mathematics

Article

Analyzing Decision-Making in Cognitive Agent Simulations Using Generalized Linear Mixed-Effects Models

Shengkun Xie [1,*], Chong Gan [1,2] and Anna T. Lawniczak [2]

1 Global Management Studies, Ted Rogers School of Management, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; ganc@uoguelph.ca
2 Mathematics and Statistics, University of Guelph, Guelph, ON N1G 2W1, Canada; alawnicz@uoguelph.ca
* Correspondence: shengkun.xie@torontomu.ca; Tel.: +1-416-979-5000 (ext. 543474)

Abstract: Enhancing model interpretability remains an ongoing challenge in predictive modelling, especially when applied to simulation data from complex systems. Investigating the influence and effects of design factors within computer simulations of complex systems requires assessing variable importance through statistical models. These models are crucial for capturing the relationships between factors and response variables. This study focuses on understanding functional patterns and their magnitudes of influence regarding designed factors affecting cognitive agent decision-making in a cellular automaton-based highway crossing simulation. We aim to identify the most influential design factors in the complex system simulation model to better understand the relationship between the decision outcomes and the designed factors. We apply Generalized Linear Mixed-Effects Models to explain the significant functional connections between designed factors and response variables, specifically quantifying variable importance. Our analysis demonstrates the practicality and effectiveness of the proposed models and methodologies for analyzing data from complex systems. The findings offer a deeper understanding of the connections between design factors and their resulting responses, facilitating a greater understanding of the underlying dynamics and contributing to the fields of applied mathematics, simulation modelling, and computation.

Keywords: generalized linear mixed-effects models; cellular automaton; cognitive agents; agent-based simulations; complex systems; variable importance measures

MSC: 68Q80; 68T42; 68T09

check for updates

Citation: Xie, S.; Gan, C.; Lawniczak, A.T. Analyzing Decision-Making in Cognitive Agent Simulations Using Generalized Linear Mixed-Effects Models. *Mathematics* **2024**, *12*, 3768. https://doi.org/10.3390/math12233768

Academic Editors: Yu-Wang Chen, Mi Zhou and Tao Wen

Received: 27 September 2024
Revised: 25 November 2024
Accepted: 25 November 2024
Published: 29 November 2024

1. Introduction

Simulation approaches offer significant advantages in studying complex systems, providing insights across various real-world applications [1–13]. Traditional models based on ordinary and partial differential equations often fail to capture the nonlinear relationships and multifaceted interactions characteristic of complex systems [1–11,13,14]. While ordinary and partial differential equation models work well for simpler, linear dynamics, they struggle to capture emergent behaviours resulting from large numbers of interacting entities and their interdependencies. Individually based simulation models (e.g., multi-agent-based simulation models, cellular automata, and lattice gas cellular automata models), provide flexibility by incorporating details of local interactions and relationships, allowing them to capture the emergent properties and nonlinear dynamics of complex systems. This makes individually based simulation models indispensable for studying the dynamics of natural and man-made complex systems. Examples of natural complex systems demonstrating intricate interactions among individuals include fish schooling, wolf pack behaviour, and bird flocking. Similarly, complex engineering systems like robot swarms, vehicular traffic on highways, and pedestrian movement in crowded cities involve many interacting entities and individually based simulation models have been used extensively to study their dynamics.

In artificial intelligence and robotics, technological advancements have driven numerous robot applications across industries [15,16], such as manufacturing [17–22], automation [23,24], mining [25,26], surgery [27–30], and aerospace [31,32]. Considering maintenance and update costs is essential for robotics applications [33]. In some circumstances, swarms of simple robots may offer advantages over complex robots, as they are more cost-effective to produce and easier to repair. When a robot breaks in a swarm, it can be easily replaced, reducing downtime and costs. Robot swarms may also provide greater reliability, robustness, and flexibility [34]. A robot swarm can often complete a task successfully even if some of its members fail, unlike scenarios where the malfunction of a single complex robot disrupts operations. Thus, even with individual failures, overall performance remains largely stable, ensuring uninterrupted operations. Additionally, autonomous robot swarms may engage in collective learning and sharing experiences to improve decision-making and accelerate task execution [35–39]. This highlights the importance of studying robot swarms as complex systems in which robots are modelled by autonomous cognitive agents.

A cognitive agent is an autonomous agent capable of performing cognitive actions, i.e., a sequence of activities such as *perceiving* information in the environment and from other agents, *reasoning* about this information using existing knowledge, *judging* it with prior knowledge, responding as needed to other cognitive agents or to the external environment, and *learning* by updating or augmenting knowledge when new information is obtained [40,41]. In these papers, the authors explore the minimal intelligence requirements for microbots, experimenting with very primitive cognitive agents modelled using *biomimicry*. The cognitive agents apply *observational learning*, a type of *social learning* where they "*imitate what works and don't do what doesn't work*" [42]. This research has motivated further studies on how simple cognitive agents, using basic decision-making algorithms and building their knowledge bases by observing previous cognitive agents' crossing attempts, can learn to cross a highway modelled with cellular automata [43–45]. In this research, highway traffic is used as an example of a dynamically changing environment in which simple learning algorithms are explored. The simulation models of the considered complex system depend on numerous factors, each with multiple levels. Therefore, applying various statistical and machine learning methods is necessary to understand how these factors and levels influence the system's behaviour.

This paper is the continuation of these investigations. It follows up a theoretical investigation into the principles of *social learning* and *biomimicry* applied to simple cognitive agents [40–50]. The study conducted thus far focuses on understanding the fundamental mechanisms of learning and adaptation in these agents and does not involve the development or deployment of any physical or technological systems. The concepts explored in this research may have potential applications in various engineering and social science fields, including microbots, nanobots, or autonomous vehicles (AVs) in engineering or financial markets, risk analysis, and management, where psychological drivers play a pivotal role in decision-making processes. It is important to note that any practical engineering implementations of this research would require further work, technological advancements, and ethical considerations. Specifically, any application of swarm nanotechnology in military contexts would necessitate the development of advanced communication and control systems that are robust, secure, and capable of operating independently of existing infrastructure such as 5G or the Internet of Things. These technologies rely on centralized networks and communication protocols that could be vulnerable to disruption or interception. To ensure the effectiveness and security of swarm microbots in military applications, it would be crucial to develop decentralized communication and control mechanisms that are resilient to interference and can operate autonomously in complex environments. The presented simulation model of cognitive agents learning to cross the cellular automaton-based highway could be extended to modelling more complex traffic scenarios (e.g., intersections of multi-lane highways) and various forms of communication among agents and the infrastructure to study their effects on cognitive agents' learning abilities based on principles of *social learning* and *biomimicry*. However, any such potential extension would result in a much more complex simulation model, the dynamics of which would depend on many more factors, making it significantly more difficult to study emergent behaviours.

Complex systems often exhibit emergent behaviours and system-wide patterns arising from local interactions [1,2]. Computer simulations effectively capture these phenomena, helping researchers understand how local dynamics drive broader system behaviours [1,2,5,6]. The behaviours and dynamics of complex systems are highly dependent on input variables, with some variables playing a more critical role than others. Thus, identifying suitable predictive modelling approaches becomes crucial, as these approaches not only map relationships between input and output variables but also provide valuable insights into their impact on a complex system model's performance through simulation data analysis. In this context, adopting statistical and machine learning methodologies that balance predictive power and interpretability is essential. Applying these methodologies should facilitate an understanding of variable relationships and offer additional insights into how these relationships affect the accuracy and reliability of predictive models. By selecting effective predictive modelling approaches, researchers and practitioners can deepen their understanding of complex system dynamics, enhance prediction accuracy, and make more informed decisions based on model insights. Ultimately, this will lead to more effective applications of predictive modelling across various domains in complex systems [51–53].

1.1. Objective

The analysis of data from complex systems and their simulation models remains an active area of research, with significant efforts dedicated to developing predictive models and analytical techniques to extract meaningful insights [5,7,8,54–56]. Machine learning techniques, particularly Artificial Neural Networks (ANNs) [57–59], play a central role in capturing the nonlinear relationships between input and output variables in complex systems. However, a major challenge with these methods is their lack of interpretability [60–63], which limits our understanding of how different predictors influence outcomes and reduces our ability to assess their contributions to predictive models.

This study extends previous work on modelling and simulating cognitive agents learning to cross a cellular automaton-based highway, developed and analyzed in [45–47]. These cognitive agents model simple robots placed strategically at highway crossing points. The agents' primary objective is to learn to cross through vehicle traffic by leveraging shared and updated knowledge from previous crossing attempts. This learning process enhances the agents' decision-making, enabling them to decide whether to cross or wait. The behaviour of the agents is influenced by various simulation parameters, such as Car Creation Probability (determining the highway traffic density), the agents' propensity for risk taking (i.e., Desire), risk avoidance (i.e., Fear) related to crossing, and agents' access to a knowledge base. These factors interact in ways that lead to varying outcomes, from successful crossings to strategic waiting.

Understanding the relationships between these parameters and the resulting agents' decisions is essential for gaining insight into the agents' learning processes and overall performance. By analyzing these interactions, we can identify how cognitive agents adapt to complex traffic environments and improve their decision-making. These findings offer valuable insights for optimizing the behaviour of autonomous robots with minimal computational capabilities in real-world scenarios, advancing the development of adaptive systems capable of navigating in dynamic and uncertain environments.

1.2. Novelty

This study introduces a novel approach to measuring the importance of variables in simulation experiments through a custom-developed agent-based simulation model. The main objective is to analyze how experimental factors influence cognitive agents' ability to learn and to cross a cellular automaton-based highway. While previous research has examined this problem using traditional statistical methods [45–47,64], this study presents innovative tools and analytical techniques for processing computer simulation data. The potential of these tools and techniques is illustrated through their application to the considered cognitive agent-based simulation model. However, these tools and techniques are general and can be applied to simulation data from diverse complex system simulation models.

We propose using Generalized Linear Mixed-Effects Models (GLMMs) to individually model four key decision variables of the cognitive agent-based simulation model. We then measure the variable importance to quantify the significance of each input factor's contribution (i.e., model predictor). The novelty of this approach lies in its ability to provide variable importance measures and reveal the functional patterns and effects of model predictors, while also addressing the inherent uncertainty in measuring these effects. The findings enhance our understanding of the relationships between input and output variables, offering deeper insights into complex system behaviours. Furthermore, this study demonstrates the effectiveness of advanced statistical methods, particularly GLMMs, in capturing the effects of experimental factors, encouraging their broader use in analyzing data from complex systems.

1.3. The Outline

The remaining sections of this paper are organized as follows. In Section 2, we review and discuss previous studies and the relevant literature. Section 3 offers a brief overview of the data variables and the cognitive agent-based simulation model used in this research. It also provides a brief summary of the proposed methodologies. In Section 4, we present a summary of the main findings of our research and we thoroughly discuss its key results. In Section 5, we compare the investigated approach with the Artificial Neural Network approach. Lastly, in Section 6, we draw conclusions from our study and provide additional remarks regarding potential directions for future research.

