Analysis of the bullying behaviour through a game theory approach

Nicolas Lazzari (979086) - nicolas.lazzari2@studio.unibo.com

University of Bologna

Abstract. In this document a computational model on the bullying phenomenon is proposed. bullying is often seen as a "group process" and many researchers and policymakers share the belief that interventions against bullying should be targeted at the peer-group level rather than at individual bullies and victims. The aim of the model is that of showing how different approaches can effectively change the decision process of the peer-group and in particular that of the bully.

1 Bullying

Bullying is a subtype of aggressive behavior, in which an individual or a group of individuals repeatedly attacks, humiliates, and/or excludes a relatively powerless person[3]. While it's mainly present in primary and middle schools, bullying circumstances can also be identified on other environments such as prisons or workplaces.

Among several different countries the prevalence of bullies in schools is 11% while another 11% is the amount of victims being picked up. A of today, group is seen as a group-process where the group members are seen as having different roles in the process, driven by diverse emotions, attitudes, and motivations.

Many researchers and policymakers share the belief that interventions against bullying should be targeted at the peer-group level rather than at individual bullies and victims. If fewer children rewarded and reinforced the bully, and if the group refused to assign high status for those who bully, an important reward for bullying others would be lost.

1.1 Bullies behaviour and motivations

Different theories have been proposed trying to explain the behaviour and motivations of bullies. The most recent and accredited one is the one in which bullying actions are carried in a pursuit of dominance within the peer group. Children differs in the relative importance they attach to communal goals, such as making friends, being visible, being influential etc.

A child's probability of being involved in a bullying relationship as a bully is related to a high degree of status goals. Bullies perceive themselves as dominant and have an high ideal concerning dominance. Moreover they believe that other

peers expect them to be dominant, thus reinforcing the need to assert and show dominance.

This results in bullies being highly selective on their victims, which are usually individuals in a low-power position within the group, and attacks being carried mostly when other peers are present, namely $85 - 88\%$ of the time.

1.2 Peer group involvement in bullying

Peers presence during bullying episodes give rise to different social roles during those episodes. Olweus[2] has described the "bullying circle" in which eight different modes of reaction represent the combinations of children's attitudes to bullying (positive–neutral–indifferent–negative) and behaviors (acting vs. not acting).

In the proposed model we will make use of three different roles that children might take in these situations **bully**, **victim** and **bystander**.

Bully behaviour As previously explained, bullies seeks for approval and dominance over other children by means of harassment. Each time a bully approaches a victim its actions essentially boils down to deciding whether to bully the victim or not. These decisions depends from many different environmental and cultural factors. In the proposed model bully's decision will depend only on how the peer-group is composed, i.e. the number of children defending the victim, supporting the bully or just witnessing the act.

Bystander behaviour Bystanders, which account for most of the peer-group composition, are those individual that are not bully nor victims but can possibly interact with both in a bullying event.

In this model we will consider only three possible interactions that bystander can take when they approach a bullying situation:

- **support** the bully, e.g. by either taking an active part or just reinforcing him/her
- **defend** the victim
- **witness** the event without taking a stand in either direction

Even though most children attitude is against bullying actual defending behaviour is rare. Defenders of the victims are usually around $17 - 20\%$ while those who take an active role in supporting the bully are $20 - 29\%$ and $26 - 30\%$ don't take any part.

2 Model definition

The proposed model is developed using an evolutionary game theory approach. Three different agents, each with its own strategies, populate and interact within the environment: *victims*, *bullies* and *bystanders*. The model is deeply inspired on

the work presented in [1]. Two different games are played by bystanders against bullies.

In the first game bystanders needs to decide, in a bullying event, whether to stand up against the bully and help the victim or rather stay silent. The second game is similar to the first one but is only played in those situations in which there is at least one bully whose decision in the first game was to bully the victim. Bystanders will again decide whether to intervene by parting with the bully or with the victim.

Both games can be seen as a variation of the prisoner's dilemma game: agents have very different interests in carrying they're own decisions and the collaboration with other agent is indeed fundamental in deciding which behaviour is most profitable. We will analyze what these interests might be and try to see how the model reflects these variations.

2.1 Environment population

The number of players is defined by the user through the parameter *population*. The population composition can be changed by the parameters *victims-perc*, *bystanders-perc* and *bullies-perc*, which represents respectively the percentage of victims, bystanders and bullies. Percentages correctness (i.e. sum up to 100) is checked at runtime.

