Taylor & Francis
Taylor & Francis Group

Journal of Simulation

ISSN: 1747-7778 (Print) 1747-7786 (Online) Journal homepage: http://www.tandfonline.com/loi/tjsm20

Simulating an agent's decision-making process in black-box managerial environment: An estimation-and-optimisation approach

Zhou He, Chunling Luo, Chin-Hon Tan, Hang Wu & Bo Fan

To cite this article: Zhou He, Chunling Luo, Chin-Hon Tan, Hang Wu & Bo Fan (2018): Simulating an agent's decision-making process in black-box managerial environment: An estimation-and-optimisation approach, Journal of Simulation, DOI: 10.1080/17477778.2018.1440946

To link to this article: https://doi.org/10.1080/17477778.2018.1440946

Published online: 23 Feb 2018.

JOURNAL OF SIMULATION, 2018
https://doi.org/10.1080/17477778.2018.1440946

ORIGINAL ARTICLE

Simulating an agent's decision-making process in black-box managerial environment: An estimation-and-optimisation approach

Zhou He[a,b], Chunling Luo[b], Chin-Hon Tan[b], Hang Wu[c] and Bo Fan[d]

[a]School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China; [b]Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, Singapore; [c]School of Management, Harbin Institute of Technology, Harbin, China; [d]School of International and Public Affairs, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT
With the growing need to guide decision-making in today's complex managerial environment, researchers of the Operations Research/Management Science community have shown a considerable interest in modelling complex managerial systems using the agent-based modelling and simulation technique. This paper presents an estimation-and-optimisation (ESTOPT) architecture to simulate an agent's decision-making process in black-box managerial environment. An ESTOPT agent's behaviour is considered as a two-stage process of solving its optimisation problem, some parameters of which are uncertain and need to be estimated. In the first stage, the agent collects and records information for estimation; in the next stage, it attempts to solve the optimisation problem. The solution guides the agent's actions on the environment which, in turn, provides the agent with new information and payoff as feedback. In this paper, two agent-based models are introduced to demonstrate the implementation of the ESTOPT approach. The simulation outcomes compare favourably with both empirical and theoretical results, suggesting that the ESTOPT approach can be used to simulate an agent's decision-making process in black-box managerial environment.

ARTICLE HISTORY
Received 4 July 2017
Revised 29 January 2018
Accepted 6 February 2018

KEYWORDS
Simulation; agent-based modelling; decision-making; estimation; optimisation

1. Introduction

Agent-based modelling and simulation (ABMS) is a popular technique to understand the behaviour of a complex system. To model the system, it first develops an agent-based model (ABM), which is represented by a collection of agents. Next, simulation experiments with this ABM are conducted to evaluate various strategies for the operation of the system (Siebers, Macal, Garnett, Buxton, & Pidd, 2010). This technique stems from the disciplines of Complex Science and Computer Science, and it has been increasingly applied to investigate a wide variety of complex systems ranging from social systems, ecosystems, financial markets, and economies.

Motivated by the growing need to guide decision-making in today's complex managerial environment, researchers of the Operations Research/Management Science (OR/MS) community have shown a great interest in the application of ABMS in modelling complex managerial systems, such as supply chains, consumer marketplaces, service systems, financial market and transportation networks (Chen & Cheng, 2010; He, Cheng, Dong, & Wang, 2014; Negahban & Yilmaz, 2014; Zhang, Chan, & Ukkusuri, 2014). Compared with equation-based methods, ABMS provides a natural, flexible and powerful approach for modellers to capture the key elements of these systems, such as population heterogeneity, non-linear feedback/relationship, and complex interaction topology (e.g., social network). Rand and Rust (2011) provided a detailed comparison of ABMS and other common OR/MS methods (e.g., empirical modelling, behaviour experiments, and system dynamics), and noted that ABMS "allows the exploration of individual-level theories of behaviour, but the results can be used to examine larger scale phenomenon". They also proposed useful guidelines on when to apply ABMS and how to develop an ABM. In the above agent-based OR/MS studies, individuals (e.g., consumers) and organisations (e.g., firms) are designed as autonomous agents that adapt to and co-evolve with the dynamic complex system in which they exist. According to the complex adaptive system (CAS) theory proposed by Holland (1996), systematic phenomena and patterns emerge from the interactions among agents. Therefore, the importance of agent's behaviours raises a critical question: how to model an agent's behaviours when developing an agent-based OR/MS model (Macal, 2016)?

To answer this question, we focus on two major interdisciplinary communities which have contributed significantly to the agent-related research in the literature, i.e., the agent-based social simulation (ABSS) and agent-based computational economics (ACE) communities. Models of the ABSS, ACE and OR/MS share the same research objects, i.e., humans and firms,

CONTACT Bo Fan fanbo@sjtu.edu.cn

instead of animals, plants and robots that are studied by biologists and artificial intelligence scientists. Therefore, agent-based OR/MS studies could possibly borrow several off-the-shelf agent architectures from the ABSS and ACE communities to model agent decision-making process. However, there are two fundamental differences between ABSS/ACE and some agent-based OR/MS research topics when modelling the behaviours of humans and firms.

(1) Agent architectures in most ABSS models are generally imported from cognitive science, psychology, neurology, sociology and other domains, because these models are mainly built upon the notion that human behaviours are the responses to their own and/or other individuals' expectations, i.e., personal belief, desire, intention and social norms (Balke & Gilbert, 2014). However, these factors are often neglected in most OR/MS studies. Instead, entities considered in these studies are usually treated as rational decision-makers. For example, a player (e.g., a person or a firm) in a non-cooperative game aims to maximise its payoff; a vehicle seeks the best route to minimise the cost of travelling from its origin to destination (measured by time or distance). Compared with the human behaviours applied in ABSS, mathematical models could be more suitable for modelling the decision-making process of rational agents.
(2) ACE models typically need to create a large number of households and firms in order to simulate a regional, nationwide, or worldwide market. This might lead to extreme high computational complexity. To avoid such situation, agents' behaviours in ACE research are often over-simplified and described by discrete choice models, e.g., the NK model (see Section 2 for a brief introduction). However, some variables (e.g., price, distance) are naturally continuous and are challenging to discretise in a reasonable way.

These two issues indicate that there is a mismatch between existing agent architectures in ABSS and ACE models and the general requirements of agent-based OR/MS research, i.e., agents should make rational decisions and be able to solve complex optimisation problems. In addition to the above two aspects, there are two important factors of agent-based OR/MS research that should be considered when designing an agent's decision-making architecture.

The first factor is information availability, which is often ignored in the existing solutions. In practice, business entities tend to conceal private information for competition or negotiation. Besides, agent-based models often contain plenty of heterogeneous agents. Based on these two reasons, an agent may not be able to access full knowledge of the environment and multiple peers. Therefore, it is more reasonable to assume that each agent regards all the other agents as part of the environment. From the perspective of an agent, the environment is a black box because its internal mechanism is unknown (Ashby, 1961). In addition, the environment provides the agent with a payoff based on its previous decision(s). However, the payoff is uncertain due to the change of other agent's decisions and some stochastic factors. For example, when modelling the decision-making process of a vehicle, it is proper to consider the uncertainty caused by other vehicles' decisions, rather than assuming that it has full knowledge of the environment. Therefore, agents in black-box environment can be viewed as bounded rational agents because they have "limitations of both knowledge and computational capacity" (Simon, 1997).

The other expected feature of agent architecture is its compatibility with the existing OR/MS approaches and models, which are developed to understand and solve specific problems. For example, Dijkstra's algorithm or the travelling salesman model can be applied to plan the best route for a vehicle agent. Moreover, because of the high flexibility of the ABMS paradigm (Bonabeau, 2002, p. 7281), applicable OR/MS models can be extended by dropping some unrealistic assumptions and/or considering some "mathematics-unfriendly" elements. For example, the epidemic model can be enriched by modelling disease diffusion through different types of networks, and the simulation results can be compared with that derived from differential equation models (Rahmandad & Sterman, 2008). Therefore, an OR/MS-compatible approach can make original results comparable with those of ABMS and thus help validate the ABM.

From the above analysis, we suggest modelling agent decision-making process in black-box managerial environment is to simulate how a bounded rational agent (e.g., a firm) optimises its behaviours (e.g., by solving discrete and/or continuous optimisation problems) in response to continuing changes without knowing the full information about the environment and other peers' reactions. As reviewed and analysed in Section 2, such an architecture fulfilling these requirements, especially information availability considerations, seems to be lacking in the literature.

In this paper, we propose an ESTimation-and-OPTimisation (abbreviated as ESTOPT) architecture as a new approach to modelling agent decision-making for applicable OR/MS problems. In each time step of simulation, an agent's behaviours can be divided into two stages: in the first stage, the agent receives payoff based on its previous decision, and collects observable information as historical data which is then used to estimate uncertain information (e.g., some parameters of an optimisation problem); next, the agent attempts to search for the optimal solution of the problem with estimated parameters. The solution guides the agent's decisions and actions which in turn will affect other agents and the environment. Due to the trade-off between exploration and exploitation, a probabilistic mechanism is employed to decide the mode of agent decision-making – randomly or optimally. Arbitrary decisions are made to explore possible solution spaces (i.e., randomly); while optimal solutions iteratively improve agent's behaviours (i.e., optimally). We introduce two ABMs, a contribution game and a price war, to demonstrate the implementation of the ESTOPT approach, and conduct thousands of experiments under different scenarios with these two ABMs. The computational results suggest that ESTOPT can potentially model an agent's decision-making process in a black-box managerial environment reasonably well.

