The influence of individual 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 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 on overal group performance is investigated using a multi-agent simulation. Cops that have a lot of experience tend to stick to known and successful behavior (low learning rate), wheras new cops, having less experiences, are more likely to switch behavior after unexpected unfavourable outcomes (high learning rate), which may affect group dynamics. Some significant differences were found between learning rates; these were marginal though. A large effect was seen between different biases for strategies. An aggressive approach resulted in less casualties at the cost of mitigation speed, whereas a more peaceful approach increased the speed of mitigating civil violence at the cost of many cop and civilian casualties. Though the current simulation was rather simplistic, the results do show that multi-agent modeling is a useful tool for understanding group dynamics, and may provide us with the possibility to test and compare civil violence mitigations strategies.

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 Ukraine; 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 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 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.

A natural hypothesis is that many previous experiences ensures 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". We plan to test this hypothesis using a multi-agent simulation.

1.3 New Idea

Knowing how experience influences group behavior may be of vital importance for the success of a group's 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 favorable 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 affects the successful development of an organization [7]. Individual experience may thus have a great impact on group performance. However, this has not received much attention in a civil violence context yet.

This paper focuses 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 hostiles. Instead of the evolutionary algorithm, as used by Goh et al., a reinforcement learning strategy was used to let the individual agents learn, which has been shown to be a robust and natural way of teaching agents [2]. To be able to interpret the results more easily, all the mechanisms from the Goh et al. simulation were omitted in this simulation. Instead, a more simplified implementation was used. 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. The hostiles on the other hand have only one goal: to cause as much mayhem and despair as possible by hurting civilians. Civilians have no particular functions, but are subject to the actions taken by cops and hostiles. The simulation takes place in a 20x20 2D matrix. In every cell reside a number of civilians, hostiles, and cops. The number of civilians and hostiles is selected randomly from a normal distribution with a mean of 10 and standard deviation of 5 for both the civilians and the hostiles. A fixed number of cops (4000) is distributed randomly over the entire grid, thus on average every cell also contains 10 cops.

This initalization ensures that, on average, the amount of cops, hostiles, and neutrals is similar in every simulation, while also adding some variation in the individual cells.

2.1.1 Goals and Actions

The goal of the cops is to keep the number of casualties as low as possible. To achieve this, every cop must choose one of two actions. A cop can shoot a hostile, which has a chance to incapacitate a hostile. The second action is to to save a civilian, who is then safe from getting hurt, but this does not incapacitate any hostiles, so the threat persists. Hostiles only have one action: to hurt. Hostiles shoot with a 50/50 chance on either a cop or a civilian. An "aim" parameter ensures that every shot has a chance of being missed, for our simulation we set the aim at a 25% chance of actually hitting the desired target, and a 75% chance of hitting nothing.

		Shoot	Save
# Cops > # Hostiles	# Civilians > # Hostiles	Ua	Ub
	# Civilians < # Hostiles	Uc	Ud
# Cops < # Hostiles	# Civilians > # Hostiles	Ue	Uf
	# Civilians < # Hostiles	Ug	Uh

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

All the cops make the decision to either save or shoot once in every epoch. Based on the overall result in a cell, a reward is given to the individual agent. This reward can be either positive or negative. If, for instance, many civilians die a negative reward will be given, whilst when many hostiles were incapacitated, 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 different circumstances are based on the proportion of number of cops versus number of hostiles (Cops are in a majority or minority), and the proportion of number of civilians versus number of hostiles. This effectively produces 4 different situations, which combined with the two possible actions give eight different utility values ($U_a..U_h$).

The cops determine their desired action by selecting the action corresponding to the highest utility value given the current situation in their cell. Because the agents are not perfect, some noise is introduced in the selection of the highest value. Mistakes can thus be made, which increases the realism of the simulation.

The reward of an action can be calculated as follows:

$$R(s,a) = \frac{Incapacitations(s) + Saves(s') - Losses(s')}{Incapacitations(s') + Saves(s') + Losses(s')} \tag{1}$$

in which R is the reward of action a in situation s, $Incapacitations$ are the amount of incapacitated hostiles in the new situation s', $Saves$ the amount of saved civilians, and $Losses$ the amount of incapacitated cops and civilians. Based on these calculations the reward will lie between -1 (only cops incapacitated) and 1 (only hostiles incapacitated and neutrals saved). This translates to an outcome where only cops are incapacitated as being as bad as possible, the situation where civilians are saved and no cops or civilians hurt as being the best possible outcome, and every situation in between is rewarded accordingly. This corresponds to real

life where the objective of a cop should be to save as much civilians as possible, while also neutralizing any threats.