Victims and bystanders are placed randomly across the map at each time-step to simulate random social interactions between different agents. Bullies are randomly placed around victims to make sure that they can interact with both victims and bystanders in their range of action. This type of restrictions allow us to avoid modeling those complex situations in which bullying might be carried without the victim itself being present and rather focus on the interaction between children.

2.2 Interactions

Agents interacts locally with other agents whose distance is within a radius defined by the *communication-radius* parameter. This allows us to model the environment in such a way that only individuals nearby the victim and the bully will witness the fact. Global communication can be enabled by setting the *communication-radius* parameter to its maximum value.

Each agent will interact with other agents according to its strategy. Strategies distribution is set by the user for both bystanders and bullies to allow real-world scenario modeling.

2.3 Agent strategies

Each agent can choose from a given set of strategies: bullies can decide whether to *bully* or *don't* in the first game and whether to *keep bullying* or *back down* on the second one. Bystanders can *defend* or *witness* in the first game and *defend*

or *support* in the second one. Victims acts more as dummy agents to attract bullies and thus don't have any viable strategy.

In the model we will interpret each agent payoff as a mixture of relevance within the peer-group and personal value they have. In [1] a utility function is defined for both games and agents in order to better interpret and propose values for the payoff-matrix.

For game 1 bullies utility function is defined as

$$U^i_{bully}(B) = \beta_0 S(B, R) + \beta_1 P(B)$$

where the function S is the profit in term of group importance that a particular bully would gain by bullying, which depends on the bully itself (B) and the group response (R). Group response represents the bullies' expectation on how bystanders would play on the second game. Function P represents how much power the bully would gain from its action and it's a subjective variable, different in each individual.

Bystanders utility function is defined as

$$U^i_{bystander}(B') = \beta_2 V(B') + \beta_3 S(B', R) + \beta_4 C(B', \gamma)$$

where the function B' represents the bystander. Function S is the same as the one used by the bull. Function V represents the personal value that a bystander have i.e. the attitude towards confronting or joining the bully. Function C is the confrontation cost that a bystander gets when playing against a bully; γ represents the bully's personality, which may be an important factor in quantifying how hard the confrontation would end up being.

$\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 are weighting variables to model particular behaviour children.

In a similar way we can define utility functions $U^{ii}_{bully}(B)$ and $U^{ii}_{bystander}(B')$ for game 2.

We won't make direct use of this utility functions in the model but they represents a useful tool to gain insight and understating how the payoff matrix is structured and which kind of values should populate it.

2.4 Game

Each agent plays with other agents in its communication range and updates its payoff values according to a payoff matrix.

Game 1 payoff Game 1 payoff matrix is defined on table 1. All values will be relative to a baseline payoff b. Given the previous interpretation of personal payoff and utility functions it's easy to understand the values in the payoff matrix. If the bully decides not to bully the victim then the bullying event doesn't happen and both agents gains b value. If the bully decides to harass the victim instead then its value gain depend entirely on the bystander behaviour. If the bystander decide to witness the event than the bully is going to gain

importance, since its display of power lead an individual to not take any stand, hence gaining $2b$. The witness will suffer from this decision, since it will be in a low-power position within the group, hence gaining $f_{nts}b$. f_{nts} is used to quantify the factor of loss (or gain) for not taking a stand. In the other hand if the bystander decides to defend the victim then the bully won't be gaining any value and the defender will end up gaining $f_{ts}b$.

	Witness	Defend
Bully	$2b$, $f_{nts}b$	0, $f_{ts}b$
Don't bully	b, b	b, b

Table 1: Game 1 payoff matrix

Game 2 payoff Game 2 payoff matrix is defined on table 2. Much like the payoff matrix in table 1 all values will be relative to a baseline payoff b. Game 2 is only played by those bullies that decided to bully and encountered at least one bystander that decided to defend the victim. Note that those situations in which a bystander decided to stand against the bully but later decides to support it are allowed in the model. While this may seem contradictory at first, we recall that we include a wide spectrum of behaviours in the *support* strategy. For instance it may happen that only one bystander stands against the bully and, when confronted by the bully, the fear of turning into a victim brings the defender to back down and consequently indirectly supporting the bully.

