Exp Econ
DOI 10.1007/s10683-017-9523-6

ORIGINAL PAPER

Third-party manipulation of conflict: an experiment

Piotr Evdokimov[1] · **Umberto Garfagnini**[2]

Received: 1 September 2016 / Revised: 27 February 2017 / Accepted: 6 March 2017

Abstract We design a laboratory experiment in which an interested third party endowed with private information sends a public message to two conflicting players, who then make their choices. We find that third-party communication is not strategic. Nevertheless, a hawkish message by a third party makes hawkish behavior more likely while a dovish message makes it less likely. Moreover, how subjects respond to the message is largely unaffected by the third party's incentives. We argue that our results are consistent with a focal point interpretation in the spirit of Schelling.

Keywords Third-party communication · Experiment · Conflict game

JEL Classification C72 · C92 · D82

1 Introduction

Many conflicts in human history involved acts of third-party provocation. In the midst of World War I, Germany sent what is now known as the "Zimmermann Telegram" to Mexico. This telegram proposed "an understanding on [Germany's]

Electronic supplementary material The online version of this article (doi:10.1007/s10683-017-9523-6) contains supplementary material, which is available to authorized users.

Piotr Evdokimov acknowledges the financial support of the Asociación Mexicana de Cultura A.C. Both authors thank the seminar participants at ITAM, Andrei Gomberg, Ryan Oprea, the editor, and two anonymous referees for useful comments.

✉ Piotr Evdokimov
pevdokim@gmail.com

[1] ITAM, Mexico City, Mexico

[2] University of Surrey, Guildford, England

Published online: 16 March 2017

part that Mexico is to reconquer the lost territory in Texas, New Mexico, and Arizona," and its purpose was to provoke Mexico into conflict with the U.S., thereby delaying the U.S. from going to war with Germany in Europe. The authenticity of the intercepted telegram was publicly confirmed by the German Foreign Secretary, which made the provocation public, if ultimately unsuccessful at engaging Mexico in the war effort. Other examples of third-party provocation include the promotion of the Tutsi minority to positions of power over the Hutu by the Germans, and later the Belgians, in colonial Rwanda,[1] and the instigation of conflict between the Muslims and the Hindu by the British in India.[2]

History is also rife with examples of third-party peacemaking. The promotion of universal peace through nonviolent means such as communication was a central principle of Tolstoyism. The same principle inspired the movements led by Mahatma Gandhi, who communicated with Tolstoy, and Martin Luther King, Jr.[3] The anti-nuclear movement, which gained prominence following the 1945 bombings of Hiroshima and Nagasaki, and protests against the Vietnam War provide other prominent examples of nonviolent and successful calls for peace.

The present paper uses a controlled experiment to study whether and how conflict between two rival players can be manipulated by an interested third party through public communication. Can a third-party provocateur increase the likelihood of conflict by making a strategic provocation? Can a third-party peacemaker reduce its likelihood by making public calls for peace? Although manipulation of conflict can have far-reaching economic and political consequences, surprisingly little empirical work has investigated it in economics. Our first and primary research question is the following:

Question 1 *Can an interested third party manipulate the likelihood of conflict through public announcements?*

In theory, communication by a third party can be effective even if this third party cannot influence the payoffs of the conflicting parties directly. When the third party has private information about one of the conflicting players' incentives and the conflicting players' actions are strategic complements, as they are in many conflict situations, strategic communication can provoke one of the players into being hawkish, which in turn triggers an hawkish response from the player's opponent (e.g., Baliga and Sjöström 2012). We call this the *strategic communication hypothesis*. If the conflicting parties care about the payoffs of the third party, its mere presence could make their behavior more or less hawkish independent of

[1] Ethnic conflict between the Tutsi and the Hutu continues to this day and played a major role in contributing to the Rwandan genocide (Mamdani 2014).

[2] In a discussion of British colonial policy in India, Stewart (1951) quotes Brigadier John Coke as follows: "Our endeavor should be to uphold in full force the (fortunate for us) separation which exists between different religions and races, and not to endeavor to amalgamate them. *Divide et empera* should be the principle of the Indian Government."

[3] President Obama, in a 2010 address to the Parliament of India, has said: "I am mindful that I might not be standing before you today, as President of the United States, had it not been for Gandhi and the message he shared with America and the world." President Jimmy Carter called Martin Luther King, Jr. "the conscience of his generation."

message content (e.g., Bland and Nikiforakis 2015). We refer to this as the *third party social preferences hypothesis*.[4] Uninformative public announcements could also be influential simply through their suggestive power (e.g., Schelling 1980; Charness 2000). This is the *focal point hypothesis*. The three possibilities outlined above motivate our second research question:

Question 2 *What channels underlie the effects of provocation and peacemaking on the likelihood of conflict?*

The baseline condition in our experiment is a 2×2 conflict game in which each player has private information about his cost of being hawkish. In the first treatment, we introduce a third party *peacemaker* that is commonly known to strictly prefer all players to be dovish. In the second treatment, we introduce a third party who is commonly known to prefer Player 1 to be hawkish and Player 2 to be dovish. That is, the third party is a *provocateur* who benefits from conflict. In both treatments, the third party has private information about one of the players and is allowed to make public cheap talk announcements before players choose their actions. In the absence of focal point effects and social preferences toward the third party, the peacemaker's messages have no effect on behavior and his presence does not affect the likelihood of conflict in equilibrium. The game with the provocateur, in contrast, admits a unique informative communication equilibrium in which strategic communication leads to a higher likelihood of conflict.

We find that public cheap talk announcements by an interested third party have a statistically significant effect on behavior in both of our treatments. I.e., we answer Question 1 in the affirmative. To study the underlying channels (Question 2), we explore the message senders' communication strategies as well as the message receivers' responses to the messages. Irrespective of the message sender's identity, we find that communication in the experiment does not convey private information. This allows us to rule out the strategic communication hypothesis.

All message-sending strategies, including those we identify in the data, are consistent with uninformative equilibrium behavior on the part of the message-senders. While uninformative messages are ignored in equilibrium, we find receivers to be more hawkish following hawkish messages and more dovish following dovish messages. While these findings are inconsistent with equilibrium behavior on the part of message receivers, they are consistent with the focal point hypothesis. In particular, we suggest that third party messages focus the attention of the conflicting parties on specific courses of action. To the extent that a player is neither a dominant hawk nor a dominant dove, he is a coordination type that plays a hawkish (resp., dovish) action when the probability of his opponent playing a hawkish (resp., dovish) action is sufficiently high. When it is likely that both players are coordination types, as it is in our experiment, the messages could induce a coordinated response.

[4] In our theoretical analysis of the effect of social preferences, we focus on a special case of the model in Levine (1998) and Charness and Rabin (2002). In the text, we refer to Charness and Rabin (2002)-style social preferences as "social preferences" for short.

To investigate the effect of social preferences toward the third party, we study how the message receivers are affected by the identity of the message sender. Conditional on a dovish message being sent, we find that the receivers are no less likely to be hawkish when the message is sent by a peacemaker than when it is sent by a provocateur. This provides evidence against the third party social preferences hypothesis.[5]

Our results shed light on the channels through which manipulation of conflict can occur and highlight several behavioral regularities. In particular, they complement the theoretical work of Baliga and Sjöström (2012) by showing that third-party cheap talk communication can be influential even when it conveys no private information. The rest of our paper is structured as follows. Section 2 reviews the relevant theoretical and experimental literature. In Sect. 3, we discuss our experimental design. In Sect. 4, we present our results. Section 5 discusses the results and possible directions for future work.