To the best of our knowledge, this study is the first attempt to develop a new approach for agent-based OR/MS research problems in terms of modelling agent decision-making, which is an essential topic in ABMS but remains unexplored in the OR/MS literature. Compared with current ABMS practices in this area, the ESTOPT approach incorporates the estimation stage to help the agent process information, and it employs a probabilistic mechanism to help the agent tackle the exploration-exploitation trade-off. Hence, the ESTOPT could be very useful for bounded rational agents to make decisions in black-box environment. The two examples of ABMs explicitly described in this paper could be useful for interested readers to design, implement and validate an ABM and conduct experiments for the purpose of understanding OR/MS issues. Another contribution of this study is the discussion of three critical topics in agent-based OR/MS research, namely validating an ABM, guiding an ABM, and examining the impact of different exploration-exploitation balancing mechanisms.

The remainder of the paper is organised as follows. In Section 2, the common approaches to modelling the agent's behaviours are reviewed. Section 3 presents several typical OR/MS research topics applicable for ABMS, as well as the features of these topics. Next, the proposed ESTOPT approach is introduced in Section 4. Then, two ESTOPT-style ABMs, a contribution game and a price war, are presented in Sections 5 and 6, respectively. Next, three critical problems of agent-based OR/MS research are discussed in Section 7. Finally, Section 8 concludes the paper.

2. Literature review

Defining how agents behave is a necessary and critical component when designing ABMs (Rand & Rust, 2011). Since early studies aimed to demonstrate that complex patterns can emerge from the implementation of simple rules, most modellers followed the "Keep It Simple, Stupid" (KISS) principle, so that they were able to better understand how complexity emerged from simple interactions among agents (Axelrod, 1997). Based on this tenet, agents are programmed to perform static and simple rules, such as probabilistic, reactive IF-THEN rules. For example, each cell in the *Game of Life* (Conway, 1970), a two-state, two-dimensional cellular automaton model, only carries out three simple IF-THEN rules. However, the system achieves many patterns fluctuating between chaotic and ordered (Gardner, 1970). Other well-known initial ABMs that followed the KISS tenet include the *model of segregation* (Schelling, 1978) and *Boids* (Reynolds, 1987).

However, the KISS principle was challenged by the researchers who suggested that, although simple rules can be used to describe the reactions of some reactive entities (e.g., viruses, bacteria, and animals), the notion of simplicity limits the realism, and thus the applicability, of human-related ABMS (Edmonds & Moss, 2005; Sun, 2007). For example, the principle of KIDS – Keep It Descriptive Stupid – recommends to consider a descriptive and complex model at first, and then simplify the model only where can be justified (Edmonds & Moss, 2005). To enhance realism of simulation, some human characteristics were introduced and modelled for the purpose of mimicking human decision-making as closely as possible. A variety of sophisticated architectures have been developed for sociological studies, including the typical belief-desire-intention model and its derivatives, normative architectures that consider the influence of social norms, and other approaches inspired by cognitive, psychology and neurology research. For recent and detailed surveys of these models, we refer the reader to Adam and Gaudou (2016). However, sociological, psychological and neurological factors are often neglected in OR/MS studies. Therefore, these architectures could only be applied to model human-like entities (e.g., customers and pedestrians) in several sub-discipline (e.g., marketing and transportation) research.

Unlike sociologists who almost solely focus on modelling human decision-making process in ABMS, economists have to study how an organisation behaves in a market as well. Besides elementary formulas of microeconomics, the ACE community has imported many approaches from other domains to model firms. For example, one of the most widely used approaches is the NK model (Giannoccaro & Nair, 2016), which was proposed by Kauffman (1993) to understand and simulate biological systems. In the NK model, "N" is the number of all agents' discrete choices, while "K" stands for the average number of interdependent choices. An

NK model can be viewed as a fitness landscape mapping from combinations of discrete choices onto payoffs. Next, the model adopts trail-and-error algorithms to search for the peaks with higher fitness on the landscape. The solution of the NK model indicates optimal decisions for each agent. Since the agent's choices are assumed to be finite, the NK model can only be employed to deal with discrete problems. This shortcoming also exists in most of the ACE models that use evolutionary games and reinforcement learning, since appropriate discretisation is necessary to avoid the "curse of dimensionality" and high computational complexity (Safarzyńska & van den Bergh, 2010; Tesfatsion & Judd, 2006). However, for the OR/MS problems with continuous decision variables, it is challenging to discretise these variables and evaluate the impact of different discretisation settings due to the lack of agreed guidelines.

The agent's behaviours in OR/MS literature also evolves from passively performing simple rules to actively achieving some objectives. For example, agent behaviours in Thadakamalla, Raghavan, Kumara, and Albert (2004) are based on probabilities, rather than an OR model. In particular, different types of agents are created with different probabilities, and the links among agents are also generated randomly. In recent agent-based OR/MS research, there is a trend that researchers have paid increasing attention to discussing and modelling how firms operate under constraints (e.g., limited resources) in a complex business environment. A common solution is that the agent is designed to solve optimisation problems and thus achieve its objective (e.g., minimising cost/risk or maximising profit/benefit) under several constraints. For instance, Chan and Chan (2010) created a two-echelon supply chain with multiple suppliers and customers using ABMS. Each agent in this model attempts to reduce its own total costs, including inventory cost, backorder cost, penalty cost, etc. Moreover, in a competitive online-to-offline market built by He, Cheng, Dong, and Wang (2016), service merchants are modelled as profit-maximising agents, while customers are utility-maximising agents connected by social networks. More importantly, this online-to-offline market can be reduced to a competitive location and pricing problem. Therefore, comparison between agent-based simulation results and analytical findings serves as a theoretical validation of the model. These models not only demonstrated a wide range of ABMS applications in OR/MS research, but also motivated us to develop a new approach to modelling agent decision-making process in agent-based OR/MS studies.

In sum, although there are many ways to model an agent's decision-making process in the literature, they fail to simultaneously fulfil the following requirements in the OR/MS context – orienting to bounded rational entities, handling discrete/continuous-variable issues, considering information availability, and incorporating OR/MS models. In light of the above observations, we set out to propose the ESTOPT architecture as a new approach to modelling agent decision-making process for some applicable OR/MS problems.

3. Applicable OR/MS research topics of ABMS

In this section, we survey the literature and list several typical OR/MS research topics which are often investigated using ABMS. In addition, the features of these topics are analysed to help readers identify potential research questions applicable to employ the ESTOPT approach.

The OR/MS research areas where ABMS could be helpful include the following:

- Supply chain evolution. In reality, a supply chain could be a hierarchical system or a complex network, e.g., scale-free (Pathak, Day, Nair, Sawaya, & Kristal, 2007; Pathak, Dilts, & Biswas, 2007). ABMS is able to model any structure of supply chain due to its flexibility, allowing the modeller to examine how the supply chain evolves. For example, facing a disruption in the supply chain, firms can be modelled as agents selecting new partners to reduce the negative effects caused by the disruption. Therefore, the researcher could understand how the structure of a disrupted supply chain evolves with an individual firm's behaviours (Nair & Vidal, 2011).
- Transportation management. A transportation system is very complex because it consists of a multitude of vehicles, as well as pedestrians. These entities make adaptive decisions in real time. For example, the drivers can be viewed as agents searching for the shortest path based on current traffic situation. For the transportation managers who attempt to understand and optimise the system, controlling traffic lights is a possible way to affect the behaviours of agents. ABMS is a powerful method to help transportation management since many adaptive agents can be easily generated for simulation (Chen & Cheng, 2010).
- Service systems planning. Many mathematical models have been developed for service system planning, including queuing theory and facility location. However, feedback from service requesters are often neglected. ABMS provides the possibility for introducing non-linear feedback mechanism. For example, the service requesters may share his/her word-of-mouth regarding service quality of a provider, influencing the choices of other agents (He et al., 2016). Therefore, ABMS has been employed to generate new service concept in

health care systems (Kim & Yoon, 2014).

- Relationships in business and technology ecosystems. ABMS can be applied in this topic because it has the capability of dealing with the heterogeneous entities of different roles (e.g., firms and customers), as well as the complicated relationships such as competition, cooperation, co-evolution, co-specialisation, and knowledge acquisition (Carayannis, Provance, & Givens, 2011; Molinero, Riquelme, & Serna, 2015; Robertson & Caldart, 2009). Based on these relationships, more complicated business/technology ecosystem can be established with the help of game theory, data envelopment analysis and many OR-related methods (Carayannis, Provance, & Grigoroudis, 2016).
- Diffusion of information and innovations. The ABMS technique has been extensively applied to study this important topic in marketing research, since both individual-level heterogeneity and social network topology can be considered (Watts & Dodds, 2007). This topic is also investigated in epidemiology to reduce further transmission of diseases (Rahmandad & Sterman, 2008). These studies help decision-makers understand how network structure and the role of hubs affect diffusion (Rand & Rust, 2011).

From the above analysis, we suggest that ABMS can be very suitable to study the OR/MS issues with the following three features:

(1) When modelling agents' adaptive behaviours is required. In many managerial systems, the agent's autonomy cannot be simply neglected. To model such systems in more reasonable way, the modeller should describe how an agent behaves in different situations. One solution is to examine the goals and constraints of agents, so that their adaptive behaviours can be modelled as an optimisation problem.
(2) When the environment's topology is complex and/or dynamic. Possible topologies of a managerial system include two-dimensional lattice, hierarchy and complex networks (e.g., scale-free, small world). Modelling these topologies and their evolutions mathematically is challenging, especially when some agents enter or exit the environment. In contrast, ABMS is able to simulate complex and dynamic environment due to its flexibility.
(3) When individual-level heterogeneity and non-linear feedback should be considered. Entities in a managerial system are heterogeneous, since they could have different objectives, preferences, roles and information sets. Individual-level heterogeneity, together with complex topology, often produces non-linear feedback among agents.

For more general criteria for when to apply ABMS, we refer the reader to Macal and North (2014) and Macal (2016).

