Cognitive and neural bases of decision-making causing civilian casualties during intergroup conflict

BSE662A – Midterm Paper Review – Team Undecided Jr.

Dhruvil Doshi (190295), Rishabh Dugaye (190701), Devyanshi Singh (180237), Ananya Gupta (190128), Atur Gupta (190203), Nakul Jindal (190524)
Indian Institute of Technology, Kanpur ***March 6, 2022***

Paper Reviewed: Cognitive and neural bases of decision-making causing civilian casualties during intergroup conflict
Authors: Xiaochun Han, Shuai Zhou, Nardine Fahoum, Taoyu Wu, Tianyu Gao, Simone Shamay-Tsoory, Michele J. Gelfand, Xinhuai Wu and Shihui Han.

Background

An intergroup conflict should be and is fuelled by the intent of punishing only the combatants of the opposing group. However, whenever a conflict escalates, it gives rise to aggressive measures being taken, like military actions during a war. This results in lesser discrimination of the opposing combatants and non-combatants, as the motive behind the aggressive actions becomes bigger than the rights of the innocent. As discussed in *Eckhardt, W. Civilian deaths in wartime*, the civilians should not be considered a part of the war as they are unarmed and the notion of 'just war' also excludes the killing of civilians. However, the paper *Reynolds, J. D. Collateral damage* claims that causing indiscriminate and severe adversities to provide a strategic advantage to the inflictors, although prohibited by the International laws like *Additional Protocol I of the Fourth Geneva Convention of 1949*. The loss of human life and property due to the adoption of such an aggressive approach is termed as the conflict's 'collateral damage.' There are enough references to justify that whenever mass killings are associated with a conflict, a substantial proportion of the casualties are civilians or non-combatants. For example, according to *Roberts, A. Lives and statistics,* the ratio of civilian to military deaths during the Iraq War is estimated to range between 3:1 and 8:1. Similarly, the findings in *Voon, T. Pointing the finger* claim that hundreds of civilians were killed and thousands of civilians were wounded due to NATO's bombing intervention in the Kosovo Conflict. The collateral damage involved in such scenarios is against the humanitarian interests on moral and ethical grounds and results in destructive social consequences as well. However, we know for a fact that civilian casualties increase mostly when intergroup conflict escalates, as hypothesized in *Wolfe, R. J. & Darley, J. M. Protracted asymmetrical conflict erodes standards for avoiding civilian casualties.* Hence, this paper tries to capture the cognitive and neurological changes that occur before one decides to harm even the innocent non-combatants of the other group. One way to do this is to quantify and judge the harm preference (fuelled by the conflict) and harm avoidance (driven by humanity and laws) of people belonging to one group during different stages of conflict. So, the study of cognitive and neural mechanisms involved in punishment decision-making is entirely relevant for the underlying social concern because the decision to be made, in this case, is that of punishment or its avoidance.

Problem Statement and Experimental Tractability

Now that we have motivated ourselves with the background of the study, let us dive deeper into the primary questions the paper addresses and whether there are sophisticated experiments possible to encapture the conclusions. From the background, it's pretty apparent that most intergroup conflicts lead to `collateral damages,' that is, harm being caused to civilians or non-combatants by the opposite group.

However, prior literature does not segregate the punishment decision-making pertaining to outgroup combatants and non-combatants separately when dealing with the neurobiological aspects of it. Indeed, it is tricky to differentiate between the underlying decision-making that leads to harming outgroup combatants versus non-combatants in any intergroup conflict. This paper attempts to address this very question. More precisely, we deal with the question of understanding the unique cognitive and neural patterns involved in punishment decision-making against outgroup civilians. Moreover, the paper also addresses the question of how psychological processes evolve, leading to increased civilian casualties when the intergroup conflicts escalate.

The next question that comes to mind is whether it is possible to design sophisticated scientific experiments to explore solutions to the questions raised above. The authors deal with the problem by stating two psychological hypotheses and devising experiments for their verification. The first hypothesis proposes that `harm preference' and `harm avoidance' are two distinct psychological constructs that are responsible for different punishment decision-making tendencies towards the outgroup combatants and non-combatants respectively. Intuitively, this hypothesis only seems fair since, during any intergroup conflict, the goal of each group is to primarily combat and harm the combatants in the opposing group and prevent adversities on the civilians as much as possible. This hypothesis is not only coherent with the moralities of most groups in real life but also the protocols of the Geneva Convention of 1949. Verifying this hypothesis through experiments would quite effectively answer the primary question raised in this paper. The second hypothesis claims that the rising civilian casualties during escalating conflicts might be a result of reduced harm avoidance towards outgroup non-combatants. This hypothesis, if verified through experiments, provides a reasonable answer to the second major question raised by the authors.

To conclude this section, we finally discuss how the experiments and hypotheses devised by the authors are actually suitable for encapsulating the complex psychological issue pertaining to collateral damage during intergroup conflicts. To verify the proposed hypothesis, the authors orchestrated various experiments in which participants were subjected to conflict-like situations and asked to make decisions by choosing electric shocks as punishment for outgroup combatants. However, these punishments were also coupled with punishments to outgroup non-combatants to various degrees of intensities.

