How does experience of cops affect the successfulness of solving civil violence?

Maikel Withagen (S1867733) [a] Steven Bosch (S1861948) [a]
Robin Kramer (S1970755) [a]

[a] *University of Groningen, Faculty of Mathematics and Natural Sciences, Nijenborgh 9, 9747 AG, Groningen*

Abstract

Much research has been done to how civil violence may emerge, but less focus has been on what factors affect cop performance to control civil violence. In this paper, the effect of experience of cops, modeled by individual learning rates, on the overal group performance is investigated. No significant results were found for the success, nor for speed of resolving the issue. Because of the lack of results and some shortcomings to the model, no conclusions can be drawn.

1 Introduction

1.1 Problem

The civil war in Bosnia and Herzegovina from 1992 to 1995, the Ferguson unrest in 2014, and more recently the civil war in the Ukrain; Civil violence has been an issue in many countries for many years. There is no single cause for these riots and wars. Some were the result of cultural differences, some of the feeling of being treated injustly, and others came about due to political reasons. “Each war is as different as the society producing it" [8], and getting more insight into the development and handling of these riots is of utmost importance.

1.2 State of the Art

Much research has been done to understand how these riots emerge, including simulations of riots using a game theoretic approach [5] and social networks [4] among others. These investigations have shown that the behavior at a macroscopic level, that is, the behavior of an entire group, is the result of behavior on the microscopic level, that is, the individual agents of that group. In a more recent study, Goh and colleagues [3] also studied how macroscopic behavior emerged using a game theoretic approach in a simulation. The most important issue the authors adressed was how different events affected the individual tendency to riot and how this affected macroscopic behavior and situations.

Goh et al. included many different interactions in his simulation, such as a probability that civilians turned to active protesters, how jail time affected rehabilitation of arrested protesters, and how the amount and types of people affected individual decisions. The focus was thus mainly on the civilian and the protesters. What the authors failed to focus on, however, was how experience affected the cops’ behavior during riots. It is believed that many previous experiences ensure a more stable performance [1, 6], meaning that people switch less often between different strategies for the same goal. In other words, more experienced humans are

less affected by new experiences compared to less experiened people, which can roughly be translated to the saying "you can't learn old dogs new tricks."

1.3 New Idea

Knowing how experience influences group behavior may be of vital importance for the success of a group performance. For one individual, it is very well possible that the 'correct' course of action may result in a negative outcome in a particular situation, because the world is not deterministic. The experienced person will be resilient to such an 'accidental' negative outcome, whereas a less experienced person may be affected more severely by that negative outcome. The latter person may therefore change to a less favourable course of action in a future occurence of said situation, which may result in frailty of the group's dynamics. It has already been shown in an organizational context that individual experience may determine the success of development of the organization [7]. Individual experience may thus have a great impact on group performance. As of yet this has not received much attention in a civil violence context however.

The current paper focusses on how experience influences cop behavior. By using a simulation of a multi-agent system, we were able to compare how experienced cop agents, not experienced cop agents and a mix thereof perform against a group of agressive hostile agents. A reinforcement learning strategy was used to let the individual agents learn the optimal strategy, which has been shown to be robust and natural way of teaching agents [2]. In the following sections, the simulation and data acquisition will be descibed in more detail. Following, the results of the different simulation will be presented. The paper concludes with a discussion of the implications of the results, and some shortcomings to this paper.

2 Method

2.1 Simulation model

Three types of agents will be included in this simulation: cops, hostiles and civilians. The goal of the cops is to keep (civil and cop) casualties as low as possible. The hostiles on the other hand have only one goal: to kill as many civilians as possible to cause mayhem and despair. Civilians have no particular function, but are subject to the actions taken by cops and hostiles. The simulation takes place in a 20x20 2D matrix. In every box reside a mean of 20 civilians, 10 hostiles and 9 cops. The agents that reside in this box can only see the other agents in that box. This allows for easier computation of interactions with visible agents. In every simulation, the agents will be randomly spawned in the matrix.

2.1.1 Goals and Actions

The cops' goal is to keep casualties as low as possible. To achieve this, every cop must choose one of two actions. The cop can shoot a hostile, which eliminates a treath of future killings, but at the same time there is the risk of killing a civilian. The second action is to to save a civilian, who is then safe from being killed, but the cop has a higher risk of being killed by a hostile. Hostiles only have one action: to kill. Depending on the amount of civilians and cops, one group has a larger possibility to get killed by the hostiles. When, for example, only cops and hostiles reside in a box, the possibility that cops are targeted is 100 percent. It depends on the amount of hostiles how many civilians and cops may die.