2. Related Work

The simulation data from the considered model of cognitive agents learning to cross a cellular-automaton-based highway have previously been studied using classical statistical methodologies [45–47,64]. In a recent study by [48], canonical correlation analysis was conducted to examine the correlations between variables to ensure that no important variables would be omitted in the future research. Additionally, regression tree analysis was performed to explore the effects of model configuration parameters on agents' decisions, with a particular emphasis on the Knowledge Base Transfer parameter. The findings contributed to a better understanding of how autonomous agents learn in different traffic environments. The methodology proposed in [48] can also be applied to analyze data from other simulation models to gain deeper insights into factor effects. The regression tree analysis in [48] illustrated the input factor effects on the decision outcome variables. However, the nature of regression tree analysis limited its ability to provide quantitative information about the functional relationships between input and output variables, as well as the quantification of variable importance. To overcome this limitation, [49] modelled the same simulation data using ANNs. The findings from [49] demonstrated that ANNs effectively capture the relationships between input and output variables, outperforming classical modelling approaches such as Generalized Linear Models (GLMs). However, the non-parametric nature of ANNs allows investigation of modelling performance solely through plots of observed and predicted values. Additionally, the study in [49] addressed factor effects and their significance by measuring variable importance. The primary objective of [49] was to extend the investigation of simulation data from models of complex systems beyond classical statistical methods to advanced statistical modelling and analysis. However, the variable importance measured through the weight values obtained from ANNs offered only limited model interpretability. Therefore, in [50], the concept of explainable data analysis was introduced for a better understanding of the agents' learning performance in the cognitive agent-based simulation model studied in both [48,49]. The study in [50] utilized principal component analysis to reduce input data dimensions and then applied K-Means clustering to group the data. Two new algorithms were developed to identify the optimal number of clusters, enhancing clustering results and effectively reducing the number of treatments. The proposed algorithms showed promising results and outperformed traditional methods for determining the optimal number of clusters in K-Means clustering. The work in [50] offers a general solution to the problem of analyzing high-dimensional computer simulation data. However, the problem of variable importance was not specifically addressed in this work.

Recent advancements in complex systems analysis include the integration of machine learning techniques [65–67], such as deep learning and reinforcement learning [68], to extract deeper insights from both real-world and simulation data of these systems. For example, in study [68], deep reinforcement meta-learning and self-organizing approaches were applied for traffic signal control, demonstrating that meta-learning methods outperform classical learning methods. The deep reinforcement learning approach has proven effective in understanding how local behaviours influence global system dynamics. Furthermore, explainable AI (XAI) techniques have gained significant attention in complex systems analysis. In [69], an XAI approach was proposed for distress prediction in complex financial systems. These methods incorporating explainable AI techniques aim to provide insights into the decision-making processes of AI models, making them more transparent and interpretable. In healthcare, XAI has been explored to address the challenge of AI being perceived as a "black box", emphasizing its role in building trust and improving understanding of AI-driven decisions [70]. Research in study [71] emphasizes the need to tailor explanations to specific user types, offering an overview of XAI's recent developments, techniques, and assessment methods.

Additionally, the modelling and analysis of complex systems using XAI have been explored in [72–74]. Despite these advances, a key limitation of using explainable methods in machine learning, deep learning, and reinforcement learning, to enhance understanding of relationships between input and output variables in complex systems, lies in the trade-off between model accuracy and interpretability. Although deep learning networks and reinforcement learning agents excel at capturing patterns in data, extracting meaningful explanations for their decisions remains challenging. XAI techniques, such as feature importance analysis or surrogate models, seek to bridge this gap, but they often simplify the underlying models, potentially losing critical details essential for accurate predictions. Furthermore, the computational cost of some explainability methods, especially in deep learning, presents challenges to their scalability and practicality in real-time applications.

Transitioning from machine learning models to statistical methods for data analysis of complex systems [75–77], GLMMs have been widely adopted to investigate both the fixed and random effects of underlying factors. For instance, in [78], GLMMs were used to identify factors that influence formulaic sequence development over time, highlighting their ability to account for both systematic and individual variations. In another example, researchers in [79] used GLMMs to analyze ants' movement patterns, examining metrics such as average speed, event duration, and stopping duration between events, which revealed a power law relationship indicating predetermined movement durations. Additionally, in [80] GLMMs were applied to analyze complex systems data from the US Department of Defense, where the inclusion of random effects significantly enhanced model accuracy. This approach demonstrated that incorporating random effects leads to greater precision in the models. GLMMs have gained substantial traction among ecologists and evolutionary biologists due to their ability to handle uncertainties and their flexibility in analyzing non-normally distributed data within ecological systems. Research in this domain has explored statistical inference and outlined optimal data analysis practices tailored for scientists in these fields [81], as well as power analysis techniques specifically for GLMMs in ecology and evolution [82]. However, despite their advantages, GLMMs are best suited for modelling linear relationships. While they can incorporate random effects to manage hierarchical or clustered data, they often encounter challenges in capturing complex, nonlinear interactions between variables in complex systems.

3. Methods and Materials

3.1. Simulation Model

The model of cognitive agents learning to cross a cellular automaton (CA)-based highway was developed in [64]. Cognitive agents employ simple learning algorithms rooted in an "*observational social learning*" mechanism [42,83], where each agent learns by observing the outcomes of other agents' attempts to cross and by imitating those that succeeded. The modelling approach shares similarities with children learning to cross a road at "uncontrolled pedestrian crossings" and animals learning to cross a country road. The simulation model consists of several components: the highway, vehicles, agents, their knowledge and learning approaches, and agents' decision processes. This article aims to present a concise introduction to this model and its designed factors for simulation setups. A more detailed description of the considered simulation model and prior statistical analysis of its simulation data can be found in [45,46,48–50,64].

3.1.1. The Highway

The highway is modelled by a sequence of cells (each cell representing 7.5 m of a highway); its length and the number of lanes can be adjusted as per user specifications. In this paper, we consider a unidirectional single-lane highway. As vehicles move along the highway, the occupancy state of each cell—whether occupied or vacant—dynamically changes based on the presence of a vehicle, continuously updating over time. When a vehicle enters a cell already occupied by an agent, the agent is considered to be struck by the vehicle.

3.1.2. The Vehicles

In the model, vehicles are generated randomly at the start of the highway at each time step, based on a specified Car Creation Probability (CCP). Since each cell on the highway can only accommodate a single vehicle, any newly generated vehicle that encounters an occupied cell will wait until the cell is vacant before entering the highway. Vehicles on the highway follow the modified rules of the Nagel–Schreckenberg cellular automaton (CA) traffic model [84]. Their objective is to reach their maximum allowable speed while avoiding collisions with other vehicles. However, they do not decelerate to avoid agents crossing the highway. The traffic dynamics on the highway are influenced by two configuration parameters: Car Creation Probability (CCP) and Random Deceleration (RD). The CCP value determines the vehicle density on the highway, directly affecting the level of traffic congestion. For example, a CCP value of 0.1 indicates a low vehicle density, allowing vehicles to move freely, whereas a CCP value of 0.9 signifies a very high vehicle density, resulting in full congestion. The RD parameter modifies vehicle motion behaviour and can have a value of either 0 or 1. When set to 0, vehicles adhere to the Nagel–Schreckenberg traffic model rules, adjusting their speed to avoid collisions. When set to 1, vehicles randomly decelerate with a probability of 0.5 at each time step, provided that this deceleration does not cause a collision. Thus, the RD parameter introduces more randomness into vehicle motion, influencing the complexity of the traffic environment. By adjusting the CCP and RD parameters, the model can simulate a wide range of traffic scenarios to analyze their impact on the cognitive agents' learning ability to cross the highway, which is the focus of this research.

3.1.3. The Agents, Their Knowledge and Learning Approaches

In the simulation model of cognitive agents learning to cross a CA-based highway, the agents' primary objective is to successfully cross the highway, i.e., to avoid being hit by an oncoming vehicle. At every time step, an agent is generated at a specified crossing point located on one side of the highway and is placed in a queue at this crossing point. In the considered simulations, the agents are generated at the crossing point 60 cells away from the start of the highway, which corresponds to 450 m from the beginning of the highway. Once an agent reaches the edge of the highway, it is called an "active agent" and it faces three possible actions:

1. Attempt to cross the highway;
2. Wait at the crossing point;
3. Move along the highway to a neighbouring crossing point, referred to as "Horizontal Movement" in the presented simulations.

Each active agent attempts to cross the highway in two steps: (1) entering the highway and (2) exiting the highway, i.e., completing the crossing. If an active agent is struck by an oncoming vehicle during the crossing attempt, it is promptly removed from the simulation. This represents the active agent's unsuccessful crossing. Active agents make their crossing decisions based on a specific decision formula. The active agents' decision formulas will be discussed in Section 3.1.4. Additionally, the outcomes of active agents' crossing decisions are influenced by four key experimental factors/parameters of the model: "Horizontal Movement" (HM), "Fear", "Desire", and "Knowledge Base Transfer" (KBT).

The factor HM determines whether an active agent can move to one of the neighbouring cells while waiting to attempt to cross. When HM = 0, active agents must wait at their selected crossing points and cannot move. However, if HM = 1, each active agent has the option to remain at its current crossing point or to move randomly, one cell to the left or right along the highway, with a probability of 1/3 for each case. As a result, if HM = 1, multiple crossing points may emerge during the simulations, even if only one crossing point is initialized at the start of each simulation. Since the active agents use "observational social learning" (i.e., they imitate successful strategies from other agents and avoid unsuccessful ones), multiple crossing points increase the active agents' ability to observe the outcomes of crossing decisions of other agents, thereby enhancing their learning. The maximum distance that the active agents can move away from their initial crossing point is capped at five cells in our simulations.

The agents' propensity for risk taking is controlled by the factor called "Desire", while their propensity for risk aversion is controlled by the factor called "Fear". For example, if Fear = 1 and Desire = 0, an agent exhibits complete fear and no desire to cross the highway, indicating total risk aversion. We consider 25 different combinations of Fear and Desire values in our experimental setups. Each combination, alongside each level of the other factors, constitutes a treatment in our simulation experimental design, aiming to analyze the effects of these factors and their interactions.

To enable agents to learn from the outcomes of past crossing attempts, the model incorporates a knowledge-based (KB) table that records the active agents' both correct and incorrect crossing and waiting decisions made by them when attempting to cross the highway. The KB table maintains a record of these decisions for different estimated distances of oncoming vehicles (close, medium, far, and out of visual range) and their estimated speeds (slow, medium, fast, and very fast). Thus, at each time t, the information in the KB table provides a historical record of correct and incorrect crossing and waiting decisions for each combination of vehicle distance and speed, i.e., the assessments of active agents' decisions up to time t. Before making a crossing decision, an active agent accesses the KB table and, based on its knowledge of the historical records, decides whether to cross or not. The KB table entries are continuously updated throughout each simulation run, with each decision assessment leading to an update of the corresponding table entry.

The model employs two distinct learning approaches, determined by the value of the "Knowledge Base Transfer" (KBT) factor. This factor dictates whether the KB table of the initial crossing point is transferable at the end of a simulation run with a lower CCP value to the agents at the start of a simulation run with an immediately higher CCP value. When KBT = 0, the KB tables are not transferred from agents in a traffic environment with a lower CCP value to those in an environment with an immediately higher CCP value, or any other value. When KBT = 1, the KB table is always transferred at the end of a simulation run from agents in a lower CCP environment to agents in an immediately higher CCP environment at the beginning of a simulation run. This process of KB table transfer is carried out for each simulation repeat. The factor KBT also determines the initialization of KB tables. For example, when KBT = 0, the KB table is initialized as a tabula rasa, i.e., a "blank slate" represented by zeros at each table entry. During the initialization period of each simulation run, active agents attempt to cross the highway regardless of the observed distance and speed combinations until either the first successful crossing or five consecutive unsuccessful crossings occur, whichever comes first. After this initialization period, active agents make decisions based on the KB table and their decision formula. However, when KBT = 1, the KB table is initialized as a tabula rasa only for simulations with CCP = 0.1. For higher CCP values, each KB table is initialized with the information accumulated from less dense traffic environments.

3.1.4. The Agents' Decisions and the Model Simulation Loop

As discussed earlier, each active agent makes its decision to cross or not cross the highway based on its knowledge of historical records that evaluate the outcomes of crossing and waiting decisions made by other agents. The results of an active agent's decision can be classified into the following categories:

1. Correct Crossing Decision (CCD): The active agent decided to cross the highway and did so successfully.
2. Incorrect Crossing Decision (ICD): The active agent decided to cross but was struck by an oncoming vehicle.
3. Correct Waiting Decision (CWD): The active agent decided to wait, which was the correct choice, as crossing would have resulted in being hit by an oncoming vehicle.
4. Incorrect Waiting Decision (IWD): The active agent decided to wait, but if it had crossed, it would have done so successfully.

