Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

MEASURING "GROUP COHESION" TO REVEAL THE POWER OF SOCIAL RELATIONSHIPS IN TEAM PRODUCTION

SIMON GÄCHTER, CHRIS STARMER AND FABIO TUFANO*
University of Nottingham and *University of Leicester***

30 November 2022

We introduce "group cohesion" to study the economic relevance of social relationships in team production. We operationalize measurement of group cohesion, adapting the "oneness scale" from psychology. A series of experiments, including a pre-registered replication, reveals strong positive associations between group cohesion and performance assessed in weak-link coordination games, with high-cohesion groups being very likely to achieve superior equilibria. In exploratory analysis, we identify beliefs rather than social preferences as the primary mechanism through which factors proxied by group cohesion influence group performance. Our evidence provides proof-of-concept for group cohesion as a useful tool for economic research and practice. JEL Classification: C92, D91

I. Introduction

A vast array of economic and social activity occurs in groups and teams. People need to coordinate and cooperate as colleagues in the workplace, teams on sports fields, army units on the battlefield, and across a host of less formal interactions with relatives, friends, and neighbors. In this paper, we report an extensive program of conceptual and experimental research building from the arguably plausible idea that the 'qualities' of social relationships within households, firms, or other organizations, collectively constitute an important factor of production. While various strands of existing literature hypothesize that social relationships may matter for economic outcomes (see Section II below), our motivation stems from the absence of any systematic approach to *measuring* the productive value of social relationships. Our primary contribution is to develop and test a measurement tool, based on a new concept

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

of *group cohesion*, with a view to providing foundations for the study of social relationships as factors of production.

We proceed via two main steps. The first is to introduce a simple, but conceptually well-grounded, approach to characterizing *any* real group in terms of a group cohesion index, intended as a summary statistic for aspects of social relationships that matter for team production. Our second main step is to test the predictive power of the group cohesion index in a large-scale program of experiments and accompanying analysis investigating (weak-link) team production in real groups that vary in terms of their pre-existing social relationships.

Since group cohesion is a novel concept in economics, in Section III, we explain the rationale for the concept, our approach to measuring it and some of its key properties. To preview briefly, our starting point is that members of any real human group inevitably have some relationships to other group members: for example, to begin with a very simple idea which we later operationalize, some people might feel "close" to other group members, whereas others may feel quite "distant." In our approach, we use the term "group cohesion" (or sometimes just "cohesion" for brevity) to refer to the state of the aggregate closeness ties within a group. Because closeness is an essentially subjective concept, it is natural to wonder whether it can be reliably measured either for pairs of individuals or aggregated to form a meaningful group-level statistic. Our research supports positive answers to both questions. Our measurement of group cohesion is based on the well-established "oneness scale" (Cialdini et al. (1997)) whose psychometric properties we replicated successfully in previous research (Gächter, Starmer and Tufano (2015)). The oneness scale uses simple, and very portable methods to assess how close one person feels to another, based on their own self-reports. From a measurement viewpoint, our innovation is to develop an aggregate statistic, based on oneness, to characterize the set of relationships within a group. Specifically, in our experiments, we ask each group member, privately, to indicate their oneness with every other group member. The group cohesion index

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

is then calculated as an average of individual oneness ratings (full details of our measurement approach and its psychological foundations are in Section III.A).

This seemingly modest measurement innovation generates a tool with considerable predictive power: Across a set of six experiments (see Appendix Table 3 for a summary of key aspects), we demonstrate that group cohesion is strongly associated with group outcomes. We explain the main experimental setup in Section IV. A key feature is that we study the behavior of *real* groups – not artificially created ones – achieved by recruiting groups of friends to participate. Hence, we observe real closeness ties based on pre-existing sociological and psychological characteristics that are absent (by construction) in groups set up on the spot in the lab, including those using minimal group manipulations (Goette, Huffman and Meier (2012)). [1] As we will show, measured cohesion tracks tangible sociological features of the real groups we study (see Section III.B).

Our workhorse to study group outcomes is a weak-link game, chosen to model coordination problems endemic to real organizations and teams (e.g., Camerer and Weber (2013)). In our version of the weak-link game, group members simultaneously choose an "effort level". Payoffs to each group member then depend on their own effort and the lowest effort chosen by anyone (the "weakest link") in the group. The game has multiple strict Pareto-ranked Nash equilibria in material payoffs reflecting two dimensions of group success: *coordination* (matching the effort level of other group members) and *cooperation* (achieving Pareto-superior equilibria). Building on the approach of Brandts and Cooper (2006), we designed our game to be "harsh" in the sense that groups lacking pre-existing social relationships would be expected to collapse to the Pareto-worst equilibrium.

Section V presents the key behavioral patterns in our data. We identify a strong positive association between group cohesion and performance and, while the likelihood of coordinating on some equilibrium is largely independent of the level of cohesion, it is crucial for equilibrium

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

selection: low cohesion groups usually descend rapidly to the worst Pareto-ranked equilibrium; high cohesion groups typically achieve higher Pareto-ranked equilibria. We replicate these patterns via an independently conducted, pre-registered, experiment (Study 2, Table 3, Appendix). We also confirm that our results are robust to the timing of oneness measurement (before or after play of the weak-link game), eliminating the interpretation that experience of game play explains variation in cohesion.

While our results clearly establish that groups with high cohesion systematically outperform low cohesion groups, one must be cautious about causal interpretation. Ultimately, our results are correlational, but we think a plausible interpretation of our findings is that group cohesion is a summary statistic for tangible features of real groups that matter, causally, for team production (at least in the context of the weak-link settings we study). Interpreted in that way, the group cohesion index as a new tool would be much less valuable if one could predict group performance, just as well, using a small number of easily measured group characteristics; our results, however, cast doubts on the prospects for doing that.

In Section VI, we present econometric analysis showing that group cohesion is a powerful and dominant predictor of group performance even when controlling for a large range of measured group characteristics – moreover, those characteristics become insignificant as predictors of group outcomes, once cohesion is present as a regressor. In the last game period, the model predicted effects of cohesion on group outcomes are also substantial: minimally cohesive groups are almost certain to collapse to a minimum effort of 1; maximally cohesive groups almost never fall to a minimum effort of 1; and large financial incentives are needed to promote the levels of effort expected for high cohesion groups.

In Section VII, we discuss the interpretation of our results considering two main avenues. First, we consider the possibility that, because our experiments involve groups of friends, the association between effort and cohesion might be explained by subjects having planned to

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

share their earnings with participating friends. We test and discount this as a plausible account of our main findings. Second, exploiting data on participants' beliefs and social preferences gathered in our replication study, we explore the extent to which the association between group cohesion and minimum effort is mediated through beliefs or social preferences. In contrast to results found elsewhere (e.g., Chen and Chen (2011)), we find that the effects of cohesion operate mainly via the channel of beliefs with only a limited influence of social preferences.

To preview our conclusion of Section VIII, our studies establish proof of concept for group cohesion as a useful new tool of economic analysis to capture and reveal the, previously hidden, power of social relationships as factors of production.

II. Related Literature and Our Contribution

Before presenting the substance of our paper we briefly place it in the literature. In the broadest sense, we contribute to the literature on social capital (e.g., Putnam (2000); Glaeser, Laibson and Sacerdote (2002)) by tackling one of its central problems. In a recent typology of that literature, Jackson (2020) argues that "[m]easuring various forms of social capital is especially difficult as they are dependent upon relationships between people, which are often intangible and only indirectly observed" (p. 333). We demonstrate how (social) relationships can be observed and measured to provide quantitative assessments of the (psychological) quality of social network links (Goyal (2005)), thereby providing a micro-foundation of social capital. We do this by introducing the novel psychological concept of group cohesion. As we will explain, group cohesion builds on the concepts of "relationship closeness" and "oneness". These concepts are firmly established in the psychology literature (see Section III.A) but are less considered in economics, with the possible exception of "social distance" (e.g., Akerlof (1997)). In the experimental economics literature, social distance has mainly been juxtaposed to complete anonymity and manipulated experimentally by giving participants cues about the identity of other individuals, for instance, via visual identification (Bohnet and Frey (1999)) or

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

via their names (Charness and Gneezy (2008)). By contrast, we *measure* the closeness of relationships between group members and construct the concept of group cohesion on such measurements. To our knowledge, this is an entirely new approach in economics.

Another contribution is to the experimental literature on coordination games, which – following seminal papers by van Huyck, Battalio and Beil (1990; 1991) – has largely studied coordination among anonymous individuals without considering the role of social relationships. This research (see Cooper and Weber (2020) for a recent survey), highlights primarily the importance of structural features that facilitate coordination on efficient equilibria such as: communication (Cooper et al. (1992); Brandts and Cooper (2007)); leadership (Weber et al. (2001)); individual incentives (Brandts and Cooper (2006)); group size (Weber (2006)); choice of group members (Riedl, Rohde and Strobel (2016)); and organizational or societal culture (Weber and Camerer (2003); Engelmann and Normann (2010)). By studying a fixed weak-link game, we keep structural features constant and show that the socio-psychological property of group cohesion is an independent and powerful predictor of group outcomes.

We also contribute to a growing literature on the economic impact of groups and group identity (Charness and Chen (2020)). This includes studies investigating in-group favoritism (e.g., Currarini and Mengel (2016)); interactions among friends (e.g., Glaeser et al. (2000); Leider et al. (2009); Babcock et al. (2019); Chierchia, Tufano and Coricelli (2020); Gächter et al. (2022)); and the role of identity in organizations (e.g., Akerlof and Kranton (2005); Ashraf and Bandiera (2018)), including social-psychological dimensions of employment relationships (e.g., Baron and Kreps (2013)). Our work builds most directly on prior experimental work which has established the impact of group identity on behavior in various contexts including in prisoner’s dilemma and battle of the sexes games (Charness, Rigotti and Rustichini (2007)), in trust games (Hargreaves Heap and Zizzo (2009)) and in weak-link games (Chen and Chen (2011)). While the last of these comes closest to our work in studying weak-link games, relative

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

to all three studies, our work breaks new ground: we study *real groups*, not artificially constructed ones; and we do this for the novel purpose of developing and providing proof of concept for a tool to measure the quality of behaviorally relevant features of extant socio-psychological relationships within real groups.

