Alliance formation in a side-taking experiment

Peter DeScioli
Department of Political Science, Stony Brook University

Erik O. Kimbrough
Department of Economics, Simon Fraser University

Abstract

We investigate in an economic experiment how people choose sides in disputes. In an eight-player side-taking game, two disputants at a time fight over an indivisible resource and other group members choose sides. The player with more supporters wins the resource, which is worth real money. Conflicts occur spontaneously between any two individuals in the group. Players choose sides by ranking their loyalties to everyone else in the group and they automatically support the disputant they ranked higher. We manipulate participants' information about other players' loyalties and also their ability to communicate with public chat messages. We find that participants spontaneously and quickly formed alliances, and more information about loyalties caused more alliance-building. Without communication, we observe little evidence of bandwagon or egalitarian strategies, but with communication, some groups invented rank rotation schemes to equalize payoffs while choosing the same side to avoid fighting costs.
Keywords: alliances, bandwagon, egalitarian, conflict, experimental economics

Corresponding author: Peter DeScioli, Department of Political Science, Stony Brook University, Stony Brook, NY 11794-4392, Email: pdescioli@gmail.com

Acknowledgements
We thank Morimitsu Kurino, Rahmi Ilkilic, Bettina Klaus, Rob Kurzban, Ronald Peeters, Dotan Persitz, Dave Porter, and Bart Wilson for comments. We thank participants in seminars at the University of Arkansas, Simon Fraser University, the New York Area Political Psychology Meeting, and the NYU CESS Experimental Political Science Conference. We thank Ian Mark for computer programming of the experiment software. We thank Bay Mcculloch for research assistance. This research was supported by a grant from the International Foundation for Research in Experimental Economics (IFREE).

Introduction

Choosing sides is a common predicament at every scale of social and political life. People choose sides in conflicts between friends (DeScioli & Kurzban, 2009), colleagues choose sides in conflicts at the workplace (Kaukiainen et al., 2001), voters choose sides in elections (Huddy, 2013), legislators choose sides on new bills (Fowler, 2006), jurors choose sides in trials (Devine et al., 2001), and nations choose sides in international conflicts (Snyder, 1997; Walt, 1997). People face difficult tradeoffs when deciding which side, if any, to support in disputes, especially when bound by prior loyalties and obligations. Here we investigate the strategies people use to choose sides. We report an economic experiment that recreates incentive structures that occur in side-taking problems. We test for alliance, bandwagon, and egalitarian strategies in different information and communication treatments.

Conflicts arise not only from people's malicious intentions but also from external circumstances outside of their control. Many conflicts ultimately stem from limited resources such as food, shelter, territory, money, or status. As a result, conflicts can erupt unexpectedly and not only between enemies but also between close friends and within tightknit groups (Richardson, 2014). For instance, an economic crisis could force two legislators, usually close associates, into stark opposition over the government's spending priorities. When conflicts occur within a group, the problem of choosing sides is most acute. If two disputants are in the same social network, then they are likely to call on the same people for support, putting those bystanders in a bind. At this point, a bystander is pressed to choose between their divided loyalties to each side. This choice will reveal the strengths of their loyalties and hence could alter their enduring relationship with each disputant.

We design a side-taking game to extend previous models of fighting[1] to include bystanders who choose sides. The game focuses on the bystanders' perspective. In a group of players, fights occur between two players who are randomly and unexpectedly matched to compete for a resource. The remaining players are bystanders who choose sides in the conflict. A player chooses sides based on their divided loyalties to the other players, which the player assigns before a conflict occurs. A player's loyalties consist of a ranking of all other players, such that the player supports the fighter to whom they have greater loyalty. This ranking succinctly represents a player's divided loyalties to others. After bystanders choose sides based on their loyalties, the fighter who receives more supporters wins the resource, V, which is acquired with a costless threat display from their larger coalition of supporters to the smaller opposing coalition. If a tie occurs, then the fight escalates and all players pay the fighting cost, C, which represents when neither side backs down and the conflict escalates (discussed further below). This basic stage game is repeated such that all players simultaneously choose their loyalty rankings, then players observe conflicts and other players' side-taking (with full information, players observe everyone's loyalty rankings), then payoffs are resolved accordingly, and then the process repeats.

[1] These models of fighting include the hawk-dove game, war of attrition, lottery contest, and all-pay auction (Dechenaux et al. 2014; Hammerstein & Parker, 1982; Konrad, 2009; Maynard Smith, 1982; Vojnović, 2016).

Importantly, the resource *V* is indivisible, as in previous models of fighting, and is not shared with the winner's supporters. This absence of direct compensation is a defining feature of the bystander role and contrasts with models in which a coalition's winnings are shared (Mesterson-Gibbons et al., 2011; Murnighan, 1978; Ray, 2007; Riker, 1962). When winnings are easily divisible and transferrable, players typically seek the smallest winning coalition to maximize their own portion of the spoils (Riker, 1962). However, the stakes of many conflicts are not readily divisible and when a disputant does not divide the winnings, they tend to prefer as many supporters as possible. For instance, presidential candidates, lawmakers, political activists, and invaded nations often prefer maximum numbers of supporters. Last, indivisible stakes create an asymmetry between the roles of fighter and bystander in which the bystander's main problem is not capturing a portion of the winnings but instead managing their fighting costs and ongoing relationships to each side.

Another key feature of the game is that the fighting cost *C* occurs when there is a tie, with equal numbers of supporters on each side. This represents situations in which fighting is prone to escalate when the opponents are evenly matched, which is commonly observed across many different forms of conflict (Arnott & Elwood, 2009; Cooney, 1998, 2003; Dechenaux et al., 2014). In these cases, a threat from one side is not sufficient to deter the opponent, since it is unclear who is more likely to prevail in a protracted contest. Similarly, in a legislative context, evenly matched coalitions can become mired in gridlock, forcing everyone to accept the lower payoffs of the status quo because a majority cannot agree on a proposal (Baron & Ferejohn, 1989). More generally, the costs of fighting—including injuries, expenditures, delays, casualties, damaged reputations, etc.—are critically important to everyone involved in a conflict. Hence, when ties are costly, bystanders may attempt to avoid disagreements and deadlock by finding some common basis by which to coordinate their side-taking choices, such as following a leader or a legal convention (e.g., DeScioli & Kurzban, 2013; McAdams, 2008; Van Vugt, 2006).

From a game theory perspective, the side-taking game does not have a dominant strategy or unique equilibrium (citation to author's work redacted). The game is unusual in that a player's choices (rankings of loyalty) have no direct effect on their own success in conflicts; rather, a player's loyalties only directly affect whether other players win rewards. Hence, in a one-shot version of the game, a player is indifferent over their loyalties and any combination of loyalties in the group is a Nash equilibrium (for the moment, we hold aside fighting costs, which we return to below). However, in a repeated game, players may be able to apply contingent trigger strategies (Fudenberg & Tirole, 1991) that allow them to use their own loyalties to indirectly influence others' loyalties toward them (we have not yet modeled this possibility but view it as an important direction for future research).

In previous work, we formally modeled a version of the side-taking game that was analytically simpler, without fighting costs and where victory was probabilistic (rather than deterministic) in proportion to the number of supporters on each side (citation to author's work redacted). Most relevant, players could make loyalty pacts in which two players simultaneously increase their loyalties toward each other. In this case, two players can potentially improve their winnings by ranking each other higher, at the expense of the other group members who are demoted in loyalty. When players can form loyalty pacts, they have an incentive to form alliances because both players can gain greater support for future disputes with other players in the group. In this model, we found that players' loyalties are unstable because there are always individuals who can form new advantageous loyalty pacts at the expense of other players, leading to a perpetual cycling of loyalties.[2]

Then to examine the stability of alliances, we turned to a computational model in which players sequentially choose their loyalties by using side-taking heuristics. We examined an alliance-formation heuristic in which each player ranked the other players according to how those players ranked them. When the players applied this heuristic in an unlimited sequence of turns, the group consistently converged to a stable structure of loyalties, such that continued application of the alliance heuristic led to no further changes to players' loyalties. Moreover, across repeated iterations, these stable alliance structures shared a common characteristic: For a fight between a given pair of opponents in the group, the players' loyalties led to close to the same number of supporters on each side (and more equal than for random loyalties). Hence, the simulation suggests that alliance formation can lead to stable loyalties, which in turn lead to groups where conflicts are more evenly matched.

