The influence of experience on group performance during civil violence: A Multi-Agent-System Approach.

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 on how civil violence may emerge, but there has been less focus on what factors affect cop performance to control civil violence. In this paper, the effect of experience of cops, modelled by individual learning rates, on the overall group performance is investigated. No significant results were found for the success, nor for speed of resolving the issue. Moreover, the overall mean utility of either shooting and saving in all situations did not differ in a large degree, all having a 50% chance of being selected. These results suggest that both saving and shooting is of equal importance, and that individual learning rates do not have a large impact on the success of the group. Because the simulation was rather simplistic, care should be taken when trying to transfer the implications to real life.

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 different times and places. There is no single cause for these riots and wars. Some were the result of cultural differences, some of the feeling of being treated unjustly, 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 addressed was how different events affected the individual tendency to riot and how this in turn affected macroscopic behavior and situations.

Goh et al. included many different interactions in their 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 experienced 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 occurrence 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.

This paper focusses on how experience influences group behavior. By using a simulation of a multi-agent system, we were able to compare how experienced and inexperienced cop agents perform against a group of aggressive hostile agents. A reinforcement learning strategy was used to let the individual agents learn the optimal strategy, which has been shown to be a robust and natural way of teaching agents [2]. In the following sections, the simulation and data acquisition will be described in more detail. Thereafter the results of the different simulation will be presented. The paper concludes with a discussion of the implications of the results, and some critical notes to this research.

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, while taking out the hostiles. The hostiles on the other hand have only one goal: to kill as many civilians and cops 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 cell reside on average 20 civilians, 10 hostiles and 10 cops. The agents can only take actions affecting other agents within their own cell. 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 threat of future killings, or he can choose to save a civilian, who is then safe from being killed. Hostiles only have one action: to kill. A hostile kills either a cop or a civilian depending on a given probability.

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 in relation to the amount of saved civilians and killed hostiles, for instance, a negative reward will be given, whilst when many hostiles were killed in relation to the amount of killed civilians and cops, 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 which 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 certain values, depending on what type of agent we want to simulate ('trigger happy' or 'peaceful and calm'). 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 cell and σ is a random value between 0.5 and 1.5 to make the decision stochastic, because it is assumed that the cops have no perfect knowledge of their environment. 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 a cop any more, that is, if there are no more hostiles left, it can move to one of (at most) four neighboring cells in the matrix. The cop will move to the cell where he is needed most. In other words, the cops will move to where there is the biggest discrepancy between the amount of agents and the sum of hostiles and civilians. This way, the cops can always search for a place where they 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 variable 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:

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

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 successful. According to the formula, whenever 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 global success value of each run and the amount of epochs of these runs. Moreover, the global 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 success value, 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, all of 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.

The cops with a lower learning rate should adjust there behavior less quickly compared to the cops with a higher learning rate. Over time, the cops with a lower learning rate (and the

cops with decreasing learning rate) should show more stable performance. However, it will take longer to adapt to the world and learn what strategy is best.

2.2.2 Initial biases

Another point of interest is how behavior evolves over time, when bias does exist. Can old dogs 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 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 useful. 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.

The same results are hypothesised for the biased simulations, as were for the no-bias simulation. That is, over time the lower learning rate will show more stable performance and the optimal level will take longer to achieve. However, a difference may show when looking at performance. If, for example, the cops are initialized with a bias for shooting, but the best strategy is to save everyone, then the quick learning may ensure an overall better performance. However, if the cops are initialized with shooting, and shooting is the best strategy, then lower learning rate cops will perform best.

3 Results

3.1 Effects of Learning Rate

The first simulations were run with learning rate values of 0.8, 0.5, 0.2 and 0.0, corresponding to the inexperienced, mixed-experienced, experienced and non-learning cops. Following, the simulation with a decreasing learning rate, starting at 0.8 was run. The mean performance and standard deviation (SD) can be found in Table 2.

λ	# epochs (SD)	success (SD)
0.0	17.79 (1.623)	0.25 (0.036)
0.2	17.66 (2.194)	0.25 (0.036)
0.5	18.12 (2.502)	0.25 (0.036)
0.8	18.45 (2.288)	0.25 (0.036)
<0.8	17.53 (1.825)	0.25 (0.036)

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

A Repeated Measure Analysis of Variance (RM-ANOVA) showed that there was a main effect on learning rate for the amount of epochs ($F(4,4995) = 31.204$, $p < 0.001$), but not for the success value ($F(4,4995) = 2.306$, $p = 0.054$). Post-hoc pairwise comparison, using Bonferoni correction, showed that the 0.0, 0.2 and <0.8 simulations were significantly faster compared to the 0.5 and 0.8 simulation. This suggests that a(n ultimately) lower learning rate and, therefore, more stable performance, ensure better performance. Note, however, that the difference is at most 1 epoch, meaning that the effect is not very large.

