Complex Intell. Syst. (2018) 4:155–169
https://doi.org/10.1007/s40747-017-0062-8

ORIGINAL ARTICLE

Deter and protect: crime modeling with multi-agent learning

Trevor R. Caskey[1] · James S. Wasek[1] · Anna Y. Franz[1]

Received: 13 April 2017 / Accepted: 6 October 2017 / Published online: 25 October 2017

Abstract This paper presents a formal game-theoretic belief learning approach to model criminology's routine activity theory (RAT). RAT states that for a crime to occur a motivated offender (criminal) and a desirable target (victim) must meet in space and time without the presence of capable guardianship (law enforcement). The novelty in using belief learning to model the dynamics of RAT's offender, target, and guardian behaviors within an agent-based model is that the agents learn and adapt given observation of other agents' actions without knowledge of the payoffs that drove the other agents' choices. This is in contrast to other crime modeling research that has used reinforcement learning where the accumulated rewards gained from prior experiences are used to guide agent learning. This is an important distinction given the dynamics of RAT. It is the presence of the various agent types that provide opportunity for crime to occur, and not the potential for reward. Additionally, the belief learning approach presented fits the observed empirical data of case studies, producing statistically significant results with lower variance when compared to a reinforcement learning approach. Application of this new approach supports law enforcement in developing responses to crime problems and planning for the effects of displacement due to directed responses, thus deterring offenders and protecting the public through crime modeling with multi-agent learning.

Keywords Crime modeling · Agent-based modeling · Belief learning · game theory

Introduction

Problem-oriented policing (POP) is a policing approach initially proposed in the 1970s that focuses on "problem-solving" as a systematic way to understand crime and disorder [1]. Typically, POP follows the SARA model, where law enforcement (S)cans their jurisdiction for a problem, (A)nalyzes the issue, develops and deploys a (R)esponse, and then (A)ssesses the effectiveness of the response [2]. Knowledge-based approaches to identifying problems, such as CompStat (short for Computer Statistics) which analyzes crime data at an aggregate level, put emphasis on doing something quickly when crime spikes [3]. Responses are commonly implemented with a geographic focus; in practice, this is known as place-based or "hot spot" policing. Law enforcement dedicates resources to the problem area to combat the crime issue. One major concern of using this technique is that crime will merely displace to nearby areas, with the benefit to the response area coming at the expense of surrounding areas, creating a 'whack-a-mole' effect.

Research by Telep et al. [4] and Weisburd and Telep [5], which reviewed "hot spot" policing and crime displacement studies, report that displacement occurs less often than it is believed to occur. In some cases, they found a diffusion of benefit to surrounding areas. When there is evidence of displacement, the amount of crime displaced does not outweigh the benefit of crime reduced in the response area. Due to these findings, they consider "hot spot" policing to be an effective strategy. Andresen and Malleson [6] claim these prior displacement studies had measurement issues making identification of displacement difficult. The size of the targeted response areas compared to the displacement catchment areas was disproportionate and new units of analysis for identifying displacement have been recommended. Sorg et al. [7] have also noted that while there may be positive effects from

✉ Trevor R. Caskey
tcaskey@gwmail.gwu.edu

[1] The George Washington University, Washington, DC 20052, USA

"hot spot" initiatives, the effects may not last. Crime rates in their study returned to the response area within a few months after the initiative ended, depicting an inverse displacement.

This paper presents a new approach to explore this displacement problem using agent-based modeling (ABM) and game-theoretic belief learning to simulate offender (criminal), target (victim), and guardian (law enforcement) behaviors. The ABM paradigm is a generative approach to system modeling, and is an accepted approach to modeling systems with emergent behavior driven by the interactions of agents [8]. The ability to model autonomous agents in an environment and then observe the emergent macro level behavior of the system is a natural fit to the interest in the displacement problem. The use of ABM as an acceptable approach to study criminological systems has been reviewed by Birks et al. [9], Groff et al. [10] and Gerritsen [11]. The originality in using belief learning for agent logic is that the agents learn by observing opposing agents' actions without regard for the payoffs that drove the other agents' choices. This differs from reinforcement learning approaches that are guided by accumulated rewards. The beliefs learned depict where opposing agent types will likely occupy and not were positive reward signals have been experienced, providing a direct mechanism to model offender, target, and guardian dynamics.

The remainder of this paper is organized as follows: first, the underlying criminological theories are introduced along with the modeling approaches used in this work. Next, a literature review is discussed, covering areas of research that complement this work, which fills gaps in the literature. Finally, the methodology is setup and a case study is conducted, comparing the belief learning approach in this paper to a previously researched Q-learning implementation of reinforcement learning.

Background

Theoretical basis

Routine activity theory (RAT) developed by Cohen and Felson [12] is a micro-level theory that describes the most basic units needed for a crime to occur. The theory states that criminal acts require convergence in space and time of likely offenders, suitable targets and the absence of guardianship against crime [13]. RAT asserts that crime is opportunistic and easily reflected by analogy with a crime triangle in Fig. 1. Crime has the potential to occur when offenders and targets/victims meet in space and time without guardianship. Guardianship can be through the formal presence of a guardian (law enforcement) or informally through the collective presence of bystanders. The collective presence has an effect akin to a criminal not wanting to be seen committing a crime in a crowd or fear of being stopped by a bystander.

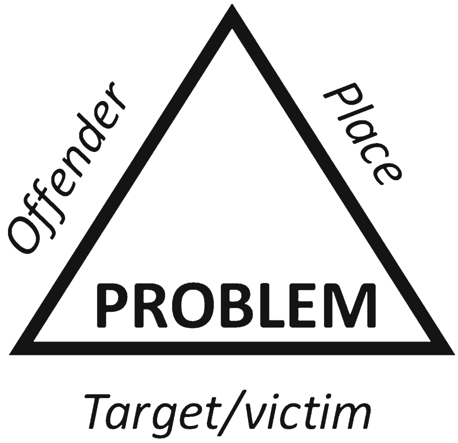


Fig. 1 Crime Triangle (from POP Center [14])

As noted by Groff [15], RAT provides a groundwork for interaction, but does not provide a framework for decision making. Cornish and Clarke's [16] rational choice theory (RCT) contends that offenders are rational, similar to non-offenders; however, they have a propensity to commit crime that sets them apart. RCT provides a perspective by which offenders are rational actors and seek optimal strategies for themselves. RAT provides a framework for interacting pieces within a criminological system and RCT provides a theory for rational offender decision-making.

Agent-based modeling

The ABM paradigm is a computational method to model complex systems through the collective interaction of autonomous entities. These entities, called agents, have characteristics and behaviors that describe how they make decisions and act due to interaction with other agents and their environment. At the most basic level, ABM consists of a set of agents and their relationships [17]. ABM is considered a "bottom-up" approach to system modeling, as it is oriented around the micro-level behaviors and interactions of heterogeneous agents. ABM can be used to explore goal directed behaviors such as completing a task in robotic autonomy [18] or to study emergent macro-level outcomes that arise from micro-level agent interactions such us workforce attrition in STEM organizations [19]. Agent behaviors are typically a set of rules to follow to make a decision given their interactions and environmental observations. Agents individually assess their situations and make their decisions using their rule set. A staple feature of the ABM paradigm is the repetitive interaction of agents. Agents can be developed to evolve and adapt, emulating some sort of learning from their interactions [17]. In general ABM's consist of [20]:

- a set of agents, their attributes and behaviors;

- a set of agent relationships and methods of interaction—an underlying topology of connectedness defines how and with whom agents interact; and
- the agents' environment—agents interact with their environment

While no set definition exists for what an agent is, the only agreed upon characteristic is autonomy, meaning the agent can function independently in the environment and in interactions with other agents [20]. This autonomy generally means agents should also be self-contained (i.e., individually identifiable, heterogeneous, discrete entities with their own attributes), have a state that varies over time, and be social (interact with other agents and/or the environment) [20].

Game theory

Game Theory is a way of thinking about strategic interactions between self-interested players [21]. It is used in many disciplines, like economics, as it is concerned with how self-interested players will behave in these strategic interactions. A game can be thought of as any interaction between two or more players where the outcomes of the interactions depend on what the players choose to do. Self-interested means the players have personal descriptions of the state of the game and they choose their actions based on this description. Each player's view of the state of the game is described through a payoff structure, called a utility function. The player's utility functions capture their attitude toward the actions available [21]. Game theory often assumes that players are rational, and will attempt to maximize their utility. Games can be played simultaneously (i.e., matching pennies and rock/paper/scissors), or sequentially (i.e., chess and poker). Simultaneously played games are considered normal form (or strategic form) games and sequentially played games are considered extensive form games. Games are generally made of three elements: players, alternatives, and payoffs. The general formulation of a normal form game is [22]:

- a finite set of players N $= \{1, 2, \ldots, n\}$, indexed by i;
- a set of alternatives for player i, $a = (a_1, a_2 \ldots, a_n)$ in $A = A_1 \times A_2 \times \cdots \times A_n$,where A is the set of alternatives for each player and a is the strategy profile which is the list of alternatives chosen by each player; and
- the utility for player i, U_i, as a function of alternatives played, which is the payoff for the different players. U_i describes how each player evaluates the different outcomes of the game.