When we add fighting costs to the side-taking game, it becomes partly a coordination game in which players should try to choose their loyalties to avoid costly ties. The previous side-taking models did not directly examine fighting costs. When ties are costly, players should try to avoid choosing loyalties that will divide supporters evenly in disputes. As in any coordination game (Schelling, 1960), there are multiple possible solutions for avoiding ties, and a player's best tactic depends on how other players choose their loyalties. Generally, players can solve coordination games if everyone conditions their choices on the same public signal, creating a correlated equilibrium (Aumann, 1974), such as driving through a green light and stopping for a red light. For choosing sides in conflicts, people could potentially use a variety of public signals: They could side with (or against) the fighter with more success in previous fights, side with the fighter with more supporters, follow the group's status hierarchy, follow a leader's decision, follow a randomizing device like a coin flip, follow legal conventions or moral rules, and so on. If most players in the group collectively apply any one of these strategies, then they can align their side-taking decisions to avoid costly ties.

These considerations point to a basic tension when people choose sides. Individuals form alliances in order to recruit supporters for their conflicts. However, when everyone pursues their own alliances, the structure of loyalties in the group tends to make costly ties more likely, creating a coordination problem. Therefore, individuals might instead choose to break from their alliances in order to pursue a coordination tactic like choosing sides based on the fighters' previous success or the status hierarchy. But of course, this will work only if other players do the same, and it comes at the potential cost of losing one's own supporters. Hence, individuals face a difficult tradeoff between pursuing their own alliances versus pursuing group-wide conventions to avoid costly collisions between factions.

Based on this basic tradeoff and multiple social science literatures about side-taking in conflict, we focus on three possible strategies for choosing sides in the present experiment. The

[2] Because victory was probabilistic in the model, this result does not directly apply to the game in the present experiment (where victory is deterministic), though the model generally illustrates why alliances can be appealing.

first decision rule is an *alliance strategy* in which an individual assigns greater loyalty to those who are more loyal to them (i.e., more likely to support them in disputes). For instance, two individuals can improve their power by forming an alliance in which they side with each other, instead of choosing sides based on other factors such as popularity, status, or moral rules. Alliance formation is observed in many social contexts from close relationships to national politics to international relations (Cooney, 1998, 2003; DeScioli & Kurzban, 2014; Snyder, 1984, 1997). For example, previous research on close friendships found strong correlations between participants' rankings of close friends and their perceptions of friends' ranks of them. Similar findings were obtained by analyzing millions of ranked friendships on the social network MySpace (DeScioli et al., 2011). As we mentioned, individuals who build alliances are prone to split into evenly matched sides in conflicts because each player supports their own loyal supporters, which tend to differ from other players' loyal supporters since loyalties are intrinsically relative and limited (DeScioli & Kurzban, 2009, 2013).

The second decision rule is a *bandwagon strategy* in which an individual assigns greater loyalty to those who receive greater loyalty from others. This essentially means giving greater loyalty and support to the more popular, influential, or powerful people. Bandwagoning is found in various forms across a wide range of conflicts from interpersonal to international (Feldman, 2003; Snyder, 1997; Van Vugt, 2006; Walt, 1997). It is closely related to hierarchical authority relationships found across cultures in which individuals support higher status against lower status individuals in conflicts (Fiske, 1992). People can potentially bandwagon based on a number of different dimensions of power, including individuals' fighting abilities, numbers of supporters, institutional roles, etc. In the present experiment, we examine bandwagoning in the specific form of giving greater loyalty to those who have received greater loyalty from others. When used collectively, a bandwagoning group can achieve consensus and avoid the costs of disagreement, gridlock, and escalated fighting—hence solving coordination problems (Van Vugt, 2006). However, bandwagoning leads to unequal payoffs, which gives disadvantaged players an incentive to form alliances to challenge the hierarchy.

The third decision rule is an *egalitarian strategy* in which individuals assign greater loyalty to those who have less power and support, which is essentially the opposite of the bandwagon strategy. The egalitarian strategy too is found in a variety of human conflicts (Boehm, 1999; Walt, 1997; Waltz, 1979). For instance, Boehm (1999) reviews ethnographic evidence that many human societies exhibit anti-hierarchical behavior in which individuals suppress status-striving in others by opposing high-status individuals. Similarly, scholars in international relations find that weaker nations often align to oppose powerful nations to balance power or threats (Walt, 1997; Waltz, 1979). When pursued collectively, an egalitarian strategy, like bandwagoning, can also potentially allow bystanders to agree on which side to support (the weaker side), and hence to avoid costly gridlock. (Although it could be difficult to implement because it requires constant shifting of loyalties as individuals alternate as winners and losers.)

Despite previous observations and theories, it is unknown whether people spontaneously apply any or all of these strategies when confronted with a real problem of choosing sides. We use an economic game with monetary payoffs to present participants with the problem of choosing sides. The game is deliberately simplified to include only the most essential elements of a side-taking problem: Two players dispute over a resource, other players choose sides, and this process repeats. By using a minimal social environment, we can investigate what strategies, if any, people use to choose sides, while excluding potentially confounding complexities such as identities, relationships, entitlements, and histories. We test for the minimal conditions under which people exhibit bandwagon, egalitarian, and alliance behaviors. This approach is analogous to previous psychological research on the minimal conditions required for the perception of agency (Heider & Simmel, 1944), group identity (Tajfel & Turner, 1979), cooperation (Axelrod, 1984), or respect for property (DeScioli & Wilson, 2011).

Moreover, in a 2x2 between-subject design, we vary players' information about loyalties and public communication. First, we manipulate whether participants have only partial information about others' loyalties, which is revealed by how they actually chose sides in fights and is displayed in the fight history (partial information treatment), or full information about loyalties, which is shown as an additional table of all players' ranked loyalties toward everyone else from the previous period (full information treatment). This allows us to test how people use better information about loyalties when they have it, namely whether they use this information to better pursue bandwagon, egalitarian, or alliance strategies. Second, we manipulate whether participants can engage in public communication by sending chat messages to the group. This allows us to examine whether participants use communication to facilitate particular side-taking strategies.

Methods

Participants and procedure

We recruited 200 participants (47% female, age: $M = 22$, $SD = 5$ years) for experimental sessions (75 minutes). Participants entered the lab and were seated at private computer terminals. They read the instructions (Appendix A) on the computer and also had a paper copy. Participants had to correctly answer two comprehension questions to proceed. Participants played the side-taking game for 20 periods. They started with a \$5.00 endowment and their earnings could increase or decrease each period depending on the outcomes of conflicts. After the game, participants were paid their earnings (M = \$18.44, SD = \$5.12). The experiment software was programmed in z-Tree (Fischbacher, 2007).

Game design and measures

We designed an eight-player side-taking game to observe how participants choose sides in other people's disputes. The game is played anonymously on a computer network. The instructions described the game to participants with language that fits the theme of choosing sides in conflicts such as "disputes," "rankings," "fighting cost," "support," and "oppose." (See Appendix A for full instructions.)

There are eight players labeled with letters A through H. Two players at a time are randomly chosen to dispute over a resource with a cash value (\$1.50). The other six players are bystanders who choose sides, and the disputant with more supporters wins the resource. (The supporters do not earn money in other players' disputes.) When the dispute is lopsided (7 vs. 1, 6 vs. 2, or 5 vs. 3), the fight is settled with a costless threat display and no one pays a fighting cost.

If a tie occurs (4 vs. 4), then the winner is determined randomly, the dispute escalates, and all players (both fighters and supporters) pay a fighting cost which is set to $0.10.[3]

Participants choose sides by ranking their loyalties to the other players. In the decision stage, players simultaneously rank their loyalties to the other seven players highest to lowest from 1st to 7th. When a fight occurs, bystanders automatically side with the player they ranked higher in the previous decision stage. In the fighting stage, there is a series of four disputes between randomly paired opponents, such that each player fights once per period and is a bystander in the other three fights. Participants see each fight presented on their screen for 15 seconds; everyone sees which two players disputed and which players supported each side. Participants also see a cumulative history showing the fighters and supporters for all past fights in a list with a scrollbar. The fight history remains on the screen permanently and is also available in the decision stage. After the fights, participants return to the decision stage to repeat the process.

