at its node  $n$ :

$$\{p \in Lp_a^{Sel}(n) / C_a^p(t) \geq SdI\}. \quad (6)$$

392 Once product  $p$  is bought, its price  $Rpr_p$  is used to update the amount  
 393 of money spent by the agent. In addition, the purchase intention of all  
 394 products belonging to the same category is set to zero in  $Lp_a$  (to avoid that  
 395 a customer purchases all products of a same category: these products could  
 396 also be located at another node).

397 For a [purchased](#) product, a purchase intention of associated products  
 398 could be strengthened by the use of some association rules. We will detail  
 399 this aspect in Subsection 4.2.3.

#### 400 3.2.5. State 5: checkout or exit of a store

401 When agent  $a$  has finished the shop exploration and/or has found all  
 402 products of its shopping list, it moves towards the exit, pays at checkouts  
 403 and leaves the shop. It plans a trajectory and chooses the checkout line  
 404 according to two criteria, the distance to reach the checkout line and the  
 405 number of agents already waiting to pay. Agent  $a$  tries to minimize its  
 406 estimation  $S_{a,c_j}$  (Equation 7) for the checkout line  $c_j$  (the number is fixed  
 407 *a priori*) [defined as](#)

$$S_{a,c_j} = (1 + Q_{c_j}) \cdot \Delta(a, c_j), \quad (7)$$

408 where  $Q_{c_j}$  is the size of the queue at the checkout line  $c_j$  and  $\Delta(a, c_j)$  is the  
 409 [Euclidean distance between the location of agent  \$a\$  and checkout line  \$c\_j\$ .](#)

410 After the selection of the checkout line, the agent moves towards the  
 411 queue and keeps a fixed distance from the last agent (the value is similar for  
 412 all agents). As in real life, the agent waits until the time of payment (the  
 413 duration is proportional to the number of purchased products), and then it  
 414 moves to the nearest exit (during this activity, the other agents continue to

415 move in this queue). After the shop’s exit, [the agent](#) is removed from the  
 416 simulation.

417 In the following, we focus on the way of evaluating the matching function  
 418 for modelling the attraction between a product and a customer.

### 419 3.3. Modelling the product-customer attraction

420 The proposed behavioural model relies on a principle of attraction be-  
 421 tween a customer and a product. In our model, the interaction between an  
 422 agent (namely  $a$ ) and a product (called  $p$ ) will depend on a matching (called  
 423  $\Phi(a, p)$ ) between these two vectors:  $R_a$  (receptiveness features for the cus-  
 424 tomer agent) and  $A_p$  (features of the product). Equation 8 characterizing  
 425 this matching, is a weighted sum of functions  $f_i$ , merging the vectors  $A_p$  and  
 426  $R_a$  in our case:

$$\Phi(a, p) = \frac{1}{4} \sum_{i=1}^4 f_i(A_p^i, R_a^i). \quad (8)$$

427 The different functions  $f_i$  are defined following the type of used data:  
 428 quantitative or qualitative information.

429 When  $A_p^i$  and  $R_a^i$  concern quantitative information (such as price, height,  
 430 etc.), the expression for  $f_i$  is defined by Equation 9. The result tends to 1  
 431 when the attractiveness and the receptiveness components are close. On the  
 432 contrary, when the difference between these parameters is significant, the  
 433 result tends to 0. For example, function  $f_1$  compares the positions in height  
 434 of the customer’s head with the product on the shelf, [while](#)  $f_2$  compares the  
 435 price level (budget) that the agent is willing to support financially with the  
 436 price of product:

$$f_i(A_p^i, R_a^i) = 1 - |A_p^i - R_a^i| \quad \text{where} \quad A_p^i, R_a^i \in [0, 1]. \quad (9)$$

437 When  $A_p^i$  and  $R_a^i$  concern qualitative or categorical information, ex-  
 438 pressed as ordinal variables (*e.g.*, special offers, brand category),  $f_i$  is de-  
 439 fined by Equation 10. For example, if agent  $a$  is absolutely not sensitive to  
 440 the brands (*i.e.*,  $R_a^3 = 0$ ), then the result of function  $f_3$  will tend to zero.  
 441 Conversely, if a customer is very [influenced by](#) the reputation of the brand  
 442 for the purchased product  $p$  (*i.e.*,  $R_a^3 = 1$ ), then he/she will be particularly  
 443 attracted to products of major brands ( $f_3 = 1$  if  $A_p^3 = 1$ ) and will reject  
 444 low-[cost](#) products. Similarly,  $f_4$  is used to combine promotions and special  
 445 offers with the receptiveness of customers.



$$f_i(A_p^i, R_a^i) = A_p^i \cdot R_a^i + \frac{(1 - R_a^i)}{2} \quad \text{where } A_p^i, R_a^i \in \{0, 0.5, 1\}. \quad (10)$$

446 To illustrate how these functions  $f_i$  are combined to determine a purchase  
 447 decision, we consider a simple example involving three customers and  
 448 two products in [the same category](#). Table 2 illustrates the motivations of  
 449 three customers. The first customer (called  $a_1$ ) is a young student with the  
 450 following features: 1.70m tall, a very limited budget, no particular interest  
 451 for prestige brands but he/she is very [interested by](#) promotions. The second  
 452 customer (called  $a_2$ ) is a senior executive (comfortable standard of living).  
 453 He/She is [also](#) 1.70m tall, very attracted by famous brands, but he/she pays  
 454 little attention to promotion. The third customer (called  $a_3$ ) is a parent,  
 455 measuring 1.65 meter. He/She wants to manage his/her budget, appreciates  
 456 brands but no more, and finally he/she enjoys promotions and special offers.  
 457 These three customers must choose between two products (Table 3). The  
 458 first product (called  $p_1$ ) comes from a high-end prestige brand, situated on  
 459 the top of the shelf, quite expensive and presented with a special offer. The  
 460 second one (called  $p_2$ ) comes from its own supermarket label proposed with  
 461 a limited price, with no promotion and situated at eye level in the shelf.

Table 2: Parameters of Receptiveness for three kinds of customers ( $a_1$ ,  $a_2$  and  $a_3$ )

Customer	Receptiveness to			
	Height ( $R_{a_i}^1$ )	Budget ( $R_{a_i}^2$ )	Brand ( $R_{a_i}^3$ )	Promotion ( $R_{a_i}^4$ )
$a_1$	0.7	0.2	0	1
$a_2$	0.7	1	1	0
$a_3$	0.65	0.6	0.5	0.5

Table 3: Parameters of Attractiveness for two different products ( $p_1$  and  $p_2$ )

Product	Attractiveness to			
	Height ( $A_{p_j}^1$ )	Price ( $A_{p_j}^2$ )	Brand ( $A_{p_j}^3$ )	Promotion ( $A_{p_j}^4$ )
$p_1$	0.9	1	1	0
$p_2$	0.66	0.7	0.5	1

462 Table 4 illustrates the different matchings  $\Phi$  for two products ( $p_1$  and  $p_2$ )  
 463 and three agents ( $a_1$ ,  $a_2$  and  $a_3$ ). We [notice](#) that, although agents  $a_1$  and

464  $a_3$  own different profiles, they are attracted to  $p_2$  ( $a_2$  prefers to select the  
 465 product  $p_1$ ). For agent  $a_1$ , the promotion proposed for  $p_2$  offsets the height  
 466 of the store shelf. For agent  $a_2$ , the brand and the price are interesting for  
 467 its features, and naturally it guides its choice for  $p_1$ . Finally for  $a_3$ , the  
 468 special offer and the height influence its decision towards product  $p_2$ .