III. Group Cohesion in Real Groups

Since group cohesion is a novel concept in economics, we devote subsection A to explaining the concept, its roots in established psychological literature and our approach to its measurement. Subsection B shows that measured group cohesion passes a basic test of construct validity in varying coherently with tangible sociological properties of real groups.

A. Group Cohesion: psychological foundations and measurement

Our study involves the development and application of a new tool: A simple and portable measure of group cohesion designed to summarize the social and psychological relationships that exist between members of any group. To this end, we build on an established literature which has developed tools to measure the nature and strength of *bilateral* relationships between pairs of individuals. This literature demonstrates that important features of possibly complex bilateral relationships can be summarized by simple measurement tools, which ask subjects to report how "close" they feel towards another focus person (Aron, Aron and Smollan (1992)). Our strategy builds on and extends this literature by assuming that important aspects of relationships that exist within groups can be summarized in terms of the *set of pairwise closeness relationships within them*. On our measure, a group will be more cohesive to the extent that its members feel, collectively, closer to one another. Since individual judgments of relationship closeness will be its foundation, we now describe the key properties of tools for measuring bilateral relationship closeness.

According to psychologists Kelley et al. (1983), relationship closeness increases with people's *frequency of interactions*, the *diversity of activities* people undertake together, and the

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

strength of influence people have on one another. In an effort to measure these determinants of relationship closeness, Berscheid, Snyder and Omoto (1989) developed a 69-item "Relationship Closeness Inventory" to assess, in detail, the frequency of interactions, diversity of jointly undertaken activities and the influence a pair exerts on each other. While the Relationship Closeness Inventory is fine-grained, it is not practical for many purposes. To provide a handy measurement technique, in a highly cited paper, Aron, et al. (1992) proposed a simple tool: the "Inclusion of the Other in the Self" (IOS) scale depicted in Fig. 1a. The IOS scale "is hypothesized to tap people's *sense* of being interconnected with another. That sense may arise from all sorts of processes, conscious or unconscious" (Aron, et al. (1992), p. 598). Essentially, it aims to measure relationship closeness without examining its detailed determinants (i.e., frequency or diversity of activities; strength of mutual influence).

Aron, et al. provide statistical evidence that the IOS scale successfully tracks key dimensions of relationship closeness: people tend to pick more overlapping pairs of circles for a given other, the more frequent or diverse their interactions, and the stronger their perceptions of mutual influence. Subsequent research, most notably by Starzyk et al. (2006), developed an 18-item "Personal Acquaintance Measure" intended for application to a wider range of relationships including acquaintances. Their measure also correlates strongly with the IOS scale. Together, these results make the IOS scale a very promising tool for our purposes. It also has the decisive advantage of being intuitive for respondents and very simple to implement.

[Figure 1 here]

Since our research relies critically on the IOS scale, in a background paper (Gächter, et al. (2015)), via a study with 772 subjects, we assessed the psychometric validity of the IOS scale for a wide range of relationships (from strangers to close friends), by testing whether we could replicate key findings in the foundational psychological research that validated the IOS scale as a reliable predictor of relationship closeness. Our results replicate, remarkably closely, the

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

correlations of the IOS scale with the Relationship Closeness Index reported by Aron et al. The IOS scale also varies coherently with the form of the relationship (lowest for acquaintances, medium for friends, and highest for close friends), with the Personal Acquaintance Measure of Starzyk, et al. and with Rubin's Loving and Liking Scale. In Gächter, et al. (2015), we also find that the principal components of the questionnaire-based measures correlate strongly (0.85) with the IOS scale. Hence, we conclude that the IOS scale is a psychologically meaningful and reliable tool for measuring bilateral relationship closeness.

In our measurement of relationship closeness, we follow Cialdini, et al. (1997) who combine the IOS scale with the "We scale," depicted in Fig. 1b. The Cialdini et al. measure is calculated as the average of responses on these two scales. They call this the "oneness scale," which they interpret as reflecting a "sense of shared, merged, or interconnected personal identities" (p. 483). In Gächter, et al. (2015), we confirmed Cialdini et al.'s claim that oneness correlates slightly better with the questionnaire-based measures than the IOS scale alone and, hence, we use the oneness scale for our analysis.

We deployed the oneness scale as follows (wider procedural details are in Section IV). Subjects participated as groups of four and each person rated three other visually identified group members, separately and privately, on the IOS and We scales as depicted in Fig. 1; group members knew they would not receive feedback about each other's ratings. Both IOS and We scale responses are scored on a scale from 1 to 7. Oneness is the average of the two measures and hence ranges from 1, "lowest oneness", to 7, "highest oneness."

Since groups contain four people, who each rate the three other members in their group, any group generates twelve bilateral oneness ratings. We construct our group cohesion index by selecting, for each group member, the oneness score for the person they rated lowest. We then compute group cohesion as the average of these four scores. Hence, our index can be thought

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

of as summarizing the minimum envelope of oneness in a group. Our results are not sensitive, however, to different ways of averaging the individual oneness reports (see Section VI.B).

B. Group Cohesion and the Sociological Properties of Naturally Occurring Groups

In much of this paper, we focus on whether or how well group cohesion predicts performance in stylized "production tasks". Before pursuing this, however, we briefly probe the validity of our measurement tool via some simple tests examining whether measured cohesion varies as expected with tangible, socio-demographic, features of the groups in our experiments.

[Figure 2 here]

The simplest approach to this exploits the procedures we used to assemble groups. To create variation in how well members of groups knew one another prior to our experiments, we recruited participants as groups of four self-selecting friends who were then either matched into new groups of four members (Non-friends, N-matching) or kept together as friends to proceed to the experimental tasks (Friends, F-matching) (see Section IV for further details). If cohesion is tracking the pre-existing relationships within groups, we should expect that already existing groups (F-matching) will tend to have higher measured group cohesion than the ones we constructed fresh (N-matching). Fig. 2 plots the distribution of measured group cohesion separately for N- and F-matching. It is evident that group cohesion tends to be higher in the F-matching groups as compared to N-matching groups (means are 3.81 and 1.84, resp.; the distributions differ according to a Mann-Whitney test, $z = 5.896$ $p < 0.001$). Note that, with measured group cohesion ranging from 1 to 5.5, there is good scope for observing its association with group behavior, if such association exists.

For a more sophisticated test of construct validity, we use individual-level data on 15 characteristics of our participants, collected via post-experimental questionnaires. The characteristics range from self-reports of relatively concrete variables (e.g., age or gender) through to more subjective self-assessments of dispositional traits (e.g., political attitudes or

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

happiness). An established literature related to "homophily" and the sociology of friendship (e.g., Baccara and Yariv (2013); Dunbar (2018); McPherson, Smith-Lovin and Cook (2001)) shows that "like-befriends-like" hence members of self-assorted groups are expected to be more similar in terms of socio-demographic characteristics than members of other groups. This is clearly true of our self-selected groups (i.e., the F-matching groups): Based on both parametric and non-parametric tests, the null of equal variance between and within F-matching groups is rejected in the expected direction at $p < 0.05$ for 11 of the 15 characteristics (see Table SM2.1 in the supplemental materials, henceforth "SM"). In contrast, no significant differences are found for N-matching groups. This analysis demonstrates that homophily is an indicator of pre-exiting relationships among group members. Hence, if group cohesion measures what we intend, we should expect that group cohesion and group homophily will be correlated. To test this prediction, we construct a simple *homophily index* that increases with the similarity of group members on each of the fifteen variables we measured to capture tangible sociological features of group members. We explain the construction of the Homophily Index in detail in Section VI, where it features as a control variable. For now, however, we note that an OLS regression of group cohesion on the homophily index produces a highly significant coefficient ($p < 0.001$) with an R^2 of 0.37. We take this as reassuring evidence that, as well as being a simple, intuitive, and portable group-level statistic capturing bilateral assessments or relationship closeness, the group cohesion index also reflects homophily within groups, consistent with the literature on the sociology of friendship.

IV. Experimental Setup

A. The Test Environment: The Weak-link Game

Our workhorse for studying team performance is the so called weak-link game. Since the seminal papers by van Huyck et al. (1990; 1991), it has been widely studied in the lab, partly because it represents a form of coordination problem endemic to organizations (Camerer and

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Weber (2013)). A classic example is workers preparing an aircraft for takeoff: the plane can only leave once the *slowest* worker has fulfilled their task (Knez and Simester (2001)).

We use a version of the weak-link game due to Brandts and Cooper (2006). In a group of four, players simultaneously choose one of five "effort levels" 1 to 5. The payoff to each player i is given by $\pi_i = 190 - 50e_i + 10b\cdot[\min(e_1, \ldots, e_4)]$ where e_i is player i's own effort, $\min(\cdot)$ is the lowest effort in the group, and b is a "bonus" rate controlling the marginal return to changes in minimum effort. In our main experiment, we set $b = 6$ mimicking Brandts and Cooper's baseline treatment. Table 1 illustrates the payoff matrix as generated by the payoff function π_i.

[Table 1 here]

Each player chooses an effort level (i.e., a row of Table 1) and their payoff then depends on their own choice and the minimum effort among all members of their group (given by the column). The key tension embodied in the weak-link game is easy to see: everyone prefers that everyone chooses maximum effort (of 5) because this is the unique social optimum, which simultaneously maximizes everyone's payoff (at 240 points); but the optimum may not be achieved because it is costly for any individual to exceed the minimum of efforts. On standard analysis, rational players will match their expectation of the minimum of others' efforts. The game has five strict Pareto-ranked equilibria on the diagonal of Table 1. Notice that the achievement of high payoffs requires elements of coordination (choosing the same effort level as other group members) and cooperation (groups achieving Pareto-superior Nash equilibria).