The decision stage (90 sec) and fight stage (60 sec) repeat for a total of 20 periods. Prior to the first period, there is an initial fighting stage in which participants' loyalty rankings are assigned arbitrarily by the experimenter and there are four fights. This is designed to begin the fight history prior to participants' ranking decisions. We use a repeated game with feedback about loyalties each period in order to examine bandwagon, egalitarian, alliance formation and other contingent strategies.

Alliance formation measure. For each ranking decision, we measure alliance formation as the correlation between a participant's ranking in period t and other players' ranks of ego (the participant themselves) in the previous period, t - 1 (DeScioli & Kurzban, 2009; DeScioli et al., 2011). (The measure uses other players' previous loyalties because with simultaneous decisions, a player only observes others' previous loyalties not current ones.) This metric varies from -1 to 1 with positive values indicating that a participant supports players who support them, and negative values indicating the opposite, the participant tends to support players who more frequently oppose them.

Bandwagon and egalitarian measure. We measure bandwagon and egalitarian strategies as the correlation between a participant's rankings in period t and other players' power in the previous period, t - 1. We calculate an individual's power as the average rank of that player by other players (excluding ego), which can vary from 1 (most power) to 7 (least power). This measure varies from -1 to 1 in which positive values indicate bandwagon decisions to support powerful individuals (those with high support from other players) and negative values indicate egalitarian decisions to oppose powerful individuals.

Experimental design and hypotheses

In a 2x2 experimental design, we manipulate whether participants: (a) have partial or full information about other players' rankings (partial / full), and (b) can send public chat messages (no-chat / chat). In each of the four treatments, we had 48 participants (6 groups of 8), except due

[3] We set the fighting cost to be relatively small compared to the prize to make it a moderate concern. It is possible that participants perceived this cost to be a weak incentive to avoid escalation. However, we note that the total social cost is substantial because the fighting cost applies to all eight players (8 x $0.10 = $0.80, 53% of the resource value), and each player could potentially suffer this cost four times each period whereas they had only one opportunity to win a resource each period.

to miscounting we unintentionally ran one additional group in the full-information, chat treatment so it had 56 participants. In the full-information treatment, participants observe a table of loyalties showing all eight players' rankings of everyone else in the previous period. In the partial-information treatment, participants do not see the table and have only partial information about loyalties from viewing the fights and fight history (which were available in all treatments). This manipulation allows us to examine how participants use additional information about others' loyalties to choose their own loyalties.

In the chat treatment, participants can send chat messages to the whole group and view other players' messages in the decision stage. We also added a reminder to help participants remember while chatting to submit their new rankings before the stage ends.[4] Participants in the chat treatment were told that they could send any messages except information about their identities, threats, side payments, or profanities; violations of these rules would disqualify them (no violations occurred).

The alliance hypothesis predicts that participants will rank other individuals according to how those individuals rank them. That is, values for the alliance measure will be significantly greater than zero. The bandwagon hypothesis predicts that participants will rank other individuals according to their overall power in the group and the egalitarian hypothesis predicts the opposite. Further, each hypothesis predicts greater use of each strategy, respectively, with greater information and communication. Finally, in the chat treatment the ability to communicate widens the strategy space by allowing players to propose and discuss agreements about their ranking decisions.

Results

Alliance, bandwagon, and egalitarian strategies in periods 11-20

We first examine alliance formation in the second half of the experiment, periods 11-20, after participants had time to learn and interact. We aggregate the data by averaging alliance values across periods 11-20 for each participant and then averaging these values across participants (Figure 1, panel A). The alliance measure differs from zero in the full-information, no-chat treatment, $t(47) = 7.36$, $p < .001$, the full-information, chat treatment, $t(55) = 6.81$, $p < .001$, and the partial-information, no-chat treatment, $t(47) = 6.48$, $p < .001$; it did not differ from zero in the partial-information, chat treatment, $t(47) = 0.40$, $p = .69$. These findings show that in three of the four treatments, participants' loyalty rankings tracked other players' ranks of ego (their rank of the participant themselves) in the previous period, whereas this did not occur in the partial information, chat treatment.

We analyzed alliance values with a 2 (full / partial) x 2 (chat / no-chat) ANOVA with averages at the individual level as the dependent variable. We found a main effect of full information such that the table of loyalties increased alliance formation, $F(1, 196) = 24.48$, $p < .001$. We found a main effect of chat, $F(1, 196) = 10.82$, $p < .01$, but importantly, chat reduced alliance formation. Last, there was a marginally significant interaction, $F(1, 196) = 3.38$, $p = .07$.

[4] The reminder for the chat treatment caused a software error that bypassed an error-checking function, allowing a few participants to submit inadmissible rankings such as two rank 5 values. This occurred in less than 1% of decisions and 0.2% of individual ranks and was not mentioned in participants' chats or comments, suggesting that it did not interfere with game play.

This analysis indicates that full information about loyalties increased alliance formation, while chat communication tended to suppress alliances particularly in the partial-information treatment.

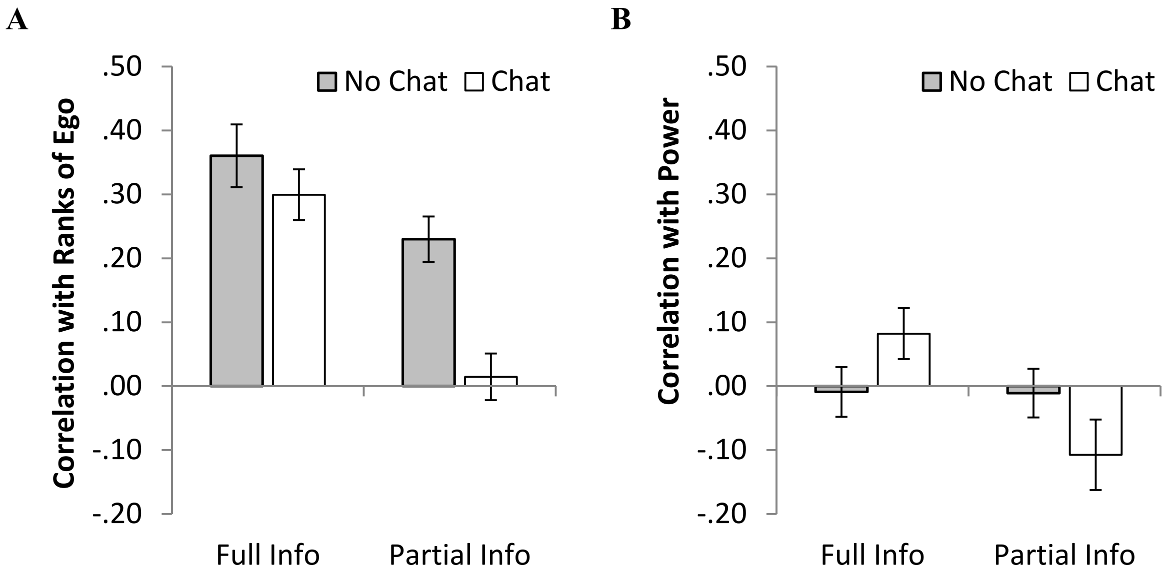


Figure 1. Mean (*SE*) values for alliance (A) and bandwagon /egalitarian (B) measures, periods 11-20. The alliance measure is the correlation between a participant's ranking and others' ranks of ego (the participant themselves). The bandwagon/egalitarian measure is the positive/negative correlation between a participant's ranking and others' power based on others' ranked loyalties.

We conducted the same analysis for bandwagon and egalitarian strategies in periods 11-20 (Figure 1, panel B). These values did not differ from zero in the full-information, no-chat treatment, $t(47) = 0.23$, $p = .82$, or the partial-information, no-chat treatment, $t(47) = 0.28$, $p = .78$. Bandwagoning was greater than zero in the full-information, chat condition, $t(55) = 2.06$, $p < .05$, and it was marginally significantly less than zero (egalitarian) in the partial-information, no-chat treatment, $t(47) = 1.95$, $p = .057$. A 2x2 ANOVA showed a main effect of full information, $F(1, 196) = 4.85$, $p < .05$, no main effect of chat, $F(1, 196) = 0.00$, $p = .95$, and a significant interaction, $F(1, 196) = 4.66$, $p = .03$. This interaction reflects that without chat, bandwagon values did not differ across full and partial-information conditions, but with chat, participants leaned toward bandwagoning when there was full information and egalitarian strategies when there was partial information. In sum, in the groups without chat, we did not find evidence for bandwagon or egalitarian strategies, but in groups with chat, there was a trend toward bandwagoning when there was full-information about loyalties, and an egalitarian effect when there was partial information. We further discuss the effects of chat below.