Indeed when a bully receives support its payoff gain is three times the baseline, the highest in all the games, however if the bully is confronted by a defender its high-power position is mined with a payoff loss of $-b$. Meanwhile bystanders gain (or lose) $f_{sb}b$ and $f_{idb}b$ respectively if they try to support or defend the victim but the bully keeps bullying. If the bully stops bullying and retracts its decision then there is a clear gain for each defender, $2b$. Each supporter will regret the decision, $f_{sdb}b$, as they could've helped other defenders. In fact there has been studies[3] in which it was shown that many children are more inclined in to helping victims but they prefer to *mind their business* avoiding any chance of turning into a victim. On the other hand backing down is a dangerous decision a bully can take: while it doesn't really impact relationships with supporters, 0 payoff, it clearly puts the defender in an powerful position, $2b$, and the bully in an uncomfortable situation ($-2b$).

	Support	Defend
Keep bullying	$3b$, $f_{sb}b$	$-b$, $f_{idb}b$
Back down	0, $f_{sdb}b$	$-2b$, $2b$

Table 2: Game 1 payoff matrix

2.5 Strategies evolution

Bullies and bystanders change strategies according to the *imitate the better realization* protocol. After each game turn each agent will copy the strategy of the agent of the same kind with highest payoff. With a probability value of p_r (*prob-revision* parameter) each agent can choose a random strategy among the ones available.

3 Model experiments

We are mostly interested in how the distribution of strategies changes across different configurations. For this very reason we won't take into account the payoff value of each agent. In the following section we'll see how well the model performs and what kind of insights we can get from its execution. Each experiment will be run for 500 steps with a population of 300 individuals (11% victims, 11% bullies and 78% bystanders) with *communication-range* $= 3$ and $b = 5$, if not explicitly stated.

Equal distribution If we start the model by setting every factor to 0 and equally distribute bystanders strategies and bullies strategies we obtain the payoff matrix in table **??** for game 1. Nash Equilibrium is in the highlighted cell and in fact, if we turn off game 1 in the model and let the simulation run we end up with the configuration in figure **??**. Bullies constantly decides to bully while bystanders distribution highly depends on the initial position of bystanders and their random movement, which brought witnessers to outnumber defenders. On figure 2 we see the result of the simulation with added noise ($p_r = 0.0067$). We can clearly see how Nash equilibrium is still obtained: whenever defenders outnumber witnessers for a consistent amount of time, e.g. after ≈ 320 steps, bullies quickly adapt. Recall that all agents communicate at all time, whenever a defender spreads its idea with other bystanders they might change their strategy if it seems more convenient. Bullies however keeps bullying

	W	D
B	10, 0	0, 0
DB	5, 5	5, 5

Table 3: Game 1 payoff matrix

until it's no longer sustainable: we can see that spikes in defender distribution doesn't always turn into lower bullying action.

A similar behaviour can be seen if we allow global communication across agent, as can be seen in figure 3. The main difference lies in how fast information spread across individuals. We see that, as long as defending victims is the mainstream behaviour, bullies sharply adapt to stop their behaviour. Since defending or witnessing doesn't really make a difference from a bystander's point of view if the bully doesn't pursue its bullying actions, bystander's behaviour sharply changes from defenders to witnessers. As soon as one bully randomly decides to start bullying again and coincidentally its payoff is the highest among all bullies then the behaviour of bullies sharply changes again until the end of the simulation.