The details of these experiments shall be discussed in later sections, but how do we convince ourselves that these simulations could indeed do justice to the problem statement in a real-world scenario? The paper lays out three arguments for that. First, Real-world intergroup conflicts are usually initiated when there is an unequal distribution of valuable resources. For instance, due to unequal distribution of land and water caused by environmental, social, and economic factors, several disputes have emerged between the farmers and pastoralists in Dafur, Sudan. Africa's longest-running war between north-south Sudan was also initiated due to the unfair distribution of oil revenue and oil reserves in the border areas. Oil revenue sharing also caused severe conflicts between the Shia and Sunni Arabs in Iraq. To ensure these factors, the authors ensure that intergroup conflicts arise due to the unfair distribution of limited resources. Second, similar to real-life intergroup conflicts, participants were made to take decisions that would punish outgroup combatants but also harms outgroup civilians to various degrees. Third, as intergroup conflicts escalate in real-world scenarios, the level of collateral damage also boosts. Experiments in the paper manipulate conflicts on three levels(low, middle, or high) for the participants. Thus by conducting statistical modelling of the choices, we observe increasing harm to outgroup non-combatants and can also differentiate between `harm preference' and `harm avoidance' tendencies towards outgroup combatants and non-combatants, respectively. These three arguments thus, reasonably support the capability of the designed simulations to mimick real-world conflicts.

Besides the laboratory simulations, to further support their claims, the authors also conducted experiments with two real-world ethnic groups, Jewish and Palestinian, during the Israeli-Palestinian conflict. Participants were also subjected to fMRI monitoring to keep track of activities in relevant brain networks(like the salience network, theory-of-mind network) alongside the punishment decision-making process. With more concrete ideas on the paper's central problem statements, their experimental tractability, and the suitability of the experiments to reasonably address the posed questions, we are ready to explore the

methodologies and results involved in greater detail in the following sections.

Results And Implications

In the first part of the study, intergroup conflict was studied in a laboratory setting. During phase 1 of the study, the participants were divided into teams of two and asked to rate their closeness with other participants in the study. The ratings obtained show the apparent fact that people feel closer to their group members than other participants. But the main point to notice is that closeness towards outgroup members in group conflict and bystanders in individual conflict is nearly the same, which shows that group formation strengthens the affection between the group members rather than weakening it for other members.

After phase 4 of the study, the results show that people prefer to harm the combatants and avoid harming non-combatants in group and personal conflicts. With an increasing level of conflict, harm preference to combatants increases. The interesting thing to note was that during group conflicts harm avoidance to non-combatants decreases but during individual conflicts, this harm avoidance do not decrease which shows that in individual conflicts people do not tend to harm the bystanders with increasing conflict level as they do not associate the bystander with their enemy. We can say that this result is robust as it was confirmed by one-sample t-tests and two sample t-tests.

In the second part of the study harm avoidance and preference was studied in a real-world context. The results obtained were not completely in coherence with the results of the laboratory study. The harm preference for combatants is very less which shows that people in the real world context are not in favour of doing harm to the combatants as well. This may be because in a real-life conflict the life of people are at stake and in such a situation people would not like to appear as a war supporters because it may cause harm to their social image. These results may be biased because when the people are actually in a conflict then their decisions would be influenced by the people around them and if the majority are in support of causing harm to the combatant, then they would also choose this but in the study, these people may be giving there actual opinions which may change due to influence of others in a real condition. The other results were in coherence with part 1 of the study.

The third part of the experiment was focused on studying what part of the brain is involved in harm preference and avoidance towards out-group combatants and non-combatants using fMRI. It was observed that a decreased harm avoidance towards non-combatants during group conflicts was correlated to decreased activity in the left middle frontal gyrus. This is not observed in the case of individual conflicts which makes it clear why harm avoidance decreased for non-combatants in group conflict but not in individual conflicts in the 1st part of the study. Further, it was observed that harm avoidance towards non-combatants was correlated to increased activity in the left precentral cortex for group conflicts and in bilateral precentral cortices for individual conflicts. Further, it was observed that increased harm preference towards outgroup combatants was correlated to increased activity in the left temporoparietal junction, left prefrontal cortex, midcingulate and bilateral occipital cortex and increased harm preference to combatants with increasing level of conflict was correlated to the right occipital cortex. The relation between decreased harm avoidance towards non-combatants and increased activity in left MFG was confirmed using moderation analyses and second level regression analyses hence providing robustness to the result.

Critical Comments

In our opinion, the results do not answer the generalized initial problem statement. However, they did find a significant correlation between the results and left frontal MFG, which is a great start for future work.

The first reason for our claim is that the experiments in the paper are not set in a real-world context nor do they represent real-world scenarios and situations accurately. The problem statement is regarding collateral damage, which is a relatively broad topic but has been reduced to a rather smaller theme on revenge analysis.