Table 4: Example of matching between Attractiveness of product and Receptiveness of customer

Customer	$f_1$		$f_2$		$f_3$		$f_4$		$\Phi(a_i, p_j)$	
	$p_1$	$p_2$	$p_1$	$p_2$	$p_1$	$p_2$	$p_1$	$p_2$	$p_1$	$p_2$
$a_1$	0.8	0.96	0.2	0.5	0.5	0.5	0	1	0.375	0.74
$a_2$	0.8	0.96	1	0.7	1	0.5	0.5	0.5	0.825	0.665
$a_3$	0.75	0.99	0.6	0.9	0.75	0.5	0.25	0.75	0.5875	0.785

469 The next section presents the validation of this model through various  
 470 simulation scenarios.

#### 471 4. Experiments and Validation

472 The previously introduced agent model was used to build a simulation  
 473 of a supermarket. The model was implemented in Java using the platform  
 474 *Repast Simphony*<sup>7</sup> [30]. In the following, we give details concerning the  
 475 real case used for our experiments (Sect. 4.1). Then, we explain how we  
 476 calibrate our agent model (Sect. 4.2). Finally, we present the experiments  
 477 and the obtained results (Sect. 4.3).

##### 478 4.1. Description of our case study

479 The store under consideration is an existing supermarket located in Cal-  
 480 ifornia (USA)<sup>8</sup>. The trading area is about 4,150 m<sup>2</sup>. The store has two  
 481 entrances, on each side of the check lanes area. It has nine departments  
 482 (Figure 3): Bakery, Meat, Pharmacy, Dairy, Produce, Frozen foods, Cook-  
 483 ies & Crackers, Deli and Groceries.

<sup>7</sup>Repast, acronym for REcursive Porous Agent Simulation Toolkit, is a tool (imple-  
 mented in Java) offering designers different features such as graphical interfaces, a be-  
 havioural model for agents and a coupling with external tools (e.g., Matlab, R).

<sup>8</sup>We would like to thank the Sorensen Associates (USA) for this case study: <http://shopperscientist.com/resources/sorensen-associates/index.html>, Accessed: 2020-01-28

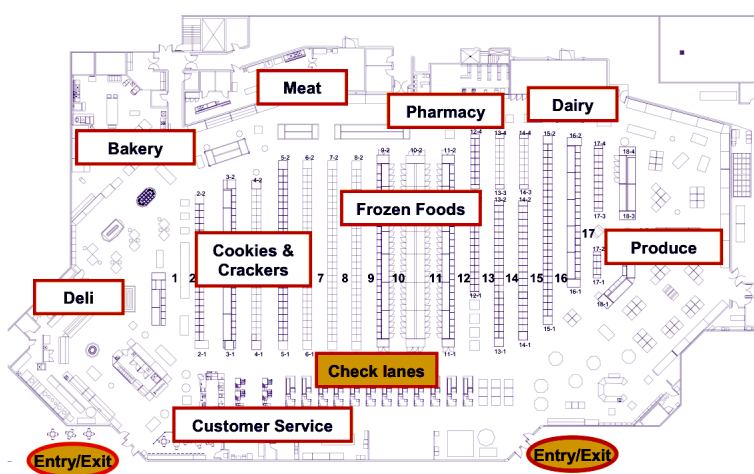


Figure 3: Major department locations

484 For this case study, we know the floor plan, a list of products available on  
 485 shelves, as well as their price, the current promotions, their categorization  
 486 (for example, granny apples belong to the subcategory “Apples”, which is  
 487 a part of the sub-category “Fresh Fruit” which is included in the category  
 488 “Produce”), their relative positioning (top-of-the-range, mid-range or low-  
 489 cost products), their location in the department and their placement on the  
 490 shelf (height).

491 In addition, we own two kinds of collected data related to the customers  
 492 of this supermarket: receipts and individual trips. Receipts were collected  
 493 during twenty-eight days. Each week, the store registers around 35,000  
 494 transactions representing about 285,000 items purchased per week (220,000  
 495 receipts). Individual trips of customers were also recorded during the same  
 496 period. Identification tracking (*PathTracker*®) tags based on Radio Fre-  
 497 quency IDentification (RFID) technology [41] were placed on approximately  
 498 250 shopping carts and baskets. 1,500 antennae were located around the  
 499 perimeter of the store to detect signals transmitted by tags at every second.  
 500 An antenna covers an area of about  $1.6m^2$ . A server captures the coordinates  
 501 of tags with regard to the wireless antennae. Each sequence of coordinates  
 502 represents the path for each customer as he/she travels throughout the store.  
 503 Unfortunately, the number of recorded trajectories (2,076 trajectories) is in-  
 504 ferior to the number of receipts since the displacements of customers have  
 505 been registered only for some shopping carts and baskets (those equipped  
 506 with RFID technology). Moreover, the trajectories are not linked to receipts

507 due to privacy reasons.

## 508 4.2. Calibration of our model

509 We propose to calibrate our model and particularly  $Lp_a$  and  $Ln_a$  with  
510 the receipts and individual trips previously described. In this context, we  
511 used data-mining techniques (4.2.1) namely clustering (4.2.2) and associa-  
512 tion rules mining (4.2.3).

### 513 4.2.1. Data-mining techniques useful for our model

514 The Data Mining techniques are widely exploited in the context of cus-  
515 tomer segmentation (see [publications](#) related to RFM<sup>9</sup> analysis based on  
516 K-means, Decision Trees, ANNs, etc.) [17]. To the best of [our](#) knowledge,  
517 in this domain, supervised techniques are mainly used in order to build a  
518 predictive analysis on possible future customer behaviours. This requires  
519 the use of the results (labels) given by the clustering tools [48]. The use  
520 of supervised tools would also imply [the need of](#) expert knowledge for the  
521 construction of [trajectories](#) (walking paths) labels, which seems [like](#) a hard  
522 task.

523 In our case, we aim to reproduce in simulation the observed behaviours,  
524 not to predict the category of a future customer. Let us recall that the  
525 link between wandering data and receipt data is not known. Even though  
526 the labels of customer profiles exist by using fidelity cards and purchase  
527 receipts, it is not possible to use them for the classification of the walking  
528 activity. For all these reasons, unsupervised methods have been preferred to  
529 calibrate our model for the real supermarket described before. Two classical  
530 techniques was used: clustering and association rules.

531 Clustering techniques are requested in order to extract a number of  
532 groups sufficiently representative of walking behaviours within the shop.  
533 Each walking pattern has to be different enough from the other groups and  
534 homogeneous within the same group [14]. These clusters will make it pos-  
535 sible to initialise the agents' characteristics, such as the predefined list of  
536 places  $Ln_a$  to be visited, and making it possible to seed the simulation with  
537 representative walking paths.