We chose this specification of a weak-link game as our baseline setup in the expectation that – in the absence of aids to cooperation (e.g., communication) – low cohesions groups, typical of those used in prior experimental implementations of this game (e.g., Brandts and Cooper (2006; 2007)), would rapidly descend to the worst equilibrium.[2]

B. Sampling Strategy and Sequence of Events

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Since our goal was to study the performance of real groups, invitations to prospective participants asked each invitee to bring three additional people who all knew each other and the invitee. Hence, participants (*n* = 260 students, "Study 1") arrived at the lab in sets of four acquaintances. Upon arrival, we assigned them to one of two matching protocols, the F-matching (47 groups) or the N-matching (18 groups). In the F-matching, each quartet of acquaintances was allocated to the *same* group ("Friends"). By contrast, in the N-matching, each set of four acquaintances was split up so that each was randomly assigned to become a member of a *different* experimental group ("Non-friends"). Thus, the only difference between the two matching protocols is that, under F-matching, group members are selected to have some prior history of social interactions with each other, whereas the N-matching aims to minimize the likelihood of prior social interaction but keeping the recruitment procedures constant. Using these two matching protocols, we achieved the desired variation in pre-existing cohesion across groups (Fig. 2).

Since our setup required participants to both provide oneness ratings of other group members and to play a (repeated) weak-link game, a very important issue is whether the experience of one type of task might affect behavior in the other. We addressed this issue in two ways. First, pilot experiments revealed that measuring oneness before the weak-link game does have some influence on minimum effort. A key question is then whether prior play of the game affects subsequently measured oneness. To test this, we ran a within-subject experiment (172 new subjects; 27 F-matching groups and 16 N-matching ones) conducted in two stages. We refer to this as our "two-week experiment" (see Table 3, Appendix). In week 1, we measured oneness and elicited various individual characteristics. In week 2, the same subjects in the same groups played the weak-link game followed by elicitation of oneness. Since relationship closeness should not change systematically over the course of one week, any systematic changes in oneness ratings would be likely due to effects of the experience of game play.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Our results show that the oneness scores are not significantly different between week 1 and week 2 (individual average ratings as observations, Wilcoxon signed ranks test, $z = -1.033$, $p = 0.302$). At the group level, the Spearman rank order correlation between week 1 and week 2 group cohesion is 0.928 ($n = 43$; $p < 0.001$). This demonstrates an encouraging degree of test-retest reliability at the level of the individual. To further test the impact of game play on oneness ratings, we regressed changes in group cohesion on average minimum effort. The coefficient on minimum effort is insignificant (ordered probit, $\beta = -0.032$, $z = -0.28$, $p = 0.783$).[3] We conclude that prior play of the weak-link game has no detectable impact on subsequent measurement of oneness. This provides strong support for the sequence where we elicit the oneness ratings, for the construction of group cohesion, after the weak-link game.

C. Procedures

In all matching conditions, each group sat at a block of four computer workstations with partitions to prevent them from seeing each other's screens and responses. Each session started with an introduction read aloud by the experimenter. After that, each group of four participants was asked to stand up – one group at a time – so that each of its members could see the other members of their group.[4] Subjects then followed computerized instructions, via their own screens. These first introduced the weak-link game followed by questions to test subjects' understanding of it. After the test, subjects played eight periods of the weak-link game. In each period, after each group member had (privately) entered their own effort level, their computer screen reported their own choice, their group's minimum, their own points for the current period, and their own accumulated points for all completed periods. Subjects knew that total accumulated points across the eight periods would be converted to cash at an exchange rate of 500 points = £1.00. For oneness measurements (elicited after game play for reasons explained above) after computerized instructions, each participant was asked to focus on each other group

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

member in turn and to respond, in sequence, to both the IOS scale and the We scale (Fig. 1) tasks. The full experimental instructions are in the supplemental materials (see section SM14).

We recruited participants via ORSEE (Greiner (2015)) and ran the experiments with z-Tree (Fischbacher (2007)) in the CeDEx lab at the University of Nottingham. Sessions lasted about one hour. Participants received task-related payoffs plus a £2.00 show-up fee (the mean payment was £7.88). Payments were made privately.

V. Associations between Group Cohesion and Weak-Link Team Production

Before presenting our primary results, we note that our experimental environment is “harsh”, as intended, in that groups whose participants have no significant history of prior social interaction tend to quickly gravitate towards the lowest ranked equilibrium of the weak-link game. Using data from the N-matching, we find that, by period 8, 90 percent of groups collapse to minimum effort = 1; only two groups do better, achieving effort levels 2 and 3, respectively.[5] These results confirm existing evidence (e.g., Brandts and Cooper (2006, 2007)) and establish that there is ample scope for improvement in cooperation in our environment, if the factors captured in the group cohesion index measure matter for team production.

A. Group Cohesion, Minimum Effort and Wasted Effort

Fig. 3 presents scatter plots of minimum effort against group cohesion with separate panels for the first and last periods of the weak-link game. Each plot also includes a line of best fit (OLS) and the 95% confidence interval. We find a significant positive association between group cohesion and effort for both periods. Medium-to-high levels of group cohesion appear necessary for selecting high effort levels (i.e., minimum effort > 3). There is also evidence of some dynamic component revealed both by the change in concentration of observations across periods and picked up by the regression line which is both steeper and more strongly significant in period 8 (see p values in note to Fig. 3).

[Figure 3 here]

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

To further examine the dynamics suggested by Fig. 3a, we separate the full set of 65 groups into three subsets of "low", "medium" and "high" cohesion groups (for details of partitions see Fig. 3 caption). Fig. 3b reveals marked differences in the dynamics by showing the time path of (average) minimum effort, separately by partition. This reveals differences in both the initial levels of and trends in minimum effort across partitions: in contrast to low and medium cohesion groups, high cohesion groups cooperate more effectively in the initial period and do not experience a decay of minimum effort over time.

Interestingly, the dynamics of "wasted effort" (i.e., the total of effort in a group above the group minimum in a particular period) seem largely independent of cohesion levels and the uniformly low rates of wasted effort by period 8 imply strong convergence on equilibrium play for all levels of cohesion.[6] As Fig. 3c shows, average wasted effort in period 1 is around 5 and collapses to about 1 by period 8. The analyses of Figs. 3b and 3c suggest that group cohesion is primarily associated with cooperation (decisions consistent with higher ranked equilibria), with relatively little connection to coordination success (group members coordinating on the same equilibrium, regardless of its ranking).

A natural question to ask is whether our results are robust to the timing of the oneness elicitation. We use the data generated by our "two-week" experiment (where oneness is also elicited one week before the weak-link experiment, see Section IV.B) to conduct a simple but informative check comparing average minimum effort across experiments (original vs two-week experiment) using the partitions for group cohesion (i.e., low, medium, and high) introduced in Fig. 3b. These tests show that for both low and high cohesion groups, the achieved levels of minimum effort are statistically indistinguishable across the two experiments. For groups with mid-range cohesion, minimum effort is somewhat higher for the two-week experiment. For both experiments, however, we identify a strong positive association between cohesion and minimum effort, regardless of the timing of the oneness elicitation. This

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

holds regardless of whether we include observations from groups with mid-range cohesion (for more details of analysis see supplemental material, Section SM4).

B. A Pre-Registered Replication

While the results presented in the last subsection are encouraging, they are also novel. Therefore, replicability is of first order importance to establish confidence in the behavioral patterns just reported (e.g., Camerer, Dreber and Johannesson (2019)). We therefore replicated the experiment and report the results in this sub-section. In the following, we sometimes use "Study 2" as a convenient label for the replication study and refer back to the original study as "Study 1". To provide a credible replication, we pre-registered the experiments[7] for Study 2 and we hired an independent contractor (the University of Birmingham Experimental Economics Laboratory (BEEL)) to implement them. We provided the experimental protocol, software, and instructions, but we were not involved in data collection. BEEL followed our original recruitment procedures but with a new subject pool from Birmingham University. The protocols and instructions were as for Study 1 except that, to probe the relationship identified in Study 1, we introduced two further sets of measurements. First, subjects' beliefs about the minimum effort in their group were elicited in each round of the weak-link game. Second, the post-experimental questionnaire included incentivized elicitation of "Social Value Orientation" (Murphy and Ackermann (2014)) as a measure of group social preferences. We discuss the details of these measures and the associated results in Section VII.

The main results of Study 2 (276 participants; 49 F-matching groups and 20 N-matching ones) are described in Figure 4. A comparison with the corresponding Fig. 3 for Study 1 reveals that, qualitatively, the results are remarkably similar.[8] Panel 4a replicates the positive relationships between group cohesion and minimum effort though with the difference that, in the replication, the relationship is strongly significant for both the first and the last period. Fig. 4b confirms the ability of higher cohesion groups to achieve and sustain higher minimum effort

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

levels over time while Fig. 4c confirms the finding that the dynamics of wasted effort are largely independent of cohesion levels.[9] In sum, the results of Study 2 confirm that group cohesion has a replicable association with cooperation in the weak-link game.

[Figure 4 here]

C. Individual-Level Effort Choice

In this sub-section, we dig down to examine the association between *individual level* effort and group cohesion using pooled data from Studies 1 and 2 (see Fig. SM6.1, for corresponding analysis separately by study). Fig. 5 shows the distribution of individual effort comparing individuals in groups with low (panel *a*) and high (panel *b*) group cohesion (these correspond with the two extreme partitions of Figs. 3 and 4). In these panels, for each period, color coding shows the distribution of efforts while the average of individual effort is indicated with a circle.

[Figure 5 here]

Notice that the time profile of average individual effort is clearly different comparing low and high cohesion groups: for low cohesion groups, it starts just above 3 and descends close to the minimum of 1 by period 8; whereas, for high cohesion groups it starts higher (close to 4) and descends less steeply converging by period 8 to an average effort level of around 3. Persistent differences in the distributions of effort are also apparent comparing low and high cohesion panels (for instance, there is markedly more incidence of efforts above 3 in the high cohesion panels). An econometric analysis also finds a highly significant positive influence of individual average oneness on individual effort choices.[10]

We further examine these dynamics by focusing on each individual's change in effort following rounds in which they delivered above minimum effort. A subject who did not choose the minimum effort in period t is modelled as having a choice between three (mutually exclusive and exhaustive) options in period $t + 1$ which we label nice, moderate, or harsh: *nice* agents deliver at least as much effort as before; *moderate* agents reduce effort but no lower

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

than the previous period minimum; *harsh* agents reduce their effort below the previous minimum. We conjectured that subjects with high average oneness ratings of their fellow group members would be more likely to be nice, with the reverse true for individuals with low average oneness ratings of their fellow group members. An ordered probit analysis (using cases where a subject did not choose the minimum effort in period *t*) shows that their reaction in $t + 1$ (coded 1, 0 or -1 for nice, moderate, or harsh) varies positively with their average oneness ratings of the other three group members ($\beta = 0.067$, $p = 0.002$, pooled for Studies 1 and 2).