Using their available utility functions, players can then compute a best response and choose an alternative. Interesting dynamics arise when games are played repeatedly providing long-term interaction between players. Players can make assessments on what they believe opposing players will do or what they think their future payoffs can be, then select a best response using this information. This repeated game playing gives rise to learning in games where players can adjust their strategies and utilities update given new observations and outcomes in previous stages of the game.

Agent learning

Agent learning can generally be summarized using the concept of stigmergy (Fig. 2). Stigmergy is a means of interaction between agents and is used to coordinate the effect of current actions to stimulate further actions.

In the case of an agent based model, agent actions (e.g. alternative selections) produce marks (e.g. indicators of presence or rewards received) in a medium (e.g. the environment). These marks, in turn, can be observed by other agents and future actions can be stimulated given these observations. The type of learning agents employ affects how the observations of the marks are used and how information is updated given actions taken. There are two main approaches that often guide agent learning, namely belief learning and reinforcement learning.

In belief learning, players learn and adapt given observation of opponent players' actions without knowledge of the payoffs that drove the other players' actions. Only the observations of opposing players' actions are used to derive a player's beliefs. Beliefs are expressed as the ratio of strategy choice counts for a given strategy to the total experience of strategies selected [23]. The belief that player i has that opposing player (denoted $-i$) will play strategy k at stage t is denoted by $B^k_{-i}(t)$. This is essentially the proportion of time that strategy k has been played by player $-i$ compared to all strategies player $-i$ has played up to stage t. Beliefs are learned and updated every iteration of the game following a weighted fictitious play process [24,25] in Eq. (1), where the belief in the current stage is the weighted combination of the belief from the previous stage and any new information from the current stage.

$$B^k_{-i}(t) = \phi B^k_{-i}(t-1) + (1-\phi) * 1_{\{a_{-i}=k\}}(t) \tag{1}$$

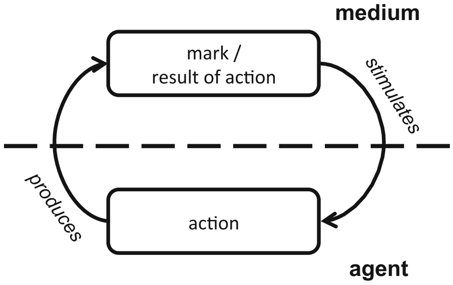


Fig. 2 Stigmergy

$1_{\{a_{-i}=k\}}(t)$ is an indicator function that equals 1 if the alternative chosen by opposing player a_{-i} during stage t is equal to strategy k, and 0 otherwise. $\phi \in [0, 1)$ is a weight parameter that depreciates prior beliefs. With $\phi = 0$, all weight would be placed on the most recent information and as ϕ approaches 1, more weight is placed on prior beliefs (*i.e.*, fictitious play). Player i would then use these beliefs in a utility function to choose an alternative.

In reinforcement learning, players learn by interacting with the environment through feedback, or reinforcement, given actions taken. Players seeks to maximize their rewards and ultimately learn through trial and error over repeated interactions with the environment. Actions that results in positive reinforcement are more likely to be taken again. Similar to training a pet by reinforcing good behavior with a treat, over time the pet learns to repeat the desired behavior. In a multi-agent setting, the reinforcement received by a player may be affected by the actions other players have taken, creating a dynamic reward environment. Action-value functions are one way for a player to find an optimal strategy. Watkins [26] introduced the popular one step reinforcement learning model, Q-Learning, which directly approximates the optimal action-value to be followed by the learning player.

$$\begin{aligned} Q(S_t, A_t) &\leftarrow Q(S_t, A_t) \\ &+ \alpha \left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t) \right] \end{aligned} \tag{2}$$

Here, the action-value function can be thought of as similar to a utility function. $Q(S_t, A_t)$ is the expected reward the player can receive for selecting action A_t when in state S_t at stage t. $\alpha \in [0, 1]$ is the learning rate of how much the previous action-value is adjusted given the reinforcement received. The new reinforcement is the reward, R_{t+1}, plus any value gained from the difference between the maximum available action-value $\max_a Q(S_{t+1}, a)$ from other alternatives and the previous action-value. $\gamma \in [0, 1]$ is the discount factor for the maximum available action-value.

In summary, reinforcement learning involves learning what actions to take by maximizing rewards. Only these accumulated rewards are used in determining the next alternative to select without knowledge of opposing player choices, even though opposing player choices may affect the reward they receive. Belief learning differs from reinforcement learning in that players learn by observing what alternatives opposing players select without regard for the utilities that drove the opposing players' choices. Players do not learn which alternatives return the best reward. In both belief and reinforcement learning, players make subsequent alternative selections using updated beliefs or expected rewards.

There are several methods by which a player may select an alternative as a best response given the information available to them. In game-theoretic terms, it is commonly assumed players are rational and seek to maximize their utility. For both belief and reinforcement learning, this would equate to players using an *argmax* function to select the alternative that returns the maximum utility. Not doing so is deemed an irrational choice in a game-theoretic sense. *argmax* is also known as a greedy or exploitive best response because players are trying to get the maximum immediate payoff. In one-shot games this can be an effective strategy. However, in repeated games, a player following the same strategy at every stage invites opposing players to exploit this predictable behavior. What players can do to mitigate this predictability is to balance their exploitation with exploration. One approach to balancing exploitation and exploration of alternatives is called an ϵ-greedy best response. In this method, ϵ proportion of time, players choose to randomly explore one of the available alternatives without regard for utility. The remaining $(1 - \epsilon)$ proportion of time, players use an *argmax* best response. Another way to avoid predictability, players can play a mixed-strategy, meaning they select an alternative based on a probability distribution. The probability an alternative may be selected is proportional to the utility it provides and the alternative that provides the maximum utility is more likely to be selected. Following a mixed-strategy keeps opponent players guessing and introduces randomized behavior into repeated games.

Related work

A variety of work has been done in using both agent-based modeling for studying criminological systems, as well as game theoretic approaches to crime and security. One collection of researchers explored statistical approaches to modeling crime within an ABM. Specifically, they were interested in repeat and near repeat victimization, which is once a location is a victim of crime, it is more likely to be victimized again. Short et al. [27] modelled offenders (which they called criminals) on a grid environment to produce hot spots. Following statistical formulas, offenders could commit crimes and move around the environment. Jones et al. [28] extended this model by incorporating guardian agents (which they called law enforcement) in the grid environment. The presence of guardian agents was used to deter offenders from committing crime. Chaturapruek et al. [29] adjusted Short et al. [27] by changing how offenders moved around the grid environment. Instead of using local information to their current location, offenders could survey the whole grid and move following a stochastic process called Lévy flights. This allowed the offenders to "jump" around the grid to more attractive spaces beyond their immediate neighbors. Camacho et al. [30] extended Jones, et al. by implementing variations of strategies by the guardian agents, finding that the best strategy was dependent on the size of the hot spots and the number of guardians used in the model to deter crime.

Tambe and a suite of researchers have done research utilizing game theoretic approaches to security [31]. Tambe's research uses algorithms to compute optimal resource allocation strategies to defend targets from attackers. The topics of this research have ranged from investigating crime in metro-rail systems for major cities [32,33], to defending critical infrastructure targets [34], and some have researched to defend wildlife from poaching [35]. The important feature in many of these security games was the use of a quantal response [36] function to derive probabilities in a mixed strategy to selecting alternatives. The quantal response can take on a logit form:

$$\pi_a(t) = \frac{e^{\lambda H_a(t)}}{\sum_{b=1}^{k} e^{\lambda H_b(t)}} \tag{3}$$

$\pi_a(t)$ is the probability of choosing alternative a at stage t, $H_a(t)$ is the player's preference for alternative a at stage t, and λ is a shape parameter that can be used to control the 'noise' in the probabilities. If $\lambda = 0$, the probabilities across the alternatives become uniformly distributed. If λ is large, the player will be more certain about the alternative it prefers more. This quantal response function in Eq. (3) is also commonly referred to as a *softmax* function or Boltzmann distribution method in agent learning literature [37,38] to make a choice following a distribution.

Groff [15] incorporated ABM and RAT to simulate offenders and targets (called civilians), and guardians (called police). Using various factors such as perceived wealth of potential targets and presence of capable guardians, the offenders could decide to commit a crime or not. The model was executed with the agents moving on a street network (represented as a network of nodes and edges) along predefined routes. By varying the amount of time and distance agents traveled from their 'home', tipping points in time and distance were found in which more crime occurred. Groff expresses that improvements to the work could be made by developing logic into guardian agents to have more realistic strategies that were not random, but directed to specific location.