All the cops, during every epoch, can either save or shoot once. Based on the overall result in that block, a reward is given to the individual agent. This reward can be either positive or

		Shoot	Save
# Cops > # Hostiles	Many civilians	Ux	Uy
	Few civilians	Ur	Ut
# Cops < # Hostiles	Many civilians	Ua	Ub
	Few civilians	Uc	Ud

Table 1: The different scenarios an agent can encounter and fictional success values for each action.

negative. If many civilians die, for instance, a negative reward will be given, whilst when many hostiles were killed, a positive reward will be given. The decision that rewards are given based on group success, as opposed to individual success, is based on the idea that in a crowded situation it is not always clear what the results of individual actions are, but that a general idea of group performance can be perceived.

The reward is used to update the utility of the particular actions in a specific circumstance. These circumstances are shown in Table 1. The cops must thus decide, based on the amount of team members, opponents and civilians what action would be most successful, that is, which action has the highest utility. Because the agents are not perfect, their judgement of the situation and the decision for an action may be erroneous. Mistakes can thus be made, which increases the realism of the simulation.

In the beginning of the simulation, all the utilities are set to equal values, such that no biases may exist. Reward of an action can be calculated as following:

$$R(s,a) = \frac{Kills(s) + Saves(s') - Losses(s')}{Kills(s') + Saves(s') + Losses(s')}$$

in which R is the reward of action a in situation s, $Kills$ are the amount of killed hostiles in the new situation s', $Saves$ the amount of saved civilians, and $Losses$ the amount of killed cops and civilians. Based on these calculations the reward will lie between -1 (only deaths) and 1 (only kills and saves). Because hostiles only have one action, no learning is necessary.

As Table 1 also shows, there is no situation in which the teams are of equal size. The cops must decide whether they have the overhand or not, which is done according to the following function:

$$\Omega = \frac{n_{cops}}{n_{cops} + n_{hostiles}} * \sigma$$

in which Ω is a value between 0 and 1, n is the amount of agents of a group in that box and σ is a random value between 0.5 and 1.5 to make the decision stochastic. It is assumed namely that the cops have no perfect knowledge of their environmnet. If Ω is higher than 0.5, the cops assume they have the overhand. The same function is used to assess the amount of civilians; The number of cops is then replaced by the amount of civilians.

When there is no threat in the area of an cop anymore, that is, if there are no more hostiles left, it can move to one of (at most) four neighboring boxes in the matrix. The cop will move to the box in which the action with the highest utility can be applied. In other words, the cops will move to where he/she will thrive best. If the cop can move to several areas, one of the areas will be picked at random. This way, the cops can always search for a place where he can be of use.

2.1.2 Learning

Many reinforcement learning algorithms, such as Q-learning, have implemented the fact that newer experiences have a reduced influence [9]. This is implemented with a learning rate vari-

able λ. High learning rates allow agents to learn more quickly compared to a low learning rate, but this also allows the agents to switch to suboptimal strategies more often. This may result in a lower performance compared to the low learning rate. Gradually changing the learning rate over time allows the function to converge to the true Q-value, that is the utility value. In the current simulation, a similar approach will be taken.

The utility will be updated according to the following function:

$$U(s,a)_{new}(t) = U(s,a)_{old} + \frac{\lambda * R(s,a)}{1+\lambda}$$

in which $U_x(t)$ is the new utility (ranging from 0 to 1) of an action a in situation s, t is the moment in time, λ is the learning rate factor ranging from 0 to 1, and $R(t)$ is the reward, as calculated before. After each epoch, in which one action could be taken by each agent, the utility functions will be updated. Following the utilities will be normalized, such that the utility of the two actions in a particular situation will sum up to one. In a way, the success of one action will also mean the discounting of another, therefore producing a stronger preference for what is successfull. According to the formula, henever there is a reward of zero, no changes in utility will occur

2.2 Experimental Design

To see the influence of learning rates on behavioral learning, 1000 runs, that is, 1000 civil wars were run. In the first four simulations the agents receive a learning rate of 0.8, 0.5, 0.2 and 0.0 respectively, which will be fixed over the 100 runs. In the final simulation, the agents will start with a high learning rate value of 0.8, which will be decreased by a factor of 0.05 every run. In other words, people gain experience on the job and will become less adaptive with every civil war.

During the simulations, we will record the general success value of each run and the amount of epochs of these runs. Moreover, the general mean utility table of every run will be saved. The simulation of one riot is finished when there are no civilians left in the area.

One simulation is considered to be more successful if the mean successvalue, which is the same as the reward function, was significantly higher compared to the other simulations. If no difference in success is found, we will look at the amount of epochs it takes to finish a simulation, that is, the speed of solving the problem. After every run, the utility tables will be remembered and used to initialize agents in the next run. This allows the agents to learn over the different civil wars.

2.2.1 No bias

In the first simulations, the focus lied on which behaviours would be most useful. Therefore, the all the agents were initialized with the default utility values, which is 0.5 for both actions in all situations. It is expected that the tendency to save will be higher when there are many civilians and that the tendency to shoot is higher when many hostiles are present. Learning this behavior should over time increase performance, which should show either in success or in speed.