VI. The Predictive Power of Group Cohesion for Minimum Effort

The combined results of the two studies presented in Section V establish a strong and replicable positive association between group cohesion and minimum effort. In this section, we probe the robustness and scale of that relationship through two sets of additional analyses.

A. Does Group Cohesion Outperform Homophily Measures as a Predictor of Effort?

This sub-section presents regression analysis assessing the power of group cohesion as a predictor of minimum effort with a particular focus on the impact of controls for homophily. Via this analysis we address an issue raised in the introduction: since we interpret group cohesion as capturing the effects of real relationships that exist between group members, could we achieve comparable or better predictive power through use of information about observable individual characteristics? The main analysis we report makes use of the homophily index, first introduced in Section III, but here we provide more details of its construction.

Table 2 reports results for three models of group-level minimum effort which feature either group cohesion or the homophily index or both as independent variables. The homophily index combines data on 15 individual characteristics that we measured for this purpose (see SM1 for details).[11] For each of these 15 variables, we construct an homophily sub-index by first coding observations for each variable into a small number of mutually exclusive categories (e.g., two genders; three nationality groups). For each variable, we then assign a homophily sub-index to

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

each group calculated as the proportion of group members associated with the highest-proportion category (e.g., suppose that, in a group, 3 members are female and 1 is male then, by definition, the gender homophily sub-index for that group is 3/4=0.75). The homophily index used in the regressions of Table 2 is then the average of the 15 sub-indices for each group. The models are estimated using standard ordered probit with clustering at group-level, since groups make multiple decisions.[12] The regressions pool data for all 8 periods with separate panels for Study 1 (Panel A), Study 2 (Panel B) and the combined data set (Panel C).

[Table 2 here]

The estimated models that include group cohesion without homophily (models 1, 4 & 7 in Table 2) show that cohesion is a stable and strongly significant predictor across the two subject pools and the pooled data set. Similarly, when the homophily index enters without cohesion, consistent with our prior expectation, we find a strongly significant association for homophily in each case (models 2, 5 & 8). Critically, however, when both variables enter together, the homophily index is never significant while group cohesion remains strongly significant and with a coefficient very similar to that in the regression without the homophily index. As robustness checks, we conducted similar analysis along two further routes either entering all 15 homophily sub-indices as separate variables in regressions alongside group cohesion, or by using the 6 main principle components of the 15 homophily sub-indices as regressors alongside group cohesion (see SM8 for details).[13] The consistent outcome of this analysis is that various homophily-inspired measures do not match the performance of the group cohesion index in predicting group minimum effort.

B. Assessing the Magnitude of Cohesion-Related Cooperation

In this sub-section, we consider the *magnitude* of the effects of group cohesion on minimum effort, observed in our data. As one approach to this, we explore the predictive power of group cohesion by regressing it on minimum effort in the last period (period 8) to generate the

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

predicted probabilities for each possible level of minimum effort, conditional on different levels of group cohesion. The results (presented in detail in SM9) demonstrate a very sizeable predicted impact of group cohesion on minimum effort as we move between the extreme points of the group cohesion scale. For example, imagine a group characterized by minimum cohesion (equal to 1): such a group is almost certain to be at minimum effort of 1 (the actual probability of minimum effort in this case is approximately 93 percent, based on pooled data from Studies 1 and 2). By contrast, a group with maximum possible group cohesion (equal to 7) is unlikely to end up at minimum effort of 1 (probability of less than 12 percent) and is predicted to achieve minimum effort of at least 3 with a probability of about 83 percent.

One might wonder how far these results depend on the specification of the group cohesion variable. Recall that we calculate group cohesion as the average of the minimum oneness ratings in a group. While this minimum "envelope" seems a natural statistic, particularly in the context of the weak-link game, there is no "special sauce" involved here: indeed, using the group average of individual oneness ratings as an alternative cohesion metric delivers very similar results (see supplemental material, Table SM10.1).

As a second approach to assessing the scale of cohesion effects, we ran a series of new experimental treatments which varied the bonus (i.e., b in the payoff function π_i of the weak-link game – see Section IV.A). In these treatments, in line with the earlier research by Brandts and Cooper (2006) and others, we recruited unrelated individuals (not groups of friends) and they completed 8 rounds of the weak-link game. The bonus rates in four between-subjects' treatments (60 subjects each) were set at 6, 14, 22 and 30, respectively (see supplemental material, Table SM3.1, for the respective payoff tables). The first two bonus levels correspond with the lowest and highest bonus payments implemented by Brandts and Cooper (2006), while the other two go substantially higher in steps of 8 (the highest more than doubles their maximum). Increasing the bonus monotonically increased the average minimum effort. At

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

bonus level 6, it was close to the minimum possible value of one and corresponds with the expected minimum effort associated with low cohesion groups (i.e., a cohesion level of approximately 3, see Fig. SM12.1). Our results show that substantial increases in the bonus, beyond those used by Brandts and Cooper, are needed to induce average minimum effort levels comparable to those associated with high cohesion (see SM12 and Fig. SM12.1 for details). For example, a bonus level of 22 in the Bonus Study produces an average minimum effort comparable to that expected from groups with a cohesion level of approximately 6. These results show that the economic value of group cohesion – or more precisely the value of the factors it proxies – is substantial, when gauged by the financial incentives needed to induce effort levels comparable to those of high cohesion groups.

VII. Towards an Explanation of the Power of Group Cohesion

Bringing real groups to the lab, as we have done, is a departure from classic lab experiments which might, initially, trouble those who presume that (at least approximate) anonymity is a *sine qua non* principle for experimental games, required to avoid the shadow of the future "infecting" strategic behavior in the lab. We aim to convince readers otherwise. A key rationale for our approach comes from the fact that real groups, and the real relationships that have developed within them, are our object of study. Yet working with real groups does create some methodological challenges and issues of interpretation, one of which we address next.

A possible interpretation of the relation between cohesion and effort is that the members of high cohesion groups – by virtue of tending to know one another – might have agreed to share their payoffs, thus changing the payoff structure of the weak-link game making cooperation easier.[14] In the post-experimental questionnaire, we asked participants whether they planned to share their earnings with other group members and whether their expectation of sharing had affected their game decisions. Our robustness tests extend the analysis of Table 2 by adding controls for self-reports of sharing. While this reached significance in some specifications, it

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

had only a very modest impact on the coefficient for group cohesion which remained strongly significant in all cases (see SM11 for details). While this is reassuring, self-reports of sharing may not be entirely reliable and they may also be partly endogenous to game play.

With these limitations in mind, we ran a further set of treatments that we call the *Sharing Study* (Table 3, Appendix). For this study, we recruited fresh participants *individually* (hence, subjects typically did not know any other participant). Subjects played the weak-link game of Table 1 (where $b = 6$) following other standard procedures used across our studies but with the distinguishing feature that, before making their game decisions, subjects were told that there was some probability that we would pool all individual earnings within each group and share them equally among group members. We implemented three versions of this protocol ($n = 60$ each) with the known probability of sharing being either 0.5, 0.8 or 1. This allows us to assess an upper bound for the impact of sharing (when sharing is certain) and its sensitivity to different levels of uncertainty associated with any potential sharing arrangement.

The treatment where sharing is certain generated an average minimum effort of 2.73 which is comparable to the expected minimum effort associated with a group cohesion of close to 5 (see SM12). While introducing a little uncertainty about sharing (by setting the sharing probability at 80%) depressed average minimum effort a little (to a value just below 2.5), when the likelihood of sharing was only 50%, average minimum effort fell dramatically to a level only slightly above 1 (see Fig. SM12.1). While this evidence does not eliminate the possibility that expectations of sharing played some role, it counts against it being a convincing explanation of the broad patterns in our data: this is so because the ceiling of the sharing effect is well below the predicted effect of maximal cohesion (=7) and because uncertainty about sharing – quite likely in any actual sharing arrangements – rapidly diminishes its impact.

The results of the Sharing Study are interesting for the further reason that the treatment where sharing is certain can be interpreted as implementing an extreme form of social preferences in

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

which each agent places the same weight on the earnings of all group members, including themselves. Viewed in this way, the results are consistent with some explanatory role for social preferences, albeit a limited one. A natural question is then, what is the relative importance of social preferences versus beliefs in mediating the impact of cohesion on effort?[15]

We offer some tentative insight to this, exploiting data on beliefs and social preferences collected as part of Study 2. Specifically, immediately after entering their effort decision for each round of the weak-link game, but before knowing what others had done, each participant entered their best guess about what would be the minimum effort in that round.[16] Then, at the end of the study, we measured participants' social preferences via a set of standard Social Value Orientation tasks: the "Social Value Orientation Slider Measure" due to Murphy, Ackermann and Handgraaf (2011).[17] We use responses to these two sets of tasks as key inputs to a decomposition analysis based on the following simultaneous equation model:

$$Min\_Effort = \alpha_1 + \beta_1\, Beliefs + \beta_2\, Social\_Preferences + \beta_3\, Grp\_Cohesion + \varepsilon_1 \quad (1)$$

$$Beliefs = \alpha_2 + \beta_4\, Social\_Preferences + \beta_5\, Grp\_Cohesion + \varepsilon_2 \quad (2)$$

$$Social\_Preferences = \alpha_3 + \beta_6\, Grp\_Cohesion + \varepsilon_3 \quad (3)$$

The first equation posits beliefs, (social) preferences and group cohesion as determinants of minimum effort. Group cohesion is treated as the unique (a priori) exogenous variable which can influence minimum effort directly (Eq. 1) and, indirectly, via beliefs (Eq. 2) or social preferences (Eq. 3).[18] In the spirit of models linking social preferences and beliefs (e.g., Dufwenberg, Gächter and Hennig-Schmidt (2011)), the model also allows social preferences to influence beliefs (Eq. 2). Although very simple from a psychological point of view, the model is presented in the spirit of a tool for assessing the relative importance of beliefs and social preferences as channels mediating the impact of group cohesion on effort, in our data.[19]

The estimated model produces significant coefficients (at 5% or 1% levels) for every β coefficient except one: specifically, we find no direct effect from social preferences to

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

minimum effort (i.e., β_2 is not significantly different from zero). Hence, group cohesion impacts minimum effort through three active channels: it operates directly (via β_3) and through its impacts on both beliefs and social preferences, though the last of these channels works entirely through the secondary effect of social preferences on beliefs. Detailed estimation results are in the supplemental material, Table SM13.1.