4. The ESTOPT approach

4.1. The architecture and mechanism

Figure 1 illustrates the architecture of the ESTOPT approach. As a complex adaptive system (CAS), a basic ESTOPT-style ABM consists of the following elements:

(1) **Environment and the agent's input**. In this architecture, an agent called "Environment" interacts with all the other agents in the following ways. (1) The environment monitors all agents. Some events and rules may be triggered and performed by the environment when some conditions related to agents are met. For example, when a firm agent is bankrupt, the environment may be programmed to remove it from the model. (2) The environment receives actions from agents. The *action* concept will be discussed later. (3) The environment provides a payoff for each agent based on many factors, such as previous action(s) of the agent, other agents' impacts, and stochastic components in the model. (4) The environment provides information for each agent, assuming that each agent only considers interested or observable information. In other words, agents in ESTOPT are bounded rational because they have limitations in accessing the full information of the environment. Therefore, the agent's input is comprised of two parts: payoff and information.
(2) **Historical data**. After observation, the agent should record related information for estimation. Without recording, the agent cannot learn anything from past experience. The most important information to be recorded could be previous actions and payoffs, which can be analyzed to guide future decision-making. Besides, the agent may need to collect some uncertain but observable information in its OR model. For example, a driver agent tries to solve its shortest path problem, in which the time needed to travel from node A to B is assumed to be unknown and dependent on other agents' path choices. In this case, the agent should memorise the travel time information of each visited edge for estimation. In practice, what kind of historical data should be collected, how to assign weights to each record, and how to aggregate information are highly problem-dependent.
(3) **Estimation**. For an ESTOPT agent, its goal of collecting and learning from historical data is to perform more accurate estimations about uncertain parameters of the OR model. Given sufficient data, many approaches can be incorporated to estimate uncertain information, such as point and interval

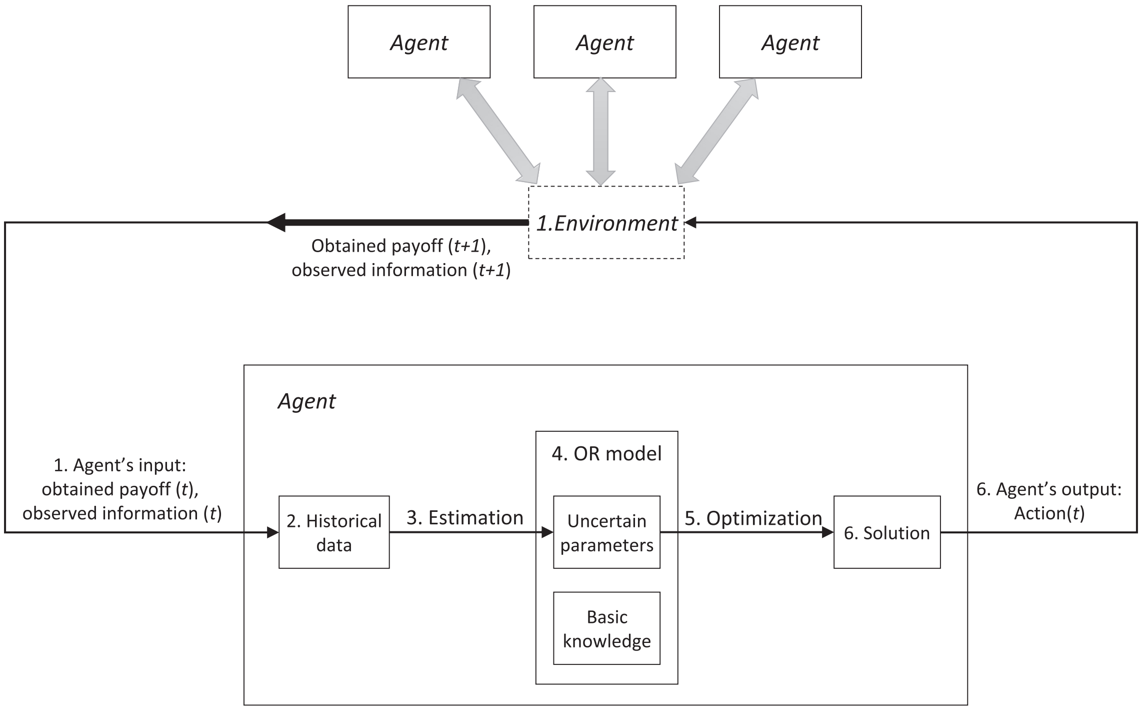

Figure 1. The architecture of the ESTOPT approach.

estimation, regression, prediction, Bayes' theorem and other methods of probability theory and statistics (see, e.g., He, Dong, & Yu, 2018). In some OR models such as robust optimisation, it could be unnecessary to estimate the precise values of parameters. In such cases, only some distribution assumptions need to be tested in this stage. Another important issue is the portion and weight of previous historical records used in estimation. For example, if the weighted moving average method is applied to forecast an indicator, we need to determine how many recent periods should be considered, and what their weights are. A discount factor can be borrowed from reinforcement learning to weigh the importance of recent practices and information. Generally, as simulation continues, more historical data will be collected, making the estimated results more stable and convincing. Therefore, the ESTOPT agents are able to learn from the past.

(4) **The OR model**. This critical component consists of two parts. The first part is basic knowledge, i.e., true and accurate information. A typical example is the structure of the OR model, which includes the objective function representing the goal of the agent, and the constraints that the agent has to consider. Typically, most attributes of the agent (e.g., individual preference, resources and predefined thresholds) are also considered as basic knowledge of the OR model, because they are either measurable, controllable or independent of external influence. For example, a firm agent knows that its inventory capacity is exactly 50 units. The second part of the OR model is the uncertain information that need to be estimated, such as other agents' choices, uncertain parameters and distribution of stochastic variables.

(5) **Optimisation**. When the OR model is well formulated and its uncertain parameters are estimated, it needs to be solved. Mathematical analytics, heuristic searching methods, and other approaches can be employed to find the optimal (or near-optimal for the sake of high computational complexity) solution for the agent.

(6) **Solution and action**. The solution embraces the new values of discrete/continuous decision variables. Hence, the agent's decisions are updated and will be sent to the environment as a new action.

To conclude, in each time step, an ESTOPT agent receives a payoff and observes information from the environment. In the first stage, collected observations are used to estimate uncertain but required information of the OR model. After estimation, the OR model is formulated and can be solved in the second stage. Finally, the solution renews the agent's actions, which will be sent to the environment for new payoff.

4.2. The exploration-exploitation trade-off

The exploration-exploitation trade-off exists in learning approaches because the agent needs to decide whether to obtain new knowledge or to use that knowledge to improve performance (Berger-Tal, Nathan, Meron, & Saltz, 2014). For example, when playing games like chess, we have two decisions: play the current best move (exploitation), or play an experimental move (exploration). Another example is oil drilling: drill at the best known location (exploitation), or drill at a new location (exploration). Similarly, the agent in ESTOPT architecture has two decision-making modes: (1) Optimally – the agent solves its OR model and performs the optimal action to enlarge its payoff. (2) Randomly – the agent makes any feasible decision to explore its solution space.

In ESTOPT architecture, we introduce a switching probability $\rho \in [0, 1]$ to determine the mode of decision-making. In each time step, the agent picks a random number from the uniform distribution $U(0, 1)$. If the random number is less than ρ, the agent acts randomly; otherwise, optimally. In the simplest case, ρ is a small number (e.g., 0.05) and remains unchanged. At the beginning of simulation, however, the information is insufficient for the agent to make optimal decisions because the agent has not explored the solution space yet. Therefore, the switching probability ρ can be a function of some changing variables. For example, $\rho(t) = 1 - \frac{1-\epsilon}{T}t$ where variable t is the current time step, constant T can be the maximum time step and constant ϵ can be 0.05 or 0. In this linearly-decreasing case, the agent only explores the solution space at first because $\rho(0) = 1$. As ρ decreases, the probability of making arbitrary decisions continues to decline since the agent has significantly investigated the possible solutions. The agent will be more likely to optimise its decisions when ρ becomes smaller. Therefore, this probabilistic mechanism is able to improve the agent's decisions at the end of simulation.

4.3. Advantages

The advantages of using the ESTOPT approach are as follows:

(1) It takes full advantages of the limited information to simulate agent decision-making process. The information availability factor is considered in the estimation stage. In this stage, available information can be fully utilised to formulate the agent's OR model. If more information becomes accessible, the ESTOPT approach can easily use it to further reduce uncertainty in decision-making process, allowing the modeller to observe and examine the influence of disclosing more information on the agent's behaviours and the model's performances.
(2) It provides a new way to solve complex OR models, the complexity of which often depends on the number of involved entities. After modelling these entities as ESTOPT agents, each entity will have its own OR model which is much simpler. Therefore, burdensome OR/MS models can be easily reduced to micro-level decentralised optimisation problems to guide agent's behaviours.
(3) It perfectly mimics the trial-and-error method, which is often used by an entity who has little knowledge about the environment in which it exists. Initially, the agent has no knowledge about the uncertain information. Hence, it tended to tentatively test many possible options, and then obtain feedback from the environment. As the simulation iteratively runs, the ESTOPT agent will be able to collect sufficient information to make better and eventually optimal or near optimal decisions. In this sense, the ESTOPT is a replication of typical trial-and-error process (Whitehead & Ballard, 1991) with low information.

In view of the above benefits, we suggest that the ESTOPT approach has great potential to model agent decision-making process for applicable OR/MS problems.

4.4. Implementation

This section contains technical details of implementing ESTOPT for the reader who is assumed to be familiar with object-oriented programming (OOP). We recommend OOP paradigm because an agent can be naturally viewed as an object according to the corresponding class with predefined attributes and methods. For example, if an ABM is composed of five similar firms and ten similar customers, only two classes – firm and customer – need to be designed. All the 15 agents are instances of these two classes, and they could be heterogeneous due to different values of some attributes. In order to generate multiple agents, we suggest the following three steps to create an ESTOPT-style class as a template.

