Using Robots to Moderate Team Conflict: The Case of Repairing Violations

Malte F. Jung
Cornell University
Ithaca, New York, USA
mfj28@cornell.edu

Nikolas Martelaro
Stanford University
Stanford, California, USA
nikmart@stanford.edu

Pamela J. Hinds
Stanford University
Stanford, California, USA
phinds@stanford.edu

ABSTRACT
We explore whether robots can positively influence conflict dynamics by repairing interpersonal violations that occur during a team-based problem-solving task. In a 2 (negative trigger: task-directed vs. personal attack) x 2 (repair: yes vs. no) between-subjects experiment (N = 57 teams, 114 participants), we studied the effect of a robot intervention on affect, perceptions of conflict, perceptions of team members' contributions, and team performance during a problem-solving task. Specifically, the robot either intervened by repairing a task-directed or personal attack by a confederate or did not intervene. Contrary to our expectations, we found that the robot's repair interventions increased the groups' awareness of conflict after the occurrence of a personal attack thereby acting against the groups' tendency to suppress the conflict. These findings suggest that repair heightened awareness of a normative violation. Overall, our results provide support for the idea that robots can aid team functioning by regulating core team processes such as conflict.

Categories and Subject Descriptors
H.1.2; H.5.3; I.2.9; J.4; K.4.3

Keywords
Human-robot teamwork; emotion regulation; repair, team performance; group problem solving; conflict..

1. INTRODUCTION
Over the last decade the idea that robots could become an integral part of work in teams developed from a promising vision [24] into a reality. Robots support teamwork across a wide range of settings covering search and rescue missions [28], minimally invasive surgery [12], and manufacturing [5]. Robots impact teamwork not only through the task-specific functions they have been developed to serve but also by changing the social dynamics of the team [14, 29]. It is this latter role of the robot that we explore.

As robots are increasingly incorporated into teams, their impact on the social and emotional life of these teams becomes ever more important. Recognizing this, scholars have begun to explore how the physical features and behaviors of robots can shape team dynamics. For example, Hinds, Roberts, and Jones [24] demonstrated that people deferred more responsibility to a robot in a team-based sorting task when more human-like attributes defined its appearance. Focusing on behavior, Mutlu and colleagues [29] demonstrated that a robot could nudge two group participants into playing the roles of addressees, bystanders, and overhearers through subtle changes in the robot's gaze behavior alone. While these two studies were not intended to demonstrate the impact of a robot's behavior or features on team effectiveness, there is initial evidence that robots can also contribute to team effectiveness. Breazeal and colleagues [7], for example, demonstrate that nonverbal social behaviors such as gaze, shifts in posture or shifts in orientation not only improve peoples' perceptions of the robot but also improve team functioning. There is also promising evidence that a robot's emotional behavior can serve an important function in improving team effectiveness. In a teamwork scenario in which one human participant interacted with two robots, Jung and colleagues [23] found that robots can improve states that are known to be crucial for a team's effectiveness such as stress and cognitive load through subtle changes in the robot's gaze behavior. Together these studies suggest that robots can be added to teams and influence a team's effectiveness through their behaviors, even when those behaviors are not specifically directed toward supporting the task.

Despite ample prior work demonstrating the important role of emotions and emotion regulation for the effectiveness of teams [4], no study has explored whether robots can improve team dynamics and team effectiveness by regulating emotions within a team. Finding emotion regulation strategies that can be employed effectively by robots can help improve human robot teamwork across a wide range of application context since emotion regulation strategies are independent of the task-context a robot is employed in. Interestingly, Barsade [3] showed that it takes only a single individual's behavior to influence the emotional dynamics and ultimately the performance of a team. In her study, a confederate enacting one of four moods (positive or negative with either high or low arousal) was able to alter the mood of an entire team, which, in turn, shaped team effectiveness. We build on this work in two ways: First, we explore whether a robot can act as an emotion regulator for a team and thereby increase performance. Second, we explore emotion regulation through enactment of specific repair behaviors rather than through enactment of diffuse mood states as done in the Barsade study.

To answer these questions, we examined a robot's effectiveness at regulating emotions in a team-based problem-solving task. In particular, we focused on the effect of regulating emotions on perceived team conflict. We focused on perceived conflict because the way that teams manage conflict has been shown to be central in predicting team effectiveness [20, 21, 9]. We conducted a laboratory experiment in which a confederate introduced negative affect into a team thus providing the opportunity for the robot to intervene with behaviors designed to regulate negative emotions. We found that robot repair interventions influenced perceptions of conflict within the team, up-regulated positive affect during personal attacks, and had a tendency to reduce

HRI'15, March 2–5, 2015, Portland, Oregon, USA.

ACM 978-1-4503-2883-8/15/03…$15.00.
http://dx.doi.org/10.1145/2696454.2696460

negative perceptions of the confederate who introduced the negative trigger. By a negative trigger we mean behaviors that are likely to cause negative reactions in others such as personal attacks and hostile remarks but also simple disagreements. Our study makes three contributions to the literature.

First, and foremost, we demonstrate that a robot can positively contribute to a process that is central for teamwork: conflict. We show that through a simple repair, a robot can make a personal and potentially harmful violation salient and up-regulate positive affect when it is most needed after personal violations. Second, we contribute to our understanding of human robot teamwork with multiple human participants. This study is one of the first mixed human robot teamwork studies in which a robot interacts with multiple human team members. Third, we contribute to our understanding of emotion regulation in teamwork. To our understanding no prior studies have explored how emotions in teams are regulated when targeted negative expressions are made and repaired.