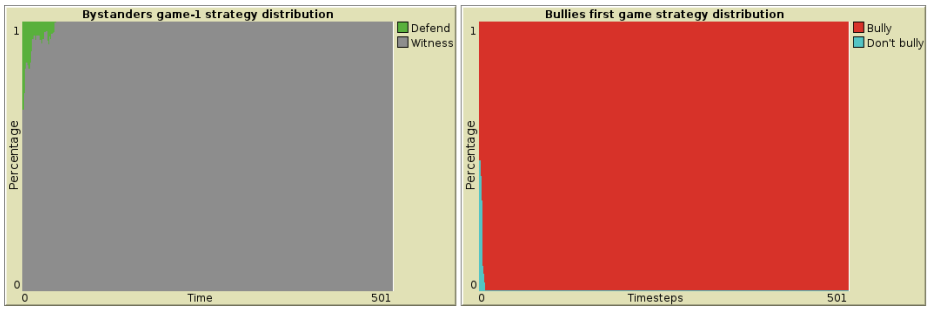

Fig. 1: Game 1 equal distribution

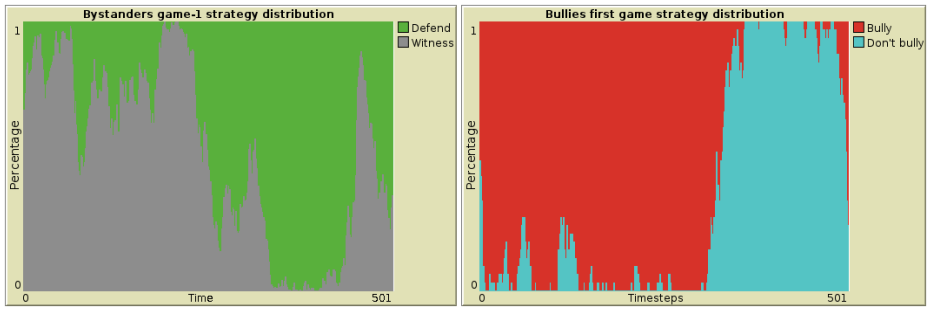

Fig. 2: Game 1 equal distribution with added noise

If we also take into consideration game 2 with the described configuration we obtain the payoff matrix in table 4 where the highlighted cells corresponds to equilibrium. Without any noise bullies are going to choose the bullying. Therefore, the bullies are going to play

	S	D
KB	15, 0	−5, 0
BD	0, 0	−10, 10

Table 4: Game 2 payoff matrix

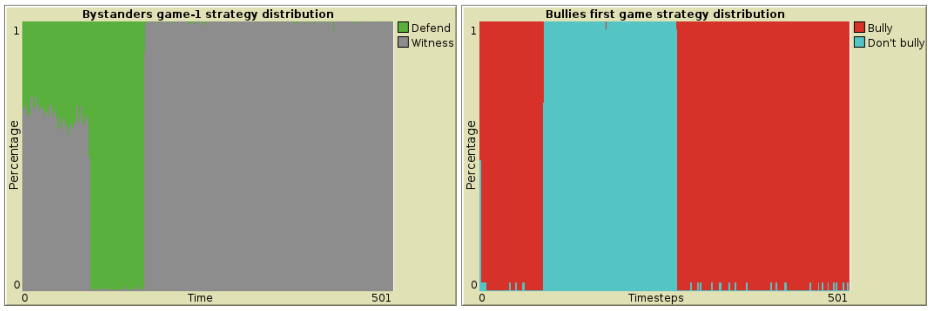

Fig. 3: Game 1 equal distribution with global communication and added noise

the second game and, given the equilibrium states, bullies are going to bully no matter what strategy bystanders decide to take. Indeed we can see this kind of behaviour happening on figure 4. Bystanders lean towards the defending the victim however, given the payoff matrix and its equilibrium points, bullies are always going to bully.

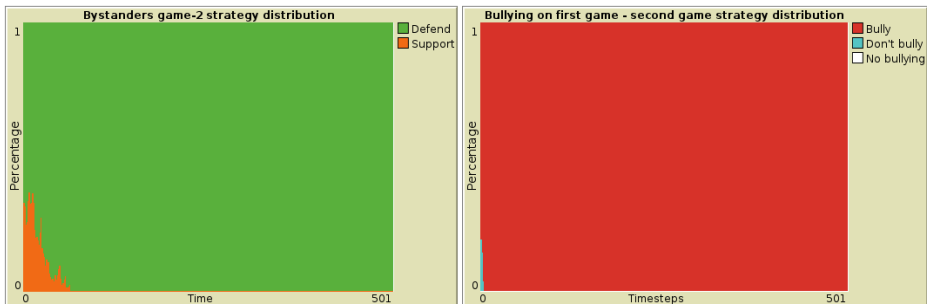

Fig. 4: Game 2 equal distribution

If we add noise in the system, as in figure 5, an interesting behaviour can be observed. Defenders, much like bullies, play both games and their payoff changes faster and differently than before. Bystanders initially decide to take a stand and defend the victim while bullies, given the equilibrium states seen before, stops bullying to avoid losing huge amount of payoff. This kind of behaviour wildly influences the second game in which only active bullies are allowed to play. From time-step ≈ 45 to time-step ≈ 325 bullies back down when confronted by bystanders. While this kind of behaviour seems to contradict equilibrium states in matrix 4 random revision probability is most likely the cause of this kind of behaviour. Indeed we can see that around ≈ 330 bullies start consistently bullying again in the first game and this results in a consistent behaviour also in the second game, where bullies start confronting bystanders again.