538 Association rules are also requested to extract complementary knowledge  
539 from cash receipts. The idea is to reinforce the purchase intention for a  
540 product according to its relationship with other products. Learnt from past  
541 purchase records, these rules help to modify, during the simulation, the

---

<sup>9</sup>Recency Frequency Monetary value

542 purchase intention  $I_a^p$  of other products  $p$  in  $Lp_a$ , for which the relationship  
 543 with the products in the basket is significant. An extension for association  
 544 rules could be possible by taking advantage of works proposed by Weng  
 545 et al. [46].

#### 546 4.2.2. Clustering on trajectories

547 As is traditionally done in machine learning, our set of trajectories has  
 548 been split in two: 1,038 trajectories were used for learning clusters, the rest  
 549 was used to validate the results.

550 Many techniques have been proposed in literature for mining trajectory  
 551 data [50]. Since our raw trajectories are time-stamped, it is possible to  
 552 compute a single matrix in which each column is labelled by a node of the  
 553 store, and each line is a recorded trajectory [25].

554 The clustering algorithm used on the resulting matrix is  $K$ -means [32],  
 555 tailored to the use of binary vectors. To fix the number  $k$  of clusters, we  
 556 tested different values between 5 and 60, and our choice is based on the ob-  
 557 tained results by analysing cluster variety and compactness (using the tool  
 558 *Silhouette* with Matlab <sup>10</sup>). In our context, the real data give interesting  
 559 results for  $15 \leq k \leq 30$ . 15 clusters are sufficient to obtain different be-  
 560 haviours; whereas 30 clusters give more details with a best precision (each  
 561 cluster **corresponds to** at least 1% of the population).

562 Finally, we choose 30 clusters (which means 30 types of agents with  
 563 different strategies and aims) for the experiments. For each cluster, we  
 564 use the centre to extract the trip points needed to initialize  $Ln_a$  in the  
 565 simulation. The cardinality of each cluster can be used to generate the  
 566 agent population. Table 5 shows the percentage representativeness of each  
 567 cluster.

Table 5: The clusters obtained and their representativeness (%)

Cluster 1	7 %
Clusters 3, 6, 14, 15, 17, 29	6 %
Clusters 4, 10, 22, 27	4 %
Clusters 8, 11, 13, 16, 19, 25, 26	3 %
Clusters 5, 12, 18, 20, 21, 23, 24, 30	2 %
Clusters 2, 7, 9, 28	1 %

<sup>10</sup><http://fr.mathworks.com/help/stats/silhouette.html>, Accessed: 2019-05-20

568 4.2.3. Information extraction from receipts

569 To initialize  $Lp_a$  of each agent, we started by sorting all the products of  
 570 the supermarket from the most sold to the less sold. The obtained ranking  
 571 allowed us to give the initial purchase intention  $I_a^p$ . Since the collected  
 572 trajectories and receipts are not linked together, we consider, for the ex-  
 573 periments presented below, that the agents are all initialized with the same  
 574  $Lp_a$ . However, during the simulation, this list will evolve according to each  
 575 agent’s specific parameters: the temporal pressure  $\eta_a(t)$  and the perceived  
 576 density  $\delta_a(t)$ . Moreover, according to  $Ln_a$  list, not all agents will purchase  
 577 the same products.

578 To dynamically modify the purchase intentions of agents ( $I_a^p$ ) for certain  
 579 products, our model includes the possibility of using association rules which  
 580 express correlations between products. These association rules can be ex-  
 581 tracted (off-line) from the 220,000 receipts. As a reminder, such rules draw a  
 582 relation between a set  $X$  corresponding to already purchased products, and  
 583 a set  $Y$  of products implied by  $X$  (if the condition  $p \in X$  is verified, then  
 584 the agent will be able to purchase products  $Y$ , and  $Y$  may be thus added to  
 585 the shopping list  $Lp_a$  for agent  $a$ ). Two metrics are associated with these  
 586 rules: a support (called  $S$ ) and a confidence (called  $C$ ). *Support* refers to  
 587 the part of receipts containing  $X$ , and *confidence* is the part of receipts con-  
 588 taining  $X$  and  $Y$ . For association rules mining, we used a classical algorithm  
 589 (Apriori) [2] by setting support  $S$  to 1 % and confidence  $C$  to 30 %.

Table 6: Examples of association rules

$X$	$Y$	$S$	$C$
Bananas	Milk	4.2 %	39.3 %
Cold meats	Bays	2.3 %	41.5 %
Apples ; Yoghurts	Bananas	1.1 %	61.8 %
Sandwich bread ; Breakfast	Milk	1.0 %	41.4 %
Fresh tomatoes	Garlic & Onions	1.9 %	35.2 %
Potatoes	Garlic & Onions	1.6 %	37.0 %
Eggs ; Milk	Sandwich bread	1.2 %	39.4 %
Bananas ; Sandwich bread	Milk	1.8 %	53.0 %
Bread ; Breakfast	Sandwich bread	1.0 %	41.0 %
Children’s products ; Sugared cereals	Yoghurt	1.0 %	30.1 %

590 Table 6 presents some association rules among 118 rules deduced by this  
 591 algorithm from our available data. These 118 rules show a strong representa-  
 592 tion of fruits and dairy products. This result may be a surprise at first sight;  
 593 but we can explain it by the categories of products, which are considered as  
 594 perishable products and are bought more often than other products. In our

595 analysis, the support of these rules relies on the rise in frequency, and not  
596 in the quantity of purchased products. Moreover, these results could also  
597 express a purchase behaviour, depending on the socio-cultural habits [16]  
598 related to the location of the store.

599 To sum up, the real data of our case study allowed us to parametrize  
600 a large part of our agent model (products list  $Lp_a$  and trajectory  $Ln_a$ ).  
601 Finally, only two parameters have been empirically set, *i.e.*, the threshold  
602 for purchase intention ( $SdI = 0.8$ ), and the initial temporal pressure (for all  
603 agents  $a$ ,  $\eta_a(0) \in [0.990, 0.999]$ ). In the experiments presented in the next  
604 section, these two parameters have been adjusted approximately to meet the  
605 real average amount of purchases.

### 606 4.3. Evaluation of our *simulation tool*

607 Four experiments were carried out for the validation of our *simulation*  
608 *tool* (Figure 4). The first two scenarios evaluate the tool based on the cur-  
609 rent real data (receipts and trajectories). The first scenario aims to compare  
610 the real data with measures obtained by our software (4.3.1), whereas the  
611 second one illustrates typical behaviours of customers (4.3.2). The last ex-  
612 periments aim to propose additional results for our tool (4.3.3), *i.e.*, to give  
613 some indications for the manager during a traffic rise, and to point out the  
614 scalability problem.

#### 615 4.3.1. Comparison between the real and simulated data

616 The evaluation describing the comparison between real and simulated  
617 data is based on objective and subjective criteria. This first experiment  
618 gives objective metrics, such as time, distance travelled and expense for the  
619 customers; and shows subjective results using heat maps. These *simula-*  
620 *tion* results are averages over 25 runs. For each run (corresponding to the  
621 simulation of a normal operational day of the supermarket), the shopper  
622 population is the same (only their arrival in the store may be different):  
623 5,000 customers following the percentage of clusters (Cf. Table 5).