[Figure 6 here]

Figure 6 summarizes decomposition analysis conducted to assess the relative contributions of these three channels. While the pie chart provides a summary of the complete decomposition for the whole model, our primary interest is in the relative sizes of the partial effects (listed on the right-hand side of Fig. 6) which decompose the total effect of group cohesion into its three constituent paths. The path from group cohesion through beliefs accounts for about 56% of the total effect of group cohesion on minimum effort. While the impact via social preferences also accounts for a non-trivial proportion (about 27%) of the total effect, this path operates only indirectly via the beliefs channel suggesting that the role of social preferences is secondary to beliefs in both scale and mechanics (i.e., no direct effect of social preferences). Finally, the direct effect from group cohesion to minimum effort accounts for 16.7% of the total effect of group cohesion. We interpret the small size of this direct effect as “good news” in the sense that the impact of the factors proxied by group cohesion can be largely explained through its influence on the familiar rational choice concepts of beliefs and preferences.

For a variety of reasons, we suggest that the results of this decomposition be treated as tentative, absent further replication or other support. For example, we note a difference between the status of our measurements of beliefs and social preferences: specifically, while elicited beliefs measure something intrinsic to the weak-link games played by our participants, the measured social preferences capture something external to the game context. This might have led to underestimation of the role of social preferences.[20] We could also measure social

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

preferences in multiple different ways and an approach combining alternative ways of measuring them (à la Gillen, Snowberg and Yariv (2019)) could be an interesting avenue for checking the robustness of our conclusions from the mediation analysis. In addition, we cannot rule out the possibility that measured social preferences were to some extent influenced by experiences in play of the weak-link games although, conditional on there being such an effect, it seems most plausible to assume that success in the weak-link game would have encouraged more generous allocations in SVO tasks. In that case, our decomposition analysis should be interpreted as identifying an upper bound on the contribution of social preferences. Notwithstanding these potential reservations, however, the fact that the lion's share of the work is done by beliefs in our data stands in distinct contrast to results based on experiments using artificially-induced groups (see Chen and Chen (2011)). At minimum, we therefore suggest that our results should unsettle any presumption that social preferences are the primary channel through which within-group relationships affect success in team production.

VIII. Conclusions

It is hard to deny that social relationships may affect many variables that naturally interest economists. An open question is how much they matter and whether economic analysis could take account of them in a sufficiently parsimonious way to render the undertaking tractable and worthwhile. The research presented in this paper sheds new, and positive, light on these issues.

In this paper, we have explored the power of group cohesion – a hitherto unobservable characteristic and potential "production factor" of any real group – as a tool for predicting strategic behavior, adopting the weak-link setting as a workhorse for proof of concept. Our previous related research has established that the oneness scale, on which our measurement of group cohesion is based, is simple to implement, highly portable and correlates extremely well with more detailed measures of personal relationships (Gächter, et al. (2015)). We used our measure of group cohesion, which is a group-level statistic of the oneness scale, to study the

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

cohesion of real groups. We showed that group cohesion varies across groups as predicted by relevant sociological and psychological literature and is stable based on test-retest measurement.

Using an extensive set of experiments involving 1160 participants and including a variety of robustness tests, benchmarking exercises, and an independent pre-registered replication, we examined the predictive power of group cohesion in the context of experimental weak-link coordination games played by real groups which vary in the extent of pre-existing social relationships among their members. Despite no possibilities for communication, high cohesion groups do much better in terms of the equilibria they achieve in weak-link games, and low cohesion groups rarely, if ever, do well. We used an econometric approach to explore possible mechanisms underpinning the association between group cohesion and group minimum effort and found that, in our model, group cohesion shapes both beliefs and social preferences but with beliefs emerging as the primary channel. We have also presented evidence that the changes in effort associated with variation in cohesion can be considered "large" in the context we have studied.

While we cannot directly extrapolate to predict the scale of comparable effects in other lab or in field contexts beyond those we have studied, our results do provide motivation for exploring such issues using our group cohesion index. On the assumption that our results do translate to the field, they have particular potential significance in the context of organizational performance (e.g., Akerlof and Kranton (2005); Ashraf and Bandiera (2018)). If group cohesion is associated with desirable team or group outcomes across a variety of organizational settings, then our tool may facilitate a wide range of productive applied research. And, for those with interests in engineering better organizational or team performance, oneness measurement techniques may be valuable for assessing the impact of interventions, including the variety of team building-activities in which so many organizations already invest.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

More generally, beyond the new evidence we have presented, we believe we have provided proof of concept for a new simple and portable tool designed to facilitate the quantitative study of social relationships as factors of team production.

Data availability

Data and analysis code are available at https://osf.io/g9u3e.

REFERENCES

Akerlof, George A., "Social Distance and Social Decisions," *Econometrica*, 65 (1997), 1005-27.

Akerlof, George A., and Rachel E. Kranton, "Identity and the Economics of Organizations," *The Journal of Economic Perspectives*, 19 (2005), 9-32.

Aron, Arthur, Elaine N. Aron, and Danny Smollan, "Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness," *Journal of Personality and Social Psychology*, 63 (1992), 596-612.

Ashraf, Nava, and Oriana Bandiera, "Social Incentives in Organizations," *Annual Review of Economics*, 10 (2018), 439-63.

Babcock, Philip, Kelly Bedard, Stefanie Fischer, and John Hartman, "Coordination and Contagion: Individual Connections and Peer Mechanisms in a Randomized Field Experiment," *Journal of Public Economics*, (2019), 104069.

Baccara, Mariagiovanna, and Leeat Yariv, "Homophily in Peer Groups," *American Economic Journal: Microeconomics*, 5 (2013), 69-96.

Baron, James N., and David M. Kreps, "Employment as an Economic and a Social Relationship," in Robert Gibbons and John Roberts ed., *Handbook of Organizational Economics*, 315-41, (Princeton and Oxford: Princeton University Press, 2013).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Berscheid, Ellen, Mark Snyder, and Allen M. Omoto, "The Relationship Closeness Inventory - Assessing the Closeness of Interpersonal Relations," *Journal of Personality and Social Psychology*, 57 (1989), 792-807.

Bohnet, Iris, and Bruno S. Frey, "Social Distance and Other-Regarding Behavior in Dictator Games: Comment," *American Economic Review*, 89 (1999), 335-39.

Brandts, Jordi, and David J. Cooper, "A Change Would Do You Good ... An Experimental Study on How to Overcome Coordination Failure in Organizations," *American Economic Review*, 96 (2006), 669-93.

Brandts, Jordi, and David J. Cooper, "It's What You Say, Not What You Pay: An Experimental Study of Manager Employee Relationships in Overcoming Coordination Failure," *Journal of the European Economic Association*, 5 (2007), 1223-68.

Camerer, Colin F., Anna Dreber, and Magnus Johannesson, "Replication and Other Practices for Improving Scientific Quality in Experimental Economics," in Arthur Schram and Aljaž Ule ed., *Handbook of Research Methods and Applications in Experimental Economics*, 83-102, (Cheltenham: Edward Elgar Publishing, 2019).

Camerer, Colin F., and Roberto A. Weber, "Experimental Organizational Economics," in Robert Gibbons and John Roberts ed., *Handbook of Organizational Economics*, (Princeton and Oxford: Princeton University Press, 2013).

Charness, Gary, and Yan Chen, "Social Identity, Group Behavior, and Teams," *Annual Review of Economics*, 12 (2020), 691-713.

Charness, Gary, and Uri Gneezy, "What's in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games," *Journal of Economic Behavior & Organization*, 68 (2008), 29-35.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Charness, Gary, Uri Gneezy, and Michael A. Kuhn, "Experimental Methods: Extra-Laboratory Experiments-Extending the Reach of Experimental Economics," *Journal of Economic Behavior & Organization*, 91 (2013), 93-100.

Charness, Gary, Luca Rigotti, and Aldo Rustichini, "Individual Behavior and Group Membership," *American Economic Review*, 97 (2007), 1340-52.

Chen, Roy, and Yan Chen, "The Potential of Social Identity for Equilibrium Selection," *American Economic Review*, 101 (2011), 2562-89.

Chierchia, Gabriele, Fabio Tufano, and Giorgio Coricelli, "The differential impact of friendship on cooperative and competitive coordination, " *Theory and Decision* 89 (2020), 423-452.

Cialdini, Robert B., Stephanie L. Brown, Brian P. Lewis, Carol Luce, and Steven L. Neuberg, "Reinterpreting the Empathy-Altruism Relationship: When One into One Equals Oneness," *Journal of Personality and Social Psychology*, 73 (1997), 481-94.

Cooper, David J., and Roberto A. Weber, "Recent Advances in Experimental Coordination Games," in Monica Capra, Rachel Croson, Mary Rigdon and Tanya Rosenblatt ed., *The Handbook of Experimental Game Theory*, (Cheltenham: Edward Elgar Publishing, 2020).

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross, "Communication in Coordination Games," *The Quarterly Journal of Economics*, 107 (1992), 739-71.

Currarini, Sergio, and Friederike Mengel, "Identity, Homophily and in-Group Bias," *European Economic Review*, 90 (2016), 40-55.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt, "The Framing of Games and the Psychology of Play," *Games and Economic Behavior*, 73 (2011), 459-78.