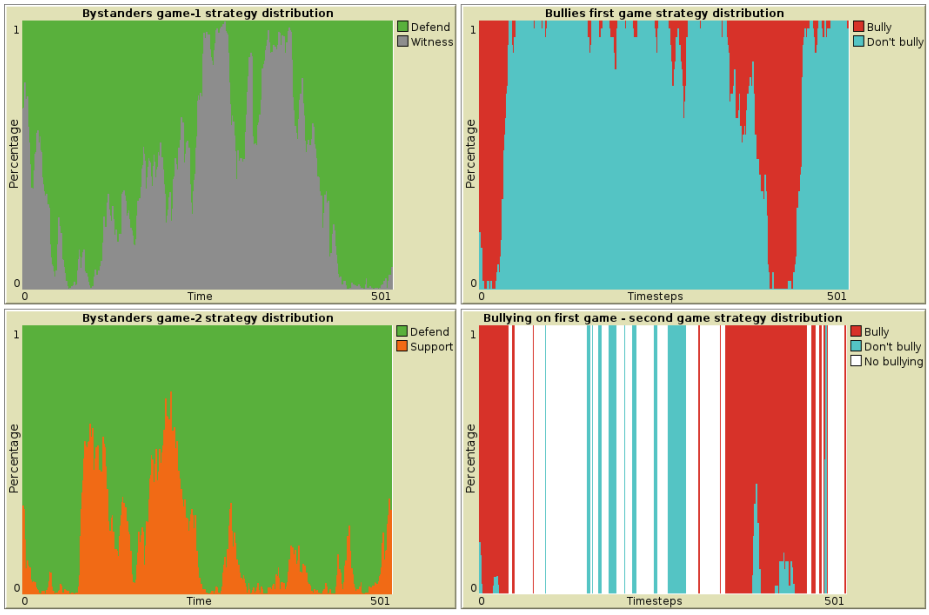

Fig. 5: Game 1 and 2 equal distribution with added noise

If we allow global communication, figure 6, then bullies mostly decides to maintain a bullying behaviour both in game 1 and game 2. This can most definitely be re-conducted to the faster communication between bullies, which don't incur into the problem of meeting only *"well-behaving"* bullies.

The bully and the defenders While the configuration discussed on the previous section is interesting from a modeling point of view it badly represents the real world. To better reflect the real world environment we will be using a different distribution of strategies, namely 17% of defenders, 29% of supporters and 54% of witnessers[3] while we'll set 100% of bullies to bullying strategy in both games.

We will change payoff factors for game 1 to better resemble real world situations[1]:

- $f_{nts} = -\frac{b}{2}$ since, as already mentioned, children are mostly against bullying behaviour from a personal value point of view. It's also easy to imagine how, without taking a stand against the bully, the difference in power within the peer-group rapidly changes and the witness could turn into a victim
- $f_{ts} = \frac{8}{5}$ to reflect the fact that, by taking a stand, bystanders will have chances to get higher personal and group value

.

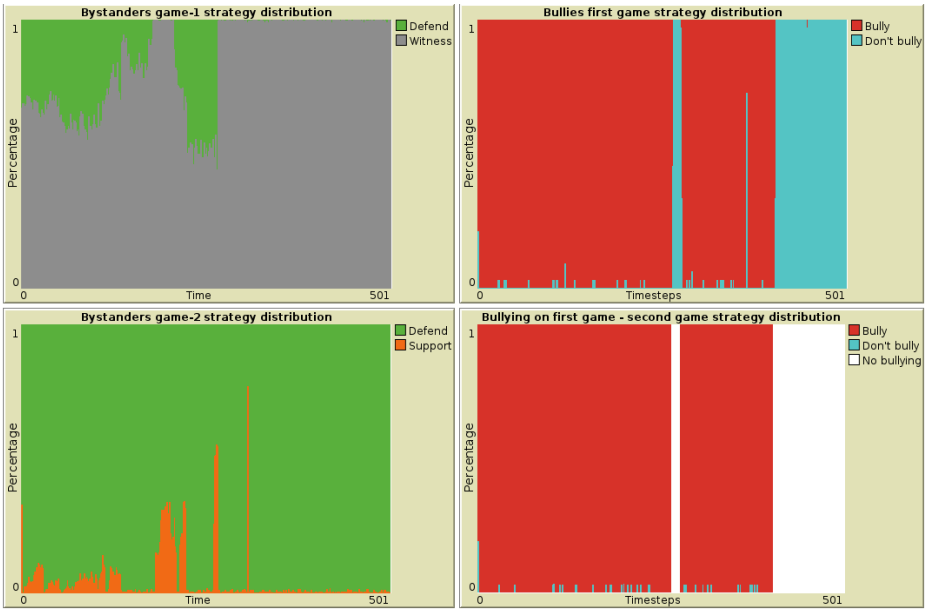

Fig. 6: Game 1 and 2 equal distribution with added noise and global communication

The obtained payoff matrix, table 5, better resembles a real-world scenario. Nash equilibrium lies on the states where bystanders ends up defending the victim and bullies decides not to bully. On figure 7 we can see that, indeed, this is achieved by the model. However it is quite unlikely that bystanders always stand up against bullies. This is a result of the f_{nts} factor being much higher than f_{ts}. We would prefer to end up on a situation in which bystanders prefer not to defend the victim but rather witness the event.