As mentioned earlier, the assessment of each decision (i.e., CCD, ICD, CWD, and IWD) is recorded as a count in the KB table of all agents waiting at the active agent's crossing point. When an active agent is deciding whether to cross, it first accesses the KB table and integrates this information into its decision-making process. Depending on the simulation setup, the active agent may use one of the following decision-making formulas: the Crossing-Based Decision Formula (cDF) or Crossing-and-Waiting-Based Decision Formula (cwDF). The Crossing-Based Decision Formula (cDF) considers only historical data on crossing decisions of active agents (i.e., CCD and ICD), while the Crossing-and-Waiting-Based Decision Formula (cwDF) includes historical data on both crossing and waiting decisions (i.e., CCD, ICD, CWD, and IWD). Each decision-making formula also takes into account the active agent's risk preferences, quantified by its Fear and Desire factors. The values of Fear and Desire factors provide insight into the active agent's inclination toward risk taking or risk aversion. Further mathematical details on the cDF and cwDF are provided in [45,46,64].

It is worth noting that the cDF incorporates only one feedback loop that assesses crossing decisions, whereas the cwDF integrates two feedback loops, one for crossing decisions and one for waiting decisions. Our simulations show that using two feedback loops in the decision-making process significantly improves active agents' performance. More agents learn to cross the highway successfully, while the number of struck agents remains steady. Additionally, the number of agents waiting in queues is reduced by the end of the simulation. Thus, when agents use the cwDF instead of cDF, they make more correct crossing decisions (CCDs) by reducing the number of incorrect waiting decisions (IWDs), while the number of incorrect crossing decisions (ICDs) remains almost unchanged.

To recap, the simulation model of cognitive agents learning to cross a CA-based highway includes several key components: the highway, vehicles, agents, their knowledge and learning approaches, and their decision-making processes. After loading the configuration and Knowledge Base files, the program executes the main simulation loop, which runs at each time step. This loop involves the following steps:

1. Generating vehicles at the start of the highway based on the Car Creation Probability.
2. Generating agents at each crossing point (CP) with their attributes of Fear and Desire.
3. Updating vehicle speeds according to the modified Nagel–Schreckenberg model.
4. Moving active agents from their CP queues onto the highway when the decision algorithm indicates they should cross.
5. Updating vehicle locations on the highway and checking whether any active agent has been hit.
6. Advancing the current time step.

Once the simulation is complete, the results are written to output files using the output functions.

3.2. *Experimental Setting and Simulated Data*

In this section, we discuss the experimental setup of the simulation model of cognitive agents learning to cross the CA-based highway. The resulting data are analyzed in this paper. The model involves many parameters/factors, which we have discussed earlier, e.g., highway length, vehicle/Car Creation Probability (CCP), vehicle Random Deceleration (RD), the selection of crossing points (CPs) at the initialization of a simulation setup, agents' "Horizontal Movement" (HM), which controls agents' ability to move to neighbouring crossing points, agents' Knowledge Base Transfer (KBT), and decision formula (DF) type, which can be cDF or cwDF.

To compare the learning performance of agents using the cDF with those using the cwDF, two data sets were generated—one for the cDF and the other for the cwDF. This was the only difference in DF parameter setup between the two data sets. In the experimental setup of the simulation model, the constant values of the parameters were as follows: (1) a single-lane highway of 120 cells in length (i.e., a stretch of 900 m of a real highway, as each cell represents 7.5 m of the highway); (2) a single CP set at cell 60 at the initial setup; (3) 1511 time steps for each simulation run; (4) 30 repetitions for each simulation setup; and (5) a representation of the KB table at each CP in a 3×4 matrix with an additional row entry. Each KB table has three groupings of distance and four groupings of speed. A vehicle is perceived as close if it is 0 to 5 cells from the CP, medium distance if 6 to 10 cells away, far if 11 to 15 cells away, and out of visual range if 16 or more cells away from the CP, regardless of its speed. This is encoded in the additional row entry of the KB table. A vehicle's speed is perceived as slow if it travels 0 to 3 cells per time step, medium speed if 4 or 5 cells per time step, fast if 6 or 7 cells per time step, and very fast if it travels 8 to 11 cells per time step. The maximum speed of a vehicle can be 11 cells per time step, which is equivalent to 99 km/h. For each decision formula, there are six common parameters/factors whose values vary in the simulation setups. These parameters are as follows:

1. CCP (Car Creation Probability): This parameter determines the density of cars' traffic and varies between the following values: 0.1, 0.3, 0.5, 0.7, and 0.9.
2. RD (Random Deceleration): If RD = 1, each car has a probability of 0.5 of randomly decreasing its speed by 1 unit; if RD = 0, this is not allowed. The RD parameter simulates erratic drivers in the model.
3. HM (Horizontal Movement): This parameter takes values of 0 or 1. It determines whether active agents can move along the highway, away from their CPs, when they decide to wait. Active agents can "move horizontally" only when HM = 1. In the simulations, an active agent can move one cell per time step, and the maximum distance from its CP is 5 cells in both directions. Thus, when HM = 1, up to 11 active agents may be making crossing decisions simultaneously. When HM = 0, active agents are not allowed to move or change their CPs.
4. KBT (Knowledge Base Transfer): This parameter takes values of 0 or 1. The KBT parameter determines if the KB table from agents at a lower CCP value is transferred at the end of a simulation run to agents in the following simulation run with an immediately higher CCP value. If KBT = 0, no transfer occurs; if KBT = 1, the KB table is transferred between simulation runs.

5. Fear Parameter: This parameter reflects an agent's risk aversion.
6. Desire Parameter: This parameter reflects an agent's propensity for risk taking.

Both the Fear and Desire parameters vary between the following values: 0.00, 0.25, 0.5, 0.75, and 1.00. These parameters are part of the active agents' decision-making formulas and influence an agent's success in learning to cross the highway.

Note that the KBT, HM, RD, and DF parameters are binary, while CCP, Fear, and Desire are categorical. The full simulation means that simulations were carried out for all the parameters' combinations described. The data analyzed in this paper come from the full simulation of the model. For each decision formula (i.e., cDF and cwDF), the data are organized into a matrix with 45,330,000 rows and 12 columns. When needed, the two matrices were merged for analysis. The 12 column headings are as follows:

- Time: In this column is recorded each time step in each simulation repeat.
- CCD, CWD, ICD, IWD: These columns, respectively, record the numbers of correct crossing decisions, correct waiting decisions, incorrect crossing decisions, and incorrect waiting decisions at each time step in each simulation repeat.
- Rep: Records the repetition number for each simulation setup.
- CCP, Fear, Desire, KBT, RD, HM: These columns record the values of the respective parameters: 5 values for CCP, 5 for Fear, 5 for Desire, 2 for KBT, 2 for RD, and 2 values for HM.

Thus, for each decision formula, 1000 different parameter value combinations were analyzed. The 45,330,000 rows in each data matrix of a respective decision formula result from multiplying 1511 time steps, 30 repetitions, and 1000 parameter value combinations. However, in this work only the data at the end of each simulation repeat was analyzed, i.e., the numbers of CCDs, ICDs, CWDs and IWDs at the final simulation step 1511.

3.3. Discussion of the Simulation Model and Its Potential Extensions

The presented model of cognitive agents learning to cross a CA-based highway was developed as an experimental platform to identify suitable minimal cognitive agents using *social learning* and *biomimicry* as approaches to learning in dynamically changing environments. It emphasized minimal entities, both in terms of storage and logical primitives, keeping in mind potential implementations in swarms of simple robots with minimal computational resources operating in isolation. Thus, the model deliberately avoided formal methods and established algorithms, instead exploring learning algorithms that are not computationally demanding. However, the presented simulation model is sufficiently general for its cognitive agents to be considered as an abstraction of an autonomous vehicle (AV) learning what to do when it suddenly encounters another moving vehicle on its path. The AV must decide whether to continue or brake/stop to avoid a collision. The AVs make their decisions solely based on the decision outcomes of other AVs encountering such situations in the past, without any communication among themselves or with infrastructure, such as road sensors. In many real-world applications, especially in developing regions or rural highways, advanced Internet of Things infrastructure, such as roadside units or networked vehicle-to-everything communication, may not be available. Thus, modelling cognitive agents that do not rely on networked inputs remains relevant, particularly in environments where 5G and Internet of Things technologies are not fully deployed or where such systems may malfunction. Nevertheless, we acknowledge the importance of incorporating networked agents in future models as Internet of Things and 5G technologies become more widespread. Such models could indeed improve cognitive agent performance, potentially reducing collisions and improving efficiency by leveraging shared data between vehicles and cognitive agents.

In models of Internet of Things-enhanced traffic environments, the cognitive agents' decision-making process would have to undergo significant changes to incorporate real-time data from roadside units and networked vehicles. The cognitive agents would receive information about oncoming vehicles' positions and speeds, allowing them to make more accurate decisions. Consequently, cognitive agents would no longer solely rely on historical knowledge of outcomes of decisions made by other agents but could incorporate real-time data into their decision-making processes, improving the accuracy of their decisions. Furthermore, incorporating the abstraction of Internet of Things technology into the cognitive agents' simulation model would enable predictive modelling, where cognitive agents could anticipate future traffic conditions based on data transmitted from networked vehicles. This would further enhance the cognitive agents' decision-making. However, such modified models would be far more complex and require significantly greater computational resources than the presented one, and the cognitive agents' performance would depend on many more factors than those considered in this work. Thus, the statistical analysis of their simulation data would be far more challenging than that of the presented model.

3.4. Effects of Parameters Using Linear Mixed-Effects Models

Expanding on the foundations of Linear Models (LMs), the Linear Mixed-Effects Model (LMM) introduces specific linear predictors that integrate random effects alongside the typical fixed effects. When all the included random effects are statistically insignificant, the model simplifies to the conventional LM. In this study, we incorporate the parameters CCP, Fear, and Desire as random effects, while KBT, HM, RD, and DF serve as fixed effects. This selection and configuration of the statistical model were motivated by the continuous nature of the CCP, Fear, and Desire factors, with specific levels chosen for simulation purposes, in contrast to the remaining factors that possess only two levels. The model coefficients estimate the effect differences between these two levels. The following description outlines the structure of this Linear Mixed-Effects Model:

$$Y^i = \alpha_0^i + \alpha_1^i \mathrm{KBT} + \alpha_2^i \mathrm{HM} + \alpha_3^i \mathrm{RD} + \alpha_4^i \mathrm{DF} + (1|\mathrm{CCP}) + (1|\mathrm{Fear}) + (1|\mathrm{Desire}) + \epsilon_i, \quad (1)$$

where the notation of $(1|\cdot)$ represents a random intercept, contributed by a level of the random-effect variable. So, there are multiple random effect intercepts from different components, and one random effect intercept is often superimposed on another to become the model intercept. In model (1), the error is assumed to be Gaussian-distributed. Since a random effect variable is assumed to have a mean of zero and a constant variance–covariance matrix in a Linear Mixed-Effects Model, we can estimate the relative effect at each level of the underlying factor and the standard deviations associated with these estimates. Therefore, the confidence interval of these estimates can be constructed. Additionally, by measuring the variability of these random effects across the levels of the underlying factor, one can further evaluate the variable importance associated with each random-effect variable.

It is important to note that, in Equation (1), the error term is assumed to follow a Gaussian distribution, which may not be appropriate for count data. Since the response variable in our study represents the total number of decisions, it is necessary to extend the LMM framework to a GLMM to better accommodate the characteristics of the count-based response.