Dunbar, R. I. M., "The Anatomy of Friendship," *Trends in Cognitive Sciences*, 22 (2018), 32-51.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Engelmann, Dirk, and Hans-Theo Normann, "Maximum Effort in the Minimum-Effort Game," *Experimental Economics*, 13 (2010), 249-59.

Fischbacher, Urs, "Z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 10 (2007), 171-78.

Gächter, Simon, Chris Starmer, Christian Thöni, Fabio Tufano, and Till O. Weber, "Social closeness can help, harm and be irrelevant in solving pure coordination problems, " *Economics Letters*, 216 (2022), 110552.

Gächter, Simon, Chris Starmer, and Fabio Tufano, "Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale," *Plos One*, 10 (2015), e0129478.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *Journal of Political Economy*, 127 (2019), 1826-63.

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter, "Measuring Trust," *Quarterly Journal of Economics*, 115 (2000), 811-46.

Glaeser, Edward L., David Laibson, and Bruce Sacerdote, "An Economic Approach to Social Capital," *Economic Journal*, 112 (2002), F437-F58.

Goette, Lorenz, David Huffman, and Stephan Meier, "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups," *American Economic Journal: Microeconomics*, 4 (2012), 101-15.

Goyal, Sanjeev, "Strong and Weak Links," *Journal of the European Economic Association*, 3 (2005), 608-16.

Greiner, Ben, "Subject Pool Recruitment Procedures: Organizing Experiments with Orsee," *Journal of the Economic Science Association*, 1 (2015), 114-25.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Hargreaves Heap, Shaun P., and Daniel J. Zizzo, “The Value of Groups,” *American Economic Review*, 99 (2009), 295-323.

Jackson, Matthew O., “A Typology of Social Capital and Associated Network Measures,” *Social Choice and Welfare*, 54 (2020), 311-36.

Kelley, Harold H. , Ellen Berscheid, Andrew Christensen, John H. Harvey, Ted L. Huston, et al., *Close Relationships*, (New York: Freeman, 1983).

Knez, Marc, and Duncan Simester, “Firm-Wide Incentives and Mutual Monitoring at Continental Airlines,” *Journal of Labor Economics*, 19 (2001), 743-72.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk, “The Formation of Prosociality: Causal Evidence on the Role of Social Environment,” *Journal of Political Economy*, 128 (2020), 434-67.

Leider, Stephen, Markus M. Möbius, Tanya Rosenblat, and Quoc-Anh Do, “Directed Altruism and Enforced Reciprocity in Social Networks,” *Quarterly Journal of Economics*, 124 (2009), 1815-51.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook, “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27 (2001), 415-44.

Murphy, Ryan O., and Kurt A. Ackermann, “Social Value Orientation: Theoretical and Measurement Issues in the Study of Social Preferences,” *Personality and Social Psychology Review*, 18 (2014), 13-41.

Murphy, Ryan O., Kurt A. Ackermann, and Michel J. J. Handgraaf, “Measuring Social Value Orientation,” *Judgment and Decision Making*, 6 (2011), 771-81.

Putnam, Robert, *Bowling Alone. The Collapse and Revival of American Community*, (New York: Simon & Schuster, 2000).

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth. 2018. “Measuring and Bounding Experimenter Demand.” *American Economic Review* 108 (11): 3266–3302.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Rabe-Hesketh, Sophia, Anders Skrondal, and Andrew Pickles, "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects," *Journal of Econometrics*, 128 (2005), 301-23.

Riedl, Arno, Ingrid M. T. Rohde, and Martin Strobel, "Efficient Coordination in Weakest-Link Games," *The Review of Economic Studies*, 83 (2016), 737-67.

Schlag, Karl H, James Tremewan, and Joël J van der Weele, "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs," *Experimental Economics*, 18 (2015), 457-90.

Starzyk, K. B., R. R. Holden, L. R. Fabrigar, and T. K. MacDonald, "The Personal Acquaintance Measure: A Tool for Appraising One's Acquaintance with Any Person," *Journal of Personality and Social Psychology*, 90 (2006), 833-47.

van Huyck, John B., Raymond C. Battalio, and Richard O. Beil, "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *American Economic Review.*, 80 (1990), 234-48.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil, "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games," *The Quarterly Journal of Economics*, 106 (1991), 885-910.

Weber, Roberto A., "Managing Growth to Achieve Efficient Coordination in Large Groups," *The American Economic Review*, 96 (2006), 114-26.

Weber, Roberto A., and Colin F. Camerer, "Cultural Conflict and Merger Failure: An Experimental Approach," *Management Science*, 49 (2003), 400-15.

Weber, Roberto A., Yuval Rottenstreich, Colin Camerer, and Marc Knez, "The Illusion of Leadership: Misattribution of Cause in Coordination Games," *Organization Science*, 12 (2001), 582-98.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

APPENDIX

TABLE 3: List of Experimental Studies

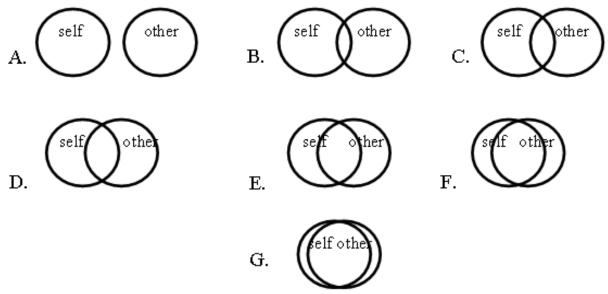
		Research Objectives	Incentive Structure	Recruitment	Random Allocation	N
I.	Study 1	Predictive Power of Group Cohesion	b=6	Friends	F- or N-matching	260
II.	Two-week Study	Construct reliability (Test-retest reliability; Task sequencing)	b=6	Friends	F- or N-matching	172
III.	Study 2	Replicating Study 1; Mediational channels: Beliefs, Social Prefs.	b=6	Friends	F- or N-matching	276
IV.	50-period Study	Long horizon	b=6	Strangers	Groups	32
V.	Share Study	To compare the cooperation enhancing effects of group cohesion with sharing rules	b=6, Pr{S}=0.5	Strangers	Groups	60
			b=6, Pr{S}=0.8	Strangers	Groups	60
			b=6, Pr{S}=1	Strangers	Groups	60
VI.	Bonus Study	To compare the cooperation enhancing effects of group cohesion with financial bonuses	b=6	Strangers	Groups	60
			b=14	Strangers	Groups	60
			b=22	Strangers	Groups	60
			b=30	Strangers	Groups	60

NOTE.— Study 2 was a pre-registered (see footnote 7) replication independently conducted at the BEEL lab (University of Birmingham, UK) by in-house experimenters. All the other studies were conducted at the CeDEx Lab (University of Nottingham, UK). Total overall sample: 1160 participants. b is the bonus rate controlling the marginal return to changes in minimum effort. *Pr{S}* stands for probability of sharing.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

Please, look at the circles diagram provided on your desk. Then, consider which of these pairs of circles best represents your connection with this person **before this experiment**. By selecting the appropriate letter below, please indicate to what extent **you and this person were connected**.

A. ☐ B. ☐ C. ☐ D. ☐ E. ☐ F. ☐ G. ☐



a. The "Inclusion of the Other in the Self" (IOS) scale

Please, select the appropriate number below to indicate to what extent, **before this experiment**, you would have used the term "**WE**" to characterize you and this person.

	1	2	3	4	5	6	7	
Not at all	☐	☐	☐	☐	☐	☐	☐	Very much so

b. The We Scale

FIG. 1.—Oneness elicitation as explained to the participants.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

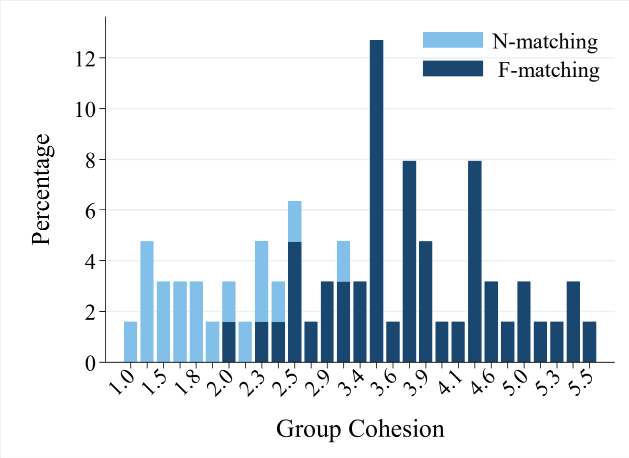
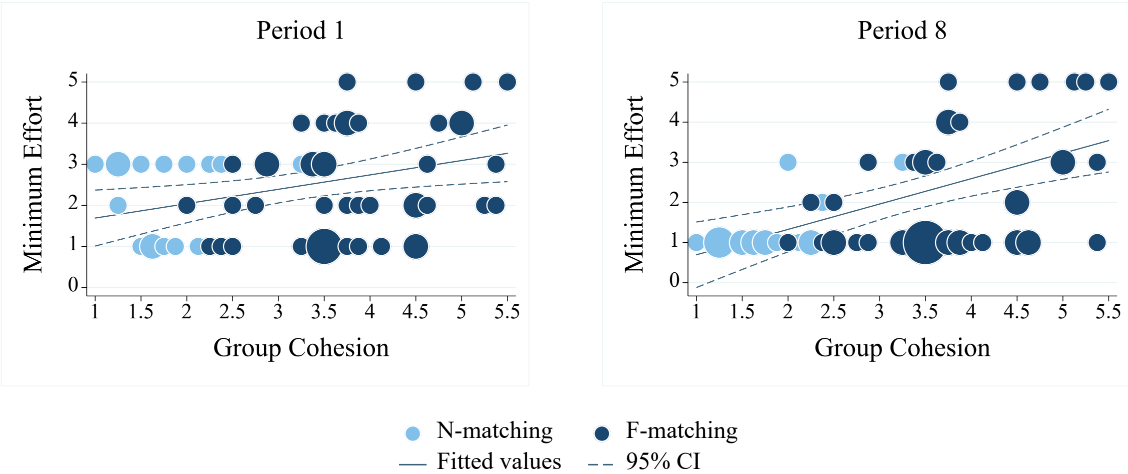
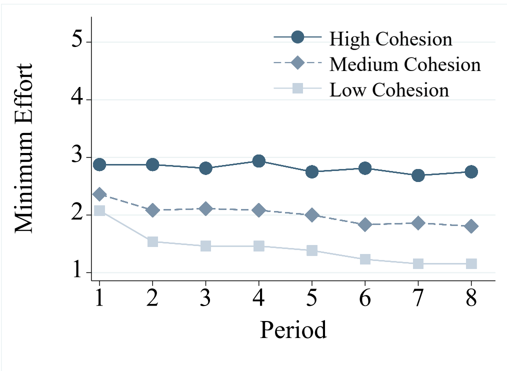


FIG. 2.—The distribution of group cohesion under F- and N-matching. The N-matching bars are stacked over the F-matching ones.