2. Background and Hypotheses

The expression and experience of negative emotions has been shown to lead to negative performance in teams [3, 35]. It is important for teams to deal effectively with the experience and expression of negative emotions and minimize their impact. Negative emotions not only impair team performance directly, but the expression of negative or even hostile behaviors likely trigger negative reactions in other team members thereby initiating a downward spiral of increasing negativity [1]. This tendency for negative behaviors to escalate has been documented in both marital [26] and work contexts [1]. It is therefore important to regulate negative emotions before a downward spiral is initiated.

A particular strategy to dampen the impact of negative behaviors and prevent negative affect from escalating is repair. Repair refers to behaviors that are intended to restore the negative impact a violating behavior has on a relationship [32]. Repair intends to dampen, limit, or even eliminate the negative impact that a conflict situation has on the emotional progression of an interaction. We therefore conceptualize repair here in line with the literature on repair in marital and interpersonal conflict as an emotion regulatory phenomenon. The ability to effectively repair negative behaviors has been found to be important in regulating marital conflict [15] and work conflict [1, 32]. Repair can take various forms. A typical type of repair is an apology, but repair can also invite others to change their behaviors. Andersson and Pearson [1], for example, describe a repair strategy of actively discouraging the behavior of the person who introduced the negative trigger. Observing groups during conflict, Jehn offers an example of this type of repair: ”Stop that; this isn't the place for that!” ([21], p. 545). Finally humor is an effective repair strategy when used by people other than the violator [13]. While there is a wide range of strategies for interpersonal emotion regulation [36] our research focuses only on the subset of repair behaviors since repair has been argued to play a particularly important role in managing conflict in marital and team interactions alike [15, 30].

2.1 Emotion Regulation through Repair by a Robot

Intervening by regulating negative emotions may offer an opportunity for robots to support team functioning. Robots are likely effective in aiding emotion regulation in teams because they are immune to interpersonal tension among team members. Several studies have already established that robots can influence the emotional experience of people through their expressive behavior. For example, in a study in which children played a game with an emotionally expressive robot, Leite and colleagues [27] found evidence that a robot could regulate the emotional experience of a child through behaviors such as humor. Another study demonstrated that participant's experience of stress during a complex task could be reduced by a robot's gaze behavior [23]. Given that repairs by people reduce negative spirals [1, 32] and that people show a tendency to respond to robot behavior in a similar way as they do to human behavior, we expect that a robot responding to a negative trigger with a repair will prevent negative affect from escalating. More specifically we hypothesize:

H1: When a team member introduces a negative trigger, teams will report feeling more positive if a robot intervenes with a repair as compared to if a robot does not intervene with a repair.

2.2 Repair and Awareness of Conflict

Teams are prone to conflict of two types: relationship and task conflict [20]. Jehn, for example, observed in one of the most influential studies on intra-group conflict, that task conflict exists when there are "disagreements among group members about the content of the tasks being performed, including differences in viewpoints, ideas, and opinions", ([20] p. 258). Relationship conflict, on the other hand, exists when there are "interpersonal incompatibilities among group members, which typically includes tension, animosity, and annoyance among members within a group", ([20] p. 258). People not only distinguish between task and relationship conflict, they also attribute different effects on performance to each conflict type [21]. Meta-analyses however suggest that both task and relationship conflict generally have negative effects on team performance [9].

Personal violations or attacks have been argued to be a particularly potent trigger of relationship conflict [32]. A negative trigger that contains a personal attack violates the victim's expectation of how he or she should be treated. An unrepaired personal violation is thought to lead to relationship conflict because it constitutes an incompatible activity [10] that lacks consensual validation of the offender and the offended and therefore leads to interpersonal tension. For a detailed review of the literature about how violations promote relationship conflict please refer to the work by Ren and Gray [32]. We therefore anticipate that personal attacks will lead to more relationship conflict.

Numerous studies on conflict have shown that task and relationship conflict typically occur in tandem [9, 11]. Therefore, even though negative triggers that contain personal violations are thought of as particularly strong in eliciting relationship conflict, they may also increase task conflict. A strong negative experience of relationship conflicts has been shown to spill over into conflicts about the task [21]. As described by Ren and Gray [32], a repair can be initiated not only by the violator or the offended party, but also by a third party who acts as a conflict resolver. Interventions issued immediately after a negative trigger are likely most effective as they prevent spiraling negativity [1]. As argued in the previous section, prior work has established that robots can change how people feel [23, 27] and that people often respond to the interventions of robots as they would to those of people [31]. Given the negative impact of personal attacks and the tendency for people to experience both types of conflict in tandem we expect that conflict experience will be particularly heightened when a negative trigger contains a personal attack and we expect that if a robot intervenes directly after a negative trigger, it can dampen the experience of both types of conflict:

H2a: More conflict (relationship and task) will be perceived in a team when a team member introduces a negative trigger that contains a personal attack vs. being only task oriented.

H2b: Less conflict (relationship and task) will be perceived in the team when a robot repairs negative triggers, especially triggers that contain personal attacks vs. being only task oriented.