624 *Microscopic metrics: time, distance travelled and expense.* From different  
625 private technical reports and audits, we can consider three criteria used to  
626 compare our results with the real case (Table 7). The first two criteria aim  
627 to analyse microscopic metrics: (i) the average time spent through the store  
628 and (ii) the average distance of their trip. The last criterion characterizes  
629 the purchase behaviour (*i.e.*, the average expense per customer).

630 We find that the behaviours appear to be fairly diversified. For example,  
631 the distances travelled vary from 108 metres (corresponding to an agent with

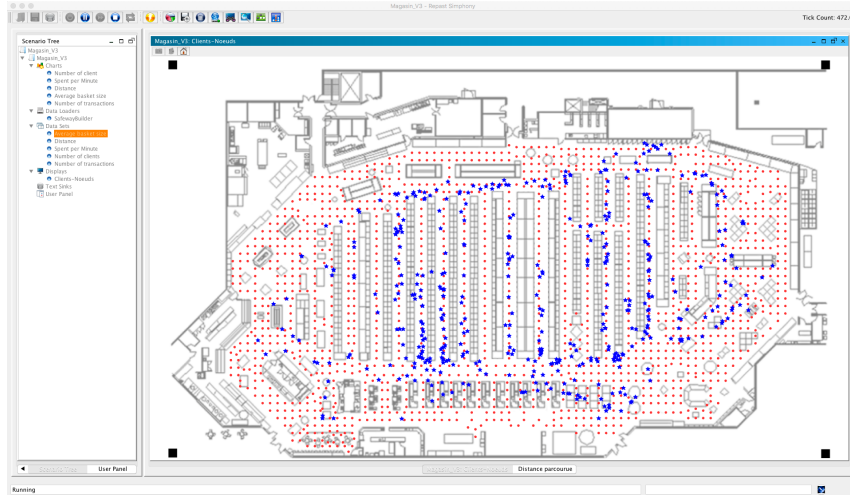


Figure 4: Illustration of our simulation tool representing the store and its customers. Red points correspond to the nodes of the graph; whereas blue stars represent the simulated shoppers.

632 a very short path and buying only few items) up to 3 km (an agent with a  
 633 important list of items to purchase). The same observation can be done for  
 634 the time spent inside the store, which can go from 2min up to more than  
 635 1h30min.

Table 7: Comparison between real data and our simulation

	Real data	Simulated Results	Difference
Time spent inside the supermarket (minutes)	23	19	21 %
Distance travelled inside the supermarket (meters)	2,067	1,533	34 %
Expense (dollars)	27.1	25.85	4.8 %

636 Considering the average expense, the results of the simulation are very  
 637 satisfying since the difference is inferior to 5%. This result tends to show  
 638 that our purchase model is relevant and realistic for customer behaviours.

639 Now, considering the two other criteria (distance and time, which are  
 640 correlated), the differences are higher. To explain this phenomenon, we  
 641 compare the real trajectory with the trajectory computed by the path plan-



642 ner implemented in our tool. Two reasons can explain this difference with  
643 a similar shape of trajectories. Firstly, the path planning algorithm used  
644 in simulation tends to produce optimal trajectories minimizing the distance  
645 (which is not always true for humans). Secondly, the distances provided by  
646 real data are slightly overestimated since the path tracker system is based  
647 on the antenna located in the middle of the aisle. For example, zigzags in  
648 an aisle cannot be captured with the system.

649 *Macroscopic comparison using heat maps.* The displacements of agents have  
650 also been evaluated from a more macroscopic view (Figure 5). The two  
651 figures present the same information concerning the traffic for this store:  
652 the main congestion points (red or orange colours), the areas with very  
653 little traffic (blue or green colours) and the areas in yellow colour with an  
654 average attendance.

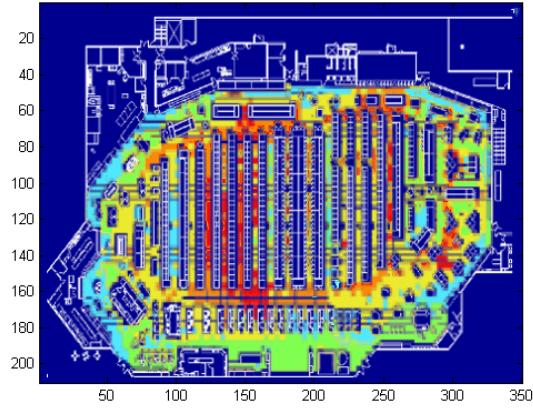
655 A heat map describing the store shelves traffic from the different au-  
656 dits is shown in Figure 5.1. This map is interesting because it results from  
657 behaviours of customers. For example, a main path which serves several  
658 shelves, has a high frequency (*i.e.*, a quasi-obliged path). Shelves including  
659 key products (which can be in a lot of carts) will also be crowded areas.  
660 Figure 5.2, obtained from simulations, presents the same information con-  
661 cerning the traffic for this store.

662 The comparison between these two figures underlines that the areas with  
663 high and very little traffic overlap. However, the areas with an average  
664 traffic may be different: the agents are autonomous and free to go (or not)  
665 through other trip points. These first results show thus that our model can  
666 produce real behaviours in terms of trajectories and purchases. In the next  
667 subsection, we go into more detail regarding the types of behaviour.

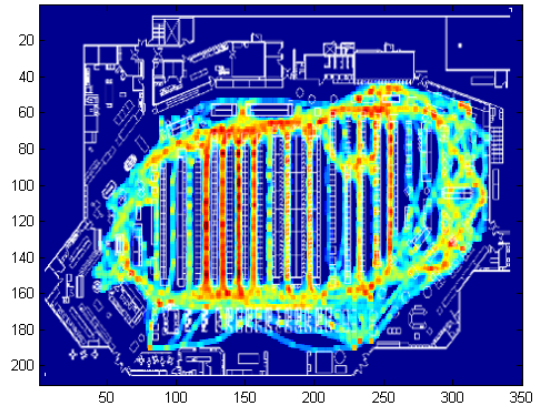
#### 668 4.3.2. Typology of customers

669 For this second experiment, the simulation results from agents belong-  
670 ing to the same cluster. Among 30 clusters, we would like to describe two  
671 examples illustrating the variability of walking behaviours: the first clus-  
672 ter describes behaviours based on a fast visit; the second cluster concerns  
673 behaviours based on a higher store exploration.

674 *Cluster 1 – Behaviours based on a fast visit at the store.* The cluster gath-  
675 ers the behaviours based on a fast visit, with very little space exploration  
676 and purchases very localized (Table 8). These purchases focus on the ar-  
677 eas between the checkout lines and the aisles end-caps (book or stationery  
678 aisles). The purchases are thus well targeted. We can imagine that these



1 Real heat map



2 Simulated heat map

Figure 5: Real and simulated heat maps

679 customers quickly come to buy their (daily) newspapers, for example. In  
 680 this context, their objective is not to purchase different products (for the  
 681 week). This cluster can be characterized by very short paths, with an aver-  
 682 age expense around \$1 and an important walking time related to reasoning  
 683 time (exceeding 60%).