3.5. Modelling the Effects of Parameters Using Generalized Linear Mixed-Effects Models

The GLMM is a versatile approach frequently employed in the analysis of complex systems. In our specific context, the responses are diverse in nature, representing various counts of decisions made by cognitive agents. Given the non-normal distribution of these responses, the applicability of the LMM is limited. A key advantage of the GLMM lies in its ability to extend the capabilities of the LMM by accommodating error distributions belonging to an exponential family. This feature increases the model's flexibility in handling various types of response distributions. The density function governing the response variables in the GLMM can be expressed as follows ([85]):

$$f(Y^i|\theta^i, \phi) = \exp\left[\frac{Y^i\theta^i - b(\theta^i)}{a(\phi)} + c(Y^i, \phi)\right], \quad (2)$$

where θ^i is the canonical parameter characterizing the location of the ith observation and ϕ is the dispersion parameter characterizing the scale. In GLMMs, instead of modelling Y^i directly, a linear predictor η^i is introduced such that the Linear Mixed-Effects Model (1) can be rewritten as

$$\eta^i \quad = \quad \alpha_0^i + \alpha_1^i \text{KBT} + \alpha_2^i \text{HM} + \alpha_3^i \text{RD} + \alpha_4^i \text{DF} + (1|\text{CCP}) + (1|\text{Fear}) + (1|\text{Desire}). \quad (3)$$

Each random intercept follows a normal distribution with a mean of zero and a specific variance, which represents the degree of heterogeneity within the population. A specific link function, denoted as $g(\cdot)$, establishes the connection between the linear predictor η^i and the expected value of Y^i:

$$g(E(Y^i)) = \eta^i. \quad (4)$$

As shown in Equation (4), the model offers flexibility by allowing transformations on the expected value of the response in Equation (1). A specific case arises when the link function $g(\cdot)$ is the identity function, implying that the linear predictor is equivalent to the expected value of the response variable, Y^i.

This comprehensive framework of the GLMMs enables us to incorporate the complex dependencies and variations observed within the data, thereby facilitating a more accurate representation of the underlying processes governing the decision-making patterns of the cognitive agents in our model.

Among the various error distributions available for GLMMs, the Tweedie distribution emerges as a particularly valuable option, adept at managing positively skewed data. Notably, the Tweedie distribution is a specific case within the broader class of exponential family distributions. Our analysis explores three key instances encapsulated within the Tweedie distribution: Poisson, Gamma, and Inverse Gaussian distributions. This investigation enables the implementation of these distributions within the GLMM framework. Furthermore, it is crucial to emphasize that we consistently employ the log link function across all three distributions to ensure uniformity and coherence across our analyses.

This precise and comprehensive approach not only allows us to address the intricacies of positively skewed data effectively but also reinforces the robustness and reliability of our analytical framework. By adopting a carefully designed and consistent approach to data modelling, we ensure the integrity and accuracy of our analytical outcomes, thereby significantly enhancing the depth and credibility of our research findings.

Various methods have been developed for parameter estimation within the GLMM framework, all rooted in the principle of maximum likelihood estimation. Let the fixed effects for KBT, HM, RD, and DF be denoted as $\boldsymbol{\alpha}$, and the random effects for CCP, Fear, and Desire as $\boldsymbol{\gamma}$. The likelihood function is formulated as follows:

$$L(\boldsymbol{\alpha}, \phi | Y) = \prod_{i=1}^{n} f(Y^i | \boldsymbol{\alpha}, \phi, \boldsymbol{\gamma}) l(\boldsymbol{\gamma}) d\boldsymbol{\gamma}, \quad (5)$$

where $l(\boldsymbol{\gamma})$ represents the distribution of the random effects. The integration process presents challenges in obtaining a closed-form solution due to the potential non-normality of the function $f(\cdot)$. A commonly adopted approach to address this complexity is the use of Taylor expansions [86]. For example, the technique of penalized quasi-likelihood incorporates both fixed and random effects within the second-order approximation of the Taylor series expansion [87]. Although effective and straightforward in the context of LMMs, this method may produce biased estimates when the response distribution deviates from normality, as observed in the case of a Poisson distribution [85].

In our study, estimates for both fixed and random effects are obtained using the *glmer* function, a tool available within the 'lme4' package for R (version 3.6.0 or higher). This function employs a numerical integration technique to approximate the likelihood of the GLMM [85]. Specifically, it employs the Laplace approximation, where the integral takes the form of $\int \exp\{l(x)\}dx$ and the integration is approximated by identifying the maximum and second derivative of $l(x)$. This method facilitates a robust and efficient estimation process, producing reliable results even when confronted with complex and non-normally distributed data. This ultimately improves the accuracy and reliability of our parameter estimates.

4. Results

In this section, we present the results of modelling and analyzing the simulation data generated by our simulation model. We begin with the simulation data obtained from different scenarios using diverse simulation inputs. The workflow of this investigation is illustrated in Figure 1.

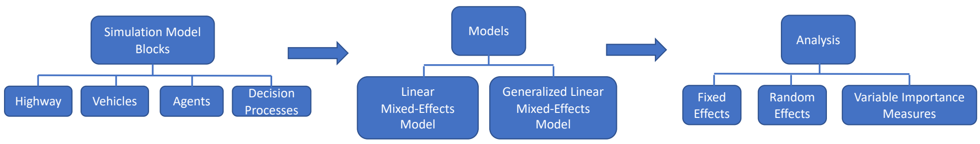


Figure 1. The workflow of the presented research, including the cognitive agent simulation model, statistical modelling, and analysis.

4.1. Results of Linear Mixed-Effects Modelling of Cognitive Agent Decisions

In the LMM, we treated the factors KBT, HM, RD, and DF as fixed effects because we assumed they had a constant relationship with any of the responses. However, a closer examination reveals that the influence of these fixed effects may vary across different response variables. This observation highlights the diverse impacts of these fixed-effect factors, corresponding to distinct decision outcomes within the population of cognitive agents. Notably, the estimates of random effects associated with the factors CCP, Fear, and Desire are determined based on the grouping of responses according to different levels of these factors. The uniformity of these random effects within each level enhances the understanding of their specific influences on decision patterns.

The model output from the LMM, as displayed in Table 1, reveals the significance of all four response variables within the models, reiterating their influential role on decision-making processes. Thus, further validation of the importance of these input factors is provided. Additionally, a notable pattern emerges, revealing that Knowledge Base Transfer leads to a significant increase in the number of CCDs, while decreasing the numbers of ICDs, CWDs, and IWDs. This suggests that Knowledge Base Transfer enhances cognitive agents' ability to learn to cross the highway, despite the fact that the transferred knowledge originates from agents who developed it in a different traffic environment than those receiving it. This finding underscores the strong connection between the use of knowledge-based strategies and the resulting decision outcomes.

Turning to the estimates of random effects, Table 2, along with the confidence interval plots in Figures 2 and 3, highlights the nuanced influence of each level of the factors CCP, Fear, and Desire on the decision-making process. These visual representations emphasize the varying impacts of different levels of these factors on the decision outcomes, establishing a clear monotonic relationship—an interpretative advantage of the LMM. The interpretability of the LMM, in contrast to the previously employed GLM and GAM, greatly enhances the understanding of the complex dynamics underlying decision-making patterns.

Furthermore, in Table 3, we measure the variable importance by calculating the standard deviation for continuous random variables and the range divided by four for discrete random variables. To maintain consistency with the settings in GLMMs, we treat random-effect variables as discrete. This is reasonable since the factors CCP, Fear, and Desire consist of multiple levels. Figure 4 presents the variable importance plots for GLMMs with Gaussian error functions. We observe that the variable importance rankings are as follows: HM, Fear, DF, KBT, CCP, Desire, and RD for CCD; Desire, HM, Fear, CCP, RD, KBT, and DF for ICD; Fear, HM, CCP, Desire, RD, KBT, and DF for CWD; and Fear, HM, KBT, DF, CCP, Desire, and RD for IWD. These results closely mirror the rankings from the previous models, where the predictors HM and Fear play key roles in cognitive agents' decision-making, and predictor Desire exhibits a strong relationship with the number of ICDs.

Table 1. The estimated linear coefficients with standard errors for the fixed-effect terms of the Linear Mixed-Effects Models (LMMs) that fit the responses CCD, ICD, CWD, and IWD, respectively.

	CCD	ICD	CWD	IWD
(Intercept)	186.55 *	3.94	66.48 ***	2389.71 ***
	(87.39)	(2.44)	(11.77)	(421.86)
KBT	176.56 ***	−2.59 ***	−16.45 ***	−1484.83 ***
	(1.62)	(0.04)	(0.33)	(16.18)
HM	622.03 ***	8.57 ***	32.94 ***	1915.08 ***
	(1.62)	(0.04)	(0.33)	(16.18)
RD	−5.63 ***	3.92 ***	−17.44 ***	35.11 *
	(1.62)	(0.04)	(0.33)	(16.18)
DF	260.97 ***	0.66 ***	−1.62 ***	−1220.68 ***
	(1.62)	(0.04)	(0.33)	(16.18)
AIC	805,479.02	366,817.18	614,138.94	1,081,307.28
BIC	805,560.03	366,898.20	614,219.96	1,081,388.30
Log Likelihood	−402,730.51	−183,399.59	−307,060.47	−540,644.64
Num. obs.	60,000	60,000	60,000	60,000
Num. groups: CCP	5	5	5	5
Num. groups: Fear	5	5	5	5
Num. groups: Desire	5	5	5	5
Var: CCP (Intercept)	2512.83	7.28	121.64	82,155.53
Var: Fear (Intercept)	34,862.53	8.32	503.70	762,263.23
Var: Desire (Intercept)	793.90	14.13	67.07	43,757.86
Var: Residual	39,550.80	26.40	1629.59	3,925,780.43

Note: *** $p < 0.001$; * $p < 0.05$.

Table 2. The estimated random-effect coefficients in Linear Mixed-Effects Models that fit the responses CCD, ICD, CWD, and IWD, respectively.

	CCD	ICD	CWD	IWD
CCP				
0.1	−87.47	3.85	18.95	502.47
0.3	5.52	1.64	−5.88	−37.13
0.5	19.43	−0.90	−7.88	−125.92
0.7	27.80	−2.03	−5.5	−168.11
0.9	34.72	−2.57	0.30	−171.31
Fear				
0	269.91	4.02	−24.57	−1277.81
0.25	61.11	1.42	−16.87	−380.34
0.5	−7.86	−0.30	−4.95	222.16
0.75	−88.89	−1.64	19.64	409.88
1	−234.27	−3.50	26.75	1026.1047
Desire				
0	−45.17	−5.73	12.56	337.32
0.25	−8.76	−1.31	3.36	52.98
0.5	11.05	0.65	−2.45	−66.47
0.75	20.09	2.43	−5.70	−150.04
1	22.80	3.96	−7.76	−173.79

Table 3. Variable importance measures for response variables CCD, ICD, CWD, and IWD based on standard deviations of the factors KBT, HM, RD, DF, CCP, Fear, and Desire considered in the Linear Mixed-Effects Models.