In Table 3 the mean amount of epochs and the mean success of the five conditions are shown for both the shoot- and save-biased cops. The first value that deserves some explanation is the speed of the shoot-biased cops that do not learn ($\lambda = 0.0$). In this situation there is 90%

λ	# epochs (SD)	success (SD)	λ	# epochs (SD)	success (SD)
0.0	350.58 (49.631)	0.30 (0.050)	0.0	9.32 (0.864)	0.15 (0.032)
0.2	19.89 (2.690)	0.39 (0.026)	0.2	13.41 (1.893)	0.17 (0.039)
0.5	20.64 (2.354)	0.37 (0.027)	0.5	14.57 (1.936)	0.17 (0.039)
0.8	20.64 (2.344)	0.35 (0.029)	0.8	14.90 (1.937)	0.17 (0.039)
<0.8	20.28 (2.145)	0.37 (0.027)	<0.8	11.55 (1.214)	0.16 (0.035)

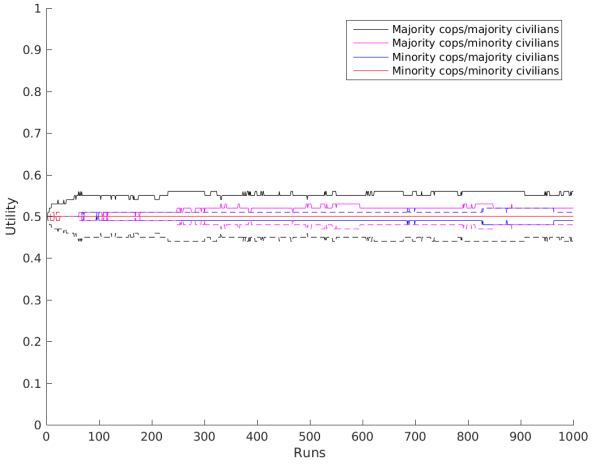
Table 3: The amount of epochs necessary to finish and the success of the simulation for the cop-teams with different levels of experience. The cops have a bias for shooting (left) and saving (right).

possibility that the cops will shoot in every step. Even when there are no more hostiles, the cops will keep shooting most of the time. Saving civilians will only occur once in ten times and will, therefore, take a very long time. In the statistical analysis, this scenario was excluded. For the shoot-biased cops a significant main effect was found on both speed ($F(3,3996) = 22.365$; $p < 0.001$) and success ($F(3,3996) = 21.587$; $p < 0.001$). Post-hoc analysis with Bonferoni correction showed that the 0.2 learning rate condition scored significantly better than the <0.8 condition on both performance measures. The <0.8 condition, in turn, scored significantly better than the 0.5 and 0.8 conditions, again suggesting that a(n ultimatly) lower learning rate and, therefore, variability, ensures a higher performance.

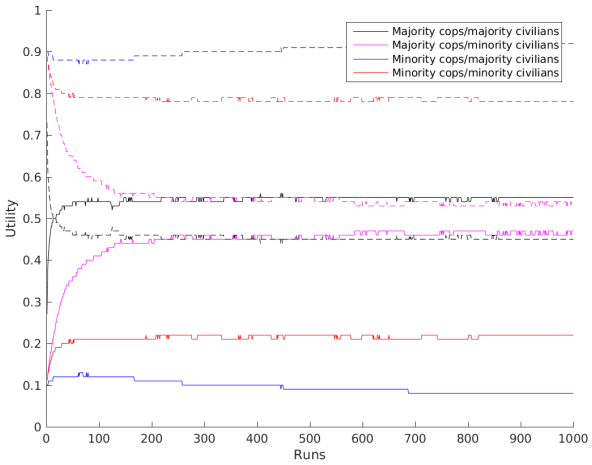
For the save-biased cops a main effect was found on speed ($F(4,4995) = 2025.756$; $p < 0.001$), but not on success ($F(4,4995) = 0.960$; $p = 0.428$). Post-hoc analyses showed that the the learning rate conditions have the following ordering in speed: 0.0, <0.8, 0.2, 0.5, 0.8, with only significant ($p<0.001$) differences. Again, less variable behavior is associated with higher speed.