2 Related literature

Schelling (1980) recognized the possible effect of third party communication,[6] arguing that "when there is no apparent focal point for agreement, ...[the third party] can create one by his power to make a dramatic suggestion." McAdams and Nadler (2005) explore Schelling's idea about third party communication in a game of conflict under complete information.[7] We depart from McAdams and Nadler in several important directions. First, players in our games have incomplete information about their opponents' incentives, which allows us to investigate strategic communication. Second, the third player in our experiment is biased, and variation in his payoffs allows us to explore the effect of social preferences on behavior. Third, the incentives in our experiment are different from those in McAdams and Nadler. While we consider a game of conflict in which the actions are strategic complements, McAdams and Nadler focus on the case of strategic substitutes.

[5] We find other evidence against third party social preferences. Exploiting the observation that communication is uninformative, we derive equilibrium predictions about the effect of social preferences in both of our treatments. All else equal, social preferences toward the third party should lead to more hawkish behavior in the presence of a provocateur and less hawkish behavior in the presence of a peacemaker, compared to the baseline conflict game. While the presence of a peacemaker made hawkish behavior less likely, the presence of the provocateur had no effect on the likelihood of conflict. Moreover, even in the presence of social preferences, uninformative communication is disregarded in equilibrium. This is not what we observe in the data.

[6] "[...] a third player with a payoff structure of his own who is given an influential role through his control over communication." Schelling (1980, p. 144)

[7] The control in McAdams and Nadler (2005) is a symmetric Hawk-Dove game with complete information which has two pure strategy Nash equilibria: (*Hawk*, *Dove*) and (*Dove*, *Hawk*). Their treatments are: (1) spinning a wheel which selects (*Hawk*, *Dove*) or (*Dove*, *Hawk*) with 50–50 chance in front of the subjects; (2) using an additional subject, called the leader, who opens an envelope containing a recommendation to either play (*Hawk*, *Dove*) or (*Dove*, *Hawk*); (3) same as (2) but with the leader publicly chosen as the highest scorer in a test.

A strand of literature employs recommended play to study equilibrium selection. Van Huyck et al. (1992) and Brandts and MacLeod (1995) investigate the effect of recommending equilibrium strategies to players. Cason and Sharma (2007) study whether private rather than public recommendations can implement a correlated equilibrium in a hawk-dove game. In these papers, recommendations are made by the experimenter whose incentives may be unknown to the subjects, while the main ingredient of our experiment is strategic recommendations by a third player with a commonly known bias.

Bland and Nikiforakis (2015) study how third party externalities affect behavior in a two player coordination game. If third party messages have no strategic content (as observed in our data), the presence of the third party in our experiment can be thought of as an externality that affects subjects through social preferences. Consistent with the findings of Bland and Nikiforakis (2015), we find that this externality does not affect subjects' behavior in the direction predicted by social preferences in the case where the active players' incentives are misaligned with those of the third party (the provocateur). However, unlike the third party in Bland and Nikiforakis (2015), the third party in our experiment is active rather than passive. Moreover, the third parties in our experiment affected subjects' behavior by virtue of their messages and not simply their presence. Galbiati and Vertova (2008) study a public goods game with exogenous obligations. Their experiment is vaguely related to ours because the obligations can be viewed as third party suggestions. However, at least in theory, some of the suggestions in our paper have equilibrium effects while those in Galbiati and Vertova (2008) are always ineffective.

3 Experimental design

The basic building block of our experiment is the two player conflict game shown in Table 1, which is adapted from Baliga and Sjöström (2012). Two players, Player 1 and Player 2, simultaneously choose one of two actions, *Hawkish* and *Dovish*. A dovish player gains 95 points against a dovish opponent, and gains only 10 points if the opponent is hawkish. Player i has to pay a cost of $c_i \geq 0$ for being hawkish. Being hawkish against a hawkish opponent results in a payoff equal to $95 - c_i$, while a gain of 10 additional points occurs if the opponent is dovish. We assume that the cost parameter c_i is private information of Player i. Thus, Player i knows c_i but does not know c_j. It is common knowledge that the costs c_1 and c_2 are independently drawn from the same uniform distribution with full support on [0, 95].[8]

We incorporate incomplete information in the experiment for several reasons. First, this is a realistic feature of many conflict situations.[9] The assumption that each player has private information about his cost of being hawkish is a convenient, albeit simplistic, way to model uncertainty. Second, incomplete information is necessary for studying the effect of strategic communication on manipulation of

[8] In the experiment, the software approximated costs up to 3 decimal digits.

[9] Does a country possess weapons of mass destruction? How much political will does a country have to engage in a conflict?

Table 1 Payoff matrix when no third party is present

		Player 2	
		Hawkish	*Dovish*
Player 1	*Hawkish*	$95 - c_1, 95 - c_2$	$105 - c_1, 10$
	Dovish	$10, 105 - c_2$	95, 95

Table 2 Payoff matrix in the presence of a peacemaker

		Player 2	
		Hawkish	*Dovish*
Player 1	*Hawkish*	$95 - c_1, 95 - c_2, \mathbf{0}$	$105 - c_1, 10, \mathbf{10}$
	Dovish	$10, 105 - c_2, \mathbf{90}$	$95, 95, \mathbf{150}$

conflict. Third, with complete information, three cases are possible depending on the realizations of cost parameters: (i) both players have a commonly known dominant strategy; (ii) one player has a dominant strategy while the other player is a coordination type; or (iii) both players are coordination types. The first two cases are uninteresting, while the latter deals with subjects' ability to coordinate on an equilibrium and has already received significant attention in the literature. Incomplete information, by contrast, generates a unique equilibrium prediction in our environment.

The parametrization of the game in Table 1 implies that players' actions are strategic complements. That is, each player's best response is increasing in the belief that the opponent is hawkish. This is a natural assumption in a game of conflict. As shown in Online Appendix A, this game has a unique Bayesian Nash equilibrium in cutoff strategies: Player i chooses the hawkish action if and only if $c_i \leq 47.5$. Note that the equilibrium cutoff strategy is also a best response to the belief that the opponent is simply randomizing over actions with equal probability.

We introduce two treatments to study Question 1. In the first treatment, equilibrium uniquely predicts manipulation of conflict to be ineffective. In the second, manipulation of conflict is theoretically possible. Specifically, the first treatment adds a third player to the baseline game, with payoffs as highlighted in bold in Table 2. For Players 1 and 2, the payoffs are identical to those in the two player game. The third party, Player 3, is a peacemaker who prefers both Player 1 and Player 2 to choose the dovish action, and his payoffs are common knowledge.

While the third party cannot take any action that affects payoffs directly, it can make public cheap talk announcements before Player 1 and 2 make their decisions. The timing of the game is as follows. First, Nature draws c_1 and c_2. Second, Player 1 and the third party observe c_1, while Player 2 observes c_2. Third, the third party sends a publicly observable cheap talk message $m \in \{Hawkish, Dovish\}$.[10] Fourth, Player 1 and Player 2 simultaneously choose their actions.