Table 8: Results for the customer’s profiles for cluster 1

	Walking time (seconds)	Reasoning time (seconds)	Distance travelled (meters)	Expense (dollars)
Minimum	28	0	106.37	0
Mean	44.91	22.08	134.71	1.08
Maximum	80	56	179.50	8

684 *Cluster 2 – Behaviour based on a higher exploration of the store.* The second  
685 cluster (including different purchase behaviours) gathers behaviours based  
686 on longer distances with a higher purchase volume (Table 9). While some  
687 customers could leave the shop without purchasing anything, for cluster 2,  
688 the average basket is around \$88.22 with expenses reaching \$99 for other  
689 customers. The longest path may reach an increase of 32%, related to  
690 the shortest one (2,600 meters), also underlining a difference between the  
691 different paths.

Table 9: Results of profiles for cluster 2

	Walking time (seconds)	Reasoning time (seconds)	Distance travelled (meters)	Expense (dollars)
Minimum	685	1,461	2,594.77	0
Mean	944	2,395.33	2,895.88	88.22
Maximum	1,342	3,315	3,429.20	99

692 More generally, a deeper analysis on  $k$  clusters shows that certain clus-  
693 ters represent behaviours of very similar and homogeneous displacements;  
694 whereas other ones could present more varied displacements. Recall that  
695 our choice was to constrain the number of clusters by finding a compromise  
696 between a reasonable number of clusters and a greater homogeneity of their  
697 walking activities.

698 The following subsection gives additional results, which may also show  
699 other possibilities for our tool.

#### 700 4.3.3. Additional results

701 We propose to investigate our model performance through two additional  
702 experiments: an increase of the traffic and the scalability problem.

703 *Traffic rise.* In this scenario, the manager would assess the impact of higher  
704 traffic on the store’s turnover. From our point of view, this experiment con-  
705 sists in increasing the number of agents inside the simulation and evaluates  
706 the impact in terms of purchases.

707 The results of this experiment show that the average shopping cart tends  
708 to decrease, although the number of customers increases (Figure 6). This  
709 result is consistent with our behavioural model, where the temporal pressure  
710 can influence the purchase behaviour. If the perceived density is higher, the  
711 temporal pressure is stronger: *i.e.*, the customers become very impatient  
712 and they want to exit the shop as soon as possible (these behaviours lead  
713 them to reduce their purchases).

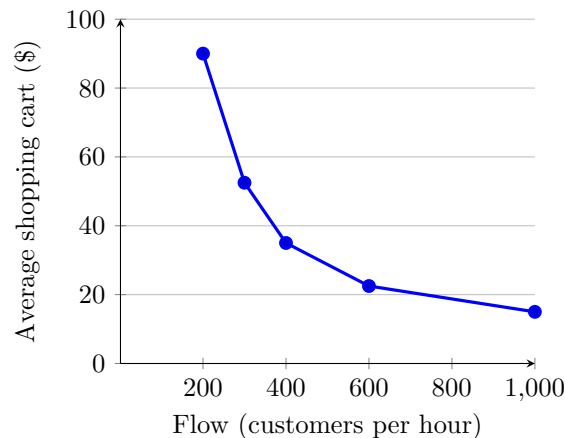


Figure 6: Trends of shopping carts according to customers flows

714 In the last experiment, we evaluate the ability of our tool to simulate a  
715 large number of agents, *i.e.*, to study the scalability of our simulation.

716 *Scalability problem.* To simulate a retail space requires to reproduce a sig-  
717 nificant number of customers. However, our model, by its nature, may be  
718 time-consuming. This experiment studies the ratio between simulated and  
719 computing times (Figure 7). Results are obtained (similar for the different  
720 experiments) on a computer equipped with a dual-core 2.66 GHz CPU and  
721 8 GB.

722 We can thus note that the simulation may be considered *real time* (*i.e.*,  
723 ratio of 1) approximately for 1,000 customers per hour. Our model is thus  
724 perfectly applicable to a medium-sized store, and in particular for the store

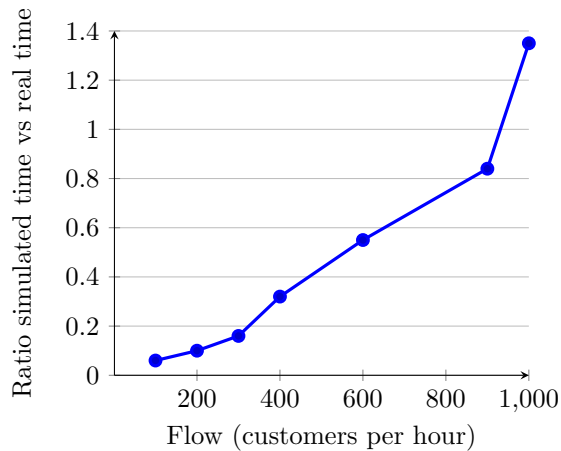


Figure 7: Ratio between simulated and computing times according to customers flows

725 being studied. In contrast, for more frequented stores or shopping malls,  
 726 we expect that it will be needed to reduce the computational [time](#) for our  
 727 approach.

728 Through the various experiments, we have shown that our model repro-  
 729 duces fairly accurately the customers travelling within the store (see heat  
 730 maps in Figure 5). The distances covered in simulation are lower, due to  
 731 the use of shortest path algorithms in our model. In terms of purchases,  
 732 the differences obtained between the reality and the simulation are less than  
 733 5%. Our model can [thus](#) reproduce the overall performance of a store: in  
 734 terms of attendance and sales. The scenarios concerning traffic rises show  
 735 that our [approach](#) could be useful to predict interesting metrics like the av-  
 736 erage expense for the customers. [However, it also highlights a drawback of](#)  
 737 [a real-time simulation by constraining the number of agents.](#)

738 In the following section, we discuss the potential and limits of our agent-  
 739 based model for the simulation of customers in a shop.

## 740 5. Discussion

741 The purpose of this section is to underline the potential of our approach  
 742 in light of our hypothesis ([Sect. 5.1](#)) and to mention thus [a few](#) limitations  
 743 ([Sect. 5.2](#)).

744 5.1. Potential of our approach

745 Although the analysis described by Badeig et al. [3] is more general, we  
746 think that it is well suited for our work. As suggested by their analysis,  
747 the design of tools based on multi-agent systems for complex applications  
748 requires four main properties (autonomy, proactiveness, context awareness  
749 and situatedness). Hereafter, some discussions on our approach are given  
750 for each property.

751 *Autonomy* suggests “the necessity to articulate three types of autonomous  
752 entities (human manager, artificial customers and simulated retail area)”.  
753 In our model, we assume that these entities are governed by different be-  
754 havioural rules that can vary from a supermarket to another depending on  
755 different socio-cultural aspects (localization of the supermarket, standard of  
756 living of customers, consumer habits, for example). We showed that the  
757 use of a data-driven approach (clustering on trajectories, association rules  
758 extracted from receipts) allows to take these socio-cultural aspects into ac-  
759 count and thus facilitates the articulation between these three entities.