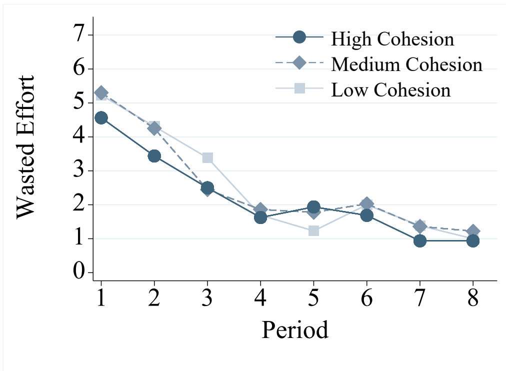
Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283



3a. The link between group cohesion and group-minimum effort in Period 1 and 8



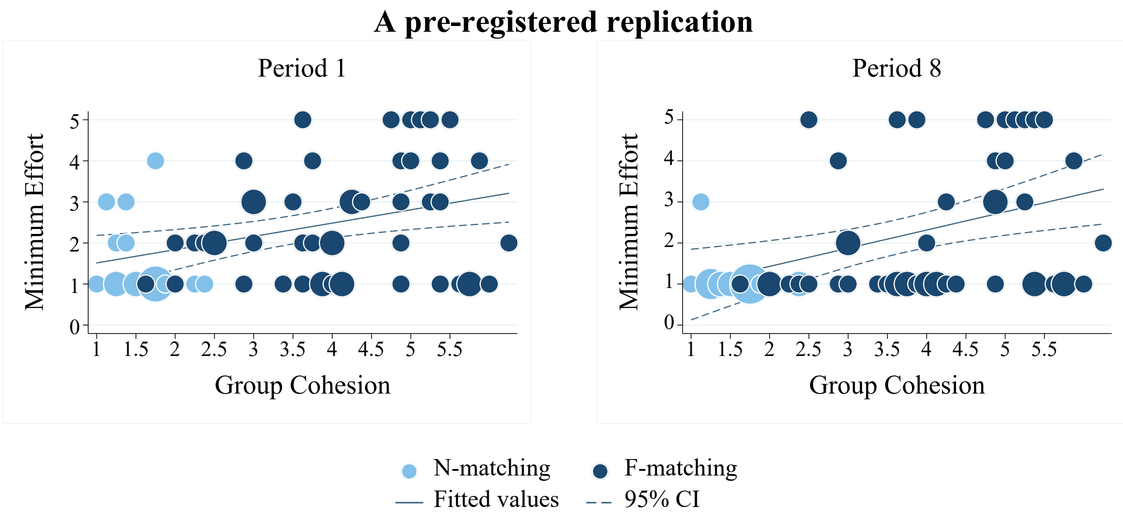
3b. Group minima across periods



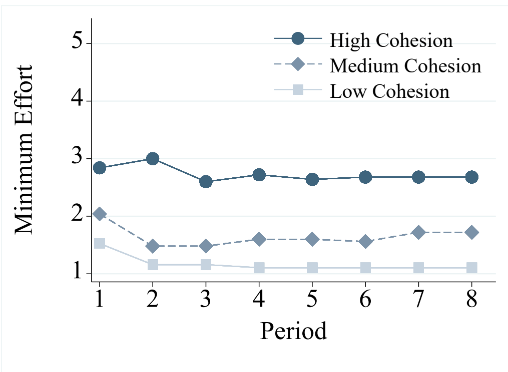
3c. Wasted effort

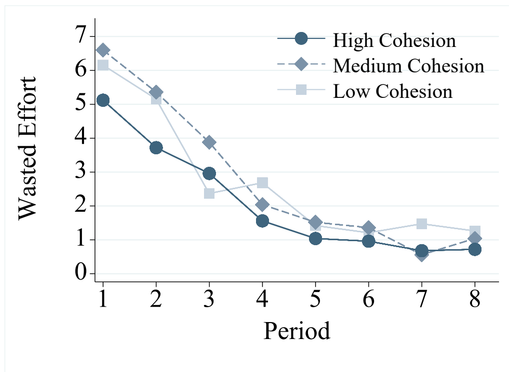
FIG. 3.—Group Cohesion, Minimum Effort and the Dynamics of Coordination. Fig. 3a: Size of symbols proportional to no. of observations; in Period 1, two N-matching observations are not displayed because they coincide with F-matching circles with coordinates (2.25, 1) and (2.5, 2); in Period 8, one N-matching observation is not displayed because it coincides with the F-matching circle at (2.5, 2). OLS Regression (65 groups), Period 1: $\beta = 0.313$ (se = 0.123, p = 0.014, $R^2 = 0.092$); Period 8 data: $\beta = 0.547$ (se = 0.123, $p < 0.001$, R^2=0.240); an ordered probit estimation generates qualitatively similar results. Fig. 3b and 3c: "Low Cohesion" Partition (13 groups): group cohesion $\in$ [1, 2]; "Medium Cohesion" Partition (36 groups): group cohesion $\in$ (2, 4]; "High Cohesion" Partition (16 groups): group cohesion $\in$ (4, 7]. Fig. 3b: average group minimum effort over time. Fig. 3c: wasted effort per period, calculated as the sum of efforts in a group above the group minimum, averaged across groups.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283



4a. The link between group cohesion and group-minimum effort in Period 1 and 8





4b. Group minima across periods

4c. Wasted effort

FIG. 4.—Study 2: pre-registered replication independently conducted at University of Birmingham. Fig. 4a: Size of symbols proportional to the number of observations; in Periods 1 and 8, three N-matching observations are not displayed as they coincide with F-matching circles at coordinates (2, 1), (2.375, 2) and (2.875, 1); OLS Regression (69 groups), Period 1: $\beta = 0.321$ (se = 0.099, p = 0.002, $R^2 = 0.135$); Period 8: $\beta = 0.405$ (se = 0.108, $p < 0.001$, R^2=0.175); ordered probit estimation generates qualitatively similar results. Fig. 4b and 4c: "Low Cohesion" Partition (19 groups): group cohesion ∈ [1, 2]; "Medium Cohesion" Partition (25 groups): group cohesion ∈ (2, 4]; "High Cohesion" Partition (25 groups): group cohesion ∈ (4, 7]. Fig. 4b: average group minimum effort over time. Fig. 4c: wasted effort per period is the sum of efforts in a group above the group minimum, averaged across groups.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

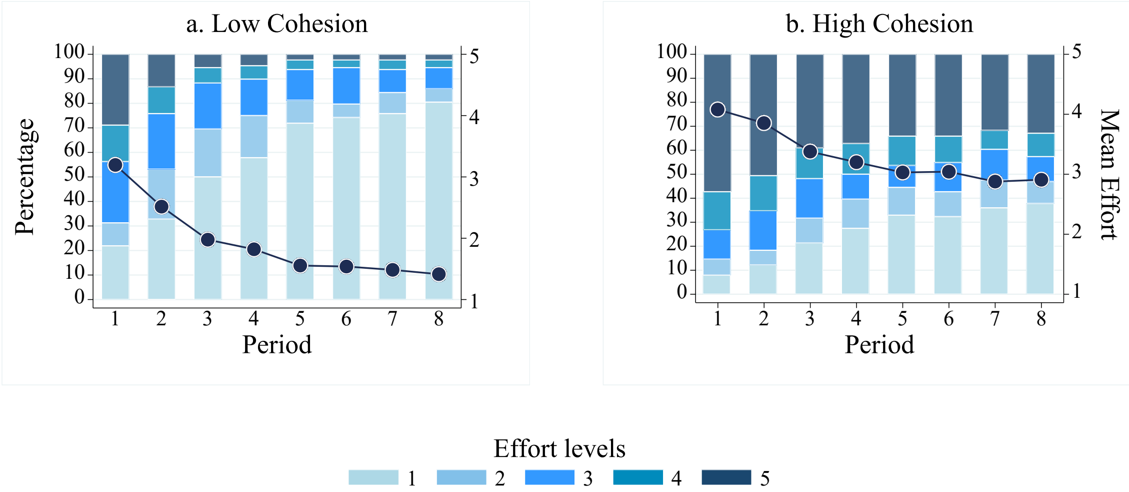
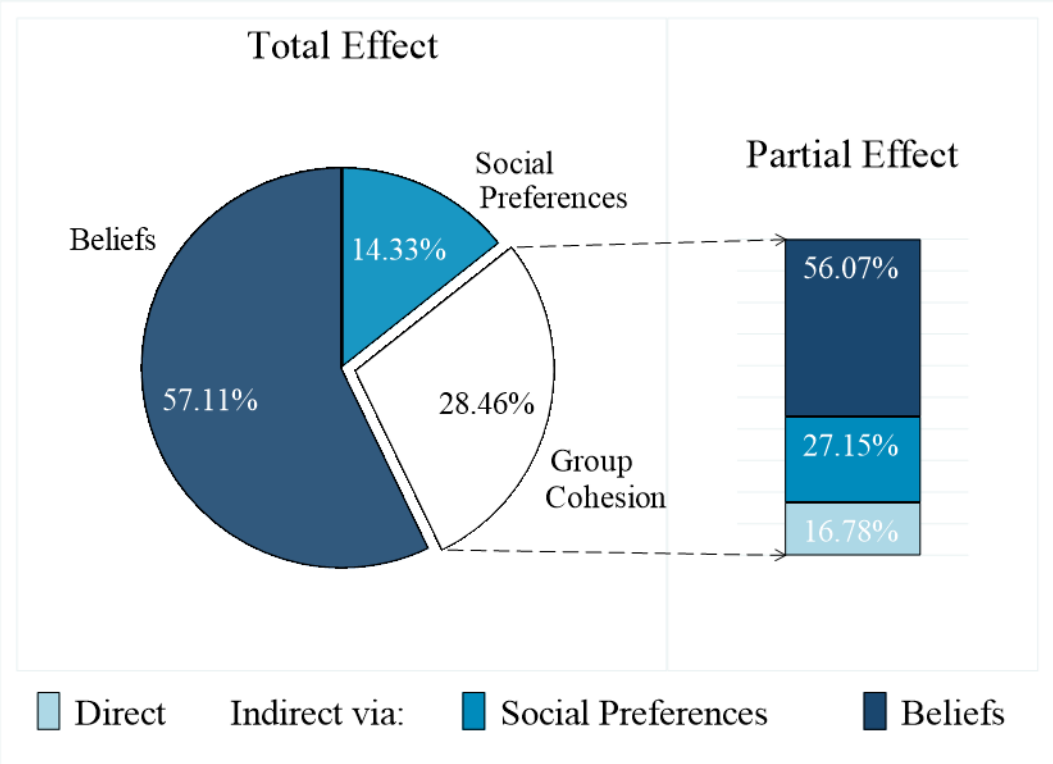