[10] Baliga and Sjöström (2012) show that it is without loss of generality to assume that the third party's message space contains only two messages, one of which makes Player 2 more likely to be hawkish.

Table 3 Payoff matrix in the presence of a provocateur

		Player 2	
		Hawkish	*Dovish*
Player 1	*Hawkish*	$95 - c_1, 95 - c_2, \mathbf{90}$	$105 - c_1, 10, \mathbf{150}$
	Dovish	$10, 105 - c_2, \mathbf{0}$	$95, 95, \mathbf{10}$

Theory predicts that communication by the peacemaker can only be uninformative. This is because the peacemaker always wants to send the message which induces dovish behavior by both players. As the only such message is the dovish message, his communication strategy does not reflect any private information. The equilibrium cutoff remains the same as in the two player game. A comparison between behavior in the two player game and the game with the peacemaker can therefore be used to address Question 1.

In the second treatment, we introduce a third player with different preferences. The payoff matrix of this new game is shown in Table 3. For Players 1 and 2, the payoffs are again identical to those in the two player game, while the payoffs of Player 3 are highlighted in **bold**. In this treatment, the third party is a provocateur who strictly prefers Player 1 to choose the hawkish action. He also prefers Player 2 to choose the dovish action regardless of what Player 1 does. The provocateur's payoffs are common knowledge. The timing of events is the same as in the treatment with a peacemaker. Thus, the provocateur makes a public cheap talk announcement after observing Player 1's cost but before Player 1 and 2 make their decisions.[11] Fourth, Player 1 and Player 2 simultaneously choose their actions.

As discussed in Online Appendix A, this three player game has a unique informative communication equilibrium[12] in which the provocateur sends hawkish messages if and only if Player 1's cost is in some intermediate range.[13] Equilibrium communication is influential and affects behavior in two ways: (1) both Player 1 and Player 2 respond to the hawkish message by choosing a hawkish action; (2) both players are more hawkish in this equilibrium than in the game without the third party, regardless of whether the provocateur actually sends a hawkish message.

[11] The public nature of communication is crucial. If the provocateur communicated privately with each player, communication would necessarily be uninformative. To see this, first note that the provocateur would always send to Player 2 the message that maximizes the probability that Player 2 will choose the dovish action, which is commonly known to be the provocateur's most preferred action by that player. As a result, private communication with Player 2 will be uninformative. Consequently, Player 1 will also ignore the provocateur's message.

[12] Recall that a babbling equilibrium always exists in cheap talk games.

[13] Intuitively, when c_1 is very high or very low, Player 1 has a strictly dominant strategy, and the provocateur cannot influence Player 1's behavior. He will therefore send the message that is more likely to induce Player 2 to choose the dovish action. When c_1 is in some intermediate range, the provocateur can potentially affect Player 1's behavior. Note that he would only want to affect this behavior in the direction of being more hawkish. In this intermediate range, the provocateur will thus use his public announcement to induce Player 1 to be hawkish by *provoking* Player 2. As actions are strategic complements, an increase in the probability that Player 2 is hawkish will indeed lead Player 1 to best respond with a more hawkish stance.

3.1 Implementation

Our experimental design employed a baseline two player condition and two treatments. In the baseline condition, subjects played the conflict game described in Table 1 in groups of two with no third party present. To keep the labeling neutral for the subjects, the hawkish action was denoted by A and the dovish action by B. At the beginning of each round, each subject observed her/his cost parameter and then chose between a hawkish and a dovish action simultaneously with his opponent. At the end of the round, the subject observed a complete summary of the information about the round, including her/his own and the opponent's chosen action, the number of points gained in the round, and the cumulative number of points gained up until (and including) that round.

In the peacemaker treatment, subjects were matched in groups of three players (Player 1, Player 2, and Player 3). Player 3 played the role of a third party peacemaker with payoffs as in Table 2. The timing of the peacemaker treatment was as follows: (i) Player 1 and Player 3 observed the realized cost of Player 1, and Player 2 observed its own cost; (ii) Player 3 sent a message m which was restricted to the set $\{A, B\}$; (iii) Player 1 and Player 2 observed the message sent by Player 3 and then simultaneously chose between A and B (a hawkish and a dovish action); (iv) Player 1, Player 2, and Player 3 observed a complete summary of the information about the round, as in the two player condition, with additional information about Player 3's message. The timing and design of the provocateur treatment were similar, with the exception that Player 3's payoff matrix was as in Table 3.

We restricted Player 3's message space to be equal to the action space available to Player 1 and Player 2. From a theoretical point of view, this restriction is without loss of generality and therefore leaves all predictions unchanged. While the restriction may nevertheless be empirically relevant, previous studies have found that subjects tend to interpret messages in cheap talk games using a natural language (see, e.g., Blume et al. 2001).

Each session of the experiment started with either 13 rounds of the two player baseline, 13 rounds of the peacemaker treatment, or 13 rounds of the provocateur treatment. This allowed us to perform a between-subjects analysis of the effect of introducing a third party into the two player game. The first round in each case was an unpaid practice round, while the remaining 12 rounds were incentivized. Subjects were randomly and anonymously matched with randomly-assigned roles in the beginning of each round. The experiment was designed with multiple rounds of play (as opposed to one-shot interactions) to facilitate learning. We allowed subjects to experience different roles to facilitate a better understanding of the incentives of every player in the game. While we did not employ a predefined rule to govern role switching, all subjects were assigned a different role at least once in each session.

Motivated by a line of research investigating how learning in one game transfers to behavior in a different, related game (see, e.g., Cooper and Kagel 2003; Rankin et al. 2000; Rick and Weber 2010), subjects interacted for an additional 13 rounds in the second half of the experiment. The subjects initially assigned to the two player condition played an additional 13 rounds (one unpaid, and 12 incentivized) of a

game with a third party, who could be either a peacemaker or a provocateur. This allowed us to assess whether previous exposure to the two player conflict game affected subjects' responsiveness to third-party messages. Such a robustness check is particularly pressing in the case of a peacemaker, whose messages in theory have no strategic content. Subjects initially assigned to one of the treatments with a third party subsequently played the baseline conflict game with no third party.[14]

While the subjects were told in advance that the experiment will be divided into two parts, they were not given the instructions for the second part of the experiment until the last round of the first part of the experiment was finished.[15] Therefore, subjects in all sessions expected something to happen in the second half. To maximize recruiting and not restrict ourselves to only 12 and 18 player-large sessions, we did not exclude subjects from participating in cases where the number of subjects showing up to a session was not divisible by either 2 or 3. When the number of subjects in a session was not divisible by 2 (for the baseline game) and 3 (for the game with a third party), some subjects were randomly chosen to sit out in every round. This was clearly explained to the subjects that participated in every session. As discussed in Sect. 4 below, we control for heterogeneity in experience in our econometric analysis and find that this has little effect on our main results.

Each session of the experiment started with subjects signing the consent forms, reading the instructions, and completing an incentivized quiz.[16] The subjects' earnings were determined as follows. Every subject was guaranteed a 30 Mexican pesos (≈US$2 at the time of the experiment) show up fee in addition to the earnings from the quiz (1 Mexican peso for each correct answer). These earnings were called the subject's "guaranteed earnings." In addition, in each (non-practice) round of the game, the decisions of each subject and his/her matched partners led the subject to gain a number of points. The subject's "additional earnings" were determined as follows:

$$Additional\ earnings = \frac{Total\ points\ gained\ during\ the\ experiment}{10}.$$

Therefore, subjects gained 10 Mexican pesos for each 100 points.