	W	D
B	10, −2.5	0, 8
DB	5, 5	5, 5

Table 5: Game 1 payoff matrix

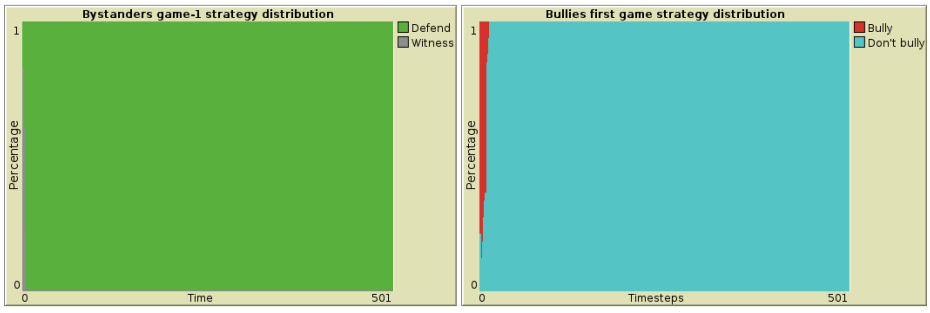

Fig. 7: Game 1 distribution

This can be achieved by obtaining a lower value in the *(B, D)* state in table 5. We can manage to so by setting $f_{ts} = -1$, thus obtaining the matrix in table 6. We can now see that nash equilibrium falls in the *(W, B)* state as we also notice from the model on figure 8.

	W	D
B	10, −2.5	0, −5
DB	5, 5	5, 5

Table 6: Game 1 payoff matrix

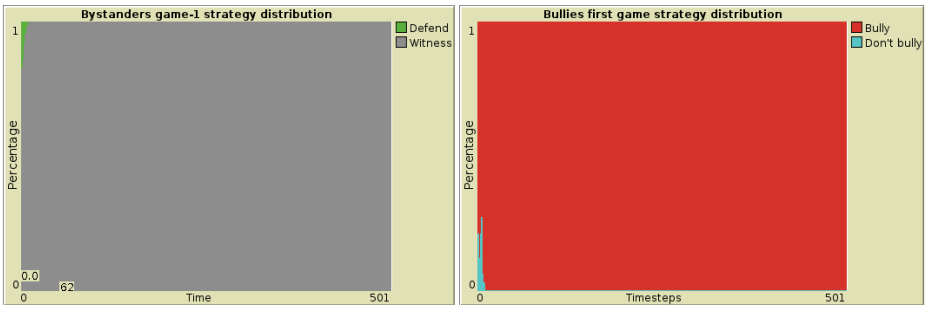

Fig. 8: Game 1 distribution

On figure 9 we can see that, even with added noise, the system remains in a quite stable configuration.

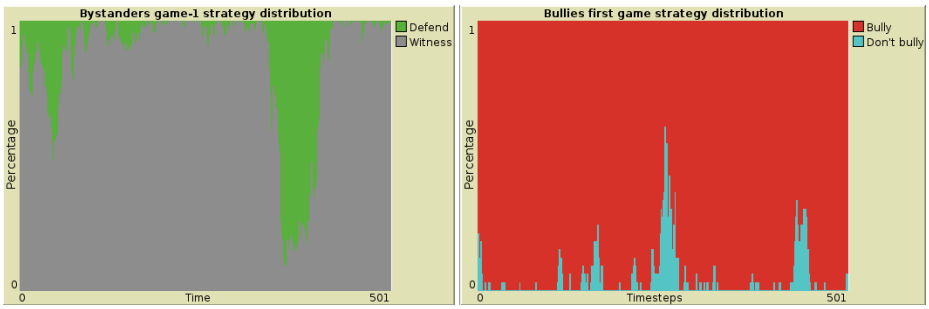

Fig. 9: Game 1 distribution with noise

Confronting the bully On the previous section we argued that it's not easy for bystanders to stand against the bully if we imagine a situations in which bystanders fear the bully and the possibility of being a victim too. However it may happen in particular circumstances that some individuals are always going to defend the victim. On table 7 and 8 we see, respectively for game 1 and game 2, the payoff matrices obtained by using the same configuration while letting individuals play the second game. Nash equilibrium can be found on those situation in which bystanders witness. Since we left $f_{sdb} = f_{sb} = f_{idb} = 0$

bystanders decision won't make any difference in the bully behaviour. In fact, as we can see from figure 10 bullies always ends up bullying the victim. In figure 11 we can see that, with $p_r = 0.0025$, the system still remains in the stable described state.