2.3 Repair, Contributions & Performance

Violating norms of social conduct and introducing a negative trigger within a team can have negative effects on perceptions of the team member who introduced the trigger as prior work has shown that it takes as little as a facial expressions for others to form global judgments about a person [25]. Additionally, conflict perceptions are likely to trigger negative emotional reactions [35] and those emotional reactions in turn can inform judgment processes [34]. It is therefore likely that, independent of their actual contribution, those who introduce negative triggers into a team will be perceived as worse contributors. Repairs are interventions that diffuse negative tensions and consequently, even when issued by a third-party, can alleviate negative feelings toward the person who was perceived as attacking others or violating social norms [32]. As a result, we hypothesize that the negative perceptions of a team member who issues a personal attack will be mitigated when the robot offers a repair.

H3: The contributions of a team member who makes a personal attack will be perceived as more positive when a robot repairs the attack than when the robot does not repair it.

As discussed earlier, negative emotions have been shown to impair team effectiveness [3]. When introduced into a team, negative emotions can be contagious and generate a downward spiral of increasingly negativity [1, 35]. Repairs, however, are interventions designed to arrest the negative spiral and reduce the impact of negative triggers. We therefore argue that when negative triggers are introduced, but repaired, the negative effects on performance will be minimized.

H4: Team performance will suffer with negative triggers, especially personal attacks, but this effect will be eliminated when a robot repairs the negative trigger.

3. THE EXPERIMENT

In a 2 (negative trigger: task only vs. task plus personal attack) x 2 (repair: yes vs. no) between-subjects experiment (N = 114 individuals, 57 teams), we studied the effects of a robot repairing negative triggers on team affect, conflict, perceptions of team members' contributions, and team performance during a problem-solving task. We recruited all participants through email lists and job postings at a research university. Teams were formed randomly based on when participants were available to attend a session. Each participant received a $15 gift certificate. Participants worked together as a team (2 participants A & B, 1 confederate C, and 1 robot, see Figure 1) to solve a "bomb defusal" puzzle based on the game Mastermind in 10 minutes. The task-trigger/no-repair condition doubled as a control since task-oriented disagreements are commonplace in teamwork.

3.1 The Robot

The robot was built on a Pioneer 3 robot base and was comprised of an OWI robot arm, an Arduino based arm-control board, and a speaker. The robot was designed to appear physically capable to scan the bomb with its arm and to interact verbally with the team. We used a Wizard of Oz set up in which the experimenter controlled the robot's movements and speech. During the task, the robot also moved back and forth about once per minute, appearing to look at each of the team members to increase the presence of the robot and help integrate it as a team member. At set times scripted utterances were triggered in which the robot referred to itself and the team as "we" to help integrate it into the team. These comments included task strategy comments and general task comments as follows:

> "We could try swapping just two of the wires."
> "I really want us to solve this in time."
> "We have about five minutes left."
> "Lets look back at the previous moves to see if we can find any patterns."

The Mac OSX "Apple Alex" voice was used at a speed slightly above normal to give a sense of urgency to each comment. Scans were initiated when participants asked the robot to "scan now." By taking 30 seconds to complete, we designed the scan procedure such that its use would constitute a significant cost on time and that it would encourage careful deliberations rather than frequent trial and error scanning. The overall interactive capabilities of the robot were left ambiguous to the participants.

3.2 Bomb Defusal Task

We designed a bomb defusal task modeled after the Mastermind game, a problem-solving task that has been used in previous studies of human-robot team interaction [2]. Seven pins were exposed on the bomb, requiring seven correct wires to be clipped to them to successfully defuse the bomb. 23 wires in nine colors were loose on the bomb and were used to create the codes. Teams were told that the bomb was unstable and that they should attempt to solve the code in less than 10 trials. The task was calibrated to be sufficiently hard so that no teams were able to solve the code within 10 minutes. This kept the task experience constant, as all teams workied on the task for the full length of the session. The time limit also created a level of stress that we intended to increase the impact of the violations.

3.3 Manipulations

3.3.1 Negative Triggers

For the negative trigger manipulation, a carefully trained confederate delivered negative triggers during the task. Depending on the experimental condition, the confederate issued two triggers either primarily directed at the task (task) or primarily directed at a team member (personal). We developed the triggers with the confederate through numerous iterations.. For personal triggers we found that adding, "You're stupid" to the beginning of a negative trigger caught the attention of participants and clearly defined the trigger as a personal attack rather than a task focused negative trigger. Telling people "You're moving too slow" and that they were "not very good at the game" were also effective at catching the attention of participants. For task-directed triggers the confederate stated, "that's not a good idea" or "I wouldn't do that." Table 1 outlines the spoken triggers and delivery times. Due to the short time frame of the task, negative triggers were delivered at specific times, 2-3 minutes and 5-6 minutes into the trial. Triggers were directed opportunistically at one of the team members. Due to the time constraint on trigger delivery, the confederate did not always violate the same team member. The triggers were presented with appropriate facial expressions and voice tone (condescension for personal triggers and frustration for task triggers) and to increase the attention from the participants the confederate aggressively pulled a wire when issuing a trigger. This physical display captured the attention of participants who were often very focused on the bomb. When not delivering a violation, the confederate acted in a neutral, passive manner and did not initiate new moves nor offer unsolicited input. Although the confederate remained silent for most of the session, he briefly answered questions directed towards him and helped with placing wires. The passive behavior helped to keep the confederate's involvement as constant as possible across conditions.