Step 1: Analyse and convert variables to attributes. For an ESTOPT agent, its variables can be divided into two groups. (1) Exogenous variables (XVs) whose values are given by the modeller and remain unchanged after agent initialisation such as customer's gender, merchant's maximum capacity, minimum profit, and other predefined thresholds. For an ESTOPT agent, its personal XVs often relate to basic knowledge. Therefore, these variables can be coded as *final* attributes, i.e., constants. (2) Endogenous variables (NVs) that need to be updated in each time step, e.g., other rival's visible attributes like their new prices, firm's recent inventory, cost, and revenue. In fact, one of main reasons for coding a class's methods is to update these NVs. Therefore, it is important to check if all NVs are refreshed by in-use methods before debugging and running. For an ESTOPT agent, its personal NVs include observed information, obtained payoff, and estimated parameters, especially decision variables.

In OOP, another property of attributes is accessibility – whether an attribute can be accessed by peers or other types of agents. In the business context, it is normal for participants to conceal sensitive information such as business secret, firm's unit cost, customer's willingness-to-pay, etc. Therefore, the model should consider whether an attribute can be accessed by other agents. If not, it should be coded as a private attribute; otherwise, it is a public field of the agent class. Note that in many OOP languages such as Java, public attributes can be programmed as private attributes and then accessed by others who perform public functions such as *public double getPrivateAttribute()*.

Step 2: Program methods. Some key methods need to be created, such as constructor, *receiveInput()*, *recordData()*, *estimateParameter()*, *solveModel()*, *takeAction()*. It is suggested that all XVs' values are defined by the constructor or a similar function, and all NVs have to be updated at least once by all other methods. In addition, we recommend that the modeller employs popular off-the-shelf software and tools to perform complex and important tasks such as estimation and optimisation. For example, the Python community has developed many open-source and powerful packages for scientific computing, such as *scipy* for general optimisation and statical analysis, *lmfit* for non-linear optimisation and curve fitting, *networkx* for analysing complex networks, and *sklearn* for machine learning. Using these packages not only accelerates the model development, but also facilitates standardising the model, so that the model and simulation results can be reproduced, validated and compared.

Similarly, methods of a class can be either private or public. Here we suggest that all methods are programmed as public functions, so that they can be called by the *Environment* agent, which will be discussed in the next step.

Step 3: Link to the *Environment* agent. We suggest to create one container-like *Environment* agent which performs the following useful tasks. Firstly, it collects all actions and changes of agents' public attributes. The collected data can be used to debug code, draw time series charts and most importantly, provide required information and corresponding payoffs for the agents. Secondly, it organises the sequence of the agents' behaviours and interactions, which are programmed as public methods of the agent class. Therefore, the *Environment* agent should be able to call related methods of all agents. Thirdly, it terminates the simulation when the model meets the stop criteria, and then reports the final results. Lastly, if it is required to model an enter and exit mechanism of agents, then the *Environment* agent needs to control the number of agents by creating and removing instances.

4.5. Modelling an OR/MS problem as an ESTOPT-embedded ABM

Modelling ESTOPT agents is just one step to develop an ABM. If an OR/MS problem is deemed to be applicable for ABMS according to the conditions summarised in Section 3, the first step is to decide the boundary of the ABM. It is infeasible to consider all aspects of all relevant entities in a complex system. The modeller should exclude unnecessary factors and focus on the most important elements associated with the research problem. Usually, the topology of the system, the general classes of agents, and the basic relationships among and within agent classes should be determined in this step.

Next, the ESTOPT approach is applied to simulate the decision-making process of each agent class. Following the three steps in Section 4.4, the agents' behaviours can be simulated and thus an ESTOPT-embedded ABM is established. However, successful ABM requires verification and validation before using. The modeller should carefully verify if the ABM is implemented *correctly*, e.g., it has no errors and bugs. In contrast, validation is more difficult since it is concerned with checking if the ABM meets the researcher's need. This will be discussed in more detail in Section 7.1.

Since an ABM often contains many random factors (e.g., the probabilistic mechanism for balancing exploration and exploitation), it should be performed many times to ensure robust outputs against the randomness. Therefore, it could be necessary to assign different but fixed random seed for each simulation, so that all experiments can be reproduced. Finally, the modeller should decide and collect the output of the ABM (e.g., performance measures) for further statistical analysis.

In the following two sections, we introduce and build two ABMs to demonstrate the implementation and performance of the ESTOPT approach. The first ABM is based on a behavioural experiment of the contribution game (Isaac, McCue, & Plott, 1985; Isaac & Walker, 1988), the result of which serve as an empirical benchmark of the ESTOPT-embedded ABM. The other ABM is compared with an artificial price war, in which the agent can access varying degrees of information about the environment. The simulation results are compared with theoretical solutions. For a realistic application of the ESTOPT approach, we refer the reader to He, Xiong, Ng, Fan, and Shoemaker (2017).

5. Model A: The contribution game

5.1. Model introduction

The contribution game, also known as the public goods game, is a classic model that has been extensively studied by sociologists, economists and scholars of public management (Isaac et al., 1985; Isaac & Walker, 1988). We choose this model for two reasons. Firstly, it is quite simple as it only contains one type of regular agent. Secondly, a recent study (Nax, Burton-Chellew, West, & Young, 2016) has recruited 236 participants to play this game with incomplete information. Therefore, the experimental results from Nax et al. (2016) can be used to validate our ESTOPT approach assuming that these participants are bounded rational players.

Suppose that there are n agents (denoted by set $\mathcal{A}$), and each agent in the game is given a finite budget B_i, $i \in \mathcal{A}$ in each time step. The agent i is required to make a non-negative contribution c_i which should be less than or equal to the given budget, i.e., $c_i \in [0, B_i]$. Next, the agent i will receive its payoff π_i from the environment according to the following rule:

$$\pi_i = B_i - c_i + \frac{r}{n}\sum_{a\in\mathcal{A}} c_a = (\frac{r}{n} - 1)c_i + B_i + \frac{r}{n} \times C_{-i}, \quad (1)$$

where r is the rate of return, and C_{-i} denotes the total contributions of all the other agents. Nax et al. (2016) noted that for all agents, the Nash equilibrium (Nash, 1951) is either $c_i = 0$ if $r < n$ (i.e., free-riding), or $c_i = B_i$ if $r > n$ (i.e., fully contributing).

Next, we describe how to convert this contribution game to an ESTOPT ABM by following the steps described in Section 4.4.

5.2. Create an ESTOPT-embedded ABM

Step 1: Analyse and convert variables to attributes. Equation (1) implies that agent i only has three variables: decision variable contribution c_i as an NV, budget B_i as an XV, and objective payoff π_i as an NV to be maximised. The other three variables belong to the *Environment* agent, including the rate of return r as an XV, the number of total agents n as an XV, and C_{-i} as an NV. According to the *standard* experiment settings (Nax et al., 2016), let budget $B_i = 40$ and contribution c_i be integers for all agents. In addition, the rate of return is public knowledge, with the value of 1.6 under Scenario A1 and the value of 6.4 under Scenario A2. However, for each agent, variables n and C_{-i} are not observable and thus need to be estimated. Table 1 reports these two classes as well as their attributes and related properties.

Step 2: Program methods. The budget is the only XV of the *Agent* class, and its value can be assigned by the class's constructor function. In each time step, the *Environment* sends payoff to the agents according to Equation (1). After receiving the input, the agent i records (c_i, π_i). Therefore, in time step t, the agent i should have $t - 1$ historical records $(X, Y) = \{(c_{i,\tau}, \pi_{i,\tau})\}_{\tau=1}^{t-1}$ to fit the following regression model:

$$Y = \left(\frac{r}{a_0} - 1\right) X + 40 + \frac{r}{a_0} a_1, \quad (2)$$

where X is the explanatory variable, Y is the dependent variable, a_0 and a_1 are parameters to be estimated, and $r = \{1.6, 6.4\}$ is a known constant. In the estimation stage, the two ABMs attempt to minimise the sum of the squared deviations after considering all previous data points that are equally weighted. We incorporate the *lmfit* package (version: 0.9.2) and use the Nelder-Mead method to fit the regression model (2). Suppose that the estimated parameters are $\hat{a_{0,t}}$ and $\hat{a_{1,t}}$, then the agent's OR model can be expressed as follows:

$$\max_{c_{i,t}} \ \pi_{i,t} = \left(\frac{r}{\hat{a_{0,t}}} - 1\right) c_{i,t} + 40 + \frac{r}{\hat{a_{0,t}}}\hat{a_{1,t}}, \quad (3)$$

$$\text{s.t.} \ \ 0 \le c_{i,t} \le 40, c_{i,t} \in Z, \quad (4)$$

$$\text{where} \ \ r \in \{1.6, 6.4\}. \quad (5)$$

It is straightforward to solve this linear model, since the optimal solution is 0 (when $r < \hat{a_{0,t}}$), or 40 (when $r > \hat{a_{0,t}}$). Here, we employ the *scipy* package (version: 0.16.0rc1) to search for the optimal $c_{i,t}$.

Step 3: Link to the *Environment* agent. We create an *Environment* agent to receive all contributions from agents and send them payoffs according to Equation (1). Using the data and the termination condition from the study by Nax et al. (2016), there are four agents involved in each simulation, and the model coded using Python 2.7.10 stops after 20 time steps. Besides, let constants ϵ and T be 0 and 20, the switching probability $\rho(t) = 1 - \frac{1-\epsilon}{T}t = 1 - t/20$ according to the linearly-decresing function discussed in Section 4.2.

To conclude, the ESTOPT elements of Model A are identified as follows. The ***Environment*** receives all contributions from agents and sends them payoffs according to Equation (1). For agent i, its **input** is the gained payoff π_i; its **solution and action** is the contribution c_i. The agent records its previous (c_i, π_i) as **historical data** to **estimate** unknown parameters in Equation (3) using regression. Its **OR model** is expressed as Equations (3) and (4), which is **optimised** by a general search algorithm.