Malleson's doctoral dissertation [39] and various associated papers [40,41] revolve around simulating burglary. Malleson utilized geospatial data about the environment, which included the locations of buildings/homes, perceived value, ease of access, and other attributes. Like Groff, offenders moved along the transportation network and as their attributes depleted (drugs, money, sleep, etc.), the offenders could select a building/home and burglarize. Malleson developed scenarios from an urban planning point of view and showed notional crime patterns. Malleson goes on to point out the lack of guardian agents present in the work and suggests adding them to the model.

Wang [42] describes in detail an agent-based approach to modeling crime. The three main elements of RAT are present: offender agents, target agents, and guardian agents (called police agents). The work used an implementation of reinforcement learning called Q-learning [26] for offender and target agents to learn and navigate a reward/cost grid. The reward/cost grid is essentially a 'mental map' used by the offender and target agents to track the history of past experiences in terms of potential reward for an offender to move to a grid cell and the cost for a target to move to a grid cell where they could potentially by victimized by an offender. However, the work did not use any learning for guardian agents and they solely moved randomly in the simulated environment. Subsequent work by Wang et al. [43] attempted to address the lack of guardian agent learning by having guardians seek the "hottest" crime locations to represent "hot spot" patrolling.

Another set of researchers, Bosse et al. [44–46], also looked at modeling crime using ABM and RAT. Their approach had the main elements of RAT: offender agents (called criminals), target agents (called passersby), and guardian agents. In their model, they used abstract representations for alternatives (i.e., Location 1,2,3). The agents could choose to operate in any of the available locations and these choices were made using the density of opposing agent types in these locations. Given that the locations themselves were the alternatives to choose from, the density of agent types in the locations represented the choices of the agent types in the previous time step. The model did not preserve the history of prior observations and used only the previous time step's information for choosing responses in the next time step.

Motivation

Several areas from the Related Work discussed provide the motivation for this research. First, this paper proposes using belief learning as the learning method used by agents. The novelty of using belief learning is that agents observe opposing agents' actions without regard for the payoffs that drove the opposing agents to make those choices, form beliefs using these observations, and then make future choices using these beliefs. Prior research by Wang [42] and Wang et al. [43] used the reinforcement learning method Q-learning. In their approach, rewards and costs are accumulated in the environment where offenders seek optimal reward and targets seek least cost (due to potential of being victimized). The dynamics of RAT posit that for crime events to occur a motivated offender and a desirable target must meet in space and time without the presence of capable guardianship. The Q-learning approach only learns rewards/costs from past crime events and assumes some global communication of these rewards and costs by the agents in the model. This is not very reasonable as the rewards from crime events is purely driven by the colocation of offender and target agents, which is dynamic and changing every stage of the model. A belief learning approach seems more reasonable as the agents directly learn in space where they believe opposing agents are likely to occupy. This assumes offenders can easily observe where targets frequent most often or where guardians patrol. Targets can observe where high crime areas are, where offenders are likely to be, or where guardians patrol. Guardians can observe where offenders are frequently or where potential targets choose to go.

Second, this paper will explicitly use the three main elements of RAT as agent types in the model and the three agent types follow the proposed belief learning method. Previous research was inconsistent in the use of the three elements of RAT as agents within the models. Malleson's [39] work did not use guardian agents in the model. Initially Short et al. [27] did not use guardian agents, however later work by Jones et al. [28] incorporated them into the model. When the three elements of RAT were incorporated as agents, again previous research was inconsistent on giving guardian agents any sort of learning process. In both Groff [15] and Wang [42], guardian agents did not follow any learning process and moved randomly. Agents in Bosse et al. [44–46] did not 'learn' over time and only responded to information from the previous time step.

Finally, agents in this paper are unique, with heterogeneous attributes that affect their interactions with other agents and the environment. Short et al. [27] and the subsequent derivative research solely used mathematical formulas to model crime in their ABM's. The agents in their models do not have any attributes that update or affect agent choices and interactions. Similar observations apply to the research by Bosse et al. [44–46]. The agents in their model do not have attributes that distinguish them apart. Agents with heterogeneous attributes and dynamic states that vary over time is an important feature of using ABM. It is this feature that distinguishes using a generative approach to modeling versus using aggregate mathematical equations that neglect agent heterogeneity [47].

Methodology

This paper presents a formal game-theoretic belief learning approach to RAT. The use of belief learning to model the dynamics of RAT's offender, target, and guardian behaviors within an agent-based model is original in that the agents learn and adapt given observation of other agents' actions without knowledge of the payoffs that drove the other agents' choices. This departs from previous research that has mostly explored statistical processes or reinforcement learning for crime modeling. This is a unique difference given the dynamics of RAT. It is the presence of the different agent types that provides opportunity for crime to occur, and not the potential for reward. *All* agents (including guardian agents) in the model use belief learning and the agents have heterogeneous attributes that affect their decision-making and the environment.

Table 1 Location attributes

Attribute	Description
$N_{\mathbf{O}}^{m}(t)$	Number of offender agents in given location at stage t
$N_{\mathbf{T}}^{m}(t)$	Number of target agents in given location at stage t
$N_{\mathbf{G}}^{m}(t)$	Number of Guardian agents in given locations at stage t
$T^{m}(t)$	Total crimes that have occurred in given location up to stage t
$C^{m}(t)$	Count of new crimes that occurred in given location at stage t

Formal setup

Consider players to be a set of n autonomous agents $\mathbf{A} := \{1, 2, \ldots, n\}$ which are partitioned into three sets, $\mathbf{O}$ for offender agent players, $\mathbf{T}$ for target agent players, and $\mathbf{G}$ for guardian agent players, such that $\mathbf{A} = \mathbf{O} \cup \mathbf{T} \cup \mathbf{G}$. The alternatives available for the agents to select and occupy is a set of k total locations $\mathbf{L} := \{m = 1, 2, \ldots, k\}$. In a repeated game, let stages be denoted by $t = \{0, 1, 2, \ldots\}$. At each stage t, agents select a location in which to operate. Each location in $\mathbf{L}$ captures the attributes in Table 1 at every stage of the game. These attributes are used by agents to perform belief learning computations. Since agents choose to operate in a location every stage, the set of locations is equivalent to the set of alternatives.

Agents

Three agent types exist in the game representing elements of RAT; offender, target, and guardian agents. Offender agents have three attributes μ, N^{s}, and N^{f}, as in Wang [42]. μ is the offender's motivation to commit a crime. N^{s} and N^{f} represent the number of successful and failed crime attempts by the offender agent. The values start at 0, increment as the model runs, and are used in updating the offender's motivation μ. These attributes are used in determining the likelihood of crime at every stage of the model. Target agents have two attributes, δ and γ. $\delta \in [0, 1]$ represents the target's desirability and is set randomly for each target agent. $\gamma \in [0, 1]$ represents the target's guardian capability to its peers and is set randomly for each target agent. Aggregated target desirability and guardian capability values within a location are used in conjunction with an offender's motivation to determine the likelihood of crime event. Guardian agents do not have any attributes. Their presence acts as a deterrent for crime and their choices of locations to operate within affect other agents' beliefs.

Belief learning

The approach to learning looked at in this study is belief learning. Agent beliefs, denoted $B_{\mathbf{O}}^{m}(t)$, $B_{\mathbf{T}}^{m}(t)$, and $B_{\mathbf{G}}^{m}(t)$, are the beliefs that offender (**O**), target (**T**), and guardian (**G**) agents occupy location m at stage t. Since the set of alternatives available to the agents is the set of locations in the model, the beliefs about player choices are the densities of agents of each type per location. The belief values are updated at every stage of the game following Eqs. (4)–(6).

$$B_{\mathbf{O}}^{m}(t) = \phi_{\mathbf{O}} B_{\mathbf{O}}^{m}(t-1) + (1-\phi_{\mathbf{O}}) \frac{N_{\mathbf{O}}^{m}(t)}{|\mathbf{O}|} \tag{4}$$

$$B_{\mathbf{T}}^{m}(t) = \phi_{\mathbf{T}} B_{\mathbf{T}}^{m}(t-1) + (1-\phi_{\mathbf{T}}) \frac{N_{\mathbf{T}}^{m}(t)}{|\mathbf{T}|} \tag{5}$$

$$B_{\mathbf{G}}^{m}(t) = \phi_{\mathbf{G}} B_{\mathbf{G}}^{m}(t-1) + (1-\phi_{\mathbf{G}}) \frac{N_{\mathbf{G}}^{m}(t)}{|\mathbf{G}|} \tag{6}$$

Beliefs are weighted combinations of beliefs from the previous stage and new information. $N_{\mathbf{O}}^{m}(t)$, $N_{\mathbf{T}}^{m}(t)$, and $N_{\mathbf{G}}^{m}(t)$ are the count of offender, target, and guardian agents, respectively, that chose to occupy location m during stage t, as shown in Table 1. $|\mathbf{O}|$, $|\mathbf{T}|$, and $|\mathbf{G}|$ are the cardinalities of the sets of offender, target, and guardian agents, respectively. These belief values for offender, target, and guardian agents are weighted by $\phi_{\mathbf{O}}$, $\phi_{\mathbf{T}}$, and $\phi_{\mathbf{G}}$, respectively. Letting $\phi = \frac{h-1}{h}$ where $h \geq 1$ is the number of stages of history to preserve, setting $h = 1$ stage of history produces $\phi = 0$, meaning no prior belief information beyond the most recent stage is preserved. If the history is set to use the current model stage $h = t + 1$, then ϕ approaches 1 as time passes, placing more weight on prior beliefs, which is pure weighted fictitious play learning. The history may be set to any fixed value, and will result in a moving average of history for the belief for the set history length. These updated belief values are then used to compute utilities using Eqs. (7)–(9) in each stage of the game.