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

$$\Omega = \frac{n_{cops}}{n_{cops} + n_{hostiles}} * \sigma \tag{2}$$

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, as it is assumed that the cops have no perfect knowledge of their environment. If Ω is higher than 0.5, the cops assume they are in majority. The same function is used to assess the amount of civilians versus hostiles; The number of cops is then replaced by the number of civilians.

2.1.2 Movement

When there are no more civilians left in the a cell, no more agents are needed there. Cops who are still present on the said cell will then move to a neighboring cell where they are needed most. That is, the cops will move (with a noise factor to ensure some random movement) 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.

2.1.3 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.

In the current simulation, a similar approach will be taken for updating our utility. However, no calculations that involve Q-values are used, as our simulation poses a highly dynamic environment that does not allow for convergence on true Q-values, as these values change per epoch and per simulation. The learning rate will be updated according to the following function:

$$\lambda_{new} = \lambda_{old} * \alpha_{\lambda} \tag{3}$$

in which α_{λ} is the update factor that is applied to the learning rate each epoch.

Subsequently, the utility will be updated according to the following function:

$$U_{x\ new} = U_{x\ old} + \frac{\lambda * R(s,a)}{1+\lambda} \tag{4}$$

in which $U_{x\ new}$ is the new utility of the cell's situation and calculated desired action combination as shown in Table 1, λ is the learning rate as calculated in Equation 3, and $R(t)$ is the reward resulting from the actions taken in the cell, as calculated with Equation 1.

Finally, to ensure that for every situation the sum of the actions is equal to one (allowing for easier calculations), the action utilities are normalized per each situation. In a way, the success of one action will then 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

2.1.4 Pseudocode

In the following code-block a simplified overview of the actions taken per simulation is shown.

```
1 Initialize the grid
2 Do
3   Determine each agent's action
4   Calculate the overall performance in each cell
5   Adjust each cell's properties (incapacitations, saves, ...)
6   Update each agent's Utility matrix, according to their
      chosen action and the overall cell success
7   Move agents(If necessary)
8 While there are still civilians present in the grid
```

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 $\lambda = 0.8$, which will be decreased by using a factor $\alpha_\lambda = 0.95$ 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 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.

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 Bonferroni 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 incapacitate- and save-biased cops. The first value that deserves some explanation is the speed of the incapacitate-biased cops that do not learn ($\lambda = 0.0$). In this situation there is 90% possibility that the cops will incapacitate a hostile in every step. Even when there are no more hostiles left, the cops will keep trying to incapacitate a hostile most of the time. Saving civilians will only occur once every ten times and will, therefore, take a very long time. In the statistical analysis, this scenario was excluded.

For the incapacitate-biased cops a significant main effect was found on speed ($F(3,3996) = 22.365$; $p < 0.001$) and success ($F(3,3996) = 21.587$; $p < 0.001$). Post-hoc analysis with Bonferroni 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 signif-

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)

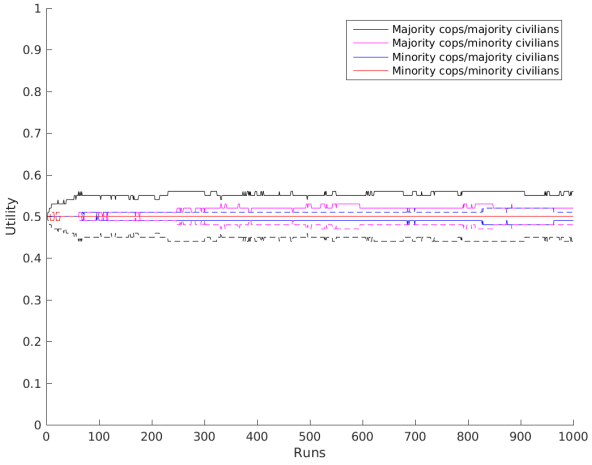
Table 3: The amount of epochs necessary to finish and the success of the simulation for the cop-teams with different levels of experience. The cops have a bias for incapacitating (left) and saving (right).

icantly better than the 0.5 and 0.8 conditions, again suggesting that in the end an ultimately lower learning rate ensures a higher and more stable performance.