Side-taking dynamics

We next examine the dynamics of side-taking strategies over time. Figure 2 (panel A) shows alliance formation across all 20 periods. The data show alliances increasing in early periods, presumably as participants learned and interacted. To analyze dynamics, we estimate two GLS regression models with and without time trends (Table 1). The dependent variable is Fisher's z-transformation of the alliance measure (which is recommended when the dependent variable is a correlation coefficient)[5]. The independent variables are full information, chat, and the interaction; to model learning effects we include 1/period and associated interactions. To control for repeated measures and group effects, we include random effects for participant and we cluster standard errors at the group level (Fréchette, 2012). We estimated an exponential rather than a linear trend for periods (with 1/period) since we expected that more learning occurs in early periods (see also Figure 1).

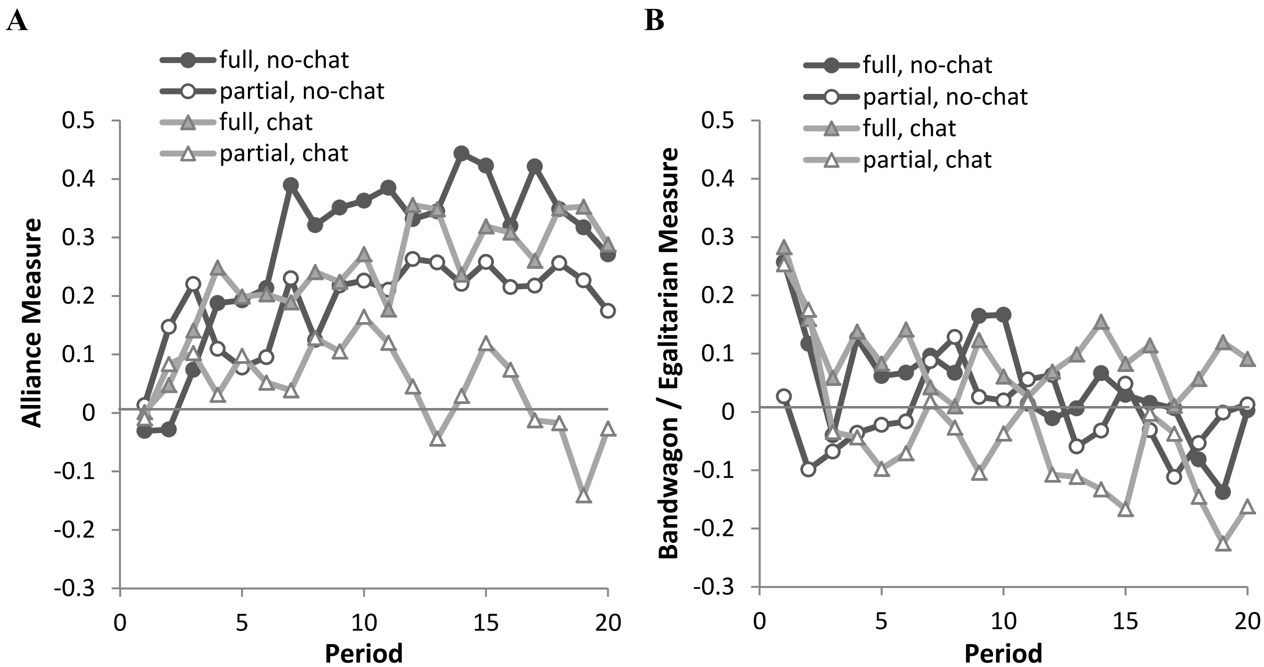


Figure 2. Mean alliance (panel A) and bandwagon/egalitarian (panel B) measures for all periods 1-20.

The regression results are shown in Table 1 (columns 1 and 2). Both models show a significant constant term, indicating that alliance values were greater than zero when there was partial information with no chat. Both models show a negative effect of chat, indicating that chat suppressed alliance formation. The effect of full information did not differ from zero in the model without period effects (column 1), but the model with period effects (column 2) showed an interaction between full information and 1/period; the negative coefficient indicates that full information about loyalties increased alliances to a greater extent as periods progressed (because 1/period decreases as periods progress).

We conducted the same analysis for bandwagon and egalitarian measures (Table 1, columns 3 and 4). Both models show that the constant term and treatment effects did not

[5] We use the Fisher transformation which is recommended because correlation coefficients are not normally distributed. We also checked and found that the regression results are qualitatively unchanged if we instead use the raw correlation coefficients.

significantly differ from zero. However, the interactions with 1/period show that as rounds progressed the treatments affected these values in the directions seen in Figures 1 and 2.

Table 1. GLS Regression Analysis of Ranking Decisions

	Alliance		Bandwagon/Egalitarian	
	(1)	(2)	(3)	(4)
Full Information	0.220	0.371*	0.076	-0.012
	(0.127)	(0.180)	(0.086)	(0.103)
Chat	-0.179*	-0.257*	-0.127	-0.314
	(0.079)	(0.113)	(0.201)	(0.296)
Full Information*Chat	0.058	0.051	0.221	0.462
	(0.154)	(0.226)	(0.235)	(0.338)
1/Period		-0.564		-0.052
		(0.295)		(0.104)
Full Information*1/Period		-1.141*		0.670*
		(0.482)		(0.278)
Chat*1/Period		0.593		1.413
		(0.366)		(0.732)
Full Information*Chat*1/Period		0.055		-1.820*
		(0.657)		(0.856)
Constant	0.233**	0.308**	0.000	0.007
	(0.052)	(0.081)	(0.045)	(0.056)
Observations	4000	4000	4000	4000
Wald Chi-Sq.	14.81	39.88	3.59	43.35

Note. Regression models of the alliance and bandwagon/egalitarian measures with and without period effects. Clustered standard errors in parentheses.
* $p < .05$, ** $p < .01$

Finally, we conducted a similar analysis of dynamics except this time we treated participants' ranking decisions as falling into discrete strategy types and then examined the frequency of types over time. For this analysis, each ranking choice was classified by whether it was most correlated with an alliance, bandwagon, or egalitarian strategy, or else fit none of these categories. The full details and analysis are presented in the Appendix. Briefly, the main results echo the analysis above, particularly that full information about loyalties increased the frequency of alliance formation and more so over time.

Best friends and enemies

We next explore which ranks were of particular focus in alliance formation. We use mutual ranks (e.g., Alice ranks Betty as number 1 and vice versa) as an indicator of a participant's attention to alliances because a choosy alliance-builder will offer their support in close proportion to how much support another player gives to them. Figure 3 shows the percentage of ranks that were mutual for ranks 1 through 7, aggregating across the second half of the game (periods 11-20). Chance is 12.5% mutual (1/8). Using chi square tests within each treatment, we found that the percentage of the time that a player's rank was mutual (yes/no) differed across ranks 1-7 (all *p*s < .001). As shown in Figure 3, rank 1 had the greatest percentage of mutual ranks within each treatment. Moreover, full information increased the overall percentage of mutual ranks in no-chat and chat conditions (chi square, *p*s < .001). Also notable, rank 7 showed greater mutual ranks in the full information treatment both with and without chat (chi square, *p*s < .001), with percentages second only to rank 1, suggesting that greater loyalty information fostered mutual enemies. In short, participants' mutual ranks suggest that their alliance strategies focused particularly on best friends (rank 1), and, when there was full information about loyalties, the next priority was on enemies (rank 7).