Next, the utilities for each agent type are computed. Similar utilities as used in [46] are used in this study. Equations (7)–(9) represent how the agents evaluate the outcome of choosing a given location.

$$U_{\mathbf{O}}^{m}(t) = \alpha_{\mathbf{O}} B_{\mathbf{T}}^{m}(t) + (1 - a_{\mathbf{O}}) \left(1 - B_{\mathbf{G}}^{m}(t)\right) \tag{7}$$

$$U_{\mathbf{T}}^{m}(t) = \alpha_{\mathbf{T}} \left(1 - B_{\mathbf{O}}^{m}(t)\right) + (1 - a_{\mathbf{T}}) B_{\mathbf{G}}^{m}(t) \tag{8}$$

$$U_{\mathbf{G}}^{m}(t) = \alpha_{\mathbf{G}} B_{\mathbf{O}}^{m}(t) + (1 - a_{\mathbf{G}}) B_{\mathbf{T}}^{m}(t) \tag{9}$$

The goal of offender agents is to commit crimes. Following RAT, a crime event can occur when offenders and targets meet in space and time without guardianship present. Using Eq. (7), offender agents seek to occupy locations they believe target agents frequently occupy but also seek to occupy locations they believe guardian agents do not occupy frequently. Offender preference to seek targets or to avoid guardians is controlled by the weight parameter $\alpha_{\mathbf{O}} \in [0, 1]$. For values of $\alpha_{\mathbf{O}}$ close to 1, offenders would place more preference on seeking targets than avoiding guardian. For values closer to 0, offenders would place more preference on avoiding guardians than seeking targets. Similarly, the goal of target agents is avoid being victimized by offenders. Using Eq. (8), target agents seek locations they believe offender agents do not occupy frequently and seek locations they believe guardian agents do occupy frequently. Target agent preference to avoid offender agents or to seek guardian agents is controlled by the parameter $\alpha_{\mathbf{T}} \in [0, 1]$. Finally, the goal of guardian agents is to deter offenders from committing crime and to protect targets. Using Eq. (9), guardian agents seek to occupy locations they believe offender agents occupy frequently and seek to occupy locations they believe target agents occupy. Guardian agent preference to seek offender agents or to seek target agents is controlled by the parameter $\alpha_{\mathbf{G}} \in [0, 1]$. These three parameters will be set to the following ($\alpha_{\mathbf{O}} = 0.5$, $\alpha_{\mathbf{T}} = 1.0$, and $\alpha_{\mathbf{G}} = 1.0$). These parameters depict that offender agents equally prefer to seek out target agents and avoid guardian agents, target agents purely seek to avoid offender agents, and guardian agents purely seek out offender agents.

Likelihood of crime

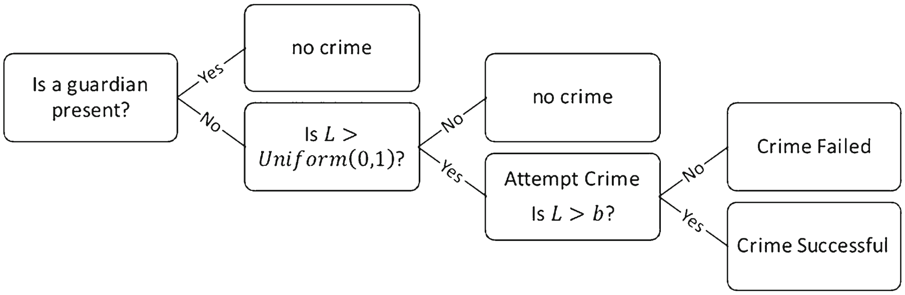
All the research discussed in the Related Work section used different approaches and mechanisms to generate crime events. In this study, at each stage of the game, offender agents can attempt a crime, but only if their motivation outweighs the guardianship present. This study uses a variation of Eck's likelihood for a crime event [12] as used by Wang et al. [43]. The formula in Eq. (10) is rooted in criminological theory and incorporates the heterogeneous attributes of agents in the model. It is not a stochastic process as use in other research. The likelihood for a crime event to occur is given by:

$$L = \frac{\mu_i \delta_m}{1 + \gamma_m} \tag{10}$$

μ_i is the motivation of offender i, δ_m is the sum of the desirability of all target agents in location m, and γ_m is the sum of the guardianship of all target agents in location m. The steps for a crime event to occur as proposed by Wang [42] are as follows (also see Fig. 3):

- If a guardian is present at a location, then no crime event occurs.
- If the likelihood in Eq. (10) returns a value larger than a random number between 0 and 1 (*i.e.*, $L > \mathit{Uniform}(0, 1)$), then a crime event is attempted

Fig. 3 Offender logic for a crime event (from Wang [42])



- If the likelihood L is larger than some minimum crime threshold $b \in [0, 1]$,then the crime is successful; else the crime attempt fails.

These steps are followed by each offender in each location at every stage of the game. The count of crimes that occurred in a location during a stage is stored in location attribute $C^m(t)$ and the total number of crimes that have occurred in the location is incremented by this value $T^m(t) = T^m(t-1) + C^m(t)$.

Adaptive offender agents

Along with following belief learning, offender agents also update their motivation to attempt crime, μ_i. The success or failure of past crime events by offenders affects their motivation to commit future crimes. This adaptability has been represented in previous research using sigmoid curves [42]. The sigmoid curve used in this study is shown in Eq. (11):

$$\mu_i = \frac{1}{1 + e^{-\left(N_i^s - N_i^f\right)}}, \tag{11}$$

where μ_i is the motivation of offender agent i, N_i^s is the number of successful crime events by the offender agent, and N_i^f is the number of failed crime attempts by the offender agent. An offenders' motivation is between the values of 0 and 1, $\mu_i \in [0, 1]$. As offender i's number of successful crime attempts outgrows the number of failures $\left(N_i^s - N_i^f\right) > 0$, μ_i approaches 1, meaning the offender is very motivated to continue to attempt to commit crime. If offender i's number of failed crime attempts outgrows the number of successful attempts, $\left(N_i^s - N_i^f\right) < 0$, μ_i approaches 0, meaning the offender is not motivated to attempt to commit crime.

Best response

In game-theoretic settings players are assumed to be rational meaning they always maximize their utility function. Offenders, targets, and guardians hardly behave this way in reality. To account for sub-optimal (or even apparent irrational alternative selection), all agents in this study will follow a mixed-strategy best response when determining which location to occupy next using a *softmax* equation similar to Eq. (3). In using *softmax*, agents are more likely to choose the location that maximizes their utility function; however, it is possible they could select a location that is not best given their utility function. At each stage of the game, offenders, targets, and guardians select which location to occupy next using Eq. (12).