	W	D
B	10, −2.5	0, −5
DB	5, 5	5, 5

Table 7: Game 1 payoff matrix

	S	D
KB	15, 0	−5, 0
BD	0, 0	−10, 10

Table 8: Game 2 payoff matrix

Fighting the bully with intervention Given the previous situation we can now imagine that some interventions are taken and bystanders are instructed that, even if it doesn't always pay out, trying to defend a victim is the right thing to do. We can obtain that by setting

- $f_{sdb} = -\frac{b}{2}$ to quantify the regret in supporting the bully when it decides to back up. This value ensures that, whenever a substantial amount of defenders is present, bystanders are going to regret not joining them.
- $f_{sb} = -\frac{b}{2}$ to quantify the fact that, whenever a bystander supports a bully, it's going to lose some power from the peer group, even by doing so the risk of becoming a victim is lower.
- $f_{idb} = b$ to quantify that, even if taking a defensive strategy didn't stop the behaviour, that's the right thing to do
- $f_{ts} = -\frac{b}{5}$ to void penalizing defenders in the game, since taking a stand but not stopping the bully is not a wrong behaviour and is not seen as so by the peer-group

Indeed if we run the simulation with the described parameters, we end up exactly with the situation we'd like: after ≈ 130 time-steps the individuals starts defending the victim and bystanders that decides to take the bullies side are not influential enough to encourage bullies in starting to bully individuals again.