To avoid confusion, we highlight here that no within-subjects comparisons across different treatments are made in our statistical analysis. That is, we never compare how the same subject behaved in a two player game and a game with a third party. Whenever we analyze our experimental data, we either focus on behavior in the first half of the experiment, or behavior in the second half of the experiment (the latter

[14] While we do not analyze behavior in the "two player after provocateur" and "two player after peacemaker" conditions, we collected this data to guarantee both consistency in the description of the experiment to the different subjects and also comparable earnings. In particular, participants in the sessions where the two player condition came first were told that the experiment will have a second half.

[15] Specifically, the instructions stated: "We will provide you with instructions for the second part at the end of the first part of the experiment."

[16] The quiz tested the subjects' understanding of the experiment with eight questions for the two player baseline and ten questions for the provocateur and peacemaker treatments. The quiz was administered only at the beginning of the first part of each session. For the second part of each session, subjects were given the new instructions with additional time to read them but no additional quiz. All the instructions for the experiment can be found in the Online Appendix.

when we study how prior experience with the two player game affected subjects' responses to their received messages).

3.2 Predictions

The model outlined above suggests the following set of predictions, which we refer to as the *strategic communication hypothesis*:

Prediction 1 *Communication is related to private information and effective at influencing the conflicting parties' behavior when the third party is a provocateur.*

Prediction 2 *Communication is unrelated to private information and has no effect on the conflicting parties' behavior when the third party is a peacemaker.*

The considerations in the introduction (Question 2) motivate the question of how introducing social preferences affect the predictions above. To answer this question, we focus on a special case of the model in Levine (1998) and Charness and Rabin (2002), where a player puts a weight λ on his own payoffs and weights $(1-\lambda)/n$ on the payoffs of other players if the number of other players is n.[17] We call this the *third party social preferences hypothesis*.

As we show in Online Appendix A.3, an increase in the degree of social preferences (lower λ) reduces the equilibrium cutoff and leads to less expected hawkish behavior in the two player baseline. Because communication is uninformative in the treatment with a peacemaker, an increase in the degree of social preferences also decreases hawkish behavior in this treatment. Thus, allowing for social preferences leaves Prediction 1 unaffected. The treatment with a provocateur gives scope for strategic communication, thereby complicating the predictions of a social preferences-based model. Focusing on the case where the provocateur's messages are not strategic, we show in the Online Appendix that an increase in the degree of social preferences increases the equilibrium cutoff and leads to more expected hawkish behavior. Thus, Prediction 2 can be generated both by a model with strategic communication and no social preferences *and* a model that allows for social preferences without strategic communication.[18]

As argued by Schelling (1980), public announcements can exert an influence simply through their suggestive power. This suggests that communication in our experiment might have an effect on behavior regardless of the identity of the message sender, and regardless of whether communication is informative in the sense of revealing private information. We refer to this as the *focal point hypothesis*.

[17] $n = 1$ without a third party and $n = 2$ when a third party is present. This formulation makes the natural assumption that players care for their opponents' payoffs equally.

[18] As explained in detail in Sect. 4.1, we find that provocateurs do not communicate strategically.

4 Results

The experiment was conducted at Instituto Tecnológico Autónomo de México in Mexico City in the Spring semester of 2015 using the software *z*-Tree (Fischbacher 2007). A total of 13 experimental sessions were conducted with an average of 15 subjects per session with no subject participating more than once. An average session lasted for 75 minutes with an individual average payment of 215 pesos (including a 30 peso show-up fee).[19]

101 subjects played the two player game in the first half of the experiment, 53 played the game with the provocateur in the first half, and 46 played the game with the peacemaker in the first half. As mentioned above, sessions that started with the two player baseline were randomly assigned into one of two possible continuation treatments in the second half. 51 of the subjects in these sessions were assigned to play the game with the provocateur and 50 to play the game with the peacemaker. The session information is summarized in Table 4.

Our main results can be summarized as follows. First, contrary to the strategic communication hypothesis, we find that neither the peacemakers' nor the provocateurs' messages were significantly affected by their observed costs. This implies that the messages could not have been used by Player 1 and Player 2 strategically. I.e., if subjects behaved according to equilibrium, the equilibrium was uninformative. Second, despite the fact that the messages were uninformative of costs, the choices of Player 1 and Player 2 were affected by what messages they received. While this result runs contrary to the uninformative equilibrium predictions, it is consistent with the predictions of the focal point hypothesis. Given that subjects' messages were uninformative of costs, social preferences toward the third party predict that the introduction of a provocateur increases the incidence of hawkish behavior while the introduction of a peacemaker diminishes it.[20] Our results reject these predictions. In particular, we find that conditional on observing a dovish message, subjects' behavior was unaffected by the message sender's identity. Moreover, introduction of a provocateur left the likelihood of hawkish behavior unaffected on average. Sections 4.1 and 4.2 describe our analysis in detail, while Sect. 5 provides a discussion and concludes.

Most of our analysis below focuses on the data collected in the first half of the experiment. When results from the second half of the experiment are discussed, this is made explicit in the text.

[19] Average earnings amounted to approximately 14 US dollars per subject at the time of the experiment. The minimum wage in Mexico is small, about 70 pesos per day. For a better reference point, consider that a 15km Uber ride from the house of one of the authors to the airport cost around 80 pesos at the time of the experiment.

[20] Recall from the previous section that both uninformative equilibrium with social preferences and informative equilibrium without social preferences predict that the introduction of a provocateur increases the incidence of hawkish behavior.

Table 4 Subjects per treatment

Two player in first half, *Provocateur* in second half	Three sessions (16, 17, 18 subjects)
Two player in first half, *Peacemaker* in second half	Three sessions (11, 18, 21 subjects)
Provocateur in first half, *Two player* in second half	Three sessions (9, 13, 13, and 18 subjects)
Peacemaker in first half, *Two player* in second half	Four sessions (14, 16, 16 subjects)

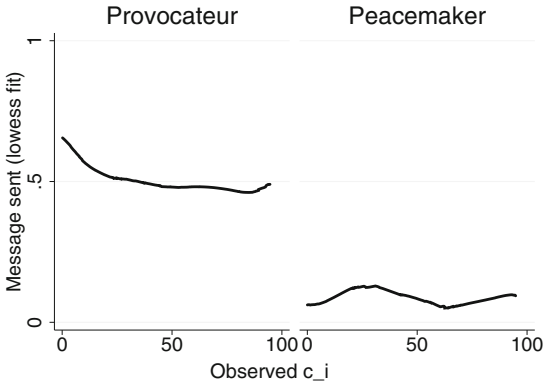
4.1 Messages

We first explore the communication strategies used by the third parties. In the first half of the experiment, 49.51% of the provocateurs' and 8.93% of the peacemakers' messages were hawkish, and these proportions were significantly different ($P<0.001$ in a Fisher's exact test).[21] We observe similar results in the treatments where subjects interacted with the peacemaker or the provocateur in the second half of the experiment, that is, following experience with the two player game. Specifically, we find no significant difference in how provocateurs ($P = 0.316$ in a Fisher's exact test) or peacemakers ($P = 0.699$ in a Fisher's exact test) sent messages before and after experience with the two player game. 54.69% of the provocateurs' and 7.29% of the peacemakers' messages were hawkish in the second half of the experiment, and, as for inexperienced subjects, these proportions were significantly different ($P<0.001$ in a Fisher's exact test).