In reality, war introduces many cognitive biases which can change the way decisions are made. People who have been involved in the actual war have a different mental state. When there are deaths, it is much more severe than what the study addresses. Further, the consequences of their actions will affect their following decisions which do not factor in the experiments. The study does not actually take the views or perform experiments on people actually involved in the war. The temporal nature

of the war is not being accounted for, i.e. when the conflict lasts for a relatively long period of time, the initial biases and issues might be altered and the morality of the stakeholders could depreciate progressively even though the intensity of the conflict is unaltered.

Another concern is the lack of cross-talk across the teams. Usually, if one side takes an action, the other side retaliates. This exchange is continual. As pointed out in the study's limitations, only one team is active since it decides the punishment, but the responses are not conveyed back in any form. Attacks in the real world take into consideration the probable retaliation from the outgroup which will further vary under extended conflict expenditure. Also, there is no parameter of self-motivation in the model. Retaliation is not the only motivation for punishment and resource allocation is a rather trivial substitute for conflicts. Political, financial, cultural and religious clashes amalgamate to form group conflicts and personal biases which are oversimplified in the experiments. Personal gain is also one of the primary motivating factors which should be accounted for. Outside influences like third party groups, the presence of a different group identity which is shared across both the conflict groups, how they are perceived by external communities and ethical constructs shared among the members of the groups do indeed factor in the decisions made and need to be accounted for.

In real group conflicts, there exists a disparity between the decision makers and the people executing the punishments which are not reflected in the manner the experiments are designed. Ingroup communication affects the ideologies and perspectives of all the people present. By allotting the decision power to an individual we dismiss the possibility of intragroup communications which might change (harsher/midler) the final action. Next, in the study, experiments were performed on only three cultures, while the problem statement wanted to address all intergroup conflicts. There can be a mix of cultures as well in groups, which can affect the decision making.

Overall, the paper was a very balanced and informative study. The authors aim to precisely differentiate the neural aspects of decision making which results in harming outgroup combatants vs non-combatants. The authors did a fascinating job of answering the statement limited to retributive decisions performed by individuals in response to group conflicts. While they are unable to generalize the harmful decisions across every domain, they accurately justify their experimental results, matching the psychological constructs and cognitive processes. They irrefutably demonstrate the decrease in harm avoidance during escalated conflicts and also correlate its neural basis by displaying the decrease in activity in the left middle frontal gyrus(MFG).

Future Prospects

Authors have hinted in the paper that people in experiment 1 had some biases owing to the fact that the procedure used in the study possibly created demand characteristics that resulted in social desirability pressures to appear moral. So the authors wished that future research help reduce these biases which is why we first begin by suggesting some changes in the current study that could be incorporated in future similar research to check the aforementioned biases.

Possible changes to the current study -

1. To take larger mini-groups than just two people. We shouldn't let a single person decide the punishment of the outgroup people by themselves. Since in the real world, these feelings towards outgroup persons get influenced by other ingroup people a lot. When people talk to other ingroup people, they discuss the harms done by outgroup people and start feeling other people's anger and misery as their own and develop a stronger feeling of hatred which is absent in experiment 1. So what we can do to consider this factor is to make larger mini-groups and let them decide as a group what punishment to give to the outgroup combatants and non-combatants.
2. Experiment 1 seemed to miss out on how anger is changed to a sense of hate when a section of people is oppressed for a significant amount of time (basically, the negative feeling becomes more potent with time). In other words, the factor of time is missing from this study. We need to make people believe that they've been mistreated for an extended period, then it would be much more similar to a real-life scenario. We believe this is very important in this study's context since hate is a much stronger feeling than anger in making people do inhumane acts towards other human beings, such as those done during the war. Chances of people choosing to

go forth on the path of evil and do utter unfair acts even on non-combatants should be much more when they feel they've been treated the same for a long time. So we think that another study including this factor should be considered, which would span for a longer time to make them feel like they've systematically been let down by the people of the other group for quite some time.

Now a bit different and more invasive follow up study to this study can be-

We can try a more invasive experiment on monkeys along the same lines as these experiments. We can make groups and mini-groups of the monkeys just like how it's done in the given study for the same purposes. This can be done by choosing monkeys from different tribes that, let's say, dislike each other a lot, to the point that they would show their anger by screaming or getting irritated if they even show the images of the other tribe's monkeys. It could also be done by taking food from one monkey and giving it to another monkey repeatedly in front of it, so they start developing a feeling of antipathy for each other. Once this is done, we can see which parts of the brain are making them angry towards the monkeys of the other tribe. After that, we can try stimulating those brain parts by sending an electrical signal and seeing if we can increase or decrease the hatred they have against each other.

Suppose this study can give some fruitful results. In that case, it might have quite revolutionary consequences because if it can be extended to different species, including humans, then we might be able to eradicate any feeling of hatred one person has for another. Of course, at the same time, it might be used for nefarious schemes as well in which people can be manipulated to hate a particular section of people and as a result can be forced to get involved in villainous acts.