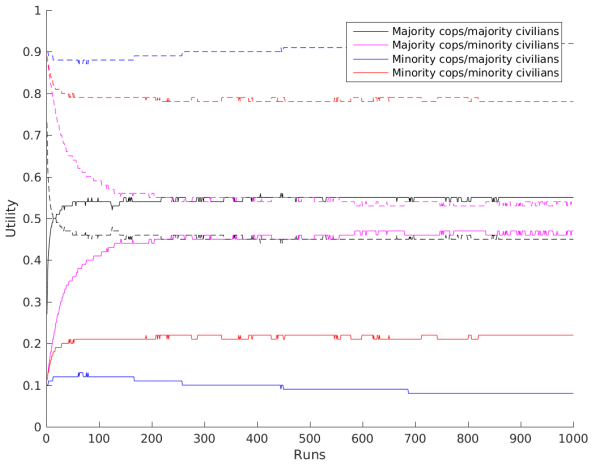
For the save-biased cops a main effect was found on speed (F(4,4995) = 2025.756; p < 0.001), but not on success (F(4,4995) = 0.960; p = 0.428). Post-hoc analyses showed that the the learning rate conditions have the following ordering in speed: 0.0, <0.8, 0.2, 0.5, 0.8, with only significant (p<0.001)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

3.2 Effects of Bias

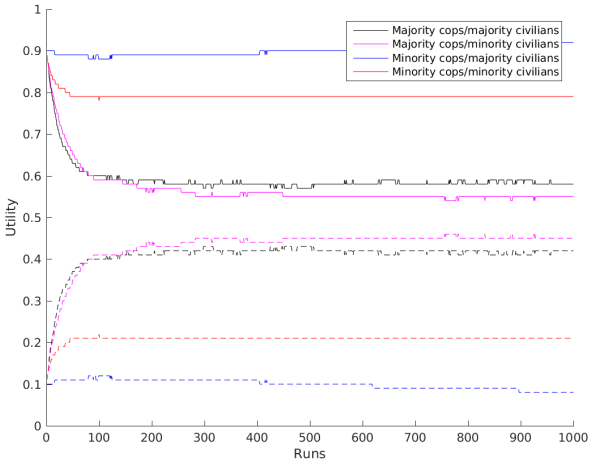
The results in Table 2 and Table 3 showed another interesting pattern. It appeared that there was an inverse relation between speed and success given the two actions. Whereas the no-bias simulation shows intermediate results ($m_{epochs} = 17.91$; $SD_{epochs} = 2.136$; $m_{success} = 0.23$; $SD_{success} = 0.075$), the incapacitate-bias simulation shows a low speed ($m_{epochs} = 20.36$; $SD_{epochs} = 2.41$) and high success ($m_{success} = 0.33$; $SD = 0.111$), as opposed to the save bias ($m_{epochs} = 12.75$; $SD_{epochs} = 2.641$; $m_{success} = 0.18$; $SD_{succes} = 0.143$). An RM-ANOVA showed that these differences were highly significant (p < 0.001). Note that for the incapacitate-bias, the 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 take out 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 amount of experience of the agents - and different biases towards an action. Results showed some significant differences in performance, though the effects were marginal. Different biases did show some clear differences in performance, showing an inverse relationship between speed and success given the two 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 were so small, suggests that the group performance of cops during mitigating civil wars does not depend on personal experience in a large degree. This can be explained by a number of factors: (1) the amount of agents, (2) the uniformity of the map and (3) randomness of agents' perception. First, averaged over the 4000 cops, success of one's preference will be canceled out by the similar success of opposing behavior. Because the cops learn based on the group success per cell, oppossing behavior in one cell can get the same reinforcement, which ultimately leads to random global behavior. Second, because the map is always completely uniform in both its structure (symmetric, identical grids) and population, agents will often . Finally

The combination of these factors caused the effect of individual learning to be negligible.

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, incapacitating hostiles greatly increased success at the cost of speed. This can be explained by the fact that the stop criterion in the simulation was set at no more civilians left in the simulation. If cops were heavily biased towards saving them, civilians would both be saved more quickly and be killed more quickly, because few hostiles would get incapacitated. This would result in a quick mitigation, but a relatively low success. On the other hand if cops were heavily biased towards incapacitating, hostiles would be removed from the simulation quickly, resulting in less civilians and cops being killed. This would result in a high success rate but an overall low speed, since it takes the cops longer to save all the civilians.

These results indicate that, instead of the composition of the team with regard to experience, 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 hurt, 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. Making the agents, that is, the cops, hostiles and civilians, more intelligent, and the matrix more city like, results can be more meaningful in the real world.

This simulation does provide potential for future researches. In this paper, the comparison was made between saving and incapacitating. 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 simulation. 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.

5 Conclusion

The effects of learning rate of individual agents on group performance during the mitigation of civil violence, 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 the real world. However, ultimately, an agent-based simulation may be of great use 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.