5.3. Simulation and results

We have performed the model 1000 times to ensure robust outputs against randomness in the agent's decision-making process. The number of simulation runs is determined arbitrarily. In practice, however, this number should be big enough to follow the "Law of large numbers" (Kolmogorov, 1950). All the 1000 independent simulations can be reproduced by assigning {0, 1, 2, . . . , 999} as random seeds. We have created 4000 agents

Table 1. Classes and attributes in Model A: The contribution game.

Class	Attribute	Type	Accessibility	Remark
Agent	c_i	NV	Private to peers only	Contribution
	B_i	XV	Private	Budget
	π_i	NV	Private	Payoff
Environment	n	XV	Private	Number of total agents
	r	XV	Public	The rate of return
	C_{-i}	NV	Private	Other agent's contributions

Table 2. Simulation, empirical and theoretical results of Model A: The contribution game.

Scenario	Data source	Mean c	% of zero-contribution	% of full-contribution
A1, $r = 1.6$	Nash equilibrium	0	100%	0%
	Nax et al. (2016), Figure 2	5	Not provided	Not provided
	ABMS	0.56	98.6%	1.4%
A2, $r = 6.4$	Nash equilibrium	40	0%	100%
	Nax et al. (2016), Figure 2	31	Not provided	Not provided
	ABMS	30.02	24.95%	75.05%

Table 3. Classes and attributes in Model B: The price war.

Class	Attribute	Type	Remark
Firm	p_i	NV	Price
	q_i	NV	Received demand
	π_i	NV	Profit
	P_i	XV	Upper bound of price
	c_i	XV	Unit cost
Market	n	XV	Number of total firms
	Q	XV	Total demand
	a, b	XV	Parameters of market preference in Equation (6)

and collected their contributions at the end of each simulation.

Table 2 reports three types of data – the theoretical Nash equilibrium, the experimental findings of Nax et al. (2016) and the simulation results of our ESTOPT-embedded ABM. As mentioned before, the Nash equilibrium is either zero-contribution when $r = 1.6 < 4$, or full-contribution when $r = 6.4 > 4$. However, such theoretical conclusion is built upon the assumption that all agents play with complete information. In both the behavioural and computational experiments, agents have to estimate uncertain information like the number of other players and their contributions. This leads to deviations from the Nash equilibrium. The critical issue faced by the agent is to decide which is larger, either r or a_0 (i.e., estimated n). It can be observed from Table 2 that, when $r = 6.4$ is slightly greater than the true $n = 4$, 24.95% of the agents would estimate that r is less than n after random initial explorations. Therefore, the mean contribution of 4000 agents under Scenario A2 is 30.02, which is very close to the 31 obtained from the behavioural experiment. This reveals that the ESTOPT approach is a promising architecture to simulate agent decision-making process when the information is limited.

The difference between the results under Scenario A1 and A2 is that, when $r = 1.6$ is much smaller than the true $n = 4$, only 1.4% of the agents have made wrong estimation, and the mean contribution of 0.56 is closer to the Nash equilibrium, 0, than to the experimental result of 5 reported by Nax et al. (2016). Here, we provide two reasons to account for this observation. Firstly, the free-riding behaviour is negatively viewed upon. Therefore, even though the participants cannot observe the contributions of others, they tend to be less selfish and are even willing to bear an affordable financial loss in practice. A similar finding has been observed in the real-life behaviour of individuals in the prisoner's dilemma where people have displayed a systemic bias towards cooperative behaviour, rather than being rational and betraying their partners (Fehr & Fischbacher, 2003). The second reason is that the investment amount under Scenario A1 is less than that under Scenario A2 (5 vs. 31). When the investment amount becomes larger, people are more likely to exhibit rational, self-interested behaviour. These two interacting reasons are why participants may contribute a small amount to the game when $r = 1.6$.

Based on the above discussion, we suggest that our ESTOPT approach is able to closely mimic rational human behaviours for similar agent-based OR/MS studies. On the other hand, as a limitation, it excludes some factors which could motivate agent to act in a less rational self-interested manner. To alleviate this shortcoming, the modeller is suggested to consider, select and integrate these factors into the ESTOPT architecture according to the research purpose.

6. Model B: The price war

6.1. Model introduction and scenario design

The above contribution game has explicitly described how to create ESTOPT agents and build a simple ABM. In this price war ABM, agents are competing for resources rather than collectively contributing. We concentrate on examining the influence of information availability on firm's decision and profit.

Suppose that there are n firms (denoted by set $\mathcal{F}$) competing for market demand by pricing, and one

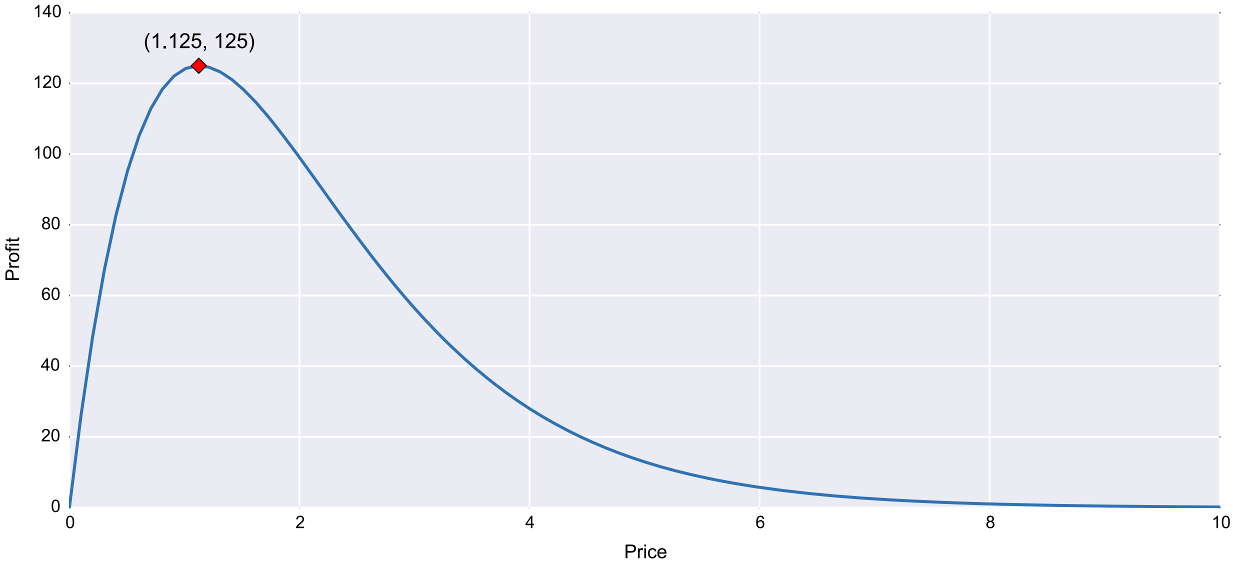

Figure 2. Scenario design for Model B: The price war.

environment-like *Market* agent that calculates corresponding demand for each agent. The attributes of these two classes are summarised in Table 3, whose information accessibility vary across designed scenarios. The *Market* agent computes the demand of firm i with price p_i, denoted by $q_i(p_i)$, according to the following rule:

$$q_i(p_i) = Q * \frac{e^{b-ap_i}}{\sum_{f \in \mathcal{F}} e^{b-ap_f}}. \tag{6}$$

Equation (6) is the classic *multinomial logit demand model* commonly used in revenue management and marketing literature (Besanko, Gupta, & Jain, 1998; Guadagni & Little, 1983). The function $b - ap_i$ associated with firm $i \in \mathcal{F}$ can be regarded as an approximation or surrogate for the "attractiveness" of its price p_i. All parameters of Equation (6) are non-negative. Hence, a firm with higher price tends to capture less demand from the market. For firm i, its goal is to maximise profit, i.e., $\pi_i = (p_i - c_i) * q_i$.

Next, we consider several scenarios from the perspective of a firm in terms of its ability to access information from its peers and the market.

Scenario B1. Each firm has no information about its peers and the market. In other words, all firms only know their own attributes and demands received from the market. To simulate firm's decision-making process under such a black-box scenario, we assume that all firms believe that there is a negative linear relationship between price and demand. Here, the OR model of firm i is stated as follows:

$$\max_{p_i} \pi_i = a_0(p_i - c_i)^2 + a_1(p_i - c_i) + a_2, \tag{7}$$

$$\text{s.t.} \quad c_i \leq p_i \leq P_i, \tag{8}$$

$$\text{where} \quad a_0 < 0,\ a_1 > 0. \tag{9}$$

The above OR model with a quadratic objective function (7) ensures that the optimal $p_i^* = c_i - 0.5a_1/a_0 > c_i$. Constraint (8) imposes a lower bound and an upper bound on firm's pricing. However, Equation (6) implies that the true price-demand relationship is not linear. Therefore, firms under Scenario B1 have made a simple but wrong assumption about the market.

Scenario B2. After conducting primary market research, all firms have gained basic knowledge about the true price-demand relationship. However, they still do not have information about other attributes of their peers and the market. Therefore, its OR model is expressed as follows:

$$\max_{p_i} \pi_i = (p_i - c_i) \times a_0 \times \frac{e^{a_1 - a_2 p_i}}{e^{a_1 - a_2 p_i} + a_3}, \tag{10}$$

$$\text{s.t.} \quad c_i \leq p_i \leq P_i, \tag{11}$$

$$\text{where} \quad a_0,\ a_1,\ a_2,\ a_3 > 0. \tag{12}$$

In Equation (10), the third component of the right-hand side copies from Equation (6), which defined the demand q_i. In fact, the uncertain a_0, a_1, a_2, a_3 under Scenario B2 correspond to the actual Q, b, a and other firms' current "attractiveness" based on their prices.