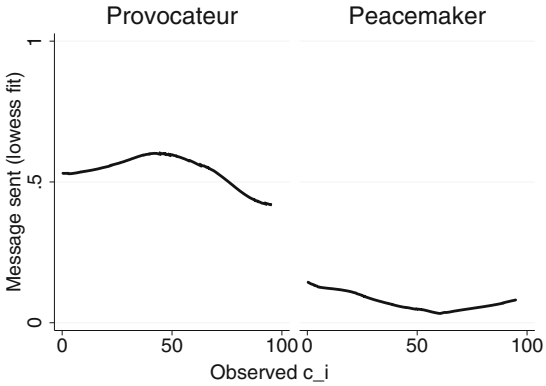
We can compare the empirical distributions of the messages observed in the experiment to those predicted by equilibrium behavior. In the experiment, $\approx 11\%$ of the provocateur's messages should have been hawkish according to the strategic communication hypothesis. Regardless of whether we focus on the first or second half of the experiment, and regardless of how we compute the standard errors, we find that provocateurs sent significantly more dovish messages than predicted by the theory ($P<0.01$ in every test we ran). Moreover, we cannot reject the hypothesis that 50% of the provocateurs' messages were hawkish ($P > 0.1$). Whereas peacemakers in theory should send only dovish messages, we find that the fraction of hawkish messages sent in the experiment was significantly greater than zero ($P<0.01$). On the other hand, because the overall fraction of hawkish messages sent by peacemakers was small (approximately 8%), such messages can plausibly be attributed to error.

For a more direct test of the strategic communication hypothesis, we need to study how private information affected subjects' messages. Figure 1 shows

[21] In every instance the results of a Fisher's exact test are reported in the text, the results are robust to estimating a logit regression with session clustered errors. I.e., none of our results rely on assuming that the observations are independent at the subject-decision level.



(a) Subjects before experience with the two player game.



(b) Subjects experienced with the two player game.

Fig. 1 Messages sent as a function of observed costs. The *bold black lines* represent estimated nonlinear relationships obtained using a locally weighted regression (lowess), with a hawkish message coded as one and a dovish message as zero. We find little relationship between provocateurs' observed costs and their sent messages before or after experience with the two player game. **a** Subjects before experience with the two player game. **b** Subjects experienced with the two player game

estimated nonlinear relationships between the third parties' observed costs and the messages they sent before and after experience with the two player game (in the figure, a hawkish message is coded as one and a dovish message as zero). The figure suggests that neither the provocateurs' nor the peacemakers' messages were strongly affected by private information (Player 1's cost of being hawkish). This observation provides evidence against the strategic communication hypothesis in the case where the third party is a provocateur. Indeed, while informative equilibrium behavior predicts that the provocateur sends a dovish message after observing costs close to the extremes of the distribution, and a hawkish message after observing costs in an intermediate range, the estimated fit for inexperienced

provocateurs suggests *less* hawkish messages close to the extremes. For provocateurs experienced with the two player game, the fit shows a non-linear trend in the right direction, but the trend is not significant: we obtain P-values greater than 0.1 on both the linear and the squared cost term in either an OLS or a logit regression of a hawkish message dummy against cost and cost squared.[22] Moreover, the lowess fit looks little like the sharply inverted-U shape predicted by equilibrium. We summarize these findings as follows:

Observation 1 *The messages of both peacemakers and provocateurs were not informative of private information.*

We interpret the results that follow in light of this observation. In particular, we henceforth assume that the equilibrium is uninformative both in the game with a peacemaker and in the game with a provocateur when discussing the theoretical predictions, including the predictions of social preferences.

4.2 Responses to messages

We first study the effect of introducing a third party on the behavior of Players 1 and 2 without controlling for message content. Taking data from the first half of the experiment, we compare the probabilities of choosing a hawkish action as a function of whether a peacemaker, a provocateur, or no third party was present. To compute the standard errors, we use a logit model in which a hawkish action dummy (=1 if a hawkish action was chosen) is regressed against a constant term, a provocateur dummy (=1 if a provocateur was present), and a peacemaker dummy (=1 if a peacemaker was present). As in all of the statistical analysis described below, we focus on non-practice rounds and cluster the standard errors by session. All of our main results are robust to different model specifications.[23]

We find that the probability of choosing the hawkish action was 36.4% in the two player game, 18.5% in the presence of a peacemaker, and 33.8% in the presence of a provocateur. Inconsistent with uninformative equilibrium behavior *without* social preferences, the presence of a peacemaker had a significant effect of reducing hawkish behavior ($P < 0.001$). Inconsistent with uninformative equilibrium behavior *with* social preferences, the effect of introducing the provocateur was not significant ($P = 0.5372$). Notice that these statistical comparisons are made between subjects, as they only make use of data from the first half of the experiment. We now provide evidence that the message receivers' behavior was consistent with using messages as focal points.

How were subjects affected by third-party messages? To answer this question, we add a hawkish message dummy (=1 if a hawkish message was sent), and a dummy representing the interaction between the peacemaker dummy and a hawkish message dummy (=1 if a hawkish message was sent in the presence of a

[22] This is true regardless of whether we use independent, clustered, or boostrapped standard errors. We also get non-significant coefficients if we pool the first and second half of the data.

[23] I.e., we get qualitatively and quantitatively similar results if we use a logit model with subject-level random effects or bootstrapped standard errors.

peacemaker) to the logit model described in the first paragraph of this section. Notice that these dummy variables allow for all possible peacemaker/provocateur and hawkish/dovish message combinations.

The results of the relevant statistical comparisons are reported in the first column of Table 5. The table reports the probabilities of choosing a hawkish action as a function of what message was sent, if any, and what the message sender's identity was. The first column does so without differentiating between message receivers in the role of Player 1 and those in the role of Player 2, while the second and third columns differentiate by player type. The stars in each cell summarize the results of a (between-subjects) statistical comparison of the probability in the cell to that in the baseline two player condition.