Table 1 –Negative Triggers (Personal & Task)

	Negative Triggers	Delivery
Personal Attacks	1. *"You're stupid, lets not use this one. Use this."* 2. *"That's not right, you are not very good at this. Use this."* [Directed at a team member. Condescending]	2-3min 5-6min
Task-Directed	1. *"Lets not use this one. Use this."* 2. *"The one isn't right. Use this"* [Directed at the task, frustrated, tense]	2-3min 5-6min

3.3.2 Repair Through the Robot

After each negative trigger, the robot delivered either one of two repair comments (repair) or one of two comments that were unrelated to the trigger and intended to be neutral (no-repair). Repairs identified the negative trigger as inappropriate and then added a normative statement to stay positive. Iterative prototyping of the repairs revealed that repair statements that simply focused on identifying the trigger or having the team stay positive were not strong enough to have an effect in the teams. We also found that some statements injected humor, which alleviated tension among team members. Thus, we opted for a more colloquial statement in order to strengthen the manipulation. The use of "Whoa, man" and "Dude" captured participants' attention and often induced laughter during our prototype interactions.

Table 2 - Robot Repair and No-Repair comments

	Spoken Repair
Repair Comment	1. *"Whoa, man, that was inappropriate. Lets stay positive."* 2. *"Dude, what the heck! Let stay positive."*
No-Repair Comment	1. *"Defusing bombs is difficult."* 2. *"There are many possible combinations."*

3.4 Procedure

All participants were invited to the lab to participate in a study about "solving problems in human-robot teams." The confederate arrived with the other participants and acted as a regular participant for the duration of the session. Upon arriving, participants filled out a consent form and a pre-session questionnaire with questions about their demographics, current mood, and past experiences with robots and artificial intelligence. Then, participants individually watched a short video tutorial about the bomb defusal task and the robot and were invited to ask any clarification questions about the task. The experimenter then led the three participants (2 subjects and 1 confederate) into another room to complete the task. The problem-solving task began with the bomb ticking down from 10:45 seconds and the robot's introduction:

> *"Hello, I am D-Bot. I will help us to solve this task. We need to work as a team to figure out the correct wire configuration. I have a powerful sensor that can scan the bomb and sense when there is flowing current in correct connections and when there is charged wire for correct colors. I also have a speech sensing and processing unit with limited capabilities. To scan, say SCAN NOW. Each scan should take 30 seconds."*

Participants then worked together with the confederate and the robot to solve the task. As the team tried various combinations and scanned them, the robot reported how many wires were either the correct color in the wrong position or the correct color in the correct position. For correctly colored wires in incorrect positions, the robot stated there were "*N* charged wires." For correctly colored wires in the correct positions, the robot would respond with "*N* current flows." The robot could not tell participants which exact positions were correct, just that there were correct wires on the board. For example, if a code included a red wire on position three and the participants placed a red wire on position three the robot would report "one current flow". If, however, the participant placed the red wire on position two the robot would respond with "one charged wire." Teams attempted to solve all seven wire positions correctly within 10 minutes.

Figure 1 - Personal negative trigger and repair.

Dependent on the experimental condition, participants were presented with different triggers and repairs. Throughout the game the robot also provided strategy comments to aid the team with solving the task. After the bomb counted down to 00:00 the experimenter announced that the task was over. The experimenter then guided the participants back to the previous room to complete a post-task questionnaire with questions about team dynamics, task perceptions, and perceptions of the robot and other team members. Finally participants were debriefed and paid.

3.5 Measures

Measures for our study were collected from the pre- and post-task questionnaires and from observations of the teams' performance.

Affect was measured using a three-item scale comprised of participants' ratings of valence on the 9-point Self-Assessment Manikin scale [6] and of how much they felt suprised and excited (α = .63). Emotion items were rated on a 9-point Likert scale ranging from "Not at all" (1) to "Very much" (9). We created a difference score by subtracting the post- from the pre-task affect scores. The resulting scores were averaged at the team level.

Perception of conflict in the group was measured by averaging Jehn's widely used and validated scales of task and relationship conflict at the team level [22]. Perception of task conflict was measured by asking participants questions such as "How frequently did you have disagreements within your team about the task you were working on?" (α = .86). Perceptions of personal conflict were measured by asking participants questions such as "How much relationship tension was there in the team?" (α = .85).

Perceived confederate contribution was measured by asking each participant to rate on a 100-point sliding scale to what degree each of the other participants and the robot contributed to the team.

Team performance on the bomb defusal task was measured by counting the number of moves the team made within 10 minutes as well as by counting number of correctly identified wires in the best configuration the team made during the 10 minutes.

As controls we measured age, gender, and prior experience with robots and artificial intelligence. However none of the controls influenced any of our results and were therefore excluded from the analysis.

4. RESULTS

53 teams comprised of 106 participants (55 men, 51 women) ages 18 - 65 (*M = 24.5, SD = 8.0*) were included in the analysis. Four teams had to be excluded from the analysis due to deviations from the protocol in the confederate's delivery of the violations resulting in 13 teams per condition except for the personal attack/neutral response condition, which had 14 teams. All hypotheses were tested at the group level with analysis of variance (ANOVA). Individual scores were averaged at the group level. To test whether group level aggregation was meaningful we calculated r_{wg} scores between team members for each self-report variable as a measure of reliability between individual participants responses. All aggregated measures showed a sufficient degree of reliability, including participant affect ($r_{wg(4)}$ = .95), personal conflict ($r_{wg(3)}$ = .77), task conflict ($r_{wg(3)}$ = .79), group contribution ($r_{wg(1)}$ = .95). Besides a significant and strong correlation between our two conflict measures none of the other self-report based measures were significantly correlated, thereby indicating discriminant validity of our measures.