	Dependent Variable:			
	CCD	**ICD**	**CWD**	**IWD**
KBT	44.141	0.648	4.113	371.209
HM	155.508	2.143	8.236	478.769
RD	1.407	0.981	4.360	8.778
DF	65.242	0.166	0.406	305.171
CCP	30.548	1.607	6.706	168.445
Fear	126.047	1.882	12.831	575.979
Desire	16.993	2.422	5.080	127.779

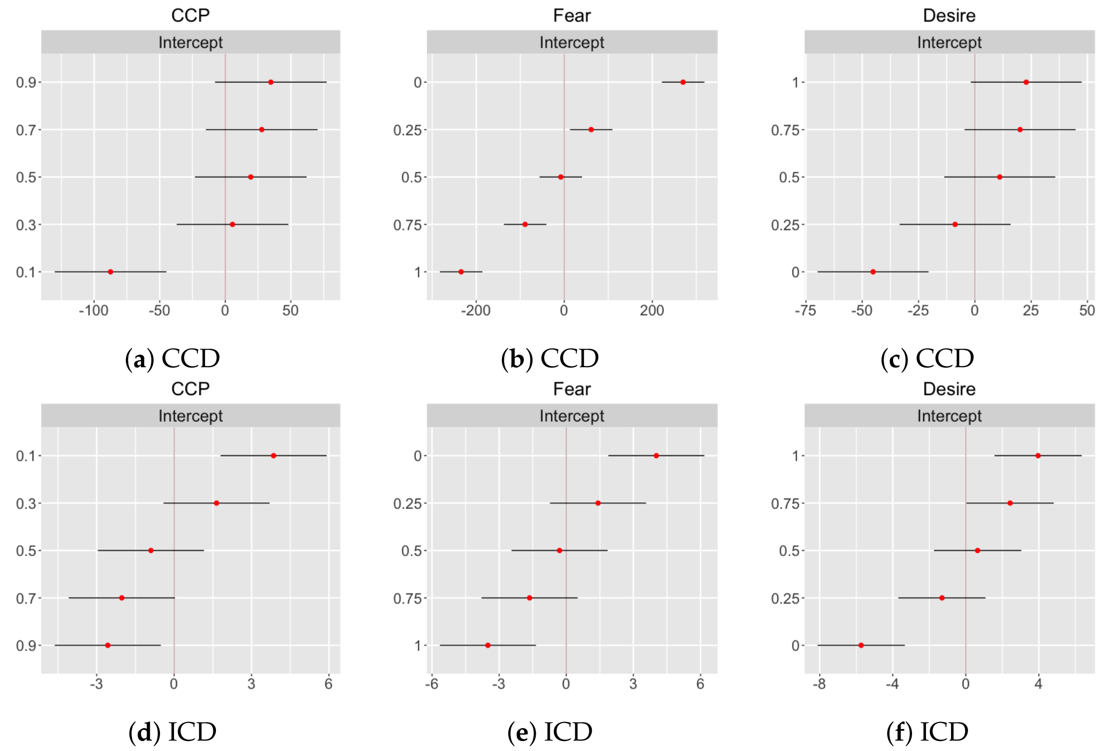


Figure 2. The plots of confidence interval estimates of random intercepts associated with the factors CCP (first column), Fear (second column), and Desire (third column), respectively, in GLMM modelling of response variables CCD (first row) and ICD (second row) with Gaussian error function.

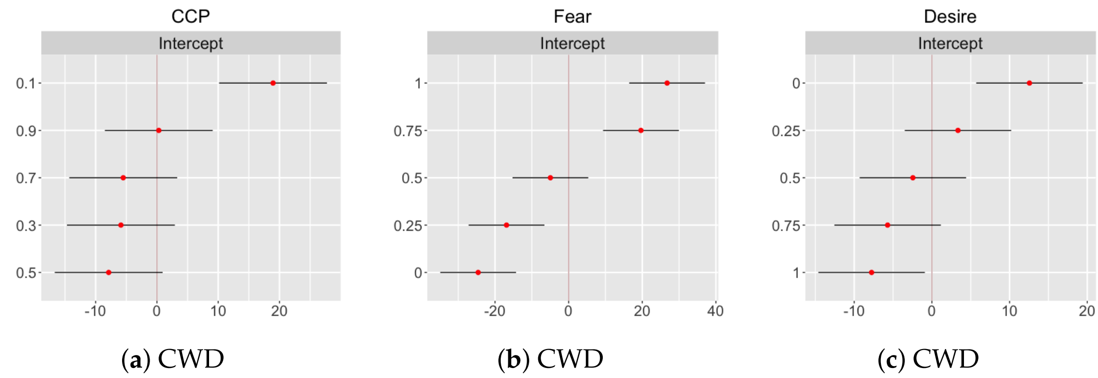
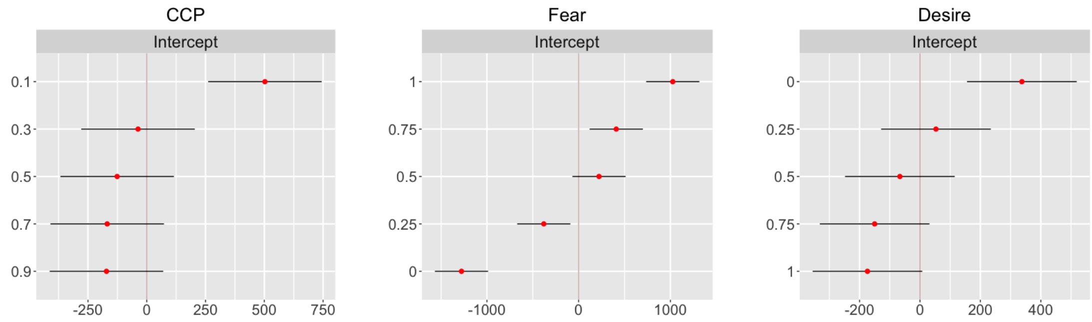
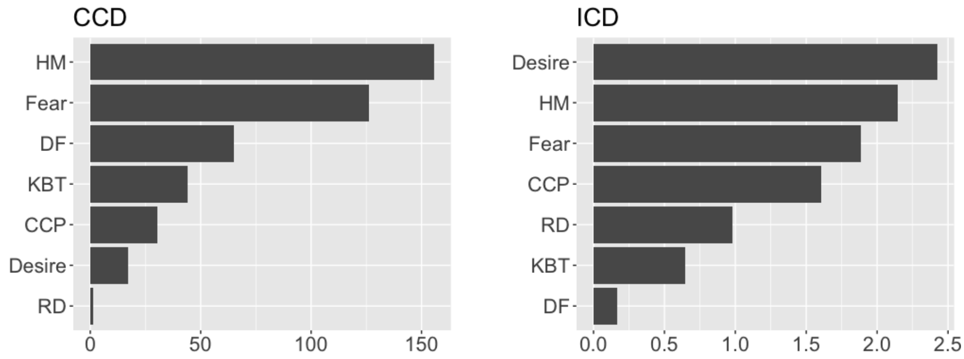


Figure 3. *Cont.*

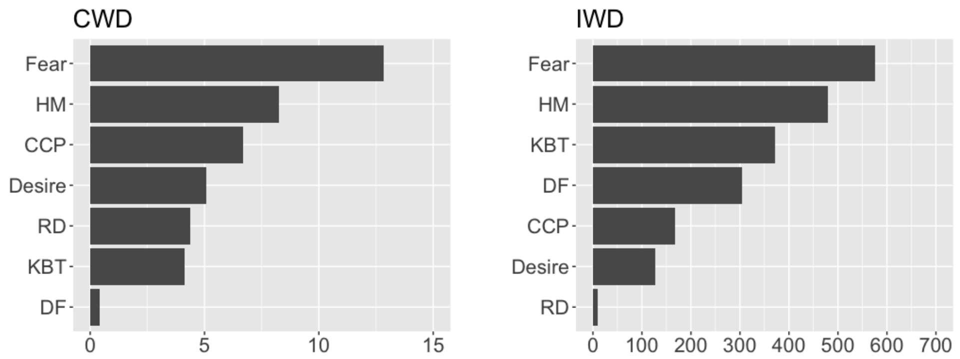


(**d**) IWD (**e**) IWD (**f**) IWD

Figure 3. The plots of confidence interval estimates of the random intercept associated with the factors CCP (first column), Fear (second column), and Desire (third column), respectively, in GLMM modelling of response variables CWD (first row) and IWD (second row) with Gaussian error function.



(**a**) CCD (**b**) ICD



(**c**) CWD (**d**) IWD

Figure 4. Variable importance measures of predictor variables in the Generalized Mixed-Effects Model with Gaussian error function, based on standard deviations, for response variables CCD, ICD, CWD, and IWD, respectively.

4.2. Results of Generalized Linear Mixed-Effects Modelling of Cognitive Agent Decisions

In this phase of the analysis, we broadened the scope of our investigation by fitting GLMMs, allowing for a comprehensive exploration and comparison of model performance across various error functions, including Poisson, Gamma, and Inverse Gaussian. The results of the fixed-effect coefficients in the GLMMs using different error functions for the considered decision types are presented in Tables 4–6. These tables display the results, respectively, for Poisson, Gamma, and Inverse Gaussian error functions. Notice that in Table 4, the presented results are for all four decision types, while in Tables 5 and 6, they are presented only for the decisions CCD, CWD, and IWD. The reason that we did not include the results for ICD in these tables is that the ICD dataset includes some data points where the number of ICDs = 0, and the Gamma and Inverse Gaussian models cannot fit data with a response value of zero. Thus, we could not apply Gamma or Inverse Gaussian models for this data set. Tables 4–6 provide information showing that each model offers valuable insights into the distinct contributions of the predictors, highlighting their respective magnitudes and directions of influence, thereby deepening our understanding of the dynamics governing the decision-making processes of cognitive agents. A closer examination of the results reveals that while the parameter estimates using Poisson and Gamma error functions show no statistically significant differences, they exhibit slight deviations from those obtained using the Inverse Gaussian error function. For example, the coefficient value of HM on CCDs is 0.93 and 1.25 under the Poisson and Gamma distributions, respectively. However, this coefficient value notably increases to 3.42 when using the Inverse Gaussian distribution, suggesting a more substantial impact of HM on correct crossing decisions within this model. Consistent with the findings from the GLM and LMM analyses, the predictors KBT, HM, RD, and DF significantly influence the decision-making process, reaffirming their critical role in types of outcomes.

Moreover, integrating Knowledge Base Transfer into the cognitive agents' process of learning significantly increases the number of their correct crossing decisions while simultaneously reducing the count of both correct and incorrect waiting decisions. In terms of model fit, the Inverse Gaussian error distribution produces the lowest AIC and BIC values, indicating a better fit compared to the Poisson and Gamma error functions. This finding underscores the importance of selecting an appropriate error function, demonstrating its critical role in achieving a more accurate and reliable representation of decision-making dynamics within the cellular automaton-based highway scenario.

The analysis of random-effect coefficients within the GLMMs, as delineated in Table 7, offers valuable insights into the intensity of the effects associated with each level of the parameters. Notably, a discernible monotonic pattern emerges across most of the random effects, highlighting a consistent trend in their influence on the decision-making processes of the cognitive agents. However, an exception to this pattern arises in the case of CCP's impact on correct waiting decisions. Specifically, as CCP increases from 0.1 to 0.7, we observe a corresponding decrease in the number of CWDs. Surprisingly, an increase in CCP from 0.7 to 0.9 correlates with an increase in the number of CWDs, representing a distinct shift in the impact of CCP on decision outcomes. The observed shift likely results from the fact that the traffic for CCP values 0.7 and 0.9 is extremely congested, making it difficult to cross the highway. Thus, waiting to cross it is correct most of the time.

This observation underscores the complex relationship between CCP and correct waiting decisions, emphasizing the dynamic interplay between these variables within the decision-making process. Such a nuanced understanding is crucial for designing effective interventions and strategies to optimize decision outcomes, particularly in scenarios where varying levels of CCP influence the agents' learning process. By exploring the implications of this unique pattern, we can gain deeper insights into the complex decision-making processes that govern the agents' behaviour.