The comparisons in the first column of the table suggest that a hawkish message was effective at producing more hawkish behavior regardless of whether it was sent by a peacemaker ($P<0.001$, comparing the first and second row) or a provocateur ($P<0.01$, comparing the first and fourth row). Similarly, a dovish message was effective at producing more dovish behavior regardless of whether it was sent by a

Table 5 Probabilities of choosing the hawkish action in different conditions of the experiment

	(1) Players 1 and 2	(2) Player 1		(3) Player 2
Two player	0.364	0.355	$\not<$	0.372
	(0.0265)	(0.0298)		(0.0275)
Hawkish msg.	0.600****	0.600**		0.600****
by peacemaker	(0.0643)	(0.102)		(0.0519)
	$\vee$****	$\vee$****		$\vee$****
Dovish msg.	0.144****	0.0980****	$<$***	0.190****
by peacemaker	(0.00690)	(0.0100)		(0.0209)
Hawkish msg.	0.470***	0.465***	$\not<$	0.475**
by provocateur	(0.0236)	(0.0272)		(0.0343)
	$\vee$****	$\vee$****		$\vee$**
Dovish msg.	0.209**	0.175****	$\not<$	0.243
by provocateur	(0.0565)	(0.0403)		(0.0752)
Observations	1920	960		960

The standard errors are computed using the logit model described in the text with observations clustered by session (in the second and third column, the explanatory variables in the logit model are interacted with a player type dummy)

The stars next to the coefficients are associated with tests of the null hypothesis of each probability being equal to that in the baseline "two player" condition. The stars next to the $\vee$ and $<$ symbols refer to across-row and across-column comparisons. The results suggest that hawkish behavior is more likely following a hawkish message and less likely following a dovish message, both overall and for Players 1 and 2 considered separately

Session-clustered standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$, **** $p<0.001$

peacemaker ($P<0.001$, comparing the first and third row) or a provocateur ($P<0.05$, comparing the first and fifth row).

Another way to assess the effectiveness of third party messages is to compare behavior after seeing a hawkish and a dovish message holding the identity of the third party fixed. We find a significant difference in responses to hawkish and dovish messages sent by a peacemaker ($P<0.001$, comparing the second and third rows) and a provocateur ($P<0.001$, comparing the fourth and fifth rows). These across-row comparisons are summarized by the $\vee$ symbol in Table 5. We conclude that message receivers responded to both acts of peacemaking and provocation, which, as discussed in Sect. 5, is consistent with a focal point interpretation, and summarize these findings as follows:

Observation 2 *Despite not being informative of costs, the messages of both provocateurs and peacemakers were effective at influencing behavior.*

Comparing the probability of choosing a hawkish action following a dovish message sent by a peacemaker and that sent by a provocateur (third and fifth rows of the first column of the table), we find no significant difference in behavior ($P = 0.2540$). This suggests that the identity of the third party had no effect on message receivers' behavior in the case a dovish message was sent. While the analogous comparison in the case of a hawkish message shows a marginally significant effect ($P<0.1$), only 15/168 (8.93%) messages sent by inexperienced peacemakers were hawkish. This suggests that any effect of the message sender's identity on behavior was small. We summarize this as follows:

Observation 3 *When a dovish message was sent, Player 3's payoffs had no significant effect on the behavior of Player 1 and Player 2.*

Observation 3 cannot be reconciled with social preferences toward the third party. Since both the provocateur's and the peacemaker's messages had no strategic content, the only difference from the point of view of the message receiver between a dovish message sent by a provocateur or a peacemaker can be attributed to the difference in payoffs of the respective third parties. That subjects were not more hawkish when the message was sent by the provocateur (who desired the hawkish outcome more) than by the peacemaker (who desired it less) suggests that they had little regard for the message senders' payoffs.[24]

In theory, Player 1 and Player 2 respond to the provocateur's messages differently (see Baliga and Sjöström (2012) and the Online Appendix). To test this prediction, we added a Player 2 dummy to the logit model above and interacted it with each of the other explanatory variables. We used the augmented model to calculate the probabilities and statistical comparisons analogous to those in the first column of Table 5 separately for Player 1 and Player 2. These are reported in the second and third columns of the same table. We find that both Player 1 and Player 2 showed significant responses to their received messages, suggesting that the overall

[24] Role switching, which occurs in our experiment, has been shown to decrease social preferences in trust games (see, e.g., Burks et al. 2003). However, it cannot explain why behavior in each of the experimental conditions is consistent with social preferences of Player 1 and Player 2 toward each other.

results are not driven by any particular player type. Specifically, looking at Player 1 showed more hawkish behavior following a hawkish than a dovish message sent by a peacemaker ($P<0.001$, comparing second and third rows of the second column) and a provocateur ($P<0.001$, comparing fourth and fifth rows). Similarly, Player 2 showed more hawkish behavior following a hawkish than a dovish message sent by a peacemaker ($P<0.001$, comparing second and third rows of the third column) and a provocateur ($P<0.05$, comparing fourth and fifth rows). These across-row statistical comparisons are summarized by the $\vee$ symbol in Table 5. While the two types of players responded differently to a dovish message sent by a peacemaker ($P<0.01$), there was no significant difference in the case of a hawkish message sent by a peacemaker ($P = 1$) or a hawkish message sent by a provocateur ($P = 0.8055$). When a dovish message was sent by a provocateur, the difference in responses was small and not statistically significant ($P = 0.1089$).[25]

4.3 Robustness checks

We now discuss the results of several robustness checks of Observations 2 and 3. First, we re-estimate the model used in the first column of Table 5 controlling for the message receiver's observed cost. The results, reported in the first column of Table 6, show that this has little effect on our main observations. Thus, when the message sender was a peacemaker, subjects were still more likely to be hawkish following a hawkish than a dovish message ($P<0.001$ comparing the estimated probabilities in the second and third rows). The same is true when the message sender was a provocateur ($P<0.01$, comparing the fourth and fifth rows). This confirms Observation 2. In line with Observation 3, the message sender's identity still shows no significant effect on how the message receiver responded to a dovish message ($P = 0.2161$).

In the second and third columns of Table 6, we control for observed costs as well as interactions between each of the dummy variables used in the model of the first column of Table 5 and a dummy for whether the observation falls in the first six or the last six of the twelve non-practice matches of the first half of the experiment. We find that the estimated probability of choosing a hawkish action does not differ across early and late matches when a hawkish message was sent by a peacemaker ($P = 0.1891$), when a hawkish message was sent by a provocateur ($P = 0.4741$), when a dovish message was sent be a peacemaker ($P = 0.1083$), or when a dovish message was sent by a provocateur ($P = 0.1749$). While we do find a marginally significant difference in how subjects behaved in the baseline two player condition in early and late matches ($P<0.1$), there is little learning overall. This suggests that observations across early and late matches of the first half of the experiment can be pooled.

[25] Controlling for costs, we find that the latter P-value becomes statistically significant ($P<0.01$); the other statistical comparisons across Player 1 and Player 2 remain qualitatively similar to those reported in the preceding paragraph. We also find that Result 3 holds separately for Player 1 and Player 2. That is, if we look at the effect of the dovish message sender's identity separately for Player 1 and Player 2, we find no significant effects ($P > 0.1$ in both cases). We discuss other robustness checks in Sect. 4.3 below.