4.1 Manipulation Checks

To verify that our manipulation for negative triggers was effective, we asked on the post-task survey if at least one person on the team made personal attacks on a 7-point scale ranging from "Not at all (1)" to "Very Much (7)". Our analysis confirmed that teams in the "personal attacks" conditions were significantly more likely to detect personal attacks (*M* = 2.7, *SD* = 1.9) over conditions in which the negative triggers were task directed (*M* = 1.6, *SD* = .9), $F(1, 49) = 9.5$, $p = .003$. Similarly, to confirm that teams detected that the robot was making repairs, we asked participants how much D-Bot responded to comments made by team members, recognized negative comments made by team members, and intervened when someone said something negative. All items were asked on a 7-point scale ("Not at all (1)", "Very Much (7)"). These three items were combined to form a 3-item scale with high reliability (α = .91). Groups in repair conditions were significantly more likely to report that the robot made repairs (*M* = 4.6, *SD* = 1.6) as compared to no-repair conditions (*M* = 2.2, *SD* = 1.1), $F(1, 49) = 74.8$, $p < .00$.

4.2 Affect

We found a significant interaction effect for violation and repair on affect, F (1,49) = 5.07, p = .029, partial η^2 = .09. When confronted with a task violation, teams on average reported experiencing more positive affect when the violation was not repaired (*M* = 0.44, *SD = 1*.34) than when it was repaired by the robot (*M* = -0.32, *SD* = 1.18). However when confronted with a personal violation, teams reported feeling better when the robot attempted to repair the violation (*M* = 0.30, *SD* = 1.41) than when it did not (*M* = -0.48, *SD* = 0.98). Therefore H1 received only mixed support as we found an effect in the predicted direction only for the personal violation conditions.

4.3 Perception of conflict

We found a significant interaction effect between repair and negative trigger type on perceived relationship conflict, F (1,49) = 7.41, p = .009, partial η^2 = .13. Identical repair utterances apparently had a different effect dependent on whether they followed a task or relationship oriented negative trigger (See Figure 3). Perceived relationship conflict was significantly higher when the robot repaired personal attacks (*M* = 3.09*, SD* = .99) than when it did not repair them (*M* = 2.04, *SD* = .88). However when the same repair utterances followed a task violation, perceptions of relationship conflict were about the same for repair (*M* = 1.82*, SD* = 0.68) and no repair (*M* = 2.01, *SD* = 0.73). Our findings were similar for the task conflict scale. We found a significant interaction effect between repair and the type of negative trigger on perceived task conflict, F (1, 49) = 4.53, p = .04, partial η^2 = .09. Specifically perceived task conflict was significantly higher when the robot repaired personal attacks vs. did not repair them (*M* = 3.69*, SD* = .86 vs. *M* = 2.64, *SD* = .89) and the levels of conflict were nearly identical when the negative triggers did not contain personal attacks whether the robot intervened with a repair or not (*M* = 2.77*, SD* = .86 vs. *M* = 2.64, *SD* = .90). Taken together, we found some support for hypothesis 2a, that teams would perceive more conflict when a team member introduced a negative trigger that contained a personal attack. This was true however only in cases when those personal attacks were followed by a repair utterance. H2b received no support..

4.4 Confederate's Contribution to the Group

In H3, we hypothesized that team members would feel more positively toward the contribution of team members who make personal attacks when a robot repaired that attack. At the team level, ANOVA results show there was only a marginally significant main effect for repair type on perception of the confederate, F (1, 49) = 3.00, p = .09, partial η^2 = .06, with teams reporting that the confederate contributed more to the team when the robot intervened with a repair (*M* = 28.8, *SD* = 21.4) as compared with when it didn't (*M* = 23.0, *SD* = 14.9) as shown in Figure 4. This shows a weak trend in support of H3.

4.5 Group Performance

Finally, we predicted that team performance would suffer under personal attacks, but that this effect would be eliminated with repairs. Despite trends in the expected directions, none of the differences were significant and H4 received no support.

5. DISCUSSION

Our study provided exciting evidence that a robot can play an influential role in shaping team conflict and regulating affect in groups. We expected that a robot's repair would decrease perceptions of conflict, yet we found something more interesting:. Perceptions of team conflict actually increased, but only when the repair followed a personal trigger. While the kinds of repair