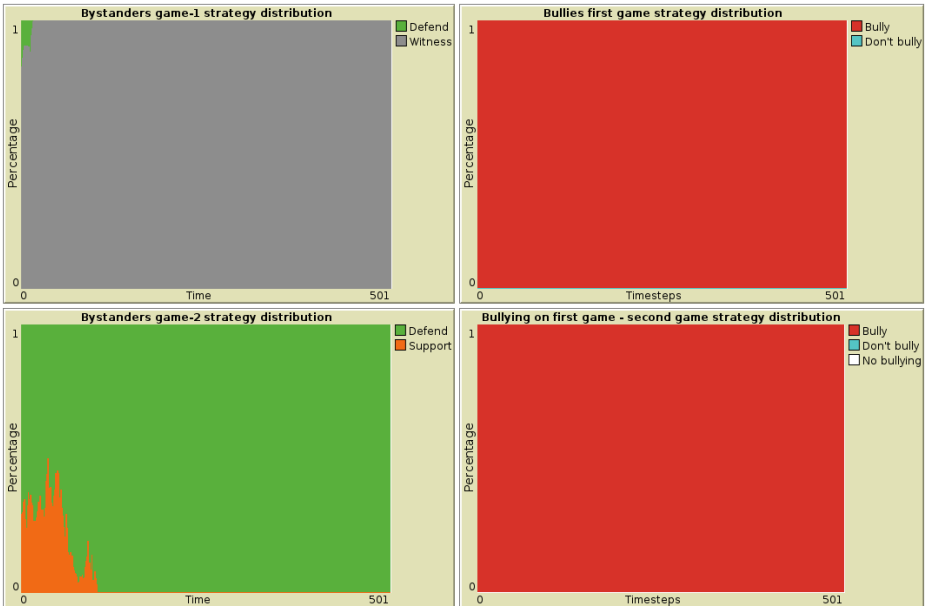

Fig. 10: Game 1 and game 2 distributions

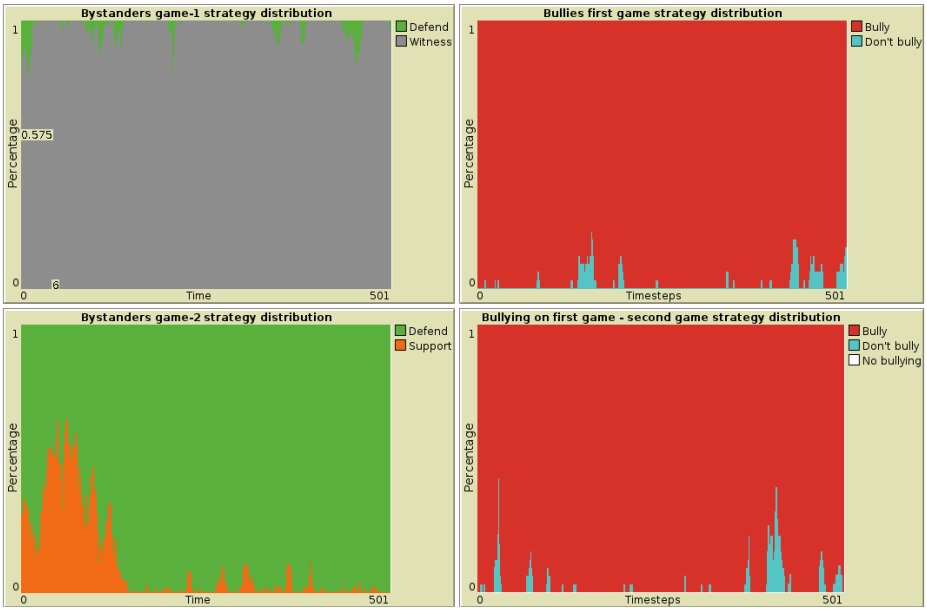

Fig. 11: Game 1 and game 2 distributions with added noise

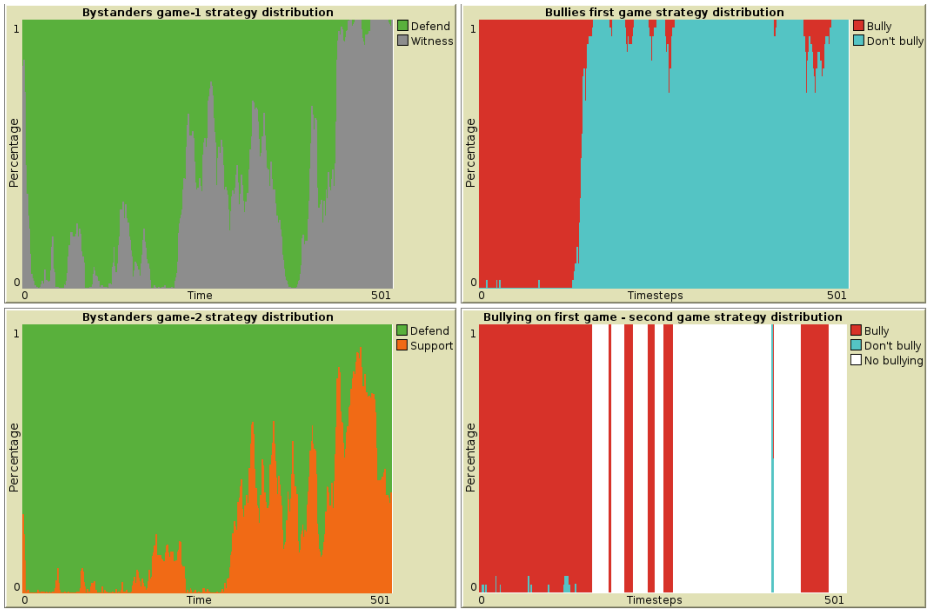

Fig. 12: Game 1 and game 2 distributions

4 Conclusion

While the complexity of the bullying phenomenon is hard to understand and to model, trying to so by means of game theory seems to be a viable approach. The developed model turned out to be capable of modeling real world scenarios but, the interaction between many different agent and the two game-phases, might bring it into unpredictable states. This is however expected when modeling such complex behaviours: individual interactions between children are difficult to identify and characterize. However, by playing with values in the model, in particular those that plays an important role in game 1 we can see how the most important part in fighting bullying is to instruct individual into defending the victims. There is much work that could be done to improve the model, starting from a more careful refactor of the payoff matrices and values to implementing different interaction patterns, such as using a network-based approach to capture the clustering principles that can be observed in bullying. Nonetheless a game-theoretic approach seems to be the best way to tackle the problem of understanding how to guide children behaviour through change in policies and norms within classrooms.

References

1. Hodgson, A.: The Game Theory Behind Bullying, Part I & II (Jan 2021), https://www.youtube.com/watch?v=RHkKJibHnWg,

https://www.youtube.com/watch?v=70YUduMOm1Q, [Online; accessed 4. Jan. 2022]
2. Olweus, D.: Peer harassment: A critical analysis and some important issues. Peer harassment in school: The plight of the vulnerable and victimized pp. 3–20 (2001)
3. Salmivalli, C.: Bullying and the peer group: A review. Aggression and violent behavior **15**(2), 112–120 (2010)