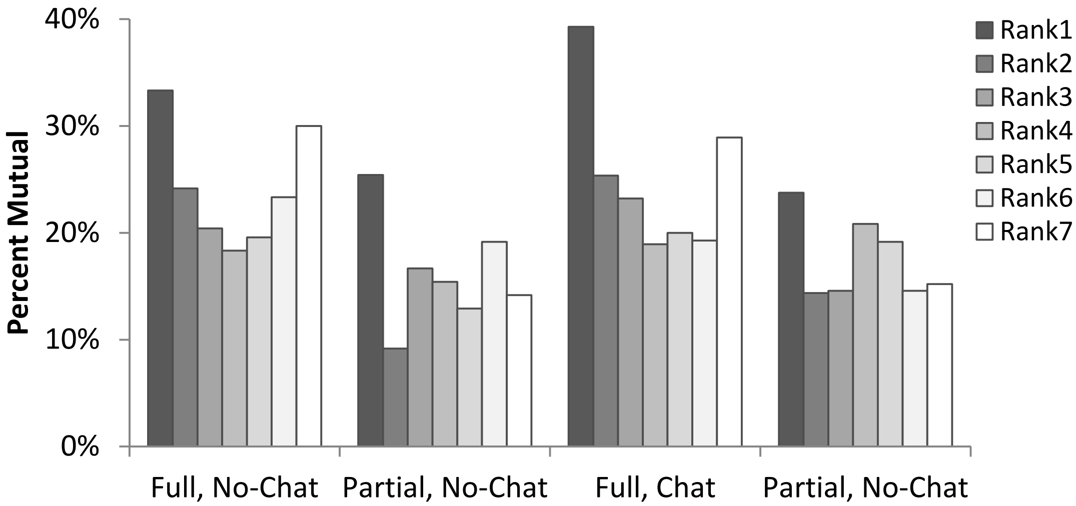


Figure 3. Percent mutual for each rank level, periods 11-20.

Chat messages and their consequences

We identified at least three strategies discussed by participants in chats. The first strategy is when two players make a pact to support each other, which occurred in half of the groups with chat. For instance, one participant wrote, "@Person B: I will keep support you if you still help me," and Person B responded, "ok, I got your back H." Participants also negotiated specific ranks such as: "1 for 1?" "2 for 2?" and "yo, F, I'll put you 3rd if you put me in your top 4." The second strategy is when multiple individuals form a team and support each other over outsiders, which was discussed in three of the thirteen groups with chat. For instance, a participant wrote, "A, B, F? Wanna like team up?" Some players tried to form a team of five, possibly realizing this was sufficient to always win against the remaining three. In one group, player E wrote, "Attention DEFGH. Let's make some money. Rank each other highest," and later, "good job DEFGH coalition." In another group, a player wrote, "OK, if we form a team of five and always support each other we will come out well."

The third strategy is rotating ranks in which all players choose the same rankings and then rotate ranks together each period, which was discussed in half of the groups with chat. Some examples were: "Let's all basically support the same people. Say EFGH for 4 rounds and then switch. We will minimize the losing costs," and, "If we all rank the same way, there will never be a tie. Then we just switch the order next round," and, "what if we all chose the same? and then shuffled every round?" and, "we all just need to agree to do the same rankings every round. I'll do abcdefgh this round. And then rotate through the list," and, "What if we all choose 4 people to support this round? Then the other 4 we will support next round, then every second round everyone will win 1.5."

These rank rotation discussions show that participants were able to use public communication to develop a novel side-taking strategy. We did not foresee this possibility but it represents a clever strategy for coordination based on a rotating schedule of ranks. More specifically, it fits theories about how groups use coordination devices to resolve disagreements and conflicts (e.g., DeScioli & Kurzban, 2013; Van Vugt, 2006).

To look more closely at the effects of rank rotation, we examined the average power differences in fights. When players in a group agree on a common ranking, they will choose the same side in fights, leading to larger power differences. At the extremes, when bystanders always agree, fights are 7 vs. 1 with a power difference of 6, and when they maximally disagree, fights are 4 vs. 4, with a power difference of 0. Figure 4 shows for all groups the mean power difference across all possible fights in the group with 95% confidence intervals. Power differences were first averaged across all (28) possible fights and then across periods in the second half of the game (to allow time for learning and discussion). Three groups with chat stand out from the rest with greater average power differences, and their 95% confidence intervals do not overlap with the other groups. These three groups discussed and agreed on rank rotation schemes in the chat messages (chat treatments, groups 4, 6, and 11). As a result, participants in these three groups reached close to full consensus about their loyalties and side-taking, with an average power difference of ~4.5, which represents lopsided fights between 7 vs. 1 and 6 vs. 2.

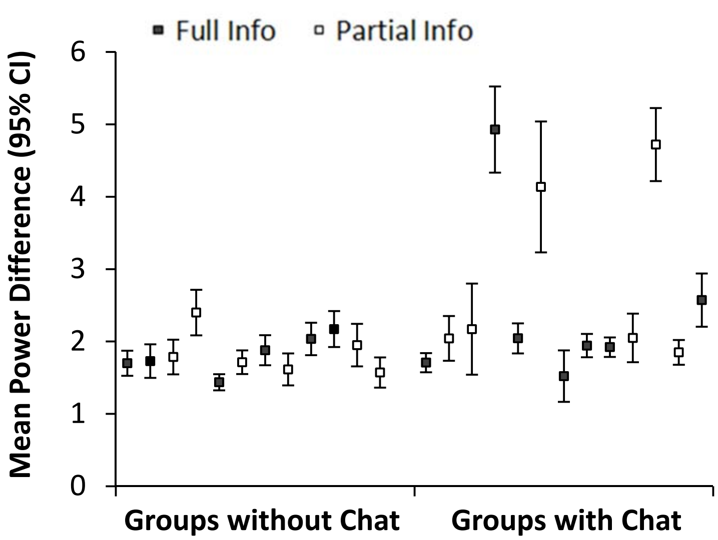


Figure 4. Mean power difference in fights, periods 11-20. Power differences were averaged across all (28) possible fights per period in each group, and then averaged across the second half of the game. Error bars are 95% confidence intervals. The three groups with greater power differences used chat to implement rank rotation schemes (see text).

Discussion

Overall, we found that participants readily formed alliances even in the minimal, abstract environment of the side-taking game. In the simplest treatments without communication, participants aligned their rankings of loyalty with other players' loyalties toward them. Moreover, when participants were provided with better information in a table of loyalties, they increasingly formed alliances and more so as periods progressed. In contrast, participants did not tend to support or oppose powerful players who were supported by others. The results of these treatments (without communication) show how people can quickly form alliances when confronted by the problem of choosing sides in conflicts.

In the treatments with public communication, we observed a different pattern of behavior. Chat communication tended to suppress alliances, particularly when participants did not see the table of loyalties; in this partial-information, chat treatment, participants showed a modest tendency toward egalitarian strategies. Looking at the chat messages, we surprisingly saw that a few groups discussed and implemented a collective rotation of ranks to even out rewards while avoiding costly ties.

We think that people's behavior in the basic side-taking game (without communication) shows a natural proclivity for alliance formation. Participants were depersonalized without names, identities, prior relationships, or previous entrenched interests. Nonetheless, a relatively simple game consisting of pairwise disputes and bystanders who choose sides was sufficient to quickly evoke alliance formation within a few rounds of the game. Moreover, participants created not only minimal groups (Tajfel & Turner, 1979) with simple boundaries but rather interlaced alliance networks of divided loyalties. To do so, they had to track information about other players' rankings in an unfamiliar context of a stylized economic game. If people form alliances even in abstract games, we suggest that these basic alliance motives are even more potent in real-world settings such as among citizens, legislators, and nations.

The fact that we saw less use of bandwagon and egalitarian strategies does not necessarily mean that people do not pursue these strategies. Instead, it shows that a minimal side-taking environment is not enough to elicit these behaviors. Indeed, when participants could communicate with chat, some groups did show a form of egalitarian behavior by rotating ranks to share the rewards. This was not the kind of egalitarian strategy that we anticipated but it is a form of turn-taking that tends to equalize rewards. Moreover, participants' rotation schemes represent an inventive strategy in which players solve a coordination problem (in this case avoiding costly ties) by synchronizing their decisions based on public messages (in this case the agreed-upon schedule of rotating ranks).