$$\pi_m(t) = \frac{e^{U^m(t)}}{\sum_{b=1}^{k} e^{U^b(t)}} \tag{12}$$

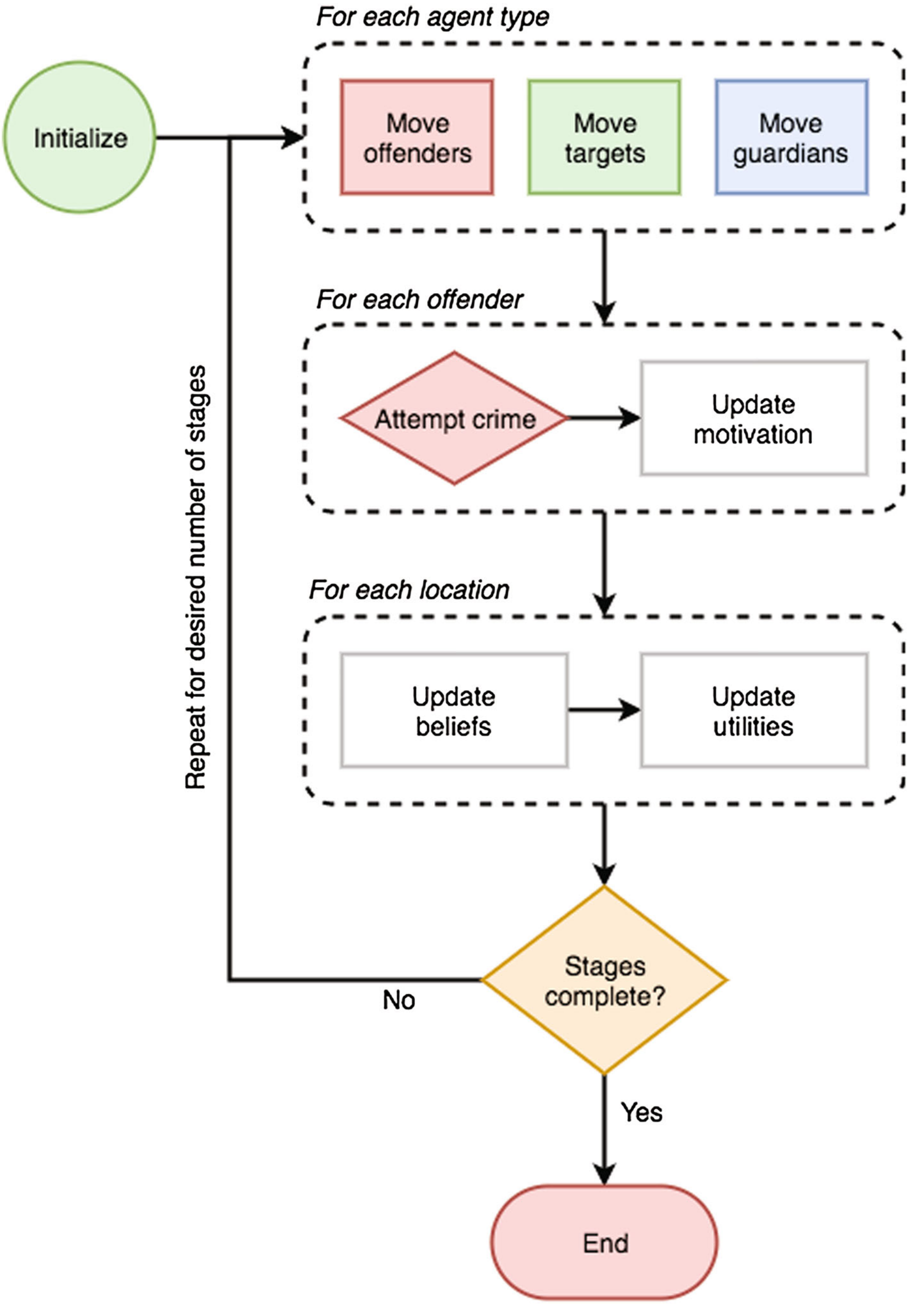
$\pi_m(t)$ is the probability of choosing location m at stage t and $U^m(t)$ is the utility a given agent type perceives for location m. $U^m(t)$ is replaced by Eqs. (7)–(9) for offender, target, and guardian agents, respectively

The use of a probability distribution to select a location introduces randomness into the model. Given this randomness, no two executions of the model will produce the same results. Therefore, multiple replications of the model should be run to return useful results.

Model algorithm

The resulting algorithm that ties all the elements of the proposed model together is as follows. Starting in the top left of Fig. 4, the model is instantiated with agents (offenders, targets, and guardians), locations to be used as alternatives, and priors of historical information to establish initial beliefs for each agent type. After the model is initialized, the first stage of the model begins. All agents (offenders, targets, and guardians) independently survey the set of locations and select a location to move to and occupy. This selection is made using the *softmax* best response in Eq. (12) and their respective utilities in Eqs. (7)–(9). After all agents have moved, crime events are computed by each offender agent using Eq. (10) and the steps depicted in Fig. 3. Given any successful or failed crime attempts, offender agents update their motivation to commit crime following Eq. (11). After crime events have been computed, each location then updates the beliefs for each agent type using Eqs. (4)–(6), respectively.

Fig. 4 Model algorithm



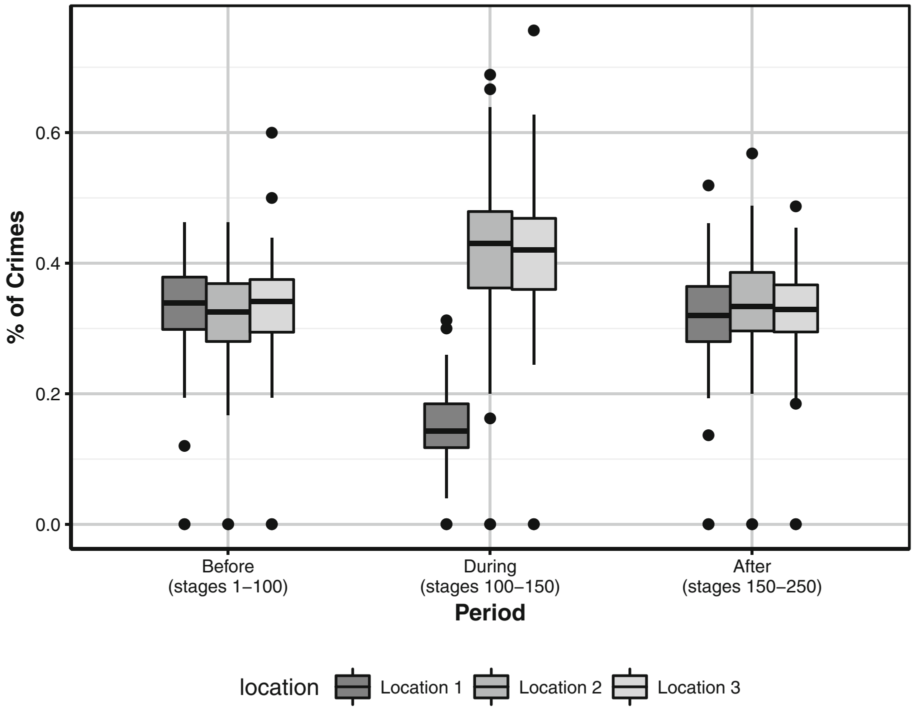
Using these updated beliefs, each location then updates the utility for each agent type using Eqs. (7)–(9), respectively. These updated utilities are now ready to be used in subsequent stages of the model. This completes one stage of the model and the model continues until the desired number of stages have been completed.

Theoretical exercise

To illustrate the behaviors of the proposed belief learning approach, consider a model setup with $k = 3$ locations, 2 offender agents, 10 target agents, 1 guardian agent, and all agents will use $h = 50$ stages of history. The model is run for 250 stages, with the 1 guardian agent having a directed patrol to Location 1 for 50% of its time starting at stage $t = 100$ and ending at stage $t = 150$. The model is run for 100 replications. Starting at stage $t = 100$, the guardian agent will spend more of its time at Location 1. Given the learning dynamics of the agents, with the guardian agent occupying Location 1 more frequently, potential crime would be prevented and offenders would be deterred from occupying the location, moving to other locations to attempt crime.

Figure 5 shows box plots of the crime proportions by location and time period for the 100 replications run. It can be seen

Fig. 5 Illustrative example of crime proportions by location



that during the initiative, the proportion of crime in Location 1 is reduced given the directed response. During the 100 stages after the initiative ends, the crime proportion increases at Location 1. This is indicative of the guardian agent not occupying Location 1 as frequently, seeking the other locations given belief history that has built up there. This in turn allows offenders to build up belief that the guardian does not occupy Location 1 as much, causing crime to return to Location 1.

Model validation and case studies

The presented belief learning approach is applied to empirical case studies along with an implementation of a Q-learning approach similar to Wang [42] and Wang et al. [43] for comparison (*In their approach, offender agents seek reward for committing crimes, target agents seek to minimize cost of being victimized, and guardians seek high crime locations*). Case studies were derived given three factors. The first being an initiative must have been completed, second the time frame in which the initiative must be known, and finally the crime data must be available from the time the initiative was carried out.

Belief learning and Q-learning model runs both follow the agent algorithm in Fig. 4 and have the same setup. The models are setup as follows: initial offender choice densities and crime rates are computed from the 911 incidence response data using the three months prior to the initiative. There are 30 target agents, 6 offender agents and 3 guardian agents used. Guardian agents are directed to spend 25% of their time in the response area during the initiatives. The models are run for 100 replications. All offenders are initialized with motivation $\mu = 0.5$, which will update independently given each offender agent's experiences. For belief learning model runs, $h = 14$ stages of beliefs are preserved. The use of $h = 14$ is to represent 14 stages (or two weeks) of reference to be used when updating beliefs. All agents will use the same history length in their belief learning. The number of locations, the overall duration of the case study, and the duration of the directed patrolling will vary depending on the case study.