Scenario B3. We further assume that the actual values of the market's attributes are public information, but other firms' information are still unobservable. Compared with Scenario B2, parameters a_0, a_1, a_2 are replaced by true Q, b, a in firm's OR model. Therefore, only one variable a_3 (i.e., other firms' current "attractiveness" based on their prices) has to be estimated by the *lmfit* package.

Scenario B4. All information has become public. Therefore, the Nash equilibrium can be calculated by solving the following equation group (13) without running the ABM.

$$\begin{cases} \frac{\partial \pi_1}{\partial p_1} = \frac{Qe^{b-ap_1}}{\sum_{f\in\mathcal{F}} e^{b-ap_f}} + \frac{a(c_1-p_1)Qe^{b-ap_1}}{\sum_{f\in\mathcal{F}} e^{b-ap_f}} - \frac{a(c_1-p_1)Qe^{2(b-ap_1)}}{\left(\sum_{f\in\mathcal{F}} e^{b-ap_f}\right)^2} = 0, \\ \dots, \\ \frac{\partial \pi_n}{\partial p_n} = \frac{Qe^{b-ap_n}}{\sum_{f\in\mathcal{F}} e^{b-ap_f}} + \frac{a(c_n-p_n)Qe^{b-ap_n}}{\sum_{f\in\mathcal{F}} e^{b-ap_f}} - \frac{a(c_n-p_n)Qe^{2(b-ap_n)}}{\left(\sum_{f\in\mathcal{F}} e^{b-ap_f}\right)^2} = 0. \end{cases} \tag{13}$$

Table 4 summarises the settings of all the scenarios in terms of accessible information (basic knowledge). From black box (B1) to white box (B4), firms are able to observe more information for their decision-making process. Therefore, this table also represents the bounded-rationality levels of the agents under different scenarios. More complex or realistic scenarios can be created for different research purposes.

The ESTOPT elements of Model B are identified as follows. The ***Market*** receives all prices from agents and sends each agent demand according to Equation (6). For firm agent i, its **input** is the gained demand q_i; its **solution and action** is the price p_i. The agent records its previous (p_i, π_i) as **historical data** to **estimate** unknown parameters in Equation (7) (or Equation (10)) using regression. Its **OR model** is expressed as Equations (7) and (8) (or Equations (10) and (11), which is **optimised** by a general search algorithm.

6.2. Simulation and results

To start the simulation, we first assign arbitrary values to the market's XVs. In particular, $n = 9$, $Q = 1000$, $a = 1$, $b = 10$. Besides, all firms are homogenised by sharing same unit cost and upper bound of price, i.e., $c_i = 0$, $P_i = 10$, $\forall i \in \mathcal{F}$. Therefore, the solution of Equation (13), also the Nash equilibrium of Scenario B4, is $p_i^* = c_i + n/(a(n-1)) = 1.125$, $\forall i \in \mathcal{F}$, and each firm's profit is $\pi^* = (p^* - c)Q/n = 125$. Figure 2 provides a curve that demonstrate the price-profit relationship for a firm whose competitors all choose 1.125 as their prices. For the other three scenarios, the simulation is also performed 1000 times and thus 27,000 firms are created in total. In each simulation, the maximum number of time step is arbitrarily set to be 100, and the switching probability $\rho(t) = 1 - t/100$. Results in the final time step (i.e., 100-th tick) are reported in Table 5 and illustrated in Figure 3.

It can be found that, firms under Scenario B1 obtained the lowest level of profit since they have misjudged the market rules. After fitting the wrong regression model (7), most firms have selected either extreme low (close to 0) or high prices (about 4), leading to the largest price standard deviation (STD) and the fewest profit. Therefore, it is essential for firms to survey the market. From Scenario B2 to B3, the uncertainty in firm's decision-making decreases as less parameters need to be estimated. Therefore, due to greater information and higher degree of market transparency, all four indicators listed in Table 5 decline monotonically and become approximate to the Nash equilibrium solution (i.e., Scenario B4). These simulation results are consistent with experiment design, implying that all the three ESTOPT-embedded ABMs are valid.

In conclusion, we have designed four scenarios and conducted thousands of experiments with the price war ABM. Simulation results under Scenario B1, B2, and B3 indicated that information accessibility plays an essential role in firms' price and profit distributions. Besides, these results are approximate to the theoretical Nash equilibrium of Scenario B4 with increasing information accessibility. Therefore, the ESTOPT-embedded ABMS paradigm is able to produce correct and meaningful findings for OR/MS research.

7. Discussion

In this section, we discuss several critical problems of agent-based OR/MS research.

7.1. Model validation

In the majority of the ABMs such as the price war ABM, the parameter settings and used data are not empirically grounded. Therefore, how can an ABM be validated?

Validation is a crucial step in modelling ABMs. However, there exist many difficulties such as randomness in simulation results, lack of standard techniques for comparing ABMs, and having a large number of parameters involved (Rand & Rust, 2011). Besides, detailed micro-level data are needed for empirical validation since ABMs are generally built in a bottom-up way. However, obtaining high-quality individual-data is very costly. Consequently, empirical ABM modelling and validation are only possible for primitive ABMs with few agents and simple rules. Compared with that for ABSS and ACE research, ABMs for OR/MS studies confront more serious issue: a firm's behavioural data, such as operations records, are very sensitive in business environment. As a result, it is difficult to obtain firm-level behavioural data. On a positive note, classic OR models established on solid mathematical ground can be used to validate similar complex ABMs. Therefore, we suggest the following validation solutions in different situations.

Table 4. Scenario design of Model B: The price war.

Scenario	Basic knowledge	Unknown information
B1	Own(q, P, c)	Equation (6), Market(n, Q, a, b), Rivals(p, q, π, P, c)
B2	Own(q, P, c), Equation (6)	Market(n, Q, a, b), Rivals(p, q, π, P, c)
B3	Own(q, P, c), Equation (6), Market(n, Q, a, b)	Rivals(p, q, π, P, c)
B4	Complete information	–

Table 5. Simulation and theoretical results of Model B: The price war.

Scenario	Mean of price	STD. of price	Mean of profit	STD. of profit
B1	1.960	1.778	21.113	34.490
B2	1.761	0.736	167.833	50.872
B3	1.215	0.015	135.039	11.354
B4	1.125	0.000	125.000	0.000

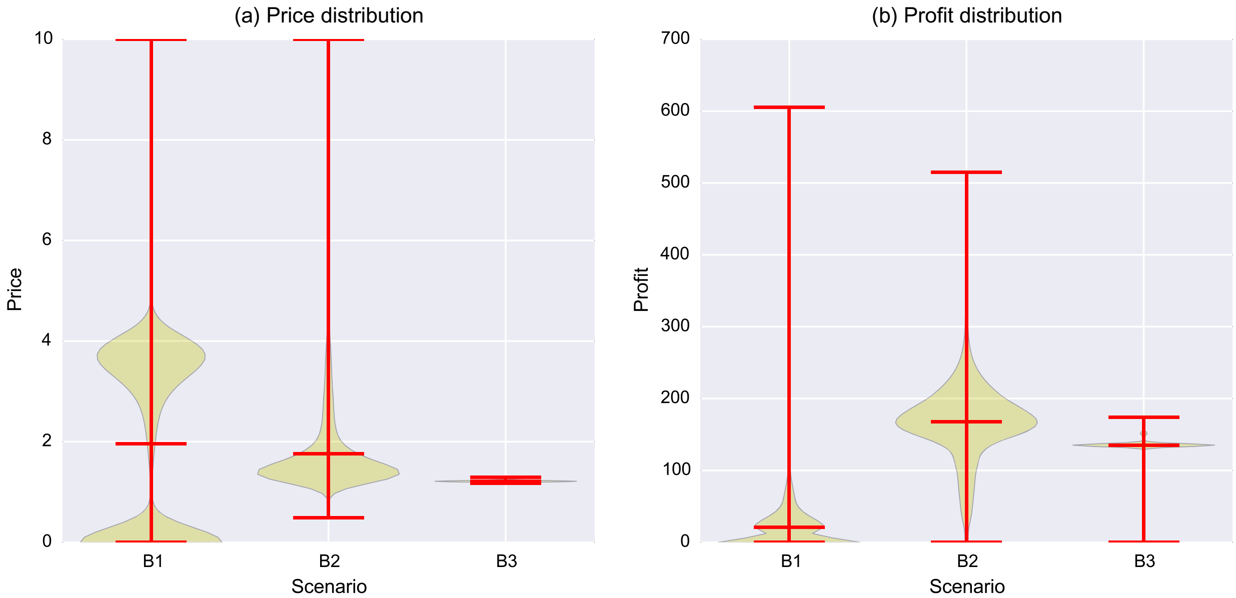

Figure 3. Price and profit distributions of 27,000 firms under four scenarios.
Notes: From top to bottom, the three points on red lines are the maximum, mean, and minimum of firm's prices and profits.

The first solution is theoretical validation if suitable empirical data are not available. In this situation, the agents' behaviours should be modelled based on solid theoretical reasoning, common sense, widely-accepted concepts, well-justified assumptions, classic models/frameworks, etc. Since classic OR models are relatively simple, ABMs can be reduced to them by removing some elements or simplifying the agents' behaviours. Therefore, the simulation results of these reduced ABMs should be in good agreement with that of the classic models. Take the price war model, for example, our simulation results are approximate to the mathematical conclusions (i.e., Nash equilibrium) after homogenising agents and enhancing agents' ability to access accurate information. However, additional parts built upon the reduced ABM can hardly be further validated without new benchmarks or empirical data. Due to this issue, we suggest that theoretical validation can only be used as a basic or preliminary approach to ABM validation.