Further, public communication suppressed alliance formation when there was partial information about loyalties. This appears surprising at first because communication is commonly thought to promote collusion. Importantly, however, these communications were public and visible to everyone in the group. Hence, a player who attempts to create an exclusive alliance runs the risk of offending others and being dropped to the bottom of their loyalty rankings. If instead players could chat privately, then their private messages might promote alliances, facilitate egalitarian opposition to previous winners, or destabilize group-wide agreements. We think this is an intriguing question for future research.[6]

Additionally, there are several other particularities of the present game, and the current findings may be sensitive to these design choices. First, participants chose sides by ranking all of the other players as opposed to choosing sides directly in each fight. We used this strategy method (e.g., Brandts & Charness, 2011) so that we could observe a player's relative loyalties to everyone instead of only two fighters at a time. However, this might have altered participants' choices, for instance if it was more difficult to foresee the implications of their ranking decisions. Second, participants had to choose sides and they were unable to opt out of a conflict. This feature helped to simplify participants' decisions for analysis; it is also designed to capture situations where the costs of opting out are prohibitive, such as when both fighters would retaliate against a bystander who abandoned them. But in many conflicts, a player can choose to opt out and this is often the best strategy; moreover, the possibility of opting out might alter how people pursue alliances and other side-taking strategies. We think this is an important direction for future work.

More generally, future research can study how additional social elements tend to evoke, amplify, or suppress different side-taking strategies. For instance, presenting a list of all players' cumulative earnings might facilitate egalitarian strategies. Or, increasing the fighting cost might cause participants to bandwagon to coordinate their decisions (Van Vugt, 2006). Ultimately, this research can help illuminate how people manage their divided loyalties. It can also provide clues about how institutions that reveal loyalties could affect the way people choose sides, along with the distribution of rewards and the propensity for escalated conflict.

References

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour, 77,* 991–1004.

Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics, 1,* 67–96.

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Baron, D. P., & Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review, 83,* 1181-1206.

Boehm, C. (1999). *Hierarchy in the forest*. Cambridge, MA: Harvard University Press.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics, 14,* 375-398.

Cooney, M. (1998). *Warriors and peacemakers: How third parties shape violence*. New York, NY: New York University Press.

Cooney, M. (2003). The privatization of violence. *Criminology, 41,* 1377–1406.

Dechenaux, E., Kovenock, D., & Sheremeta, R.M. (2014). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, forthcoming.

DeScioli, P., & Kurzban, R. (2009). The alliance hypothesis for human friendship. *PLoS ONE, 4,* e5802.

[6] We thank an anonymous reviewer for raising this point.

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin, 139,* 477-496.

DeScioli, P., Kurzban, R., Koch, E. N., & Liben-Nowell, D. (2011). Best friends: Alliances, friend ranking, and the MySpace social network. *Perspectives on Psychological Science, 6,* 6-8.

DeScioli, P., & Wilson, B. J. (2011). The territorial foundations of human property. *Evolution and Human Behavior, 32,* 297-304.

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law*, *7*, 622-727.

Feldman, S. (2003). Enforcing social conformity: A theory of authoritarianism. *Political Psychology*, *24*, 41-74.

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

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*, 689-723.

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, *14*, 456-487.

Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics*, *15*, 485-498.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: MIT Press.

Hammerstein, P., & Parker, G. A. (1982). The asymmetric war of attrition. *Journal of Theoretical Biology*, *96*, 647-682.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243-259.

Huddy, L. (2013). From group identity to political cohesion and commitment. In L. Huddy, D. O. Sears, & J. Levy (Eds.) *The Oxford Handbook of Political Psychology, 2nd Edition* (pp. 737-773). New York: Oxford University Press.

Kaukiainen, A., Salmivalli, C., Björkqvist, K., Österman, K., Lahtinen, A., Kostamo, A., & Lagerspetz, K. (2001). Overt and covert aggression in work settings in relation to the subjective well-being of employees. *Aggressive Behavior*, *27*, 360-371.

Konrad, K. A. (2009). *Strategy and dynamics in contests.* New York: Oxford University Press.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

McAdams, R. H. (2008). Beyond the prisoners' dilemma: Coordination, game theory, and law. *Southern California Law Review*, *82*, 209-258.

Mesterton-Gibbons, M., Gavrilets, S., Gravner, J., & Akçay, E. (2011). Models of coalition or alliance formation. *Journal of Theoretical Biology*, *274*, 187-204.

Murnighan, J. K. (1978). Models of coalition behavior: Game theoretic, social psychological, and political perspectives. *Psychological Bulletin*, *85*, 1130-1153.

Ray, D. (2007). *A game-theoretic perspective on coalition formation*. Oxford: Oxford University Press.

Reed, L. I., DeScioli, P., & Pinker, S. (2014). The commitment function of angry facial expressions. *Psychological Science, 25,* 1511-1517.

Richardson, D. S. (2014). Everyday aggression takes many forms. *Current Directions in Psychological Science*, *23*, 220-224.

Riker, W. H. (1962). *The theory of political coalitions*. New Haven, CT: Yale University Press.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Snyder, G. H. (1984). The security dilemma in alliance politics. *World Politics, 36,* 461–495.

Snyder, G. H. (1997). *Alliance politics.* Ithaca, NY: Cornell University Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole.

Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review*, *10*, 354-371.

Vojnović, M. (2016). *Contest theory: Incentive mechanisms and ranking methods*. New York: Cambridge University Press.

Walt, S. M. (1987). *The origins of alliance*. Cornell, NY: Cornell University Press.

Waltz, K. N. (1979). *Theory of international politics*. New York: McGraw Hill.

Online Appendix for "Alliance formation in a side-taking experiment"

A. Experiment Instructions

Instructions

This is a decision-making study. Please do not talk or communicate with any other participants during the study. If you have any questions, please raise your hand and a monitor will come to answer your questions privately.

In this study, you will participate in an interaction with seven other participants. The decisions that you make in this study are anonymous and the other participants will not know who made which decisions. Your decisions will affect your own payment and other people's payments, so please carefully read these instructions.

The interaction has a total of eight participants who will be referred to as Person A, Person B, Person C, Person D, Person E, Person F, Person G, and Person H. Each participant will begin with **$5.00** and their earnings can increase or decrease depending on decisions made in the interaction.

In each round there will be four **disputes** between two individuals, and everyone will participate in one dispute. For example, the disputes could be A vs. B, C vs. D, E vs. F, and G vs. H. The exact pairings will be determined randomly.

In each dispute, two individuals will dispute over a **Resource** worth **$1.50**. The winner keeps the resource and the loser receives $0. An individual wins a dispute by having more supporters than their opponent. In each two-person dispute, the other six people will each support one individual and oppose the other individual. The individual with more supporters wins (individuals always supports themselves). If a tie occurs (4 vs. 4), then the winner will be decided randomly, and <u>all players</u> will pay a **Fighting Cost** of **10 cents**.

Participants will choose who to support and who to oppose by ranking everyone else in the group highest to lowest from 1st to 7th. When a dispute occurs, individuals will automatically support the person who is ranked higher in their list and oppose the lower-ranked person. For example, you will **support** the person you ranked 3rd when they dispute with the people you ranked 4th through 7th, and you will **oppose** the person you ranked 3rd when they dispute with the people you ranked 1st or 2nd. (Individuals always support themselves and rank themselves "0".)

In the initial round, everyone's rankings will be determined randomly (no one will make any decisions) and then there will be 4 disputes. You will see which disputes occurred, who everyone else supported, and who won the resource.

In the first round with decisions, you will have the opportunity to change your rankings. This will change who you will support in other people's disputes. You will have 90 seconds to make your decision and this will be shown on a timer at the top of the screen. After everyone completes their rankings, another series of four disputes will occur and you will observe the outcomes.

The next rounds will proceed in the same way for a total of 21 rounds. First, everyone decides on their rankings. Next, four disputes occur and participants observe the results.

At the end of the experiment, each person will be paid the total value of all of their Resources minus the sum of their Fighting Costs. Please raise your hand if you have any questions.

Comprehension questions

1. If there is a tie, who pays the fighting cost?
(a) All participants
(b) The two people competing for the item

2. How many disputes are there in each round? [textbox]

Ranking Instructions (permanently displayed onscreen)

You have 90 seconds to make your ranking decision and the time remaining is shown on the timer above.
Rank the other seven participants from 1^{st} to 7^{th}. Please make sure that each rank is used only once so that each rank 1, 2, 3, 4, 5, 6, and 7 are used exactly once. Important: Please make sure that your rank of yourself is always "0".
Your ranking will determine who you support. You will automatically support the person who is ranked higher in your list. For example, you will **support** the person ranked 3^{rd} when they dispute with the people you ranked 4^{th} through 7^{th}, and you will **oppose** the person you ranked 3^{rd} when they dispute with the people you ranked 1^{st} or 2^{nd}.