To evaluate simulated results, root-mean-square error (RMSE) is applied to compare belief learning to Q-learning. RMSE is used when measuring the difference between model predicted values and observed values. RMSE assesses the accuracy of models by comparing error at an aggregate level. The smaller the value of RMSE, the less difference there is between the modeled results and the observed values. RMSE is computed using the observed empirical crime proportions compared to the modeled proportion of crimes for each time period for each location. RMSE is calculated in Eq. (13)

$$\mathrm{RMSE}_m = \sqrt{\frac{\sum_{x=1}^{n} \left(y_{x,m} - \hat{y}_{x,m}\right)^2}{n}} \tag{13}$$

where $y_{x,m}$ is the observed crime proportion at study area m during time period x, $\hat{y}_{x,m}$ is the modeled crime proportion for study area m during time period x, and n is the number of time periods. Following the concerns of Andresen and Malleson [6] regarding the study areas having varying sizes, prior to calculating the RMSE, all crime counts for each time period are standardized by the area (km^2) of the study areas.

The relative proportion of crime for each time period is then computed across the study areas to derive observed crime proportions.

Case study 1

In 2010 the Seattle, WA Police Department (SPD) executed an initiative called "Safer Union" in attempt to disrupt an open air drug market and to reduce overall crime between 20th and 25th Ave on East Union Street (centered around 23rd and Union) in SPD's East Precinct [48]. "Safer Union" was carried out over a 90-day period starting in October 2010. The study areas of interest comprised of the response area centered around 23rd and Union Street, five concentric one-block buffers around the response area, and a control area all within the East Precinct. The case study is setup with $k = 7$ locations (Response area, five x 1 block buffer areas, and a control area). The case study is played over 270 stages, with 1 stage representing 1 day. The initiative lasted 3 months (90 days) and follow on affects where assessed during the subsequent 6 months (180 days). The case study starts at stage $t = 0$ with guardian agents having a directed patrol in the Response Area for 25% of their time for 90 stages ending at stage $t = 90$. Data was pulled from Seattle's Open Data site [49]. 911 incident response data was queried for four time periods: three months prior to the initiative (July–Sept 2010), three months during the initiative (Oct–Dec 2010), 3 months post-initiative (Jan–Mar 2011) and 3–6 months post-initiative (Apr–Jun 2011).

The box plots in Fig. 6 show simulated proportions of crime for each time period across the study areas by learning method. An initial observation to note is the spread of the results for the belief learning scenario compared to the Q-learning scenario. The belief learning results are tight and compact around the median while the Q-learning results are widely spread and skewed with many outliers present. This graphically shows belief learning has a smaller variance around its results than that of Q-learning. We can see that both scenarios do in fact reduce crime during the initiative in the response area. After the directed patrolling ends, the proportion of crime in the response area increases while the proportion of crime in the other study areas show a slight decrease, shifting lower on the plot. This observation is supported by Sorg's finding [7] that the positive effect of reducing crime during an initiative does not endure long after the directed patrolling ending.

Table 2 "Safer Union" Average RMSE of the Study Areas for each learning method

Study area	Belief learning	Q-learning
Response area	0.1237	0.1393
Control	0.0267	0.0625
Buffer 1	0.0855	0.0970
Buffer 2	0.0327	0.0710
Buffer 3	0.0330	0.0649
Buffer 4	0.0354	0.0681
Buffer 5	0.0316	0.0665

The box plots in Fig. 6 do not show which learning method is a closer fit to the observed empirical data. RMSE was computed on each replication across the time periods for both learning methods. The average RMSE for each study area by learning method is in Table 2. In general, both learning methods have low RMSE values across each study area. However, there is difference between the learning methods.

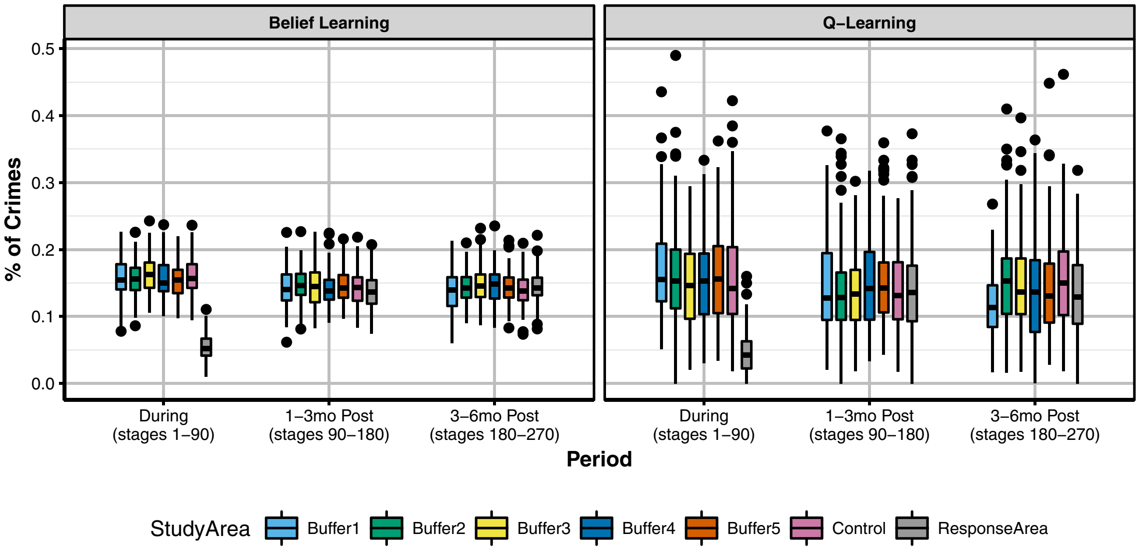


Fig. 6 "Safer Union" box plots of simulated crime proportions (by time period and across study areas) for each learning method

Belief learning has lower average RMSE values (meaning a closer fit) than Q-learning for all the study areas. This indicates that belief learning is a closer fit to the observed data than Q-learning. The box plots of the distribution of RMSE values by learning method in Fig. 7 show the variance in the belief learning results are much smaller. The box plots graphically show the mean and median RMSE values for belief learning are lower than Q-learning. Paired t-tests comparing these RMSE distributions for each study area confirm the RMSE for belief learning is statistically significant lower than Q-learning for all study areas at 99% confidence. F-tests comparing RMSE variances also confirm belief learning is statistically significant lower than Q-learning for all study areas at 99% confidence.

Case study 2

In 2013, the city of Austin, TX received grant funding for an initiative called "Restore Rundberg". This was an area of the city that accounted for a large proportion of their crime. The whole "Restore Rundberg" initiative comprised of both community oriented projects to restore this area of the city along with focused policing efforts at three hot spots. One hot spot was centered around Rundberg Lane and the Interregional Highway. "Restore Rundberg" policing efforts were carried out over 15 months, starting April 2014 thru June 2015. The study areas of interest consist of the response area centered around Rundberg Lane and the Interregional Highway, 4 buffer areas around the response area, and a control area. The case study is setup with $k = 6$ locations (Response Area, 4 buffer areas, and the control area). The case study is run over 640 stages, with 1 stage representing 1 day. The initiative lasted 15 months (456 days) and follow on affects were assessed for the subsequent 6 months (stages 457–640). The case study starts at stage $t = 0$ with guardian agents having a directed patrol in the Response Area for 25% of their time for 456 stages ending at stage $t = 456$. Data was pulled from Austin's Open Data site [50]. 911 incident response data was queried for four time periods: three months prior to the initiative (Jan-Mar 2014), 15 months during the initiative (Apr 2014–Jun 2015), 3 months post-initiative (Jul–Sep 2015) and 3–6 months post-initiative (Oct–Dec 2015).

Table 3 "Restore Rundberg" Average RMSE of the Study Areas for each learning method

Study area	Belief learning	Q-learning
Response area	0.1068	0.1818
Control	0.1923	0.2546
Buffer 1	0.0662	0.1695
Buffer 2	0.1102	0.1825
Buffer 3	0.0891	0.1659
Buffer 4	0.0453	0.1605

Table 3 contains the average RMSE for each study area by learning method. Similar to the "Safer Union" case study, belief learning has lower average RMSE values (meaning a closer fit) than Q-learning for all the study areas. This indicates that belief learning is a closer fit to the observed data of "Restore Rundberg" than Q-learning. The box plots of the distribution of RMSE values by learning method in Fig. 8 also shows that the variance in the belief learning results are much smaller. The box plots graphically show the mean and median RMSE values for belief learning are lower than Q-learning. Again, similar to "Safer Union" results, paired t-tests comparing these RMSE distributions for each study area confirm the RMSE for belief learning is statistically significant lower than Q-learning for all study areas at 99% confidence. F-tests comparing RMSE variances again confirm belief learning is statistically significant lower than Q-learning for all study areas at 99% confidence.