760 *Proactiveness* is “the way to maintain a consistent coupling between pro-  
761 cessing structures and dynamic environments”. In our model, the proactiv-  
762 ity is expressed through the product-customer attraction which allows an  
763 agent to modify its trip inside the supermarket. Note that the temporal  
764 pressure can also influence this trip. Other facets could be achieved consid-  
765 ering groups of customers. A typical example is the purchase decision inside  
766 a family: what is the influence of children when a family goes shopping at  
767 the supermarket [45]? To address this issue, we think that our model should  
768 be extended with a decision mechanism able to take into account all individ-  
769 ual attractions for a group. Many options are available: for example, the use  
770 of an algorithm to produce a global attraction that merges how the group  
771 is interested (or not) by a product, the use of a task allocation algorithm  
772 to decide which member of the group is in charge of the purchase decision  
773 depending on the type of product (children for breakfast cereals and adults  
774 for wine as an example), etc. Sakas et al. [38] propose the concept of roles  
775 for these different agents (for example, a friend or a child from the family  
776 may influence the buyer), which could also change the purchases.

777 *Context-awareness* is the way “to the design of systems in which there  
778 is a need to constantly adapt to **evolving** situations that may be hard to un-  
779 derstand”. Among the possible evolutions that can impact a store, we think  
780 that it is difficult to really understand the possible evolutions of customer  
781 behaviours: How do their needs and desires evolve according to the seasons?  
782 How do they adapt their purchase depending on their financial situation?

783 To take these evolutions into account inside the simulation, it would be nec-  
784 essary to continuously assimilate real data in our agent model. To achieve  
785 this, we could consider dynamic and incremental machine learning tech-  
786 niques that could let agent clusters evolve. In the SHOPPER model, Ruiz  
787 et al. [36] propose another solution to define customer behaviours more real-  
788 istically. They use a Bayesian model which is based on the probabilities for  
789 buying a new product related to already purchased products. SHOPPER  
790 also considers important characteristics for purchase intention (price, inter-  
791 action between products, seasonal effects). We think that this model could  
792 replace association rules and thus describe agents with “on-line” behaviours  
793 in terms of purchase intention.

794 *Situatedness* is “supported by continuous interactions with physical el-  
795 ements in the environment like traces as a result of their activity”. The  
796 records based on cameras allow managers [to have a better knowledge of](#)  
797 [their customers](#). Although this property is not currently exploited, we can  
798 imagine that we could propose an in-depth analysis, in which different activ-  
799 ities could be evaluated. For example, Zhang et al. [49] use a video tracking  
800 system in combination with a hierarchical Bayes framework to observe the  
801 social influence of the environment (salesperson contact, interactions be-  
802 tween shoppers) over different activities: exploration of the store, produc-  
803 tion interaction, purchase, etc. The results of the simulation would be more  
804 interesting for the customers: our approach could thus measure the different  
805 parameters such as the walking time (a real speed of each agent, and not an  
806 average speed), the reasoning time for the purchases, the temporal pressure,  
807 etc.

## 808 5.2. Limits

809 We propose to focus essentially on three limits, *i.e.*, the choice of k-value  
810 for clusters of customers, the consistency of data structure and the capacity  
811 to simulate in real time [a very large shopping mall](#).

812 Firstly, some [publications](#) [26] have also described the machine learning  
813 techniques for determining *meta-clusters* based on transactions involving  
814 customers and products. The authors improve thus the clustering profile of  
815 the customers who purchase a product. However, we think that although  
816 their idea may be useful for our works, their objective is not focused on  
817 the retail simulation; but on the way of defining “good clusters” for the  
818 customers.

819 Secondly, unlike other existing models that are based only on receipts,  
820 our customer model is relatively generic and can be applied to stores other  
821 than supermarkets. For example, in the case of clothing stores, where each

822 entry does not necessarily imply a checkout, the use of trajectories makes it  
823 possible to obtain a much more precise and fair store visit. The counterpart  
824 is that our model is more complicated to configure. Indeed, it is always  
825 necessary that the data structures  $L_n$  and  $I_a^p(0)$  are consistent for each  
826 agent created in the simulation. For example, it is not conceivable to put  
827 a high initial purchase intention on fruits and vegetables if **the** list  $L_n$  does  
828 not contain any nodes of the fruits and vegetables department.

829 **Finally**, as seen in the previous section, our simulation tool currently  
830 struggles to exceed 1,000 customers per hour. To overcome this limitation,  
831 we can plan to use a hybrid multi-level approach. The idea is to make  
832 our model work simultaneously (at a microscopic level) and a macroscopic  
833 level [31]. This kind of approach has been already applied **successfully** for  
834 road traffic simulation [11], and we think it will be relevant in our context.  
835 Another idea is based on a more *active* environment, which may allow agents  
836 to reduce their cognitive tasks, called affordances [21, 22]. An affordance in  
837 this context, could be seen as an attraction of a product for the customer.

## 838 **6. Conclusion and future works**

839 In this paper, we have presented an approach to *mimic* the main cus-  
840 tomer activities (the walking and the purchase) inside a store. The resulting  
841 simulation can help store managers: to study the different flows, to identify  
842 the critical areas of congestion or to predict the impact for a new location  
843 of products, in order to increase the shop profitability and to improve the  
844 service for customers. In our approach, each simulated customer attempts  
845 not only to interact with those present in the store and its environment, but  
846 also to purchase different products.

847 We have presented the state of the art from different studies, particu-  
848 larly **from** two domains: psychological/marketing and multi-agent systems.  
849 These studies focused on **a** customer behaviour model based on actoms, such  
850 as mobility, stop/halt, prehension and purchase acts. The approaches inves-  
851 tigated different aspects of marketing (for example, the purchase intentions,  
852 interactions between employees and customers), but, to our knowledge, few  
853 **of them** really deal with these two aspects: purchase behaviour and walking  
854 activity in the store.

855 **Because of** these different considerations, our model is based on a multi-  
856 agent approach, **integrating** the representation of a store and the character-  
857 istics of products and customers (agents). The proposed model based on  
858 these actoms has an essential part, **namely** the modelling of the possible  
859 attractions between products and agents. To design them, we introduce a



860 matching function between different parameters, the attractiveness (for the  
861 products) and the receptiveness (for the agents).

862 This model was validated by simulating a real store, and by comparing  
863 the real shoppers and the simulated agents. This comparison confirmed the  
864 relevance of our approach. In particular, our tool is able to study the dif-  
865 ferent behaviours of customers and the resulting flows, to visualize a heat  
866 map of the store, to test new scenarios and to predict the impact of changes  
867 in the organization of the retail space. In this paper, we described four  
868 experiments. The first scenario showed that the results in terms of times,  
869 distances and expenses were very close to real data. Moreover, the conges-  
870 tion areas are easily visible by heat maps and are good *Key Performance*  
871 *Indicators* for store managers (they may thus change the organization of  
872 their store to avoid these critical areas). The second scenario evaluated the  
873 same simulations resulting from agents belonging to the same cluster. It  
874 is thus possible to observe the typology of customers by their specific be-  
875 haviours (the managers may thus propose different promotions well suited  
876 for these customers). The third experiment involved a traffic rise and evalu-  
877 ated the impact in terms of purchases (or expenses) for an increasing number  
878 of customers. In the last experiment about scalability aspects, we showed  
879 that the version of this tool was able to generate realistic scenarios in terms  
880 of simulated time.