Dispute History (permanently displayed onscreen)

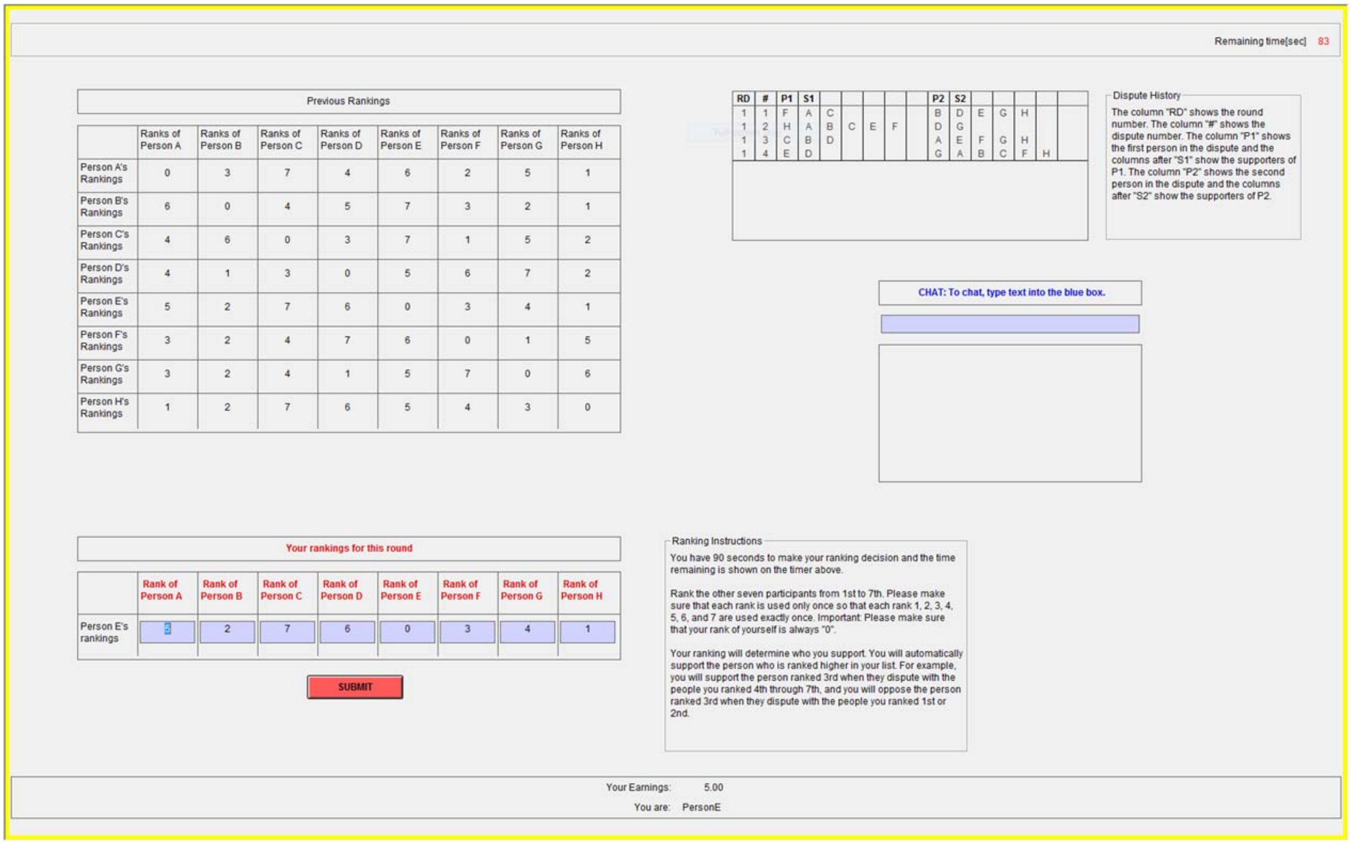
The column "RD" shows the round number. The column "#" shows the dispute number. The column "P1" shows the first person in the dispute and the columns after "S1" show the supporters of P1. The column "P2" shows the second person in the dispute and the columns after "S2" show the supporters of P2.

Additional Chat Instructions (read aloud only for groups with chat)

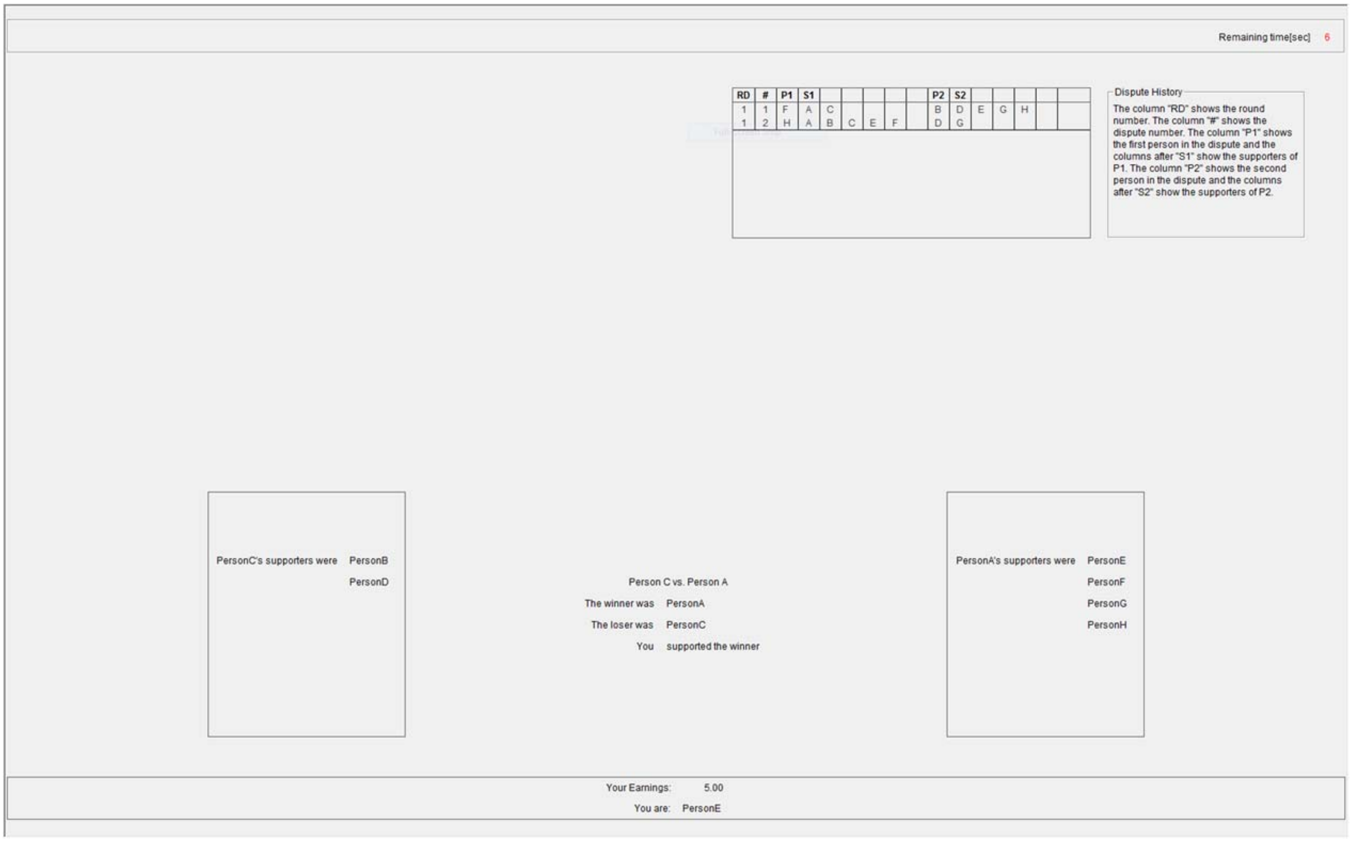
During the study, you will have the ability to chat with other participants. All participants and the experimenters will be able to see your chat messages. You are free to discuss all aspects of the interaction, with the following exceptions: You may not reveal your name or any personal information. You may not discuss side payments, make threats, or engage in any profanity or inappropriate language, including shorthand for profanities or implied profanities. We will monitor the chat box and any use of inappropriate language or identifying information will disqualify you from the study and you will forfeit all of your earnings from the study. So please remain respectful toward other participants in your communications at all times.