The second solution is empirical validation when some real-world data can be collected. Official statistical surveys, industry reports, academic papers, media news, publications, and other data sources only provide aggregated data. For ABMS, such data can be used to determine the initial settings of experiments, calibrate the values of agents' XVs, and finally evaluate the fitness of simulation results. We have performed and demonstrated empirical validation when building an agent-based contribution game model in Section 5. However, using aggregated data to initialise the simulation and to assess the final outputs could be problematic. For example, when should the simulation stop? The length of simulation matters since it directly affects the final results. In some ABMs, a time step may correspond to either an hour, a day, a week or a month. But in many ABMs, there could be no link between the simulation tick and time period in reality. This might lead to two consequences. Firstly, without time calibration, it is difficult to compare and align simulation results

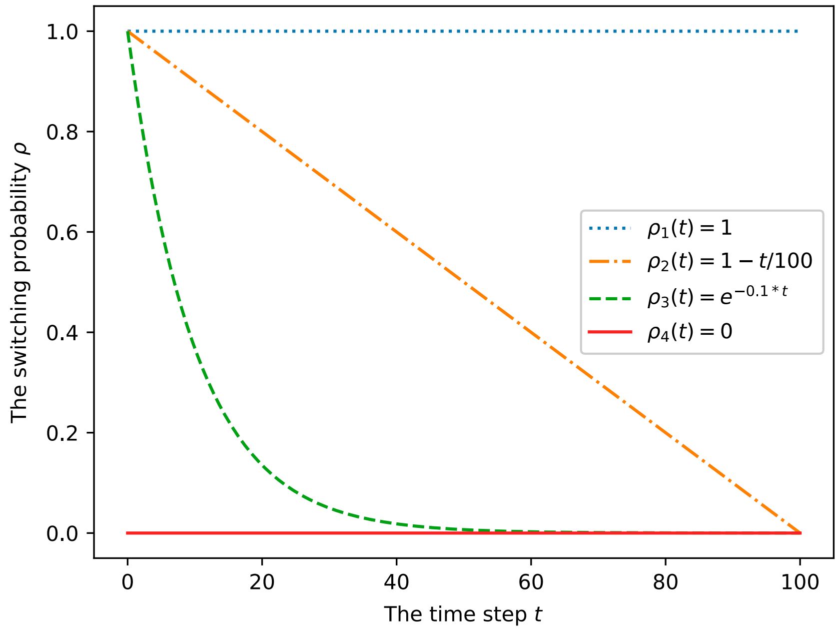

Figure 4. Four different functions of the switching probability $\rho(t)$. In time step t, if a random number $\in [0, 1]$ is smaller than $\rho(t)$, the agent will make a random decision.
Note: Generally, the agent becomes less exploratory as the function changes from ρ_1 to ρ_4.

with empirical time series data. Secondly, there is no appropriate choice of the maximum time step, which often serves as a criterion of simulation termination. These two issues require in-depth research.

The further and possibly the ultimate validation method entails massive temporal-spatial individual-level data. Based on such "big data", entities could be near-perfectly classified, learned, predicted, and finally replicated by agents. Currently, this tendency appears in some research and practice, where temporal-spatial individual-level data are collected from behavioural experiments, mobile/wearable devices, ubiquitous sensors, the Internet of things, etc (Kim, Ok, Kumara, & Yee, 2010; Reaidy, Gunasekaran, & Spalanzani, 2015). In this sense, the goal of the ABMS technique is to blur the boundary between simulation and reality.

7.2. From understanding to guiding

Many agent-based OR/MS studies attempt to understand complex systems by changing some XVs of the ABM, and examining how changed parameters affect predefined performance indicators of the system. Based on the relationship between these parameters and model outputs, is it possible for decision makers to guide the evolution of the system?

Before answering this question, we first introduce two concepts associated with the term *agent*: multi-agent system (MAS) and agent-based modelling. The first concept stemmed from the discipline of distributed artificial intelligence, which attempts to design smart agents (e.g., robots), unite them as a MAS, and solve specific practical or engineering problems. A MAS is usually hierarchical, where agents may compete, negotiate, and interact with one another in order to accomplish a certain task that a solo agent cannot achieve. Therefore, a leader agent is often created to be responsible for allocating resources and coordinating the other agents in the presence of conflicts. Optimisation methods are commonly involved when structuring agent behaviour and improving system performance. The second concept – agent-based modelling – underpins all of the models discussed in this paper. It originates from complex adaptive system (CAS) theory proposed by Holland (1996), a sub-domain of the complex systems research. As mentioned in Section 1, the goal of ABMS is to search for explanatory insight into the collective behaviour of agents. Unlike the MAS with clear overall objectives, a CAS is more decentralised so that none of the agents is able to control the whole system. For example, in an open market where multiple autonomous firms are competing for customers' orders, none of them – firms and customers – can play a decisive role in the competition evolution.

In sum, the agents in a MAS can be well controlled as the model is created to solve problems; while an ABM is designed to understand the behaviour of a complex system, and thus the modeller should not attempt to fully control the agents in a CAS. However, the stakeholder of agent-based OR/MS research could be interested in how to configure, regulate and guide the ABM without destroying an agent's autonomy. Is there an approach to extending an ABM for solving specific problems optimally? The answer could be simulation-based optimisation.

Table 6. Simulation results of Model B with different functions of the switching probability ρ.

Switching probability	Indicator	Scenario B1 (3 parameters)	Scenario B2 (4 parameters)	Scenario B3 (1 parameter)
ρ_1	Price	4.981 (2.888)	4.981 (2.888)	4.981 (2.888)
	Profit	166.259 (276.728)	166.259 (276.728)	166.259 (276.728)
ρ_2	Price	1.960(1.778)	1.761 (0.736)	1.215 (0.015)
	Profit	21.113 (34.490)	167.833 (50.872)	135.039 (11.354)
ρ_3	Price	3.401 (2.064)	1.878 (1.518)	1.140 (0.004)
	Profit	102.203 (175.967)	155.944 (57.951)	126.677 (0.304)
ρ_4	Price	2.800 (2.131)	2.053 (1.834)	1.128 (0.003)
	Profit	79.024 (166.596)	171.868 (94.902)	125.37 (0.174)

Note: Each data cell contains the mean and STD (placed in parentheses) of agents' average price or profit.

For example, in a regulated competitive market with private investment (e.g., a waste treatment market), the policy-makers have to respect the rights of private companies to pursue a reasonable profit. On the other hand, the regulated market is expected to fulfill some targets predefined by the public sector. In other words, after creating an ABM of the market and evaluating the impact of changing one XV (i.e., a simple policy), the regulator seeks to optimally adjust multiple XVs (i.e., a mixed policy) and thus achieves an objective with some constraints.

In this case, the ABM can be viewed as a black stochastic box. Given the same model, simulation results vary across random seeds, and cannot be expressed using equations. However, the relationship between XVs and simulation results truly exists. Therefore, we can employ simulation-based optimisation methods to search for the optimal XVs which can meet all given requirements. In particular, the XVs are decision variables that can be changed by the regulator directly or indirectly; while the predefined requirements are treated as the objective and constraints. Many heuristic search algorithms (e.g., genetic algorithms) can be used to address the problem, although they have to be performed many times for the sake of solution robustness.

To conclude, the modeller should respect the autonomy of agents in a CAS. Under certain circumstances that all agents can be guided by a powerful party (e.g., the government in a regulated market, or the core firm in a supply chain), the modeller is suggested to integrate the ABM into a simulation-based optimisation problem, so that optimal configuration of the ABM can be found using heuristic search algorithms.

7.3. The impact of different exploration-exploitation balancing mechanisms

We are interested in the impact of different mechanisms which are used to balance exploration and exploitation. Four different functions of the switching probability $\rho(t)$ (denoted by ρ_1 to ρ_4) are considered, as shown in Figure 4. In time step t, if a random number $\in [0, 1]$ is smaller than $\rho(t)$, the agent will make a random decision. The maximum time step is still 100 because we use Model B as the test-bed. Therefore, ρ_2 is the benchmark function which has already been applied in previous experiments. Compared with ρ_2, function ρ_3 implies that the agent will be less likely to make random decisions because it drops exponentially. The extreme cases are ρ_1 and ρ_4, in which the agent only performs exploration or exploitation. Note that the agent with ρ_4 has to make several random decisions at the beginning of simulation due to the need of solving regression problems. In particular, the agent has to record 3/4/1 random attempts according to the number of parameters to be estimated under Scenario B1/B2/B3, respectively. To conclude, the agent becomes less exploratory as the function changes from ρ_1 to ρ_4.

Table 6 reports the simulation results, from which the following findings can be observed: (1) When the agent only performs exploration, the simulation results are identical (as the random seeds are fixed) and thus meaningless. This finding implies that it is important to consider exploitation in designing agent decision-making process. (2) Under Scenario B2, the STDs of agents' average price and profit increase as the function changes from ρ_2 to ρ_4, indicating that agents get trapped in distant local optima. (3) In contrast, agents under Scenario B3 have deceasing STDs as agents become less exploratory. With ρ_4, their average price and profit are very close to the Nash equilibrium, i.e., 1.125(0.000) and 125(0.000). The reason is probably the number of parameters to be estimated. If the agent's OR model has many unknown parameters, it is challenging to search for the global optimum. Hence, the exploration-exploitation balancing mechanism should allow the agent to explore more in simulation. When the agent has few parameters to be estimated, the switching probability ρ should be small to help the agent refine its decisions.

8. Conclusion

Following many sociologists and economists, an increasing number of OR/MS scientists have noticed the importance of understanding how humans and organisations behave in different situations, which is difficult to be captured in traditional mathematical methods. Moreover, we suggest that some applicable OR/MS problems have four specific requirements in terms of agent architecture: orienting to bounded rational entities, handling discrete/continuous-variable issues, considering information availability, and incorporating OR/MS models. However, existing agent architectures in the areas of the ABSS and ACE barely meet these requirements.