Table 6 Robustness checks

	(1) Contr. for cost	(2) Matches 1-6		(3) Matches 7-12	(4) No sit-outs	(5) Prev. not sender		(6) Prev. sender	(7) After two player game
Two player	0.359	0.347	$<$	0.370	0.390	0.359		-	-
	(0.0281)	(0.0299)		(0.0275)	(0.0215)	(0.0281)		-	-
Hawkish msg. by peacem.	0.573****	0.636****	$\not>$	0.491	0.522	0.542****	$\not<$	0.621***	0.300
	(0.0427)	(0.0743)		(0.0924)	(0.157)	(0.0246)		(0.0769)	(0.095)
	$\vee$****	$\vee$****		$\vee$****	$\vee$***	$\vee$***		$\vee$****	$\not\vee$
Dovish msg. by peacem.	0.150****	0.169****	$\not>$	0.129****	0.153****	0.150****	$\not<$	0.151****	0.291
	(0.0105)	(0.0110)		(0.0205)	(0.0168)	(0.0210)		(0.0180)	(0.0424)
Hawkish msg. by prov.	0.461***	0.489***	$\not>$	0.441	0.439	0.444**	$\not<$	0.494****	0.490
	(0.0218)	(0.0336)		(0.0424)	(0.0277)	(0.0334)		(0.0233)	(0.0245)
	$\vee$***	$\vee$****		$\vee$*	$\vee$***	$\vee$**		$\vee$****	$\vee$****
Dovish msg. by prov.	0.232*	0.221*	$\not<$	0.246*	0.191***	0.231	$\not<$	0.234**	0.369
	(0.0651)	(0.0678)		(0.0624)	(0.0615)	(0.0791)		(0.0456)	(0.040)
Observations	1920	1920		1920	1315	1668		252	768

The results in Table 5 are little affected by controlling for players' observed costs (first column; observed costs are also controlled for in the other columns). There is little difference in how subjects behaved in early and late matches of the first half of the experiment (comparisons across second and third columns). Subjects respond differently to hawkish and dovish messages if the analysis is restricted to subjects who never sat out (fourth column). There is little difference in how subjects responded to messages as a function of whether they acted as a message sender in the previous round (comparisons across fifth and sixth columns). Subjects respond differently to hawkish and dovish messages sent by a provocateur if the analysis is restricted to subjects experienced with the two player game (last column)

Session-clustered standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$, **** $p<0.001$

Recall that our experimental design involved some subjects sitting out in some matches. The fourth column of Table 6 repeats the analysis in the first column of the same table focusing on subjects for which this never happened (146 out of 200, or 73% of the subjects). Our main results hold when analysis is restricted to these subjects. Thus, message receivers responded differently to hawkish and dovish messages sent by a peacemaker ($P < 0.01$; second and fourth rows) and hawkish and dovish messages sent by a provocateur ($P < 0.01$; third and fifth rows). This is in line with Observation 2. Consistent with Observation 3, the message sender's identity had no significant effect on behavior in the case a dovish message was sent ($P = 0.5473$; fourth and fifth rows). These results suggest that our main results were not strongly affected by subjects sitting out.

In the fifth and sixth columns of the table, we control for observed costs as well as interactions between each of the dummy variables used in the model of the first column and a dummy variable for whether the message receiver acted as a message sender in the previous match. We find that the estimated probability of choosing a hawkish action does not differ across subjects who did and did not previously act as message sender ($P > 0.1$ in all cases). This suggests that role switching did not have a substantial effect on subjects' behavior.

Our final set of robustness checks focuses on subjects experienced with the two player game. In the last column of Table 6, we study subjects in the treatment with a provocateur *in the second half of the experiment* (51 subjects) as well as subjects in the treatment with a peacemaker *in the second half of the experiment* (50 subjects), using the same model as that in the first column of the table. Notice that all subjects analyzed in the last column played 13 rounds of the two player game in the first half of the experiment. Notice also that the "two player" cell in the first row of the last column of the table is left empty. This is because we do not have a "two player after two player" condition in the experiment, and hence no baseline to which behavior in the "provocateur after two player" and "peacemaker after two player" conditions can be compared. We can nevertheless study how message receivers experienced with the two player game responded to third-party messages, holding the identity of the message sender fixed.

We find that subjects experienced with the two player game were significantly more likely to be hawkish following a hawkish than a dovish message by a provocateur ($P < 0.001$, comparing the fourth and fifth rows). They did not, however, show significantly different responses to hawkish and dovish messages sent by a peacemaker ($P = 0.8631$, comparing the second and third rows). This is partially consistent with Observation 2. In particular, subjects experienced with the two player game responded to the provocateur's messages despite the fact that these messages were uninformative. Consistent with Observation 3, we find no significant difference in how subjects responded to a dovish message sent by a peacemaker or a provocateur ($P = 0.1874$; third and fifth rows). One interpretation of these findings is that behavior of subjects experienced with the two player game was closer to equilibrium than that of subjects in the first half of the experiment.[26]

[26] Recall from Sect. 3 that the peacemaker's message has no effect on behavior in equilibrium.

Overall, the robustness checks in Table 6 suggest that Observations 2 and 3 are robust to controlling for costs and different kinds of experience. Subjects responded to uninformative messages and showed little difference in how they responded to dovish messages sent by peacemakers and provocateurs.

5 Discussion

Although the most informative equilibrium (MIE) criterion is commonly used in theoretical work, it is inconsistent with the observation that provocateurs' communication strategies are not significantly related to private information (Observation 1). Uninformative equilibrium also cannot explain our results, since it is inconsistent with the observation that the messages of both peacemakers and provocateurs influenced the behavior of Player 1 and Player 2 (Observation 2). The third-party social preferences hypothesis is also inconsistent both with the observation that subjects were not significantly affected by the presence of the provocateur and the observation that behavior after receiving a dovish message was not significantly affected by whether this message was sent by a provocateur or a peacemaker (Observation 3).

One interpretation of our findings is that subjects interpret Player 3's message as an unbiased recommendation. For example, Charness (2000) found that self-serving cheap talk is influential and helps players coordinate on efficient outcomes.[27] While in our experiment the messages are sent by an interested third party that responds to incentives, these messages may still focus the attention of Player 1 and Player 2 on specific actions. If a player expects his opponent to follow the message with sufficiently high probability, the message could serve as a coordination device when the player is a coordination type.[28] If coordination types are sufficiently likely, the messages could be influential on average.[29] If uninformative messages are influential, the peacemaker should still only send dovish messages, while a provocateur should send more hawkish than dovish messages relative to the equilibrium predictions. This is what we observe in the data.

The explanation above leaves open the question of why the provocateur was equally likely to send a dovish and a hawkish message, instead of sending the hawkish message with a greater probability. One possibility is that a 50/50 distribution over the messages was perceived as relatively unbiased by the provocateur, who found it attractive for that reason.[30] While we do find that Player 1

[27] Charness (2000) employs a complete information version of the stag hunt game and finds that when one of the players sends a message *before* the actions are chosen, senders tend to send the dovish message most of the time and, conditional on receiving the dovish message, choose the dovish action with very high probability.

[28] A player is a coordination type when his cost is between 10 and 85, in which case the hawkish action is a best response to a hawkish action and similarly for the dovish action.

[29] The probability that a player is a coordination type is $\approx 0.789 \left(= \frac{85-10}{95}\right)$ in all treatments of our experiment.

[30] Similarly, a strategy that does not depend on Player 3's observed costs might have appeared less biased than a strategy that does.