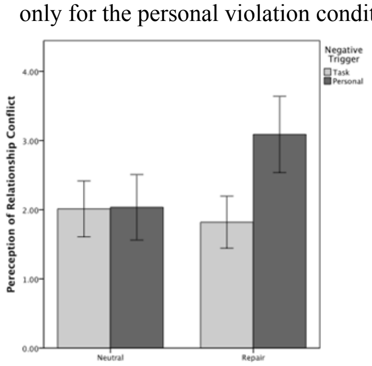

Figure 3 – Conflict Perception

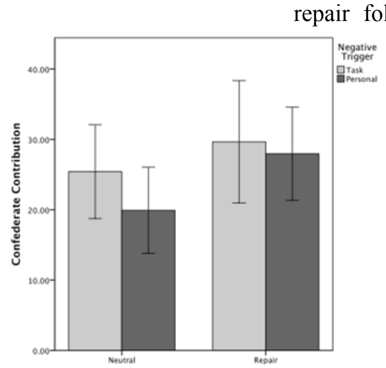

Figure 4 - Contribution of Confederate

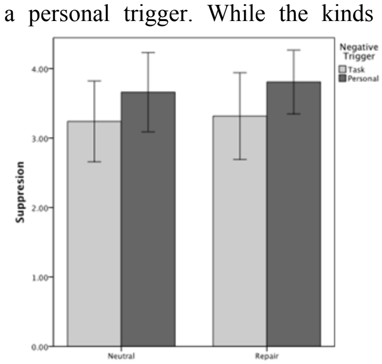

Figure 5 – Emotion Suppression

utterances we used might induce conflict on their own (in a way, the robot scolds the confederate and thus the repair constitutes a form of conflict between the robot and the confederate), it is unlikely that this happened in our case as we only saw conflict perceptions increase when the repairs followed a person-oriented violation. We did not see this increase in conflict perceptions when the exact same repair utterances followed task-oriented violations. It appears therefore that a more plausible explanation for this finding is that the robot was calling attention to the personal attack, making it impossible for the teams to ignore or dismiss it. In some sense, the robot made public that there was a normative violation by the confederate, thus rendering it more powerful .

A brief analysis of some of the video records revealed that participants reactions to the same repair utterances depended on whether the repair was preceded by a task or person oriented violation. After repairs of personal triggers participants often laughed in a somewhat mocking manner. Several looked toward the confederate seemingly in shock, waiting for a response. Expecting a response (most likely an apology) from the confederate is in line with theory that posits that violations make self-repair and after that other-initiated self-repair relevant as next turns [33]. The fact that the confederate did not respond despite the attention given from the robot might explain why some participants verbally responded to the confederate. For example, one participant mentioned, "It doesn't like you very much." Other participants did not say anything but did look discerningly toward the confederate.. In comparison, the task-oriented violations might not have been perceived as a violation to begin with and might therefore not have made an apology relevant. Calling attention to them might not have increase conflict perceptions as the robot's utterances might have been perceived as more funny rather than scolding. This is in line with our observations that after repairs of task violations, people often laughed in a confused manner. Participants looked at each other as if trying to figure out what the robot was doing, saying something like "What?" or "Huh?" Other participants appeared to not even recognize the response made by the robot, staying focused on the discussion around the task.

The possibility that a robot could prevent violations from being overlooked or suppressed is important because confronting violations more openly rather than suppressing them can benefit teamwork over time. Previous research showed that at the group level suppressing conflict can lead to group-think [19] and negatively impact team performance [8]. At the individual level, suppressing the expression of negative feelings also has been shown to have direct negative consequences for a person's health and interpersonal relations [16, 18]. In fact, we have evidence from this study that the personal violation led to an increased suppression of negative emotions within the group, F (1,52) = 4.65, p = .04, partial η^2 = .09 (see figure 5). We measured suppression through a four-item scale adapted from [18] (α = 0.83). Participants rated questions such as "When I am feeling *negative* emotions, I make sure not to express them" on a 7-point Likert type scale ranging from totally disagree to totally agree. Repairs did not change a group's tendency for suppression, which might have been due to the short duration of the task, and follow up research will have to examine the potential long-term effects on suppression.

A direct examination of the video-records supports the idea that repairs directed attention towards the violations but that they were not successful in stimulating an explicit engagement with the violation rather than suppressing it in most cases through nervous laughter. Some participants made subtle facial expressions of contempt but the most typical reaction by participants was to exchange looks and laugh nervously as in the following example in which a personal violation was followed by a repair through the robot (RP11: 00:04:08 - 00:04:20):

Participant A: *Scan now.*
Confederate: No, *you're stupid, it's not gonna work with just…*
Robot: *Wow, man! That was inappropriate!*
Participant A: [laughs]
Participant B: [looks at confederate in shock]
Robot: *Let's stay positive.*
Participant A: [inaudible comment
Participant B: [looks at confederate]

We found two instances in which a participant explicitly addressed a robot's repair 1) (RP10: 00:06:36 – 00:06:54)

Robot: *Dude! What the heck! Let's keep positive.*
Participant A: [looks down and smiles]
Participant B: *It doesn't like you.*

2) (RP17: 00:2:22 – 00:03:50).

Robot: *Wow, man! That was inappropriate! Let's stay positive.*
Participant B: [laughs] *Alright… Scan now… Haha, the robot thought you were inappropriate.*

In most cases, as with these examples, it seemed as if the violations were clearly noticed in the team but teams then tried to move on as quickly as possible. Often a participant's suggestion for the next move came immediately after the robot's repair seemingly suggesting that attention should be focused on solving the task rather than focusing on anything else.

5.1 Performance and Perceptions Thereof

We did not find any support for our hypothesis that performance would be implicated through personal violations and that repairs might alleviate the negative performance impact of those violations. Our hypothesis was based on theory that highlights the importance of the quality of interpersonal relationships for performance and in theories about the role of positive affect in improving problem solving. However, while the violations might have hurt the relationship with the confederate, they might have unintentionally improved the relationship between the other members of the team. Whenever a violation occurred, we often noticed participants exchanging looks, or smile at each other. In other words the violations might have tightened the bond between the participants and isolated the violator. Additionally, we were not able to demonstrate the desired increase in positive affect through the repairs. In preliminary trials we found that listeners responded with genuine laughter to the repairs that were deliberately non-robot like. However, in the actual study, genuine laughter was mostly absent in favor of tense giggling. The awkwardness of the violations in the context of the actual task might have led participants to appraise the repairs differently than during our prototyping.