B. Screenshots of the game interface

The participant's display during the decision stage when assigning ranks:



The participant's display during the fight stage when observing disputes:



C. Supplementary results

Strategy types over time

We also analyzed participants' rankings of loyalty as reflecting different possible types of strategy: alliance, bandwagon, egalitarian, or none (if it did not fit the others). We analyze changes in the distribution of participants' strategy types across treatments and over time. We conduct a multinomial logistic regression in which the dependent variable is a participant's type in a given period, classified based on their rankings. We classified each participant's decision based on the strategy it was most highly correlated with: the type was alliance when the correlation with others' ranks of ego was greatest; the type was bandwagon when the correlation with other players' power was greatest; the type was egalitarian when the absolute value of the negative correlation with other players' power was greatest; if the correlation with all three types was very low below a minimal threshold set at .2, then the type was classified as "none". Figure A1 shows the resulting distribution of types over periods by treatment.

We conduct two analyses. In the first, we regress the type indicator variable on treatment dummies. In the second, we also include an exponential time trend (1/period), and interactions between treatment and time (as in Table 1 in the main text), and in both cases, we cluster standard errors at the group level. Table A1 reports the estimated coefficients for each strategy type in columns, with none as the excluded category. Significant coefficients on the treatment dummies show differences in the relative frequency of each strategy type across treatments. Since the time trend is exponential, a negative coefficient indicates that the frequency of a strategy is increasing over time (since 1/period is larger in earlier periods), and a positive coefficient indicates decreasing over time.

In model 1, we see that full information increases the relative frequency of both alliance building and bandwagon types. The chat treatment reduces the frequency of alliance building, but further increases bandwagoning. Model 2, which allows for variation over time, tells a more complex story. For the alliance type (first column), full information significantly increases the relative frequency of participants who form alliances. This frequency grows over time in all treatments, as revealed by a negative and significant coefficient on the exponential time trend, and this is even more pronounced in the full-information treatment, as revealed in the negative and significant full-information*1/period interaction. For the bandwagon type (second column), the only significant effect is for the full-information treatment, possibly because the table of loyalties conveys more information about others' power. For the egalitarian type (third column), the coefficients show that the relative frequency of the egalitarian strategy increased over time when there was no table of loyalties and participants could chat (partial-information, chat condition) but not otherwise (see also Figure A1). This likely reflects the groups that used chat to discuss rank rotation schemes, which we discuss in the section about participants' chat messages in the main text.

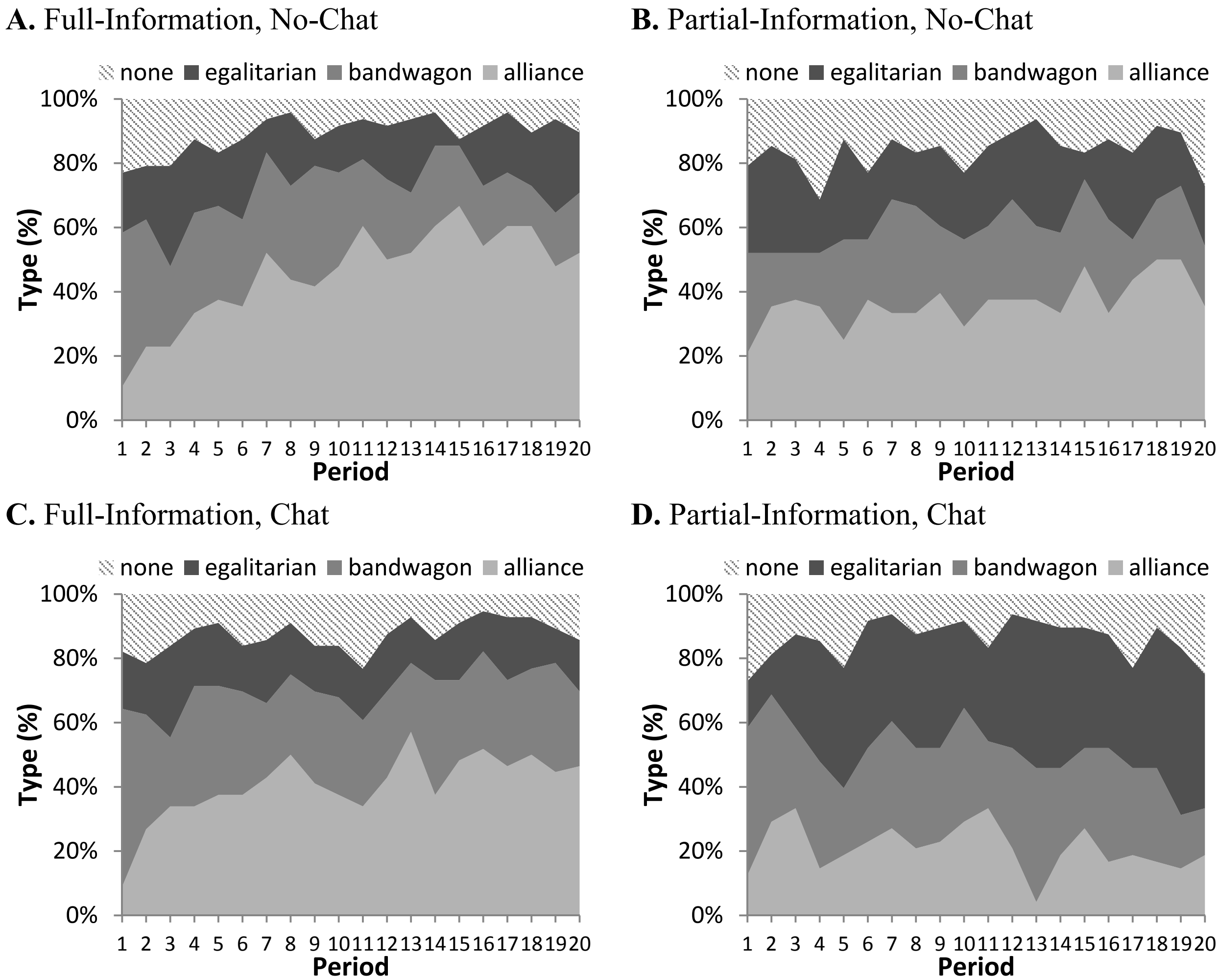


Figure A1. The frequencies of strategy types over periods for each treatment (panels A-D).

Table A1. Regression Analysis of Type (relative to None)

	Model 1			Model 2		
	Alliance	Bandwagon	Egalitarian	Alliance	Bandwagon	Egalitarian
Full Info	0.634*	0.499**	0.144	1.295***	0.543*	0.497
	(0.296)	(0.186)	(0.219)	(0.317)	(0.277)	(0.279)
Chat	-0.411*	0.360*	0.563	-0.486	0.298	1.051
	(0.201)	(0.171)	(0.429)	(0.273)	(0.285)	(0.612)
Full Info*Chat	0.109	-0.395	-0.803	-0.224	-0.492	-1.553*
	(0.376)	(0.282)	(0.480)	(0.466)	(0.409)	(0.713)
1/Period				-2.041**	-0.824	-0.103
				(0.744)	(0.748)	(0.806)
Full Info*1/Period				-5.050***	-0.101	-2.212
				(0.983)	(1.064)	(1.634)
Chat*1/Period				0.697	0.478	-3.808*
				(1.201)	(1.164)	(1.929)
Full Info*Chat*1/Period				2.281	0.452	5.400*
				(1.620)	(1.507)	(2.634)
Constant	0.814***	0.384***	0.357**	1.081***	0.498**	0.372***
	(0.156)	(0.110)	(0.132)	(0.192)	(0.184)	(0.107)
N Obs.		4000			4000	

Note. Regression models of strategy type: alliance, bandwagon, egalitarian, or none. Clustered standard errors in parentheses.

* $p < .05$, ** $p < .01$, *** $p < .001$