Table 7 Treatment effects on individual payoffs and comparison with the prediction of the uninformative communication equilibrium

	(1)	(2)	(3)	(4)	(5)	(6)
	Players 1 and 2			Player 3		
	Π^{theory}	Π^{data}	$\Pi^{data} - \Pi^{theory}$	Π^{theory}	Π^{data}	$\Pi^{data} - \Pi^{theory}$
Prov.	−0.1517	−1.142	−0.9905	−4.6289*	−73.96****	−69.3312****
	(1.3162)	(2.917)	(3.7936)	(2.2468)	(3.863)	(5.8155)
Peacem.	1.3559	11.99****	10.63****			
	(1.5729)	(1.818)	(2.3346)			
Second half	−2.9034**	−6.859**	−3.9552	1.5931	7.874	6.2806
	(1.3311)	(2.533)	(3.4309)	(9.0316)	(6.395)	(6.9586)
Second half	−0.0935	−3.318	−3.2243	−10.7226	−33.60***	−22.8729**
× Peacem.	(2.7632)	(4.907)	(4.2090)	(9.3446)	(9.187)	(8.5789)
Constant	65.444****	68.10****	2.6525	67.62****	119.1****	51.4881****
	(0.7854)	(1.532)	(1.9793)	(0.877)	(0.745)	(1.1105)
Observations	2688	2688	2688	756	756	756

Player 1 and 2 earned as predicted by uninformative equilibrium both in the two player baseline and in the provocateur treatment, but they earned more than predicted in the presence of a peacemaker; Player 3 earned less than predicted in the role of a provocateur and more than predicted in the role of a peacemaker

Session-clustered standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$, **** $p<0.001$

and Player 2 responded to dovish messages differently, it's possible that both types of players would have responded less to a provocateur they perceived as more biased against them.

Our main observations (1–3) are reflected in our analysis of players' payoffs. In Table 7, we regress subjects' payoffs on third-party dummies and a dummy for experience with the two player game (=1 for observations in the second half of the experiment). In the first three columns of the table, we study the payoffs of Player 1 and Player 2, while in the last three columns we study the payoffs of the third party. Assuming that the equilibrium is uninformative, we use the predicted payoffs (computed using the equilibrium strategies), the observed payoffs, and the difference between the predicted and the observed payoffs as dependent variables.

Uninformative equilibrium without social preferences predicts that the payoffs of Player 1 and Player 2 are unaffected by the presence of a third party. This can be seen in the first column of Table 7.[31] In the data, the presence of the peacemaker improved inexperienced subjects' payoffs ($P<0.001$), while the presence of the provocateur left them unaffected. These results are consistent with our findings in Sect. 4.2 that, not controlling for messages, the provocateur had no significant affect on subjects' behavior while the peacemaker made them less hawkish. Taking the

[31] To the extent we observe treatment effects, these effects are insignificant.

difference of observed payoffs and those predicted by uninformative equilibrium, we find that inexperienced subjects in the role of Player 1 and Player 2 did better than predicted when the third party was a peacemaker ($P < 0.001$) but no better or worse than predicted when the third party was a provocateur ($P = 0.7940$).[32] The payoffs of Player 3 are analyzed in the last three columns of the table with the peacemaker treatment serving as the baseline. Both before and after experience with the two player game, the peacemaker did better while the provocateur did worse than predicted (largest $P < 0.05$) by uninformative equilibrium. That the provocateur did worse than predicted is inconsistent with the predictions of social preferences toward the third party.

Risk aversion is an unlikely explanation for our results. While it is generally difficult to formulate predictions in cheap talk games with risk averse agents, we can numerically solve for the unique equilibrium of the baseline game in the presence of risk aversion. The analysis conducted in the Online Appendix for CRRA utility functions shows that an increase in the coefficient of risk aversion increases the equilibrium cutoff. This implies that hawkish behavior should be ex-ante more likely as players become more risk averse compared to the case of risk neutrality. This is refuted by the data.

While our data is consistent with Player 1 and Player 2 having social preferences toward each other,[33] it is unclear why a player would exhibit social preferences toward some of the other players in the game but not others. For example, in Fehr and Schmidt (1999) monetary payoffs of other players are treated symmetrically. An alternative interpretation of our data is that Player 1 and Player 2 learn to coordinate on their welfare-maximizing, sometimes off-equilibrium, outcome. In the context of complete information games, coordination on efficiency has been widely documented in experiments [see, e.g., Rankin et al. (2000)].

Our work is clearly just a first step in understanding the mechanisms behind strategic manipulation of conflict. Could private communication also be influential even if it is ineffective in theory? How robust is the suggestive power of third party messages? How robust is our finding that the provocateur's messages are not strategic? Our experimental design is such that the space of messages is coarse. We motivated this feature by the theoretical result in Baliga and Sjöström (2012) that only two messages, which can be labeled as hawkish and dovish, are sent in equilibrium of the three player game even if the message space is unrestricted. In principle, it is plausible that with a finer message space subjects in the laboratory coordinate on a different language. E.g., if the message space is made identical to the space of Player 1's possible costs, it's possible that the message senders in the experiment over- rather than under-communicate. These questions should be investigated in future research.

[32] Following experience with the two player game, subjects in the role of Player 1 and Player 2 did worse than predicted in the presence of the provocateur ($P < 0.1$) and better than predicted in the presence of the peacemaker ($P < 0.01$).

[33] Recall that subjects were less hawkish than predicted in every treatment of the experiment.

References

Baliga, S., & Sjöström, T. (2012). The strategy of manipulating conflict. *The American Economic Review*, *102*(6), 2897–2922.

Bland, J., & Nikiforakis, N. (2015). Coordination with third-party externalities. *European Economic Review*, *80*, 1–15.

Blume, A., DeJong, D. V., Kim, Y. G., & Sprinkle, G. B. (2001). Evolution of communication with partial common interest. *Games and Economic Behavior*, *37*, 79–120.

Brandts, J., & MacLeod, W. B. (1995). Equilibrium selection in experimental games with recommended play. *Games and Economic Behavior*, *11*(1), 36–63.

Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization*, *51*(2), 195–216.

Cason, T. N., & Sharma, T. (2007). Recommended play and correlated equilibria: An experimental study. *Economic Theory*, *33*(1), 11–27.

Charness, G. (2000). Self-serving cheap talk: A test of Aumann's conjecture. *Games and Economic Behavior*, *33*, 177–194.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *177*, 817–869.

Cooper, D., & Kagel, J. H. (2003). Lessons learned: Generalizing learning across games. *The American Economic Review*, *93*(2), 202–207.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*, 171–178.

Galbiati, R., & Vertova, P. (2008). Obligations and cooperative behaviour in public good games. *Games and Economic Behavior*, *64*(1), 146–170.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593–622.

Mamdani, M. (2014). *When victims become killers: Colonialism, nativism, and the genocide in Rwanda*. Princeton, NJ: Princeton University Press.

McAdams, R. H., & Nadler, J. (2005). Testing the focal point theory of legal compliance: The effect of third-party expression in an experimental hawk/dove game. *Journal of Empirical Legal Studies*, *2*(1), 87–123.

Rankin, F. W., Van Huyck, J. B., & Battalio, R. C. (2000). Strategic similarity and emergent conventions: Evidence from similar stag hunt games. *Games and Economic Behavior*, *32*(2), 315–337.

Rick, S., & Weber, R. A. (2010). Meaningful learning and transfer of learning in games played repeatedly without feedback. *Games and Economic Behavior*, *68*, 716–730.

Schelling, T. C. (1980). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Stewart, N. (1951). Divide and rule: British policy in indian history. *Science and Society*, *15*(1), 49–57.

Van Huyck, J. B., Gillette, A. B., & Battalio, R. C. (1992). Credible assignments in coordination games. *Games and Economic Behavior*, *4*(4), 606–626.