FIG. 5.—Study 1 and 2 combined: distribution of individual efforts over periods. Panel a: "Low Cohesion" Partition (32 groups): group cohesion ∈ [1, 2]; Panel b: "High Cohesion" Partition (41 groups): group cohesion ∈ (4, 7]. The bars represent the percentage of each effort level ranging from 1 to 5. The y-axes show the relevant percentages. The connected dots represent mean efforts (individual level and measured on the secondary y-axes). Supplemental Material SM6 provides further analysis for all three partitions, separated by study.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283



Decomposition of Total/Partial Effect on Minimum Effort

FIG. 6.—Study 2: modelling how group cohesion affects minimum effort. The panel reports the decomposition of the total/partial effect on minimum effort based on estimates derived from estimation of equations 1-3 above.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

TABLE 1

THE PAYOFFS (IN POINTS) FOR THE WEAK-LINK GAME

		Minimum Effort				
		1	2	3	4	5
Effort by Player *i*	1	200				
	2	150	210			
	3	100	160	220		
	4	50	110	170	230	
	5	0	60	120	180	240

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

TABLE 2

ORDERED PROBIT REGRESSIONS OF MINIMUM EFFORT ON GROUP COHESION AND HOMOPHILY

Panel A - Study 1

Dep. var.: Min. Effort	(1)		(2)		(3)	
Group cohesion	0.448***	(0.105)			0.484***	(0.136)
Homophily index			3.871**	(1.809)	-1.038	(2.315)
Log-likelihood	-644.2		-681.6		-643.6	
# level 1 (resp. 2) units	520 (65)		520 (65)		520 (65)	

Panel B - Study 2: Pre-registered replication independently conducted at the BEEL Lab.

Dep. var.: Min. Effort	(4)		(5)		(6)	
Group cohesion	0.388***	(0.099)			0.325**	(0.128)
Homophily index			6.400***	(2.468)	2.640	(2.883)
Log-likelihood	-569.0		-592.8		-565.3	
# level 1 (resp. 2) units	552 (69)		552 (69)		552 (69)	

Panel C - Study 1 and 2 combined

Dep. var.: Min. Effort	(7)		(8)		(9)	
Group cohesion	0.414***	(0.074)			0.391***	(0.095)
Homophily index			5.025**	(1.486)	0.777	(1.803)
Study 2 (dummy var.)	-0.342*	(0.191)	-0.303	(0.195)	-0.351*	(0.191)
Log-likelihood	-1231.1		-1295.7		-1230.4	
# level 1 (resp. 2) units	1072 (134)		1072 (134)		1072 (134)	

NOTES.—Data from Periods 1 to 8. Variables are at group level. Variable definition and construction are in the supplemental material, section SM1. Period dummies (always included, relative to Period 1) are significantly negative (at p<0.05). Controls for individual effects: group-level clustering. Robust standard errors in parentheses. *** p<0.01, ** p <0.05 * p<0.1.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

---

* Corresponding authors (chris.starmer@nottingham.ac.uk, fabio.tufano@leicester.ac.uk).
We are grateful to Robert Cialdini and John K. Maner for sharing experimental tools. We thank Antonio Aréchar, Michalis Drouvelis, Francesco Fallucchi, Ernesto Gavassa Perez, Jose Vicente Guinot Saporta, Natalia Montinari, Daniele Nosenzo, and Simone Quercia for assistance with experiments. We have benefited greatly from discussions with: Abigail Barr, Jordi Brandts, Leonardo Bursztyn, Colin Camerer, Gary Charness, Yan Chen, David Cooper, Vincent Crawford, Robin Cubitt, Enrique Fatas, Sourafel Girma, John Hillas, John List, Muriel Niederle, Pietro Ortoleva, Carla Rampichini, Al Roth, Marie-Claire Villeval, Roberto Weber, Frans van Winden and Leat Yariv. We thank participants at various conferences, workshops, and seminars. This work was supported by the Economic and Social Research Council [grant numbers ES/K002201/1, ES/P008976/1], the European Research Council [grant numbers ERC-AdG 295707 COOPERATION, ERC-AdG 101020453 PRINCIPLES], and the Italian Ministry of Education. The research was approved by the School of Economics Research Ethics Committee (Univ. of Nottingham). The authors declare no relevant conflict of interest.

** Gächter and Starmer: School of Economics, Univ. of Nottingham, University Park, Nottingham NG7 2RD, UK. Gächter is also affiliated with IZA Bonn and CESifo Munich. Tufano: School of Business, Univ. of Leicester, 266 London Road, Leicester LE2 1RQ, UK.

[1] In the taxonomy of Charness et al. (2013), our experiments would classify as "extra-laboratory experiments". However, a more apt label could be "field-in-the-lab experiment" because we bring naturally occurring groups of friends into a laboratory setting.

[2] Relative to other weak-link settings, this one is "harsh" in the sense defined in SM3.

[3] We explored various other specifications involving the change in minimum effort between period 1 and 8; the initial minimum effort level; all effort levels; a variable representing the period (to capture a time trend) plus interactions between the period and effort levels. None of them revealed any systematic change in group cohesion in response to game play.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

[4] It was essential for our design that subjects knew who their other group members were and, in particular, that subjects in N-groups realised that they were not grouped with their friends. Hence, in verbal instructions we asked them to "pay attention to the composition of their group" (see oral instructions in SM14). This instruction formed a brief part of the overall instructions, given some time in advance of decisions and we did not provide any signal for how subjects should take account of group membership. A reviewer suggested that this instruction might foster an experimenter demand effect. While we cannot definitively reject such a possibility, studies of experimenter effects suggest that their scale is generally modest (e.g., de Quidt et al., 2018). Nevertheless, direct evidence on this point from further research could be useful.

[5] A long-time horizon does not help low cohesion groups escape cooperation failure. We tested this with 32 fresh participants, recruited individually, who played the game of Table 1 for *50 periods* in 8 fixed groups of four anonymous members (see Appendix). Six groups were trapped in the Pareto-worst equilibrium by period 4; one by period 10; and one by period 22.

[6] We find only a weakly significant relationship between (average) group level wasted effort and group cohesion (Spearman's $\rho = -0.227$, $p = 0.069$; $n = 65$).

[7] See https://www.socialscienceregistry.org/trials/3566 (Reg. no. AEARCTR-0003566). Note that we collected one fewer group in the F-matching than planned due to no-shows.

[8] Study 2 also closely replicates the evidence that the cohesion index varies coherently with tangible characteristic of the groups (See Fig. SM5.1 and Table SM2.1).

[9] As for Study 1, we find only a weakly significant relationship between (average) Study-2 group level wasted effort and group cohesion (Spearman's $\rho = -0.209$, $p = 0.085$; $n = 69$).

[10] In a nested random model (GLLAMM, Rabe-Hesketh et al., 2005) individual effort increases with the mean oneness rating of others in their group (β=0.106; $p<0.001$; Study 1 and 2 combined). Period dummies are negative ($p<0.01$); the oneness ratings' standard deviation is positively signed and significant at 5-percent level (β=0.132; p= 0.022; Study 1 and 2

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01283

combined). Ordered probit analysis (clustered on individuals) confirms these conclusions with the only exception that the oneness ratings' standard deviation is insignificant.

[11] The 15 variables are: gender; age; field of study; nationality; no. of siblings; income; city size; no. of cohabitees; monthly budget; extent of self-finance; no. of club/group memberships; religiousness; political attitude; current happiness; future happiness.

[12] We reach consistent conclusions if instead we account for interdependence of observations by estimating nested random models using GLLAMM (for details see Table SM7.1).

[13] We are grateful to an anonymous referee for suggesting these robustness checks.

[14] While the Nash equilibria are unchanged, the risks of cooperating are substantially reduced in groups committed to "full sharing" of payoffs, making cooperation easier to achieve.

[15] In practice, it will be difficult to separate these roles clearly. For example, if groups with higher cohesion care more about each other's payoffs, in theory this reduces strategic risk, which in turn supports the expectation of higher effort levels within a group.

[16] In line with Schlag, Tremewan and van der Weele (2015), p. 484, we use non-incentivized belief elicitation because ours were fresh subjects with no clear incentive to misreport, facing a straightforward elicitation task embedded in a multi-task experiment in which hedging could otherwise have been a problem. See the supplemental material, section SM14.c, for details.

[17] Each participant made 15 dictator style allocation decisions for an identified recipient from their group. The participant knew that one of the other two group members would make allocations to them (hence eliminating reciprocity considerations) but they did not know which one. See the supplemental material, section SM14.c, for further details.

[18] The analysis is conducted at group level. We use the average of the individual beliefs in each group and the average of the individual social value orientations in each group.

[19] The approach is similar in spirit to the mediation analysis reported in Kosse et al. (2020).

[20] We are grateful to an anonymous referee for highlighting this possibility.