881 **Future work related to our approach** is currently based on the four re-  
882 quirements (namely autonomy, proactiveness, context-awareness and situat-  
883 edness) introduced for the analysis of complex applications based on multi-  
884 agent models. Our approach is constrained by the obtained real data (in  
885 our study, the real data from paths and receipts are not directly associated),  
886 and it is necessary to consider real data more precisely. The autonomy could  
887 be improved for the typology of customers, for example, by **including results**  
888 concerning meta-clusters. Proactiveness may be seen in **situations** where a  
889 group (*e.g.*, a family) and not a set of independent customers, decides to  
890 purchase certain products. We think that the tool should be able to de-  
891 tect these relationships. Situatedness requires additional data concerning  
892 **customer behaviour** (*e.g.*, halting times in front of a product, questionnaires  
893 giving some explanations about their behaviours and their choices about the  
894 purchases). To complete the model, we investigate the adaptation of two  
895 approaches (multi-level or affordance models) for context awareness.

896 **Acknowledgments**

897 This research was supported by the competitiveness cluster PICOM. It  
898 was funded by the French Ministry of Industry and the region Hauts-de-  
899 France. We would like to thank the company Sorensen Associates for giving  
900 us access to real store data. The authors would like to thank Christophe Kol-  
901 ski (LAMIH, UPHF), Sanda Lefteriu (IMT Lille Douai) and Marius Silaghi  
902 (Florida Institute of Technology, FL, USA) for their useful comments on  
903 this paper. We thank also the anonymous reviewers for their constructive  
904 remarks.

905 **References**

- 906 [1] Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules  
907 between sets of items in large databases. In: Proceedings of the 1993  
908 ACM SIGMOD International Conference on Management of Data. SIG-  
909 MOD '93. ACM, New York, NY, USA, pp. 207–216.
- 910 [2] Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association  
911 rules in large databases. In: Proceedings of the 20th International Con-  
912 ference on Very Large Data Bases (VLDB). pp. 487–499.
- 913 [3] Badeig, F., Adam, E., Mandiau, R., Garbay, C., 2016. Analyzing multi-  
914 agent approaches for the design of advanced interactive and collabora-  
915 tive systems. *Journal of Ambient Intelligence and Smart Environments*  
916 *JAISE* 8 (3), 325–346.
- 917 [4] Bell, D., Mgbemena, C., 2018. Data-driven agent-based exploration of  
918 customer behavior. *Simulation* 94 (3), 195–212.
- 919 [5] Bonnin, G., 1998. L'observation des comportements au point de vente :  
920 vers une étude de la valorisation de la visite au magasin. Ph.D. thesis,  
921 Université de Bourgogne, France.
- 922 [6] Campanella, M. C., 21st Nov. 2016. Microscopic modelling of walking  
923 behaviour. Ph.D. thesis, Technische Universiteit Delft.
- 924 [7] Chandon, P., Hutchinson, J. W., Bradlow, E. T., Young, S. H., 2009.  
925 Does in-store marketing work? effects of the number and position of  
926 shelf facings on brand attention and evaluation at the point of purchase.  
927 *Journal of Marketing* 73 (6), 1–17.

- 928 [8] Chen, Y. S., Chang, C. H., 2012. Enhance green purchase intentions:  
929 The roles of green perceived value, green perceived risk, and green trust.  
930 *Management Decision* 50 (3), 502–520.
- 931 [9] DiMicco, J. M., Maes, P., Greenwald, A., Jul 2003. Learning curve: A  
932 simulation-based approach to dynamic pricing. *Electronic Commerce*  
933 *Research* 3 (3), 245–276.
- 934 [10] Donovan, R. J., Rossiter, J. R., Marcoolyn, G., Nesdale, A., 1994. Store  
935 atmosphere and purchasing behavior. *Journal of Retailing* 70 (3), 283  
936 – 294.
- 937 [11] Espi e, S., Gattuso, D., Galante, F., 2006. Hybrid traffic model cou-  
938 pling macro-and behavioral microsimulation. In: 85th Annual Meeting  
939 Transportation Research Board.
- 940 [12] Grewal, D., Krishnan, R., Baker, J., Borin, N., 1998. The effect of store  
941 name, brand name and price discounts on consumers’ evaluations and  
942 purchase intentions. *Journal of Retailing* 74 (3), 331 – 352, research  
943 Perspective on Retail Pricing.
- 944 [13] Guttman, R., Moukas, A., Maes, P., July 1998. Agent-mediated elec-  
945 tronic commerce: A survey. *The Knowledge Engineering Review* 13 (2),  
946 147–159.
- 947 [14] Han, J., Kamber, M., Pei, J., 2011. *Data mining: concepts and tech-*  
948 *niques* (3rd Edition). Elsevier.
- 949 [15] Hart, P. E., Nilsson, N. J., Raphael, B., 1968. A formal basis for the  
950 heuristic determination of minimum cost paths. *IEEE Transactions on*  
951 *Systems Science and Cybernetics* 4 (2), 100–107.
- 952 [16] Henry, W. A., 1976. Cultural values do correlate with consumer behav-  
953 ior. *Journal of marketing research* 13 (2), 121–127.
- 954 [17] Hiziroglu, A., 2013. Soft computing applications in customer segmen-  
955 tation: State-of-art review and critique. *Expert Systems with Applica-*  
956 *tions* 40 (16), 6491 – 6507.
- 957 [18] Jia, H., Zhou, S., Allaway, A. W., 2018. Understanding the evolution  
958 of consumer psychology research: A bibliometric and network analysis.  
959 *Journal of Consumer Behavior* 17 (5), 491–502.