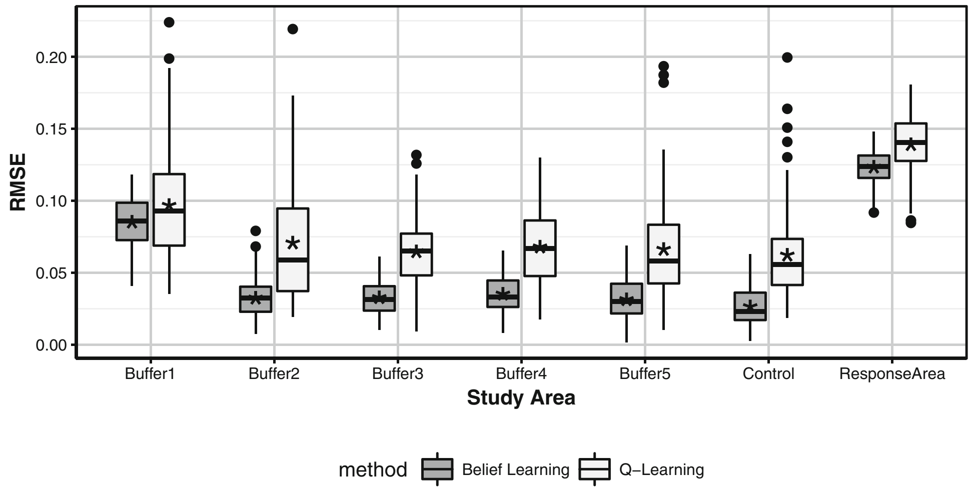


Fig. 7 "Safer Union" Box plots for RMSE of simulated crime proportions across study areas for belief learning and Q-learning. Solid lines in boxplots represent the median and the * represents the mean of distributions. (differences in mean and variance are statistically significant at 99% confidence for all study areas)

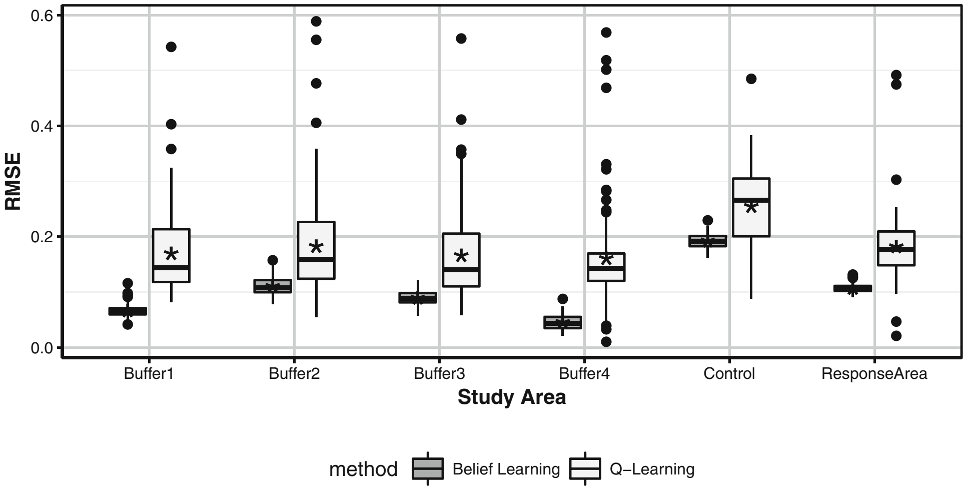


Fig. 8 "Restore Rundberg" Box plots for RMSE of simulated crime proportions across study areas for belief learning and Q-learning. Solid lines in boxplots represent the median and the * represent the mean of distributions. (differences in mean and variance are statistically significant at 99% confidence for all study areas)

Case study 3

Starting in 2014, the city of Minneapolis, MN started a recurring summertime initiative within its wards called JET (Joint Enforcement Teams) to have directed patrolling and enforcement to high crime areas. In 2015, one targeted area was in North Minneapolis covering the blocks within Lowry Ave. N. to 35th Ave. N. and Bryant Ave. N. to James Ave. N. The initiative was carried out over 90 days, starting in June 2015 and ending in August 2015. The study areas of interest comprised of the response area, 3 concentric one-block buffers around the response area, and a control area all within the North Ward. The case study is setup with $k = 5$ locations (Response area, 3 × 1 block buffer areas, and a control area). The case study is played over 270 stages, with 1 stage representing 1 day. The initiative lasted 3 months (90 days) and follow on affects where assessed during the subsequent 6 months (180 days). The case study starts at stage $t = 0$ with guardian agents having a directed patrol in the Response Area for 25% of their time for 90 stages ending at stage $t = 90$. Data was pulled from Minneapolis' Open Data site [51]. 911 incident response data was queried for four time periods: 3 months prior to the initiative (Mar–May 2015), three months during the initiative (Jun–Aug 2015), 3 months post-initiative (Sept–Nov 2011), and 3–6 months post-initiative (Dec 2015–Feb 2016).

The average RMSE for each study area by learning method is shown in Table 4. The average values are slightly higher than the RMSE results seen for "Safer Union" and "Restore Rundberg" case studies. However, there is still a difference between the learning methods. Belief learning has lower average RMSE values than Q-learning for all the study areas. This indicates that belief learning is a closer fit to the observed data than Q-learning. The box plots of the distribution of RMSE values by learning method in Fig. 9 shows, again, the variance in the belief learning results are much smaller. The box plots show the mean and median RMSE values for belief learning are lower than Q-learning. Paired t-tests comparing these RMSE distributions for each study area again confirm the RMSE for belief learning is statistically significant lower than Q-learning for all study areas at 99% confidence except for Buffer 1 and Buffer 3 which are slightly less at 90% confidence. F-tests comparing RMSE variances show, again, belief learning is statistically significant lower than Q-learning for all study areas at 99% confidence.

Table 4 Minneapolis, MN Case Study Average RMSE of the Study Areas for each learning method

Study area	Belief learning	Q-learning
Response area	0.0927	0.1106
Control	0.1725	0.1869
Buffer 1	0.131	0.1397
Buffer 2	0.1513	0.1772
Buffer 3	0.1256	0.1344

Discussion

As this new approach demonstrates in the case studies, belief learning fits the observed empirical data better and produces results with lower variance when compared to a Q-learning approach. These results offer evidence that using belief learning within an ABM to model crime dynamics produces tighter and more accurate results than Q-learning. This finding is due to the dynamic environment of the model and the agents' goals as prescribed by RAT. The learning goals for offender, target, and guardian agents is to commit crime, avoid victimization, and to prevent crime, respectively. The attributes of the locations are changing every stage of the model given the feedback of information and agent interaction with the environment. Q-learning only looks at the reward reinforcement received in the locations. These rewards are dependent on what agents occupy a given loca-

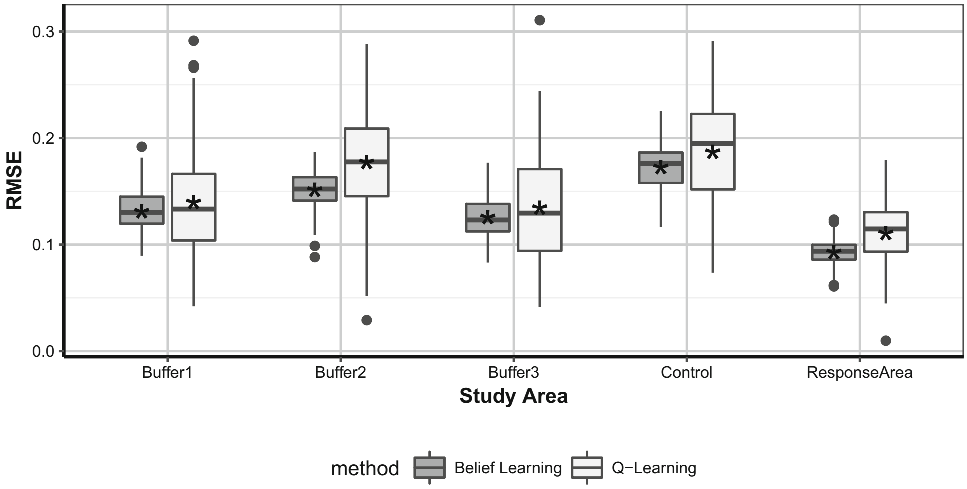


Fig. 9 Minneapolis, MN Case Study Box plots for RMSE of simulated crime proportions across study areas for belief learning and Q-learning. Solid lines in box plots represent the median and the * represent the mean of distributions. (differences in mean and 0.1 variance are statistically significant at 99% confidence for all study areas except Buffer 1 and Buffer 3 at 90%)

tion. By only responding to previously earned rewards, agents are ignoring the dynamic nature of how rewards are generated in each location during each stage of the model. This can lead to wildly varying results as seen by the spread and variance of the results in Figs. 7, 8, 9. In contrast, the belief learning approach put forth in this paper looks at a moving average history of beliefs of what locations are selected by opposing agents. As a byproduct of learning where opposing agents choose to occupy, if motivated offenders occupy the same location as targets without a guardian present, then crimes may occur. If a guardian occupies a location with offenders, then crimes are deterred and targets are protected.