The evaluation of variable importance for each predictor within the GLMM framework was conducted using flatness, measured by computing the standard deviation of each response variable. Table 8 summarizes the comprehensive variable importance scores, with the corresponding variable importance plots depicted in Figure 5. A thorough examination of the results reveals that the order of predictor importance in the Poisson GLMM consistently aligns with that observed in the LMM. However, an intriguing shift becomes apparent when considering the rankings of predictors within the Gamma GLMM and Inverse Gaussian GLMM. This notable discrepancy is observed in the varying rankings of predictors for each type of decision, highlighting the influence of the choice of error functions on assessing variable importance. For instance, when utilizing the Inverse Gaussian error function, the ranking of predictors for CCD shifts to the factors Fear, HM, DF, CCP, KBT, RD, and Desire, signalling a distinct hierarchy of influence compared to the other error functions. Despite these discrepancies, it is crucial to note that the most influential predictors remain consistent across the different error functions, with HM and Fear consistently emerging as dominant factors affecting the decision-making processes of the cognitive agents. The role of HM as an influential predictor is easier to understand than that of Fear. Recall that when HM = 1, multiple crossing points may emerge during the simulations, even if only one crossing point was initialized at the start of each simulation. Since the agents use "*observational social learning*" (i.e., they imitate successful strategies from other agents and avoid unsuccessful ones), multiple crossing points increase the agents' ability to observe the outcomes of crossing decisions of many more other agents compared to when only one crossing point is allowed (i.e., when HM = 0), thereby enhancing the cognitive agents' learning. This analysis provides a comprehensive understanding of the differential impacts of the error functions on the ranking of predictor importance and underscores the robust and stable influence of certain key predictors, such as HM and Fear, on agents' decision outcomes. Furthermore, this analysis highlights the role of these predictors and their significance in influencing the overall performance of the cognitive agents. Moving forward, a focused investigation into the main factors, namely HM, Fear, and Desire, can yield a more thorough understanding of the agents' learning and decision-making processes. This paves the way for developing improved strategies to enhance the cognitive agents' overall performance.

Table 4. Fixed-effect coefficients with standard errors for the Generalized Linear Mixed-Effects Model with Poisson error function.

	Poisson			
	CCD	**ICD**	**CWD**	**IWD**
(Intercept)	5.64 ***	1.29 ***	4.08 ***	7.42 ***
	(0.10)	(0.27)	(0.16)	(0.09)
KBT	0.25 ***	−0.28 ***	−0.25 ***	−0.77 ***
	(0.00)	(0.00)	(0.00)	(0.00)
HM	0.93 ***	1.01 ***	0.52 ***	1.04 ***
	(0.00)	(0.00)	(0.00)	(0.00)
RD	−0.01 ***	0.43 ***	−0.27 ***	0.02 ***
	(0.00)	(0.00)	(0.00)	(0.00)
DF	0.37 ***	0.07 ***	−0.02 ***	−0.63 ***
	(0.00)	(0.00)	(0.00)	(0.00)
AIC	8,574,896.31	393,983.95	1,181,136.66	45,225,415.52
BIC	8,574,968.33	394,055.97	1,181,208.68	45,225,487.54
Log Likelihood	−4,287,440.16	−196,983.97	−590,560.33	−22,612,699.76
Num. obs.	60,000	60,000	60,000	60,000
Num. groups: CCP	5	5	5	5
Num. groups: Fear	5	5	5	5
Num. groups: Desire	5	5	5	5
Var: CCP (Intercept)	0.01	0.08	0.02	0.02
Var: Fear (Intercept)	0.06	0.10	0.10	0.25
Var: Desire (Intercept)	0.00	0.24	0.01	0.01

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 5. Fixed-effect coefficients with standard errors for the Generalized Linear Mixed-Effects Model with Gamma error function.

	Gamma		
	CCD	**CWD**	**IWD**
(Intercept)	5.16 ***	4.17 ***	7.51 ***
	(0.18)	(0.14)	(0.17)
KBT	0.39 ***	−0.18 ***	−0.55 ***
	(0.01)	(0.00)	(0.00)

Table 5. *Cont.*

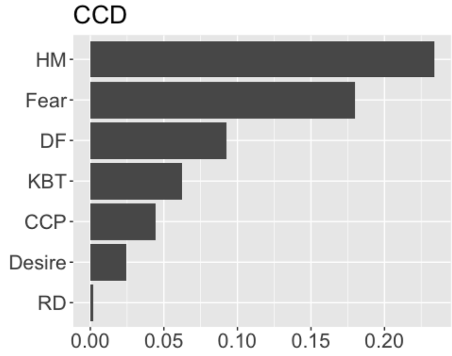
	Gamma		
	CCD	**CWD**	**IWD**
HM	1.25 ***	0.35 ***	0.71 ***
	(0.01)	(0.00)	(0.00)
RD	−0.01	−0.32 ***	0.02 ***
	(0.01)	(0.00)	(0.00)
DF	0.72 ***	−0.01 **	−0.53 ***
	(0.01)	(0.00)	(0.00)
AIC	872,177.97	552,567.86	967,423.69
BIC	872,258.99	552,648.87	967,504.71
Log Likelihood	−436,079.99	−276,274.93	−483,702.85
Num. obs.	60,000	60,000	60,000
Num. groups: CCP	5	5	5
Num. groups: Fear	5	5	5
Num. groups: Desire	5	5	5
Var: CCP (Intercept)	0.00	0.00	0.01
Var: Fear (Intercept)	0.07	0.02	0.05
Var: Desire (Intercept)	0.00	0.00	0.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

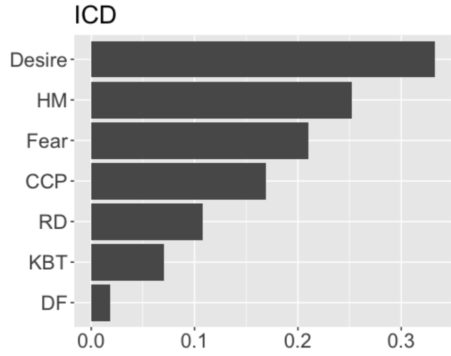
Table 6. Fixed-effect coefficients with standard errors for the Generalized Linear Mixed-Effects Model with Inverse Gaussian error function.

	Inverse Gaussian		
	CCD	**CWD**	**IWD**
(Intercept)	5.39 ***	4.33 ***	7.63 ***
	(0.97)	(0.24)	(0.26)
KBT	1.29 ***	−0.13 ***	−0.41 ***
	(0.01)	(0.00)	(0.00)
HM	3.42 ***	0.21 ***	0.51 ***
	(0.03)	(0.00)	(0.00)
RD	0.36 ***	−0.37 ***	0.02 ***
	(0.01)	(0.00)	(0.00)
DF	2.18 ***	−0.00	−0.48 ***
	(0.02)	(0.00)	(0.00)
AIC	958,219.28	545,801.35	945,278.79
BIC	958,300.30	545,882.37	945,359.81
Log Likelihood	−479,100.64	−272,891.68	−472,630.40
Num. obs.	60,000	60,000	60,000
Num. groups: CCP	5	5	5
Num. groups: Fear	5	5	5
Num. groups: Desire	5	5	5
Var: CCP (Intercept)	0.01	0.00	0.00
Var: Fear (Intercept)	0.04	0.00	0.00
Var: Desire (Intercept)	0.00	0.00	0.00

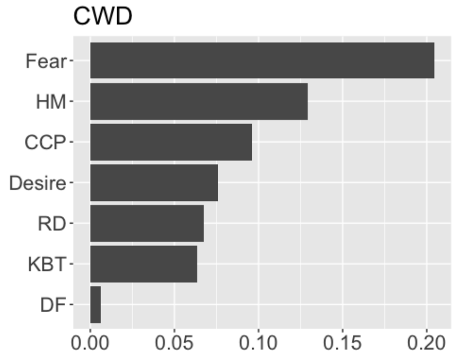
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.



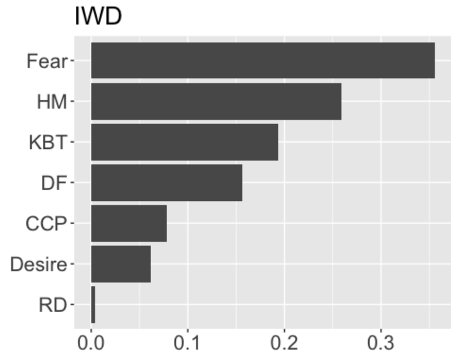
(**a**) CCD



(**b**) ICD



(**c**) CWD



(**d**) IWD

Figure 5. Variable importance measures of predictor variables in the Generalized Mixed-Effects Model with Poisson error function, based on standard deviations, for response variables CCD, ICD, CWD, and IWD, respectively.

Table 7. The estimated random-effect coefficients in Generalized Mixed-Effects Models.

	Poisson				**Gamma**			**Inverse Gaussian**		
	CCD	**ICD**	**CWD**	**IWD**	**CCD**	**CWD**	**IWD**	**CCD**	**CWD**	**IWD**
CCP										
0.1	−0.13	0.38	0.27	0.23	−0.07	0.18	0.26	0.70	0.11	0.19
0.3	0.01	0.20	−0.08	−0.01	−0.00	−0.08	−0.03	−0.08	−0.08	−0.08
0.5	0.03	−0.07	−0.12	−0.06	0.01	−0.09	−0.06	−0.51	−0.09	−0.10
0.7	0.04	−0.22	−0.08	−0.08	0.02	−0.05	−0.08	−0.70	−0.05	−0.11
0.9	0.05	−0.29	0.01	−0.08	0.05	0.03	−0.09	−0.76	0.02	−0.12
Fear										
0	0.35	0.40	−0.42	−0.91	0.67	−0.36	−0.60	1.30	−0.42	−0.54
0.25	0.11	0.18	−0.25	−0.11	0.12	−0.21	−0.13	−1.03	−0.27	−0.22
0.5	0.02	0.01	−0.03	0.21	−0.02	−0.02	0.08	−1.34	−0.08	−0.06
0.75	−0.11	−0.16	0.31	0.29	−0.16	0.26	0.21	−1.57	0.19	0.07
1	−0.37	−0.44	0.39	0.52	−0.61	0.33	0.44	−2.68	0.27	0.30
Desire										
0	−0.07	−0.88	0.18	0.16	−0.08	0.17	0.10	−0.11	0.15	0.05
0.25	−0.01	−0.06	0.06	0.03	−0.02	0.05	0.01	−0.03	0.02	−0.01
0.5	0.02	0.16	−0.03	−0.03	0.02	−0.03	−0.02	0.03	−0.06	−0.03
0.75	0.03	0.33	−0.09	−0.07	0.04	−0.08	−0.04	0.03	−0.11	−0.04
1	0.03	0.45	−0.12	−0.09	0.05	−0.11	−0.05	0.01	−0.13	−0.04

Table 8. Variable importance measures for Generalized Mixed-Effects Models.

	Poisson				Gamma			Inverse Gaussian		
	CCD	**ICD**	**CWD**	**IWD**	**CCD**	**CWD**	**IWD**	**CCD**	**CWD**	**IWD**
KBT	0.06	0.07	0.06	0.19	0.10	0.05	0.14	0.32	0.03	0.10
HM	0.23	0.25	0.13	0.26	0.31	0.09	0.18	0.86	0.05	0.13
RD	0.00	0.11	0.07	0.00	0.00	0.08	0.00	0.09	0.09	0.00
DF	0.09	0.02	0.01	0.16	0.18	0.00	0.13	0.54	0.00	0.12
CCP	0.04	0.17	0.10	0.08	0.03	0.07	0.09	0.36	0.05	0.08
Fear	0.18	0.21	0.20	0.36	0.32	0.17	0.26	1.00	0.17	0.21
Desire	0.02	0.33	0.08	0.06	0.03	0.07	0.04	0.03	0.07	0.02

5. Comparison with Artificial Neural Network Model Approach

In our previous work [49], we employed a predictive modelling approach using Artificial Neural Networks (ANNs). This section presents a comparative analysis to evaluate the variable importance derived from both GLMMs and ANNs. While both methods provide valuable insights into the influence of different predictors on the behaviour of cognitive agents, they differ significantly in terms of interpretability, sensitivity, and computational complexity. A key distinction between the two studies lies in the input simulation data: in [49], a single decision formula was used, namely, the Crossing-and-Waiting-Based Decision Formula (cwDF), whereas in this study, we analyze data from the full simulation model, i.e., data obtained from both decision formulas, cDF and cwDF.