It is interesting to note that, although calling attention to the negative trigger in the personal attack condition heightened perceived conflict, it also had a tendency to improve perceptions of the confederate who issued the negative trigger. These results support the idea that the robot's repair behavior may have generated an empathic response toward the confederate. This interpretation is in line with Ren and Gray's [32] assertion that the restoration of conflict depends on successful signaling towards the offender that a violation has occurred. Although we have no significant results that speak to team effectiveness, we anticipate that with a longer task, having a robot repair personal attacks could lead to more empathic group dynamics, which has been shown to improve team effectiveness [3, 35].

5.2 Limitations

As this is the first study to explore a robot's role in regulating conflict in groups, several limitations have to be noted. First, the immense logistic effort required to run this study made it difficult to have large numbers of teams. It would be useful to replicate this study with larger numbers of teams, especially to evaluate the perceptions of the confederate and team performance. Second, despite our best efforts at scripting the violations, carefully training the confederate through numerous trials, and setting fixed time-frames for both negative triggers, we noticed slight deviations between groups in the delivery of the violations when we reviewed the video records. In some of the "personal violation" cases the confederate called the participants "slow" instead of "stupid" and we assume that this might have weakened our results as the "slow" trigger ultimately might not have had as much negative impact as we had intended. In addition, while we had intended to always deliver violations towards the participant sitting in position A, this did not work in practice. Violations had to be directed in an opportunistic fashion as some participants did not participate actively and it was not possible to disagree with a team member who did not say anything. In order to keep the violation at the desired time, the confederate had to direct the violation to the participant who commented at the appropriate time. Ideally, the confederate would have delivered the triggers to only one team member as we expect the receiver of a trigger might react differently from the on-looking team member. Finally, the second "no repair" statement "diffusing bombs is difficult" might not have been as neutral as intended and might have been perceived as slightly positive through its empathetic tone. This might have made it more difficult to detect differences between the repair and no-repair conditions. While these deviations might have weakened the impact of the manipulations, and might explain some of our non-significant findings, our manipulation checks show the overall impact was as intended.

5.3 Open questions and future research

In 1972 Janis famously argued that groups fail at decision-making when potentially conflicting perspectives are ignored in favor of maintaining good team dynamics [19]. We found that a robot made it hard for teams to ignore conflict. Future research can address a robot's capabilities to mitigate groupthink by making hidden conflict salient and regulating group emotions.

Here we examined only one type of repair and an exciting avenue for future studies could be to compare a wider range and potentially more impactful types of repair and other emotion regulation strategies that robots could employ. For example, apologies have been shown to be highly effective forms of repair [30] and motivating a violator to perform a self-repair through an apology could be a more effective strategy for a robot to positively influence conflict dynamics. A framework that can be used to distinguish repair and other emotion regulation strategies more generally is the Process Model of Emotion Regulation by Gross [17]. This highly influential framework distinguishes five types of emotion regulation strategies based on the stage at which they intervene in a general emotion generation process. We list the five strategies in the table below and adapted the framework to distinguish different repair strategies. Previous research has found that the use of response-focused regulation such as suppression of emotion comes at high costs for a person's health and interpersonal relationships [16]. Since we found a tendency within teams to use suppression as a regulation strategy, future research could explore if a robot can move a team towards using more healthy, antecedent-focused emotion regulation strategies for example by inviting team members to discuss violations more openly. A key question here would be to explore whether such robot-driven emotion regulation should be done explicitly (i.e. by setting norms or by giving specific instructions) or implicitly by example or by re-directing attention.

Table 4: Emotion Regulatory Repair Strategies

	Description	Example
Situation Selection	Repairing a violation by avoiding a negative stimulus.	"I think you should try to avoid this topic as it seems to generate a lot of stress. "
Situation Modification	Repairing a violation by changing a negative stimulus.	"Why don't you leave the room for a moment and take a breath"
Attention Deployment	Repair by focusing attention away from a negative stimulus e.g. through a joke or direct advice.	"If something annoys you, try not to pay attention to it."
Cognitive Change	Repair by helping participants see the negative stimulus in a more positive light.	"Being for a long time in this small room here often makes people stressed."
Response Modulation	Repair by advising others to not show the negative impact of a stimulus.	"If you feel annoyed, try to not express it."

6. Conclusion

In this research, we set out to understand whether a robot can act as an emotion regulator for a team and positively influence conflict dynamics. We demonstrated that a robot can actively aid conflict regulation. This ability to aid teamwork extends the range of more task oriented team processes robots are typically employed to support.

Another important contribution of our research is that we explored emotion regulation through specific and targeted repair behaviors rather than through altering mood states as done by others [2], an intervention less well suited to robots. We found that repairs were, in fact, a powerful means of heightening awareness of personal attacks and to some degree improving perceptions of the offending team member, but that repairs of task-directed negative triggers backfired. This may be because the robot's repair was perceived as extreme relative to the negative trigger and/or that the negative trigger itself was not perceived as inappropriate. It may also be possible that such an intervention by a robot was less accepted than it would have been if made by a person. In addition, this calling out of the group member making the attack may in itself be seen as a personal negative trigger towards that person. More research is needed to tease apart these possible explanations and better understand if there is any role for a robot to play in diffusing task-directed negative triggers. We hope this study motivates others to explore how robots can contribute to teamwork by regulating emotions and repairing conflict.