Conclusions

This paper introduces an original model for using ABM and game-theoretic belief learning to explore RAT's offender, target, and guardian dynamics and potential displacement behaviors. Agents learn and adapt given observation of other agents' actions without knowledge of the payoffs that drove the other agents' choices. This differs from previous research that has mostly explored statistical processes or reinforcement learning for crime modeling. Based on the dynamics of RAT, it is the presence of the different agent types that provides opportunity for crime to occur, and not the potential for reward. With an implementation of the model to represent an actual "hot spot" initiative, belief learning generated results that fit the empirical data of the case study. The RMSE results of the belief learning approach are statistically significant with less variance when compared to a Q-learning approach. Additionally, the collection of crime counts for during and after "hot spot" directed patrolling supported displacement observations and the inverse displacement of crime returning to the response area after the directed patrol ended. The game-theoretic belief learning approach presented in this paper offers a method to interdict opportunity, deter offenders and protect the public. It is believed that application of this new approach supports law enforcement in planning for the effects of displacement to other locations and assists in developing responses to maintain displacement from certain locations.

Future research should be done in applying this approach in different agent environments, for example in a GIS environment or a transportation network. This would enrich the versatility of the approach beyond using a set of alternatives. Last, follow-on work should examine differing inputs, such as restricting the information used by the agents to compute the best responses. For example, using bounded rationality to investigate seemingly irrational choices by agents. This study provides a set of equations and steps for agent interaction as a starting point to explore these questions further.

References

1. Goldstein H, Susmilch CE (1981) The problem-oriented approach to improving police service. Dev Probl Polic Ser p 1
2. Eck JE, Spelman W (1987) Problem solving: problem-oriented policing in newport news. US. Department of Justice-National Institute of Justice, Washington, DC
3. Willis JJ, Mastrofski SD, Weisburd D (2007) Making sense of COMPSTAT?: a theory-based analysis of organizational change in three police departments. Law Soc Rev 41:147–188
4. Telep CW, Weisburd D, Gill CE et al (2014) Displacement of crime and diffusion of crime control benefits in large-scale geographic areas: a systematic review. J Exp Criminol 10:515–548
5. Weisburd D, Telep CW (2014) Hot spots policing: what we know and what we need to know. J Contemp Crim Justice 30:200–220

6. Andresen MA, Malleson N (2014) Police foot patrol and crime displacement: a local analysis. J Contemp Crim Justice 30:186–199
7. Sorg ET, Haberman CP, Ratcliffe JH, Groff ER (2013) Foot patrol in violent crime hot spots: the longitudinal impact of deterrence and posttreatment effects of displacement. Criminology 51:65–101
8. Epstein JM (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press, Washington D.C
9. Birks DJ, Donkin S, Wellsmith M (2008) Synthesis over analysis: towards an ontology for volume crime simulation. In: Artificial crime analysis systems: using computer simulations and geographic information systems. Information Science Reference, Hershey, PA, pp 160–192
10. Groff ER, Mazerolle L (2008) Simulated experiments and their potential role in criminology and criminal justice. J Exp Criminol 4:187–193
11. Gerritsen C (2015) Agent-based modelling as a research tool for criminological research. Crime Sci 4:2
12. Eck JE (1995) Examining routine activity theory: a review of two books. Justice Q 12:783–797
13. Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608
14. POP Center (2015) The problem analysis triangle. http://www.popcenter.org/about/?p=triangle. Accessed 1 Jul 2015
15. Groff ER (2007) Simulation for theory testing and experimentation: an example using routine activity theory and street robbery. J Quant Criminol 23:75–103
16. Clarke RV, Cornish DB (1985) Modeling offenders' decisions: a framework for research and policy. Crime Justice 6:147–185
17. Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci USA 99(Suppl 3):7280–7287
18. Amalraj W, Prahlad A, Chen K, Dipti T (2016) Interoperable multiagent framework for unmanned aerial/ground vehicles?: towards robot autonomy. Complex Intell Syst 2:45–59
19. Iammartino R, Bischoff J, Willy C, Shapiro P (2016) Emergence in the U.S. Science, technology, engineering, and mathematics (STEM) workforce: an agent-based model of worker attrition and group size in high-density STEM organizations. Complex Intell Syst 2:23–34
20. Macal CM, North MJ (2010) Tutorial on agent-based modeling and simulation. J Simul 4:151–162
21. Shoham Y, Leyton-Brown K (2009) Multiagent systems: algorithmic, game-theoretic, and logical foundations. Cambridge University Press, Cambridge
22. Jackson MO (2011) A brief introduction to the basics of game theory
23. Ho TH (2009) Individual learning in games. In: Durlauf SN, Blume LE (eds) New Palgrave Dictionary of Economics. Behavioural and experimental economics. Palgrave Macmillan, Basingstoke, pp 157–165
24. Brown GW (1951) Iterative solution of games by fictitious play. Act Anal Prod Alloc 13:374–376
25. Fudenberg D, Levine DK (1998) The theory of learning in games. MIT Press, Cambridge
26. Watkins C (1989) Learning from delayed rewards. (Doctoral Dissertation). King's College
27. Short MB, D'Orsogna MR, Pasour VB et al (2008) A statistical model of criminal behavior. Math Model Methods Appl Sci 18:1249–1267
28. Jones PA, Brantingham PJ, Chayes LR (2010) Statistical models of criminal behavior: the effects of law enforcement actions. Math Model Methods Appl Sci 20:1397–1423
29. Chaturapruek S, Breslau J, Yazdi D et al (2013) Crime modeling with Lévy flights. J Appl Math 73:1703–1720
30. Camacho A, Lee HRL, Smith LM (2016) Modelling policing strategies for departments with limited resources. Eur J Appl Math 27:479–501
31. Tambe M, Jain M, Pita JA, Jiang AX (2012) Game theory for security: Key algorithmic principles, deployed systems, lessons learned. In: 50th annual allerton conference on communication, control, and computing, pp 1822–1829
32. Zhang C, Jiang AX, Short MB, et al (2013) Modeling crime diffusion and crime suppression on transportation networks: an initial report. In: AAAI Fall Symp
33. Zhang C, Jiang A, Short M et al (2014) Defending against opportunistic criminals: new game-theoretic frameworks and algorithms. Decis Game Theory Secur SE 1(8840):3–22
34. An B, Shieh E, Yang R et al (2012) PROTECT—a deployed game-theoretic system for strategic security allocation for the United States coast guard. AI Mag 33:96–110
35. Fang F, Nguyen TH, Pickles R et al (2016) Deploying PAWS?: field optimization of the protection assistant for wildlife security. In: Proc Twenty-Eighth Innov Appl Artif Intell Conf
36. Mckelvey RD, Palfrey TR (1995) Quantal response equilibria for normal form games. Games Econ Behav 10:6–38
37. Sutton R, Barto A (2014) Reinforcement learning?: an introduction. MIT Press, Cambridge
38. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
39. Malleson N (2010) Agent-Based Modelling of Burglary. (Doctoral Dissertation). University of Leeds
40. Malleson N, See L, Evans A, Heppenstall A (2010) Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. Simulation 88:50–71
41. Malleson N, Birkin M (2012) Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. Comput Environ Urban Syst p 36
42. Wang X (2005) Spatial adaptive crime simulation with the RA/CA/ABM computational laboratory. (Doctoral Dissertation). University of Cincinnati
43. Wang N, Liu L, Eck JE (2014) Analyzing crime displacement with a simulation approach. Environ Plan B Plan Des 41:359–374
44. Bosse T, Gerritsen C (2010) Social simulation and analysis of the dynamics of criminal hot spots. J Artif Soc Soc Simul p 13
45. Bosse T, Gerritsen C (2010) A model-based reasoning approach to prevent crime. In: Magnani L, Carnielli W, Pizzi C (eds) Model Reason Sci Technol. Springer, Berlin, pp 159–177
46. Bosse T, Gerritsen C, Hoogendoorn M et al (2011) Agent-based vs. population-based simulation of displacement of crime: a comparative study. Web Intell Agent Syst 9:147–160
47. Kyun J, Hiroki S, Seung S, Choi R (2016) A state equation for the Schelling's segregation model. Complex Intell Syst 2:35–43
48. Thomas TA (2013) Quantifying crime displacement after a hot-spot intervention. (Thesis). University of Washington
49. SPD (2014) 911 incident response data. https://data.seattle.gov
50. APD 911 Incident response data. https://data.austintexas.gov
51. MPD (2015) 911 Incident response data. http://opendata.minneapolismn.gov/

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.