The GLMMs results reveal that the predictors HM and Fear consistently rank as the most influential factors across all four decision types, i.e., CCD, ICD, CWD, and IWD. These rankings remain relatively consistent across the considered error functions, i.e., Gaussian, Poisson, Gamma, and Inverse Gaussian, although notable shifts are observed in the relative importance of other predictors, such as CCP, Desire, and KBT, depending on the error structure used. For instance, in the Gamma and Inverse Gaussian GLMMs, Fear takes precedence over Desire, indicating that the choice of error function can significantly affect the ranking of variables. This flexibility underscores one of the key strengths of GLMMs: their ability to model a variety of response distributions, facilitating a better understanding of predictor effects under different statistical assumptions. This adaptability has been particularly valuable when modelling complex decision-making processes, where predictor impacts may vary across contexts or response types.

In contrast, the ANN models consistently identified CCP and KBT as the most important factors, with KBT showing different influences on crossing and waiting decisions depending on the number of hidden layers in the network. For instance, in the multi-layer ANN, KBT exerts a positive effect on crossing decisions but negatively impacts waiting decisions. However, in the single-hidden-layer ANN model, the effect of KBT is less clearly visible, highlighting a limitation of ANN in terms of interpretability. The use of the Olden function to calculate variable importance in the ANN framework provides a measure of influence but does not offer insight into the direction of the effects. This contrasts with the GLMM approach, where both the magnitude and the direction of the predictor effects are evident, offering a more comprehensive understanding of how variables influence cognitive agents' behaviour.

Another point of divergence between the two methods lies in the assessment of variable sensitivity. In GLMMs, sensitivity is closely tied to the variability of the predictors across different response types. For instance, while HM and Fear are highly sensitive across all decision types, factors like CCP and Desire exhibit more decision-specific effects. This detailed sensitivity analysis provides deeper insights into the functional roles of different predictors, particularly in scenarios where small changes in predictor values can significantly affect outcomes. ANN models, on the other hand, offer less clarity in this regard. While ANNs successfully identify CCP and KBT as dominant predictors, they offer limited insight into the sensitivity of variables and how small fluctuations in input values might influence the outcomes.

6. Conclusions

We applied Generalized Linear Mixed-Effects Models (GLMMs) to explain the significant functional connections between designed factors and response variables, specifically quantifying variable importance in a complex simulation model of cognitive agents learning to cross a CA-based highway. Our analysis demonstrated the practicality and effectiveness of these statistical models and methodologies for analyzing data from complex simulation models. The findings of this study provide deeper insights into the connections between design factors and their corresponding responses, enhancing our understanding of the underlying dynamics and contributing to the fields of swarm robotics, applied mathematics, simulation modelling, and computational analysis.

Implementing GLMMs to analyze simulation data of complex systems represents a critical milestone, integrating fixed and random effects into our analytical framework. While the results regarding variable importance were generally consistent across methodologies, subtle variations in the rankings of influential factors emerged. Although the definitive hierarchy of these factors remains ambiguous, our focus shifted toward uncovering underlying trends and conducting nuanced analyses to unravel the patterns that govern the learning dynamics of cognitive agents. This comprehensive approach significantly enriches our understanding of how multi-level design factors contribute to overall learning outcomes within simulated complex environments. Our study introduces an adaptable analytical framework applicable to diverse investigations involving complex systems and computer simulations while offering critical insights to enhance model interpretability, particularly in additive or mixed-effects models. This approach departs from the conventional practice of solely assessing statistical significance, addressing potential biases and fostering a more holistic and robust analysis of simulation data.

The development of a simulation framework integrating the study of Desire (propensity for risk taking) and Fear (propensity for risk avoidance) factors in decision-making reveals potential applications extending beyond technical optimization, such as in robot swarms (by enhancing decision-making strategies in multi-agent systems), to domains of risk analysis and management. By comprehensively examining how risk taking (Desire) and risk avoidance (Fear), across various levels, influence outcomes in our system, we gain valuable insights that align with not only robot swarms but also research domains where psychological drivers play a pivotal role, such as financial markets. The statistical methodology presented in this paper, applied to investigating psychological drivers in decision-making, has broad implications for understanding risk-related behaviours and decision-making processes. It may provide businesses with tools to better anticipate and manage risks through proactive, data-driven strategies.

In the future, our research trajectory will focus on refining interpretability within more complex machine learning models, particularly those grounded in decision tree-based methodologies. By leveraging the capabilities of these advanced statistical techniques, we aim to deepen our understanding of the dynamics governing decision-making processes within complex systems. This work marks a significant advancement in assessing variable importance in simulation-based experiments of complex systems, including decision-making scenarios. It offers fresh insights into analyzing simulation data for engineering, natural, and social systems. The framework, particularly through its emphasis on risk taking (Desire) and risk avoidance (Fear), not only enriches research on robot swarms but also inspires businesses to explore innovative methods of studying consumer behaviour. By understanding how psychological drivers influence decisions, businesses can better identify consumer patterns, anticipate market trends, and make strategic decisions in areas such as product development and customer engagement. The broad applicability of this approach across diverse domains of computer simulation in engineering and social sciences underscores the critical role of advanced statistical techniques in modelling and deciphering the complexities inherent in complex system dynamics.

Author Contributions: Conceptualization, S.X.; Methodology, S.X. and C.G.; Software, C.G. and A.T.L.; Validation, S.X. and A.T.L.; Formal analysis, S.X. and C.G.; Writing—original draft, S.X.; Writing—review & editing, S.X. and A.T.L.; Visualization, C.G.; Supervision, A.T.L.; Project administration, A.T.L.; Funding acquisition, A.T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Natural Sciences and Engineering Research Council of Canada (NSERC).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: A.T. Lawniczak acknowledges financial support from The Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflicts of Interest: The authors declare no conflicts of interest.

List of Abbreviations

ANNs	Artificial Neural Networks
XAI	Explainable AI
LMs	Linear Models
LMMs	Linear Mixed-Effects Models
GLMs	Generalized Linear Models
GLMMs	Generalized Linear Mixed-Effects Models
CA	Cellular Automaton
CCP	Car Creation Probability
RD	Random Deceleration
HM	Horizontal Movement
KBT	Knowledge Base Transfer
CCD	Correct Crossing Decision
ICD	Incorrect Crossing Decision
CWD	Correct Waiting Decision
IWD	Incorrect Waiting Decision
cDF	Crossing-based Decision Formula
cwDF	Crossing-and-Waiting-Based Decision Formula
DF	Decision Formula
AV	Autonomous Vehicle

References

1. Bar-Yam, Y. *Dynamics of Complex Systems*; Reading, Addison-Wesley: Boston, MA, USA, 1997.
2. Flake, G.W. *The Computational Beauty of Nature*; A Bradford Book; MIT Press: Cambridge, MA, USA, 1998.
3. Lempert, R.J. A new decision sciences for complex systems. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7309–7313. [CrossRef]
4. Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*; The MIT Press: Cambridge, MA, USA, 2007.
5. Boccara, N. *Modeling Complex Systems* Springer: New York, NY, USA, 2010.
6. Hoekstra, A.G.; Kroc, J.; Sloot, P.M. *Simulating Complex Systems by Cellular Automata*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
7. Schadschneider, A.; Chowdhury, D.; Nishinari, K. *Stochastic Transport in Complex Systems: From Molecules to Vehicles*; Elsevier: Amsterdam, The Netherlands, 2010.
8. Treiber, M.; Kesting, A. *Traffic Flow Dynamics: Data, Models and Simulation*; Springer: Berlin/Heidelberg, Germany, 2012.
9. Heath, B.; Hill, R. Some insights into the emergence of agent-based modeling. In *Agent-Based Modeling and Simulation*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 32–44.
10. Wilensky, U.; Rand, W. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*; MIT Press: Cambridge, MA, USA, 2015.
11. Sayama, H. *Introduction to the Modeling and Analysis of Complex Systems*; Open SUNY Textbooks: Geneseo, NY, USA, 2015.
12. Magnani, L.; Bertolotti, T. *Springer Handbook of Model-Based Science*; Springer: Berlin/Heidelberg, Germany, 2017.
13. Vlachas, P.R.; Arampatzis, G.; Uhler, C.; Koumoutsakos, P. Multiscale simulations of complex systems by learning their effective dynamics. *Nat. Mach. Intell.* **2022**, *4*, 359–366. [CrossRef]

14. Karaca, Y.; Baleanu, D.; Zhang, Y.D.; Gervasi, O.; Moonis, M. *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems*; Academic Press: Cambridge, MA, USA, 2022.
15. Campilho, R.D.; Silva, F.J. Industrial Process Improvement by Automation and Robotics. *Machines* **2023**, *11*, 1011. [CrossRef]
16. Soori, M.; Arezoo, B.; Dastres, R. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cogn. Robot.* **2023**, *3*, 54–70. [CrossRef]
17. Mikolajczyk, T. Manufacturing using robot. *Adv. Mater. Res.* **2012**, *463*, 1643–1646. [CrossRef]
18. Goel, R.; Gupta, P. Robotics and industry 4.0. In *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 157–169.
19. Arinez, J.F.; Chang, Q.; Gao, R.X.; Xu, C.; Zhang, J. Artificial intelligence in advanced manufacturing: Current status and future outlook. *J. Manuf. Sci. Eng.* **2020**, *142*, 110804. [CrossRef]
20. Huang, Z.; Shen, Y.; Li, J.; Fey, M.; Brecher, C. A survey on AI-driven digital twins in industry 4.0: Smart manufacturing and advanced robotics. *Sensors* **2021**, *21*, 6340. [CrossRef] [PubMed]
21. Pillai, R.; Sivathanu, B.; Mariani, M.; Rana, N.P.; Yang, B.; Dwivedi, Y.K. Adoption of AI-empowered industrial robots in auto component manufacturing companies. *Prod. Plan. Control* **2022**, *33*, 1517–1533. [CrossRef]
22. Chryssolouris, G.; Alexopoulos, K.; Arkouli, Z. Artificial intelligence in manufacturing equipment, automation, and robots. In *A Perspective on Artificial Intelligence in Manufacturing*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 41–78.
23. Holland, J.; Kingston, L.; McCarthy, C.; Armstrong, E.; O'Dwyer, P.; Merz, F.; McConnell, M. Service robots in the healthcare sector. *Robotics* **2021**, *10*, 47. [CrossRef]
24. Ribeiro, J.; Lima, R.; Eckhardt, T.; Paiva, S. Robotic process automation and artificial intelligence in industry 4.0—A literature review. *Procedia Comput. Sci.* **2021**, *181*, 51–58. [CrossRef]
25. Marshall, J.A.; Bonchis, A.; Nebot, E.; Scheding, S. Robotics in mining. In *Springer Handbook of Robotics*; Springer: Cham, Switzerland, 2016; pp. 1549–1576.
26. Corrigan, C.C.; Ikonnikova, S.A. A review of the use of AI in the mining industry: Insights and ethical considerations for multi-objective optimization. *Extr. Ind. Soc.* **2024**, *17*, 101440. [CrossRef]
27. Hussain, A.; Malik, A.; Halim, M.U.; Ali, A.M. The use of robotics in surgery: A review. *Int. J. Clin. Pract.* **2014**, *68*, 1376–1382. [CrossRef] [PubMed]
28. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc. J.* **2019**, *6*, 94–98. [CrossRef] [PubMed]
29. Bajwa, J.; Munir, U.; Nori, A.; Williams, B. Artificial intelligence in healthcare: Transforming the practice of medicine. *Future Healthc. J.* **2021**, *8*, e188–e194. [CrossRef]
30. Secinaro, S.; Calandra, D.; Secinaro, A.; Muthurangu, V.; Biancone, P. The role of artificial intelligence in healthcare: A structured literature review. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 1–23. [CrossRef]
31. Seeni, A.; Schäfer, B.; Hirzinger, G. Robot mobility systems for planetary surface exploration–state-of-the-art and future outlook: A literature survey. *Aerosp. Technol. Adv.* **2010**, *492*, 189–208.
32. Zaoui, A.; Tchuente, D.; Wamba, S.F.; Kamsu-Foguem, B. Impact of artificial intelligence on aeronautics: An industry-wide review. *J. Eng. Technol. Manag.* **2024**, *71*, 101800. [CrossRef]
33. Hehenberger, P.; Vogel-Heuser, B.; Bradley, D.; Eynard, B.; Tomiyama, T.; Achiche, S. Design, modelling, simulation and integration of cyber physical systems: Methods and applications. *Comput. Ind.* **2016**, *82*, 273–289. [CrossRef]
34. Bonabeau, E.; Dorigo, M.; Theraulaz, G. *Swarm Intelligence: From Natural to Artificial Systems*; Number 1; Oxford University Press: Oxford, UK, 1999.
35. Navarro, I.; Matía, F. An introduction to swarm robotics. *ISRN Robot.* **2013**, *2013*, 608164. [CrossRef]
36. Khan, M.M.; Kasmarik, K.; Barlow, M. Autonomous detection of collective behaviours in swarms. *Swarm Evol. Comput.* **2020**, *57*, 100715. [CrossRef]
37. Bredeche, N.; Fontbonne, N. Social learning in swarm robotics. *Philos. Trans. R. Soc. B* **2022**, *377*, 20200309. [CrossRef] [PubMed]
38. Duan, H.; Huo, M.; Fan, Y. From animal collective behaviors to swarm robotic cooperation. *Natl. Sci. Rev.* **2023**, *10*, nwad040. [CrossRef] [PubMed]
39. Blais, M.A.; Akhloufi, M.A. Reinforcement learning for swarm robotics: An overview of applications, algorithms and simulators. *Cogn. Robot.* **2023**, *3*, 226–256. [CrossRef]
40. Lawniczak, A.T.; Di Stefano, B.N.; Ernst, J.B. Software implementation of population of cognitive agents learning to cross a highway. In *Cellular Automata: 11th International Conference on Cellular Automata for Research and Industry, ACRI 2014, Krakow, Poland, 22–25 September 2014, Proceedings 11*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 688–697.
41. Lawniczak, A.T.; Ernst, J.B.; Di Stefano, B.N. Biomimicry Based Learning Outcomes of Simple Cognitive Agents. *Procedia Comput. Sci.* **2014**, *36*, 508–514. [CrossRef]
42. Alonso, E.; d'Inverno, M.; Kudenko, D.; Luck, M.; Noble, J. Learning in multi-agent systems. *Knowl. Eng. Rev.* **2001**, *16*, 277–284. [CrossRef]
43. Lawniczak, A.T.; Ly, L.; Yu, F. Success rate of creatures crossing a highway as a function of model parameters. *Procedia Comput. Sci.* **2016**, *80*, 542–553. [CrossRef]
44. Lawniczak, A.T.; Ly, L.; Yu, F. Effects of Simulation Parameters on Naïve Creatures Learning to Safely Cross a Highway on Bimodal Threshold Nature of Success. *Procedia Comput. Sci.* **2016**, *80*, 2382–2386. [CrossRef]