3.2 Effects of Bias

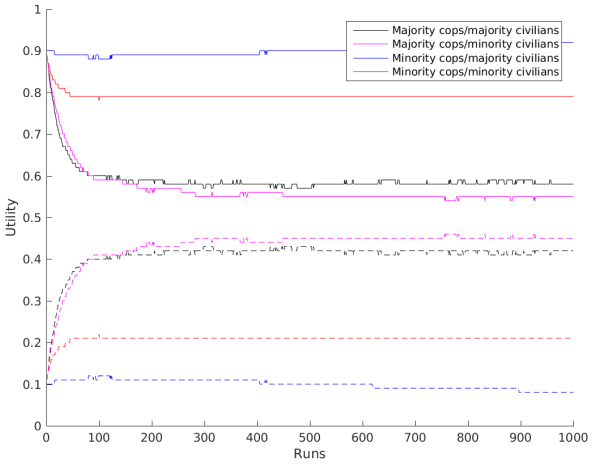
The results in Table 2 and Table 3 showed another interesting result. It appeared that there was an inverse relation between speed and success. Whereas the no-bias simulation shows intermediate results ($m_{epochs} = 17.91(SD = 2.136)$; $m_{success} = 0.23(SD = 0.075)$), the shoot-bias simulation shows a low speed ($m_{epochs} = 20.36(SD = 2.41)$) and high success ($m_{success} = 0.33(SD = 0.111)$, as opposed to the save bias ($m_{epochs} = 12.75(SD = 2.641)$; $m_{success} = 0.18(SD = 0.143)$). A RM-ANOVA showed that these differences were highly significant ($p < 0.001$). Note that the shoot-bias no learning cops were again excluded. These results suggest that a bias for an action has a large effect on the success of mitigating civil violence.



(a) Utility for no-biased cops



(b) Utility for shoot-biased cops



(c) Utility for save-biased cops

Figure 1: Convergence of utilities for all bias simulations, $\lambda = 0.5$. Dashed lines represent shooting, and solid lines saving.

In Figure 1, we show how the biases evolve over time with a medium learning rate ($\lambda = 0.5$). In this figure, the global mean utility for the different actions in the different situations is depicted for all three bias conditions. As is true for all other learning-cop groups, the utilities will converge towards a value of 0.5, though some bias will still exist. Individual agents do show preferences in particular situations, but averaged over all cops, a mixed strategy of saving and shooting appears to be the preferred approach. When the cops are outnumbered by hostiles (the red and blue lines), it does show little changes regardless of the bias. This may indicate that little is learned because the hostiles kill the cops rather quickly.

4 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 simulations were run using different learning rates - this way simulating the age or amount of experience of the agents - and different biases towards an action. Results showed some significant differences in performance, though the actual effect was rather small. Different biases did show some clear differences in performance, showing an inverse relationship between speed and success given the three actions. Moreover, a preference for actions do indeed affect performance in both speed and success.

The fact that the differences in performance between different learning rates was so small, suggests that the group performance of cops during mitigating civil wars does not depend on experience in a large degree. The initialized biases, however, did result in big changes in performances. Whereas a bias for saving greatly sped up the mitigation (at the cost of success), shooting greatly increased success (at the cost of speed). Instead of the composition of the team with regard to experience, these results indicate that the focus should lie on strategy selection, based on the goal at hand.

Note that the simulation was an extremely simplified approximation of reality. With hostiles only able to kill, and the inability for hostiles and civilians to walk through the city, the simulation can clearly be considered a low-fidelity simulation. Moreover, the cops only have one decision to make, and will only walk when all hostiles are gone in his area.

This simulation does provide potential for future researches. In this paper, the comparison was made between saving and shooting. By implementing more realistic behavior, it is possible to compare the results of different strategies or combinations thereof. Moreover, new strategies may be designed and tested by use of a multi-agent system. Showing what strategies show the most potential and how individual cops affect these strategies could ultimately result in better mitigation of civil wars and better training programs. Finally making the agents, that is, the cops, hostiles and civilians, more intelligent, and the matrix more city like would greatly improve the informativeness of the results.

5 Conclusion

The effects of learning rate of individual agents on group performance during the mitigation of civil wars, together with the effects of bias for particular strategies, was researched using a multi-agent system approach. Though only minor effects of learning rate were found, a bias for strategy did affect performance drastically. Some improvements to the simulation are necessary to allow the result to be translated to reality, but, ultimately, an agent-based simulation may be of great importance for understanding the individual agents' influence on performance, and for the assessment of strategy success.

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.