- 960 [19] Karimi, S., 2013. A purchase decision-making process model of online  
961 consumers and its influential factor: a cross sector analysis. Ph.D. the-  
962 sis, University of Manchester, Manchester Business School.
- 963 [20] Kitazawa, M., Yamada, T., Takahashi, M., Terano, T., Nov 2013. How  
964 do customers buy them at a supermarket? behavior analysis from real  
965 observation and agent simulation. In: Proc. Second Asian Conference  
966 on Information Systems (ACIS2013). pp. 217–222.
- 967 [21] Klügl, F., 2016. Using the affordance concept for model design in agent-  
968 based simulation. *Ann. Math. Artif. Intell.* 78 (1), 21–44.
- 969 [22] Ksontini, F., Mandiau, R., Guessoum, Z., Espié, S., 2015. Affordance-  
970 based agent model for road traffic simulation. *Autonomous Agents and  
971 Multi-Agent Systems* 29 (5), 821–849.
- 972 [23] Ladwein, R., 1999. Le comportement du consommateur et de l’acheteur.  
973 Collection Gestion. Economica.
- 974 [24] Lam, S. Y., Vandenbosch, M., Hulland, J., Pearce, M., 2001. Evaluating  
975 promotions in shopping environments: Decomposing sales response into  
976 attraction, conversion, and spending effects. *Marketing Science* 20 (2),  
977 194–215.
- 978 [25] Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., Ma, W.-Y., 2008. Min-  
979 ing user similarity based on location history. In: Proceedings of the  
980 16th ACM SIGSPATIAL International Conference on Advances in Ge-  
981 ographic Information Systems. GIS ’08. pp. 34:1–34:10.
- 982 [26] Lingras, P., Elagamy, A., Ammar, A., Elouedi, Z., 2014. Iterative meta-  
983 clustering through granular hierarchy of supermarket customers and  
984 products. *Information Sciences* 257, 14–31.
- 985 [27] Liu, C., 2016. Worldwide retail Ecommerce sales: the eMarketer Fore-  
986 cast for 2016. Tech. rep., eMarketer.
- 987 [28] Mathieu, P., Panzoli, D., Picault, S., 2011. Format-store: a multi-agent  
988 based approach to experiential learning. In: Third International Con-  
989 ference on Games and Virtual Worlds for Serious Applications (VS-  
990 GAMES). pp. 120–127.
- 991 [29] Moulin, B., Chaker, W., Perron, J., Pelletier, P., Hogan, J., Gbei, E.,  
992 2003. Mags project: Multi-agent geosimulation and crowd simulation.

- 993 In: Spatial Information Theory. Foundations of Geographic Information  
994 Science. Springer, pp. 151–168.
- 995 [30] North, M. J., Collier, N. T., VoVos, J., Jan. 2006. Experiences creat-  
996 ing three implementations of the repast agent modeling toolkit. ACM  
997 Transactions on Modeling and Computer Simulation 16 (1), 1–25.
- 998 [31] North, M. J., Macal, C. M., St. Aubin, J., Thimmapuram, P., Bragen,  
999 M., Hahn, J., Karr, J., Brigham, N., Lacy, M. E., Hampton, D., May  
1000 2010. Multiscale agent-based consumer market modeling. Complexity  
1001 15 (5), 37–47.
- 1002 [32] Ordonez, C., 2003. Clustering binary data streams with K-means. In:  
1003 Proceedings of the 8th ACM SIGMOD workshop on Research issues in  
1004 data mining and knowledge discovery. pp. 12–19.
- 1005 [33] Parunak, H. V. D., Savit, R., Riolo, R. L., 1998. Agent-based mod-  
1006 eling vs. equation-based modeling: A case study and users’ guide. In:  
1007 Multi-Agent Systems and Agent-Based Simulation, First International  
1008 Workshop, MABS. Vol. 1534 of Lecture Notes in Computer Science.  
1009 Springer, pp. 10–25.
- 1010 [34] Reynolds, C. W., 1999. Steering behaviors for autonomous characters.  
1011 In: Game developers conference. pp. 763–782.
- 1012 [35] Richter, Y., Yom-Tov, E., Slonim, N., Apr. 2010. Predicting customer  
1013 churn in mobile networks through analysis of social groups. In: Proceed-  
1014 ings of the SIAM International Conference on Data Mining, SDM 2010,  
1015 April 29 - May 1, 2010, Columbus, Ohio, USA. SIAM, pp. 732–741.
- 1016 [36] Ruiz, F. J. R., Athey, S., Blei, D. M., 2018. SHOPPER: A probabilistic  
1017 model of consumer choice with substitutes and complements. CoRR  
1018 abs/1711.03560.  
1019 URL <http://arxiv.org/abs/1711.03560>
- 1020 [37] Said, L. B., Bouron, T., Drogoul, A., 2002. Agent-based interaction  
1021 analysis of consumer behavior. In: Proceeding of the First Interna-  
1022 tional Joint Conference on Autonomous Agents & Multiagent Systems  
1023 (AAMAS). pp. 184–190.
- 1024 [38] Sakas, D. P., Drivas, I. C., Kavoura, N., 2015. Simulation model for  
1025 commercial success of customer behaviour. Procedia Economics and  
1026 Finance 24, 598–605.

- 1027 [39] Schmitt, J., 2009. Parcours, déplacements et actions face au rayon:  
1028 mieux comprendre le comportement physique du consommateur en ma-  
1029 gasin pour mieux comprendre ses achats. Ph.D. thesis, HEC Paris,  
1030 France.
- 1031 [40] Schwaiger, A., Stahmer, B., 2003. SimMarket: Multiagent-based cus-  
1032 tomer simulation and decision support for category management. In:  
1033 Schillo, M., Klusch, M., Muller, J., Tianfield, H. (Eds.), Multiagent  
1034 System Technologies. Vol. 2831 of Lecture Notes in Computer Science.  
1035 Springer Berlin / Heidelberg, pp. 1094–1096.
- 1036 [41] Sorensen, H., 2003. The science of shopping. *Marketing Research* 15,  
1037 30–35.
- 1038 [42] Suri, R., Monroe, K. B., 2003. The effects of time constraints on con-  
1039 sumers’ judgments of prices and products. *Journal of consumer research*  
1040 30 (1), 92–104.
- 1041 [43] Sylvain, A., Doniec, A., Mandiau, R., Lecoeuche, S., 2014. Purchase  
1042 intention based model for a behavioural simulation of sale space. In:  
1043 Proceedings of the IEEE/WIC/ACM International Joint Conferences  
1044 on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). pp.  
1045 318–324.
- 1046 [44] Terano, T., Kishimoto, A., Takahashi, T., Yamada, T., Takahashi, M.,  
1047 2009. Agent-based in-store simulator for analyzing customer behaviors  
1048 in a super-market. In: *Knowledge-Based and Intelligent Information  
1049 and Engineering Systems*. Springer Berlin Heidelberg, pp. 244–251.
- 1050 [45] Thomson, E. S., Laing, A. W., McKee, L., 2007. Family purchase deci-  
1051 sion making: Exploring child influence behaviour. *Journal of Consumer  
1052 Behaviour* 6 (4), 182–202.
- 1053 [46] Weng, S., Liu, S., Wu, T., 2011. Applying bayesian network and associ-  
1054 ation rule analysis for product recommendation. *International Journal  
1055 of Electronic Business Management* 9 (2), 149–159.
- 1056 [47] Wooldridge, M., 2009. *An Introduction to MultiAgent Systems*. John  
1057 Wiley.
- 1058 [48] Yao, Z., Sarlin, P., Eklund, T., Back, B., Jul 2014. Combining visual  
1059 customer segmentation and response modeling. *Neural Computing and  
1060 Applications* 25 (1), 123–134.

- 1061 [49] Zhang, X., Li, S., Burke, R. R., Leykin, A., 2014. An examination of  
1062 social influence on shopper behavior using video tracking data. *Journal*  
1063 *of Marketing* 78 (5), 24–41.
- 1064 [50] Zheng, Y., May 2015. Trajectory data mining: An overview. *ACM*  
1065 *Trans. Intell. Syst. Technol.* 6 (3), 29:1–29:41.