7. Acknowledgments

The authors thank our reviewers and gratefully acknowledge a grant by the Office of Naval Research that supported the work presented here (#N000140710749).

8. REFERENCES

1. Andersson, L. M., & Pearson, C. M.. Tit for tat? The spiraling effect of incivility in the workplace. *Academy of management review*,*24*, 3 (1999), 452-471.
2. Bartneck, C., Van Der Hoek, M., Mubin, O., & Al Mahmud, A.. Daisy, Daisy, give me your answer do!: Switching off a robot. *In Proc. HRI 2007,* ACM Press (2007), 217-222.

3. Barsade, S. G.. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly 47*, 4 (2002), 644-675.
4. Barsade, S. G., & Gibson, D. E. Why does affect matter in organizations?. *The Academy of Management Perspectives 21*, 1 (2007), 36-59.
5. Baxter (2014). www.rethinkrobotics.com/ products/baxter
6. Bradley, M., Lang, P. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavioral Theory and Experimental Psychology*, 25, 1 (1994), 49-59
7. Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. *In Proc. IROS 2005„* 708-713.
8. De Dreu, C. K. Productive conflict: The importance of conflict management and conflict issue. *Using conflict in organizations*, (1997), 9-22.
9. De Dreu, C. K., & Weingart, L. R. Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of applied Psychology 88*, 4 (2003), 741-749.
10. Deutsch, M. Conflicts: Productive and destructive. *Journal of Social Issues 25*, 1 (1969), 7-42.
11. De Wit, F. R., Greer, L. L., & Jehn, K. A. The paradox of intragroup conflict: a meta-analysis. *Journal of Applied Psychology 97*, 2 (2012), 360-390.
12. Duysburgh, P., Elprama, S. A., & Jacobs, A. Exploring the social-technological gap in telesurgery: collaboration within distributed or teams. In *Proc. CSCW 2014,* ACM Press (2014), 1537-1548.
13. Eisenhardt, K. M., Kahwajy, J. L., & Bourgeois, L. J.. How management teams can have a good fight. *Harvard Business Review* 75, 4 (1997), 77–85.
14. Fincannon, T., Barnes, L. E., Murphy, R. R., & Riddle, D. L. (2004). Evidence of the need for social intelligence in rescue robots. In *Proceedings of IROS 2004. IEEE/RSJ* 2, 1089-1095.
15. Gottman, J. M. (1994). *What predicts divorce?* Hillsdale NJ: Lawrence Earlbaum Associates.
16. Gross, J. J. Emotion regulation: Affective, cognitive, and social consequences. Psychophysiology 39, 3 (2002), 281-291
17. Gross, J. J. Emotion regulation: Taking stock and moving forward. *Emotion 13,* 3 (2013), 359-365.
18. Gross, J. J., & John, O. P. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology 85,* 2 (2003), 348-362.
19. Janis, I. L. (1972) *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Oxford, England: Houghton Mifflin.
20. Jehn, K. A.. A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative science quarterly* (1995), 256-282.
21. Jehn, K. A. A qualitative analysis of conflict types and dimensions in organizational groups. *Administrative science quarterly* (1997), 530-557.
22. Jehn, K. A. & Mannix, E. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *The Academy of Management 44*, 2 (2001), 238-251.
23. Jung, M. F., Lee, J. J., DePalma, N., Adalgeirsson, S. O., Hinds, P. J., & Breazeal, C. Engaging robots: easing complex human-robot teamwork using backchanneling. In *Proc. CSCW 2013, ACM Press (2013),* 1555-1566.
24. Hinds, P. J., Roberts, T. L., & Jones, H. Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction 19*, 1 (2004), 151-181.
25. Knutson, B. Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior 20*, 3 (1996), 165-182.
26. Levenson, R. W., & Gottman, J. M. Physiological and affective predictors of change in relationship satisfaction. *Journal of personality and social psychology 49*, 1 (1985), 85.
27. Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In *Proc. HRI 2012,* ACM Press (2012), 367-374.
28. Murphy, R.R. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews.* 34, 2 (2004), 138-153.
29. Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proc. HRI 2009, ACM Press (2009),* 61-68.
30. Ohbuchi, K.-I., Kameda, M., & Agarie, N. Apology as aggression control: its role in mediating appraisal of and response to harm. *Journal of Personality and Social Psychology* 56, 2 (1989), 219-227.
31. Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. Cambridge university press.
32. Ren, H., & Gray, B. Repairing relationship conflict: How violation types and culture influence the effectiveness of restoration rituals. *Academy of Management Review* 34, 1 (2008), 105–126.
33. Schegloff, E. A. When 'others' initiate repair. *Applied linguistics 21*, 2 (2000), 205-243
34. Schwarz, Norbert, Higgins, E. Tory; .(1990). Feelings as information: Informational and motivational functions of affective states. *Handbook of motivation and cognition, Vol. 2.* , (pp. 527-561). New York, NY: Guilford Press
35. Weingart, L., Behfar, K., Bendersky, C., Todorova, G., & Jehn, K. The directness and oppositional intensity of conflict expression. *Academy of Management Review*, (2014)..
36. Zaki, J., & Williams, W. C. Interpersonal emotion regulation. *Emotion 13*, 5 (2013), 803-810.