45. Lawniczak, A.T.; Yu, F. Comparison of Agents' Performance in Learning to Cross a Highway for Two Decisions Formulas. In Proceedings of the International Conference on Agents and Artificial Intelligence, Porto, Portugal, 24–26 February 2017; Scitepress: Setúbal, Portugal, 2017; Volume 2, pp. 208–219.
46. Lawniczak, A.T.; Yu, F. Decisions and success of heterogeneous population of agents in learning to cross a highway. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–10.
47. Lawniczak, A.T.; Yu, F. Analysis of rates of agents' decisions in learning to cross a highway in populations with risk takers and risk avoiders. In *Cellular Automata, Proceedings of the 13th International Conference on Cellular Automata for Research and Industry, ACRI 2018, Como, Italy, 17–21 September 2018, Proceedings 13*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 177–184.
48. Xie, S.; Lawniczak, A.T.; Gan, C. Modeling and Analysis of Autonomous Agents' Decisions in Learning to Cross a Cellular Automaton-Based Highway. *Computation* **2019**, *7*, 53. [CrossRef]
49. Xie, S.; Lawniczak, A.T.; Hao, J. Modelling Autonomous Agents' Decisions in Learning to Cross a Cellular Automaton-Based Highway via Artificial Neural Networks. *Computation* **2020**, *8*, 64. [CrossRef]
50. Xie, S.; Lawniczak, A.T.; Gan, C. Optimal number of clusters in explainable data analysis of agent-based simulation experiments. *J. Comput. Sci.* **2022**, *62*, 101685. [CrossRef]
51. Qi, D.; Majda, A.J. Using machine learning to predict extreme events in complex systems. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 52–59. [CrossRef] [PubMed]
52. Qin, J.; Hu, F.; Liu, Y.; Witherell, P.; Wang, C.C.; Rosen, D.W.; Simpson, T.W.; Lu, Y.; Tang, Q. Research and application of machine learning for additive manufacturing. *Addit. Manuf.* **2022**, *52*, 102691. [CrossRef]
53. Xue, T.; Li, X.; Chen, X.; Chen, L.; Han, Z. Machine learning phases in swarming systems. *Mach. Learn. Sci. Technol.* **2023**, *4*, 015028. [CrossRef]
54. Marwan, N.; Romano, M.C.; Thiel, M.; Kurths, J. Recurrence plots for the analysis of complex systems. *Phys. Rep.* **2007**, *438*, 237–329. [CrossRef]
55. Seely, A.J.; Macklem, P.T. Complex systems and the technology of variability analysis. *Crit. Care* **2004**, *8*, 1–18. [CrossRef]
56. Salnikov, V.; Cassese, D.; Lambiotte, R. Simplicial complexes and complex systems. *Eur. J. Phys.* **2018**, *40*, 014001. [CrossRef]
57. Goh, A.T. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* **1995**, *9*, 143–151. [CrossRef]
58. Almeida, J.S. Predictive non-linear modeling of complex data by artificial neural networks. *Curr. Opin. Biotechnol.* **2002**, *13*, 72–76. [CrossRef]
59. Alanis, A.Y.; Arana-Daniel, N.; Lopez-Franco, C. *Artificial Neural Networks for Engineering Applications*; Academic Press: Cambridge, MA, USA, 2019.
60. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining explanations: An overview of interpretability of machine learning. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 80–89.
61. Bibal, A.; Frénay, B. Interpretability of machine learning models and representations: An introduction. In Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 27–29 April 2016; CIACO: San Jose, CA, USA, 2016; pp. 77–82.
62. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine learning interpretability: A survey on methods and metrics. *Electronics* **2019**, *8*, 832. [CrossRef]
63. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **2020**, *23*, 18. [CrossRef] [PubMed]
64. Lawniczak, A.T.; Yu, F. Cognitive Agents Success in Learning to Cross a CA Based Highway Comparison for Two Decision Formulas. *Procedia Comput. Sci.* **2017**, *108*, 2443–2447. [CrossRef]
65. Hernández, Q.; Badías, A.; Chinesta, F.; Cueto, E. Port-metriplectic neural networks: Thermodynamics-informed machine learning of complex physical systems. *Comput. Mech.* **2023**, *72*, 553–561. [CrossRef]
66. Mari, C.; Mari, E. Occam's razor, machine learning and stochastic modeling of complex systems: The case of the Italian energy market. *Qual. Quant.* **2023**, *58*, 1093–1111. [CrossRef]
67. Raman, R.; Gupta, N.; Jeppu, Y. Framework for Formal Verification of Machine Learning Based Complex System-of-Systems. *Insight* **2023**, *26*, 91–102. [CrossRef]
68. Korecki, M. Deep reinforcement meta-learning and self-organization in complex systems: Applications to traffic signal control. *Entropy* **2023**, *25*, 982. [CrossRef] [PubMed]
69. Zhang, Z.; Wu, C.; Qu, S.; Chen, X. An explainable artificial intelligence approach for financial distress prediction. *Inf. Process. Manag.* **2022**, *59*, 102988. [CrossRef]
70. Hulsen, T. Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare. *AI* **2023**, *4*, 652–666. [CrossRef]
71. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [CrossRef]
72. Borrego-Díaz, J.; Galán Páez, J. Knowledge representation for explainable artificial intelligence: Modeling foundations from complex systems. *Complex Intell. Syst.* **2022**, *8*, 1579–1601. [CrossRef]

73. Kikutsuji, T.; Mori, Y.; Okazaki, K.i.; Mori, T.; Kim, K.; Matubayasi, N. Explaining reaction coordinates of alanine dipeptide isomerization obtained from deep neural networks using Explainable Artificial Intelligence (XAI). *J. Chem. Phys.* **2022**, *156*. [CrossRef]
74. Homafar, A.; Nasiri, H.; Chelgani, S.C. Modeling coking coal indexes by SHAP-XGBoost: Explainable artificial intelligence method. *Fuel Commun.* **2022**, *13*, 100078. [CrossRef]
75. Yin, J.; Ren, X.; Liu, R.; Tang, T.; Su, S. Quantitative analysis for resilience-based urban rail systems: A hybrid knowledge-based and data-driven approach. *Reliab. Eng. Syst. Saf.* **2022**, *219*, 108183. [CrossRef]
76. Li, Y.F.; Huang, H.Z.; Mi, J.; Peng, W.; Han, X. Reliability analysis of multi-state systems with common cause failures based on Bayesian network and fuzzy probability. *Ann. Oper. Res.* **2022**, *311*, 195–209. [CrossRef]
77. Roca-Flores, E.; Naumis, G.G.; Madrigal-Solís, E.; Fraedrich, K. Typhoon complexity: Northwest Pacific tropical cyclone season complex systems analysis. *Int. J. Mod. Phys. C* **2023**, *34*, 2350045. [CrossRef]
78. Duan, S.; Shi, Z. A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. *Lang. Teach. Res.* **2024**, *28*, 497–530. [CrossRef]
79. Hunt, E.R.; Baddeley, R.J.; Worley, A.; Sendova-Franks, A.B.; Franks, N.R. Ants determine their next move at rest: Motor planning and causality in complex systems. *R. Soc. Open Sci.* **2016**, *3*, 150534. [CrossRef]
80. Kim, J.M.; Li, C.; Ha, I.D. Generalised linear mixed logit and probit models applied to US Army and Navy data. *Int. J. Product. Qual. Manag.* **2020**, *30*, 126–142. [CrossRef]
81. Bolker, B.M.; Brooks, M.E.; Clark, C.J.; Geange, S.W.; Poulsen, J.R.; Stevens, M.H.H.; White, J.S.S. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **2009**, *24*, 127–135. [CrossRef] [PubMed]
82. Johnson, P.C.; Barry, S.J.; Ferguson, H.M.; Müller, P. Power analysis for generalized linear mixed models in ecology and evolution. *Methods Ecol. Evol.* **2015**, *6*, 133–142. [CrossRef] [PubMed]
83. Nehaniv, C.L.; Dautenhahn, K. *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*; Cambridge University Press: Cambridge, UK, 2009.
84. Nagel, K.; Schreckenberg, M. A cellular automaton model for freeway traffic. *J. Phys. I* **1992**, *2*, 2221–2229. [CrossRef]
85. Faraway, J.J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*; CRC Press: Boca Raton, FL, USA, 2016.
86. Hedeker, D. Generalized linear mixed models. In *Encyclopedia of Statistics in Behavioral Science*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
87. Goldstein, H.; Rasbash, J. Improved approximations for multilevel models with binary responses. *J. R. Stat. Soc. Ser. A Stat. Soc.* **1996**, *159*, 505–513. [CrossRef]