This paper presents an estimation-and-optimisation (ESTOPT) architecture to model agent decision-making process in the context of OR/MS. The two-stage ESTOPT treats an agent's behaviour as a process of solving its OR problem, some parameters of which are uncertain and need to be estimated. In the first stage, the ESTOPT agent collects and records information for estimation; in the next stage, it attempts to solve the OR problem. A probabilistic mechanism is embedded in the ESTOPT to balance exploration and exploitation. The solution guides the agent's actions on the environment which in turn, provides the agent with new information and payoff as feedback. We introduce two ABMs to demonstrate the implementation of the ESTOPT approach, and conduct thousands of experiments under different scenarios with these two ABMs. The computational results suggest that the ESTOPT approach can be used to simulate an agent's decision-making process in black-box managerial environment.

As one of the first attempts to discuss agent-based OR/MS research paradigm, we do not claim that the ESTOPT should be viewed as the best solution to modelling agent's behaviours in all OR/MS studies. Indeed, there is a consensus among ABMS scholars that the choice of modelling approaches highly depends on the context, the goal of the simulation, and the various parameters (Edmonds & Moss, 2005; Epstein, 2006). Hence, although we believe that the ESTOPT can be perfectly adapted in some applicable OR/MS problems, the simulation could greatly benefit from employing other modelling approaches (e.g., the belief-desire-intention architecture) in other cases (e.g., if emotions and norms of customers need to be considered in marketing research). To conclude, the proposed ESTOPT is expected to be improved, extended, customised, integrated, and applied by interested readers.

ORCID

Zhou He http://orcid.org/0000-0001-6288-7215

Acknowledgements

The authors greatly appreciate the editor and anonymous referees for their comments, which helped to improve this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is funded by the National Key Research and Development Program of China [grant number 2016YEF 0122300], the National Science Foundation of China [grant number 71371122], National Social Science Foundation of China [grant number 14ZDB152], Inter-discipline Foundation of Shanghai Jiao Tong University [grant number 16JXZD02], and the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

Adam, C., & Gaudou, B. (2016). BDI agents in social simulations: A survey. *Knowledge Engineering Review, 31*(3), 207–238.

Ashby, W. R. (1961). *The black box, chapter 6* (pp. 86–117). London: Chapman & Hall.

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.

Balke, T., & Gilbert, N. (2014). How do agents make decisions? A survey. *Journal of Artificial Societies and Social Simulation*, *17*(4), 13. Retrieved from http://jasss.soc.surrey.ac.uk/17/4/13.html

Berger-Tal, O., Nathan, J., Meron, E., & Saltz, D. (2014). The exploration-exploitation dilemma: A multidisciplinary framework. *PloS One, 9*(4), e95693.

Besanko, D., Gupta, S., & Jain, D. (1998). Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science, 44*(11-part-1), 1533–1547.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences, 99*(suppl 3), 7280–7287.

Carayannis, E. G., Provance, M., & Givens, N. (2011). Knowledge arbitrage, serendipity, and acquisition formality: Their effects on sustainable entrepreneurial activity in regions. *IEEE Transactions on Engineering Management, 58*(3), 564–577.

Carayannis, E. G., Provance, M., & Grigoroudis, E. (2016). Entrepreneurship ecosystems: An agent-based simulation approach. *Journal of Technology Transfer, 41*(3), 631–653.

Chan, H. K., & Chan, F. T. S. (2010). Comparative study of adaptability and flexibility in distributed manufacturing supply chains. *Decision Support Systems, 48*(2), 331–341.

Chen, B., & Cheng, H. H. (2010). A review of the applications of agent technology in traffic and transportation systems. *IEEE Transactions on Intelligent Transportation Systems, 11*(2), 485–497.

Conway, J. (1970). The game of life. *Scientific American, 223*(4), 120–123.

Edmonds, B., & Moss, S. (2005). From KISS to KIDS – An 'anti-simplistic' modelling approach. In P. Davidsson, B. Logan, & K. Takadama (Eds.), *Multi-agent and multi-agent-based simulation* (Volume 3415 of Lecture Notes in Computer Science, pp. 130–144). Berlin: Springer.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton, NJ: Princeton University Press.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American, 223*(4), 120–123.
Giannoccaro, I., & Nair, A. (2016). Examining the roles of product complexity and manager behavior on product design decisions: An agent-based study using NK simulation. *IEEE Transactions on Engineering Management, 63*(2), 237–247.
Guadagni, P. M., & Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2*(3), 203–238.
He, Z., Cheng, T. C. E., Dong, J., & Wang, S. (2014). Evolutionary location and pricing strategies in competitive hierarchical distribution systems: A spatial agent-based model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(7), 822–833.
He, Z., Cheng, T. C. E., Dong, J., & Wang, S. (2016). Evolutionary location and pricing strategies for service merchants in competitive O2O markets. *European Journal of Operational Research, 254*(2), 595–609.
He, Z., Dong, J., & Yu, L. (2018). An agent-based model for investigating the impact of distorted supply-demand information on China's resale housing market. *Journal of Computational Science, 25*, 1–15. doi:10.1016/j.jocs.2018.01.002.
He, Z., Xiong, J., Ng, T. S., Fan, B., & Shoemaker, C. A. (2017). Managing competitive municipal solid waste treatment systems: An agent-based approach. *European Journal of Operational Research, 263*(3), 1063–1077.
Holland, J. (1996). *Hidden order: How adaptation builds complexity*. Redwood City, CA: Addison-Wesley.
Isaac, R. M., McCue, K. F., & Plott, C. R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics, 26*(1), 51–74.
Isaac, R. M., & Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics, 103*(1), 179–199.
Kauffman, S. A. (1993). *The origins of order: Self organization and selection in evolution*. New York, NY: Oxford University Press.
Kim, J., Ok, C.-S., Kumara, S., & Yee, S.-T. (2010). A market-based approach for dynamic vehicle deployment planning using radio frequency identification (rfid) information. *International Journal of Production Economics, 128*(1), 235–247.
Kim, S., & Yoon, B. (2014). A systematic approach for new service concept generation: Application of agent-based simulation. *Expert Systems with Applications, 41*(6), 2793–2806.
Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. Oxford: Chelsca.
Macal, C., & North, M. (2014). Introductory tutorial: Agent-based modeling and simulation. In *Proceedings of the 2014 Winter Simulation Conference* (pp. 6–20). IEEE Press.
Macal, C. M. (2016). Everything you need to know about agent-based modelling and simulation. *Journal of Simulation, 10*(2), 144–156.
Molinero, X., Riquelme, F., & Serna, M. (2015). Cooperation through social influence. *European Journal of Operational Research, 242*(3), 960–974.
Nair, A., & Vidal, J. M. (2011). Supply network topology and robustness against disruptions – An investigation using multi-agent model. *International Journal of Production Research, 49*(5), 1391–1404.
Nash, J. (1951). Non-cooperative games. *Annals of Mathematics, 54*(2), 286–295.
Nax, H. H., Burton-Chellew, M. N., West, S. A., & Young, H. P. (2016). Learning in a black box. *Journal of Economic Behavior & Organization, 127*, 1–15.
Negahban, A., & Yilmaz, L. (2014). Agent-based simulation applications in marketing research: An integrated review. *Journal of Simulation, 8*(2), 129–142.
Pathak, S. D., Day, J. M., Nair, A., Sawaya, W. J., & Kristal, M. M. (2007). Complexity and adaptivity in supply networks: Building supply network theory using a complex adaptive systems perspective. *Decision Sciences, 38*(4), 547–580.
Pathak, S. D., Dilts, D. M., & Biswas, G. (2007). On the evolutionary dynamics of supply network topologies. *IEEE Transactions on Engineering Management, 54*(4), 662–672.
Rahmandad, H., & Sterman, J. (2008). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science, 54*(5), 998–1014.
Rand, W., & Rust, R. T. (2011). Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing, 28*(3), 181–193.
Reaidy, P. J., Gunasekaran, A., & Spalanzani, A. (2015). Bottom-up approach based on internet of things for order fulfillment in a collaborative warehousing environment. *International Journal of Production Economics, 159*, 29–40.
Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, volume 21 of SIGGRAPH '87* (pp. 25–34). New York, NY: ACM.
Robertson, D. A., & Caldart, A. A. (2009). *The dynamics of strategy: Mastering strategic landscapes of the firm*. New York, NY: Oxford University Press.
Safarzyńska, K., & van den Bergh, J. C. (2010). Evolutionary models in economics: A survey of methods and building blocks. *Journal of Evolutionary Economics, 20*(3), 329–373.
Schelling, T. C. (1978). *Micromotives and macrobehavior*. New York, NY: W.W. Norton.
Siebers, P.-O., Macal, C. M., Garnett, J., Buxton, D., & Pidd, M. (2010). Discrete-event simulation is dead, long live agent-based simulation!. *Journal of Simulation, 4*(3), 204–210.
Simon, H. (1997). *Behavioral economics and bounded rationality* (pp. 267–298). Cambridge, MA: MIT Press.
Sun, R. (2007). Cognitive social simulation incorporating cognitive architectures. *IEEE Intelligent Systems, 22*(5), 33–39.
Tesfatsion, L., & Judd, K. L. (2006). *Handbook of computational economics: Agent-based computational economics* (Vol. 2). Amsterdam: Elsevier.
Thadakamalla, H. P., Raghavan, U. N., Kumara, S., & Albert, R. (2004). Survivability of multiagent-based supply networks: A topological perspective. *IEEE Intelligent Systems, 19*(5), 24–31.
Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research, 34*(4), 441–458.
Whitehead, S. D., & Ballard, D. H. (1991). Learning to perceive and act by trial and error. *Machine Learning, 7*(1), 45–83.
Zhang, B., Chan, W., & Ukkusuri, S. V. (2014). On the modelling of transportation evacuation: An agent-based discrete-event hybrid-space approach. *Journal of Simulation, 8*(4), 259–270.