2.2.2 Initial biases

Another point of interest is how behavior evolves over time, when bias does exist. Can old dog learn new tricks, and how quickly could that happen. To investigate this, a preference of either shooting or saving was given by setting the utility to 0.9 for one action, and 0.1 for the

λ	mean # epochs	Success
0.8	55.6	0.218
0.5	46.8	0.230
0.2	58.2	0.210
<1.0	39.2	0.235

Table 2: The amount of epochs necessary to finish and the success of the simulation for the cop-teams with different levels of experience.

other. When a bias for shooting exists, there will be no more hostiles to shoot after a while. It is expected that after some iterations, the cops learn that saving is more usefull. Similarly, when cops only save, it is expected that shooting will receive a higher utility over time. For the cops having a lower learning rate this will most likely happen more gradually compared to the cops having a higher learning rate.

It is expected that the cops with the medium learning rate or the decreasing learning rate will show higher performance compared to the cops with a higher and lower learning rate. Although the group with the higher learning rate with change behavior more quickly, there performance is also more fragile. The groups with the lower learning rate, on the other hand, will not adapt as much to learn the better strategy.

3 Results

The first three simulations were run with learning rate values of 0.8, 0.5, and 0.2, corresponding to the inexperienced, mixed-experienced and experienced cops. Following, the simulation with a decreasing learning rate, starting at 1.0 was run. The mean success value can be found in Table 2.

A Repeated Measure ANOVA showed that there was no main effect of the different simulation on both the success ($F(5,25) = 387$; $p = 0.853$) or the amount of epochs ($F(5,24) = 1.396$; $p = 0.261$). Post-hoc pairwise comparison, using Bonferoni Correction, confirmed this, showing only non-significant differences between the different simulations.

Discussion In this paper a multi-agent system approach was taken to identify the effects of experience on learning during a group task, that is, riot control. Several simulation were run using different learning rate - this way simulating the age or amount of experience of the agents - after which the success of the riot control was determined. Results showed no significant differences between the different simulation, suggesting that experience is not a factor that affects learning and group performance.

Some shortcomings to this research method do deserve mentioning. First of all, only few simulations were run. Of the four different scenarios, only five simulations were run. Only few results could be obtained, which does not allow for much comparison and interpretations. Having more simulations might improve the statistical significance of the results.

Another issue encountered in this multi-agent system was the fact that the learned behavior was not transferred to new simulations. Because of the brevity of the individual simulations, agents were not capable of learning a lot, which might have explained the lack of differences in the results as well. Making the simulations longer and transferring the learned behavior to new simulations, might improve the agents' capability to learn and differences to arise.

The implication of the results found must be taken with a grain of salt. The simulation is a greatly oversimplified representation of reality. The hostiles only have one goal, which is to kill, whereas the agents only decision is between saving and shooting. Moreover, the agent will

only go away if every hostile is taken care of in his area. Also, civilians just disappeared from the city when they were saved. This is obviously not what really happens. Goh et al. [3] had improved the realism of his simulation by including many factors that may happen during a civil war. A combination of their simulation with the current proposed model may allow some transfer from the model to reality, and may therefore allow us to learn from the results.

4 Conclusion

The effects of learning rate of individual agents on group performance during civil war control was researched using a multi-agent system approach. Due to some shortcomings of the model that was used and the lack of results no actual lessons could be learned though. In future studies, many improvements should be made to the model before we can transfer these results to reality.

References

[1] John R Anderson and Richard King. *How can the human mind occur in the physical universe?* Oxford University Press, 2007.

[2] Caroline Claus and Craig Boutilier. The dynamics of reinforcement learning in cooperative multiagent systems. In *AAAI/IAAI*, pages 746–752, 1998.

[3] Chi Keong Goh, HY Quek, Kay Chen Tan, and Hussein A Abbass. Modeling civil violence: An evolutionary multi-agent, game theoretic approach. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 1624–1631. IEEE, 2006.

[4] Timothy R Gulden. Spatial and temporal patterns in civil violence: Guatemala, 1977-1986. *Politics and the Life Sciences*, pages 26–36, 2002.

[5] Roger B Myerson. Game theory: analysis of conflict. *Harvard University*, 1991.

[6] Shelley Nason and John E Laird. Soar-rl: Integrating reinforcement learning with soar. *Cognitive Systems Research*, 6(1):51–59, 2005.

[7] Ray Reagans, Linda Argote, and Daria Brooks. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management science*, 51(6):869–881, 2005.

[8] Nicholas Sambanis. Do ethnic and nonethnic civil wars have the same causes? a theoretical and empirical inquiry (part 1). *Journal of Conflict Resolution*, 45(3):259–282, 2001.

[9] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.