John Benjamins Publishing Company

This is a contribution from *Interaction Studies* 17:3

Does successful small-scale coordination help or hinder coordination at larger scales?

Seth Frey[1,2] and Robert L. Goldstone[2]
[1]Dartmouth College, USA / [2]Indiana University, USA

An individual can interact with the same set of people over many different scales simultaneously. Four people might interact as a group of four and, at the same time, in pairs and triads. What is the relationship between different parallel interaction scales, and how might those scales themselves interact?

We devised a four-player experimental game, the Modular Stag Hunt, in which participants chose not just whether to coordinate, but with whom, and at what scale. Our results reveal coordination behavior with such a strong preference for dyads that undermining pairwise coordination actually improves group-scale outcomes. We present these findings as experimental evidence for competition, as opposed to complementarity, between different possible scales of multi-player coordination. This result undermines a basic premise of approaches, like those of network science, that fail to model the interacting effects of dyadic, triadic, and group-scale structures on group outcomes.

Keywords: group structure, coordination, *n*-player games, networked games, group formation, multi-scale structure

Introduction

There are many scales at which a group can store its structure. Some groups are definable entirely in terms of the characteristics of their constituent members, at "scale 1." Others can be defined simply with the set of pairwise interactions between their members, at "scale 2." Still others can be defined satisfactorily in terms of their collective performance at a single "group-scale" task. More often, groups will store structure at many different scales, and an understanding of the scales at which they organize is important for predicting how they will behave and under what conditions they will excel.

Interaction Studies 17:3 (2016), **371–389**. DOI 10.1075/is.17.3.03fre
ISSN 1572–0373 / E-ISSN 1572–0381

In families, businesses, teams, and other social groups, different demands motivate different scales of organization, and groups that perform many functions may be organized at a different scale for each. Businesses have long organized themselves on the premise that the use of modular subgroups will improve organization-scale performance. However, an academic department might be concerned that its complementary subspecialties actually undermine department-scale cohesion. Teachers may separate the assigned seats of befriended students to facilitate cohesion at the classroom-scale (Maldonado, Klemmer, & Pea, 2009). And pairwise friendships formed at informal work gatherings seem as likely to improve as to undermine cohesion at the scale of an office.

Unfortunately, we lack not only statistical tools for characterizing multi-scale structure, but clear concepts for theorizing about it without conflicting intuitions. Take the following example. You are the headwaiter at a restaurant and you have just taken reservations for two separate parties of four. One consists of two couples, or perhaps two pairs of siblings, and the other of four mutually unacquainted peers. Which party is most likely to solve the coordination problem of getting exactly four people in the same place at the appointed time? In the first group, members of each pair already have a habit and history of coordinating with each other; two subparties of scale 2 may have an easier time coordinating than four scale-1 "parties" of individuals. However, if one of the goals of restaurant dining is to not dine alone, then the members of each couple, being paired already, have less intrinsic motivation than the four individuals to meet as a larger group. So, will the couples' experiences with scale-2 coordination help, hinder, or have no relation to the higher-scale task of coordinating on a table of four? Existing theory about groups makes no clear prediction, and the authors have found that colleagues who place themselves in the position of the headwaiter can convince themselves of either outcome.

The question of multi-scale structure has direct implications for theories like network science. Community detection is a branch of network science with widespread applications in domains as diverse as computer science, sociology, and biology (Klamt, Haus, & Theis, 2009; Kun, Boza, & Scheuring, 2010; Newman, 2004; Newman & Girvan, 2004; Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002). The goal of community detection is to use the set of dyads defining a group's network relations as a proxy measure of higher-scale order. This is a sensible approach because pairwise relations are often easier to capture measure than higher-order relations. However, the implicit causal relation in community detection is that higher-order structure has a unidirectional influence on structure at lower scales. What happens if such a relationship is systematically reciprocated, if a change in lower-scale relations (e.g. defining a group as a pair of couples) has a causal effect on higher-scale order (e.g. the ability of that group to act as a unit)? If scale-2 changes affect scale-4 order, then they can't be an independent measure of it.

Our experiment, on a game we call the Modular Stag Hunt, tests whether there might be a causal effect, either positive or negative, of subgroups on group performance at larger scales. The questions we raise are analogous to some in the social psychological literature on intergroup conflict (Hornsey, 2008; Wenzel, Mummendey, & Waldzus, 2007), where the identification of a New Yorker with a Californian is influenced by how much either are identified as American. However, where social psychology uses the tools of identity theory, the current work relies more on behavioral game theory, investigating decision making, stability, and efficiency in experimental environments with groups that organize in the lab.

Generalizing the study of coordination across scales

The most common approach for articulating questions of group structure in game-theoretic terms is to create a network with an agent at each node and define two-player interactions over its dyads. The behavior in each of these pairwise games may then serve as an input to a network measure of clustering or community structure. This basic outline describes some of the very first laboratory experiments on human groups, which were experiments about group structure articulated in terms of sets of dyadic relations (Bavelas, 1950; 1952; Christie, Luce, & Macy, 1952; Glanzer & Glaser, 1961; Guetzkow & Simon, 1955; Leavitt, 1951). Contemporary networked games continue to infer the existence of intermediate-scale coordination (internal structure) using patterns of coordination at scale-2, both in the lab (Ahn, Esarey, & Scholz, 2009; Ahn, Isaac, & Salmon, 2008; Berninghaus, Ehrhart, & Ott, 2010; Goldstone, Roberts, Mason, & Gureckis, 2010; Goldstone, Wisdom, Roberts, & Frey, 2013; Kearns, Suri, & Montfort, 2006; Mason, Jones, & Goldstone, 2008; Mullen, Johnson, & Salas, 1991), and in simulation (Fu, Hauert, Nowak, & Wang, 2008; Kun et al., 2010; Skyrms, 2004; Skyrms & Pemantle, 2000).

There are also non-network game-theoretic approaches to group structure in the lab, mostly from behavioral game theory (Camerer & Weber, 2007). Research in this tradition has shown that incremental increases in group size can aid coordination at higher scales, while larger growth spurts can hinder it (Camerer & Weber, 2008; Weber, 2006; Weber & Camerer, 2003), and more recent work is exploring additional mechanisms for facilitating increases in the scales at which groups coordinate and cooperate (Riedl, Rohde, & Strobel, 2016; T. C. Salmon & Weber, 2016). In our own work on multi-player Stag Hunts, participants in groups of 3–6 chose interactively and repeatedly from many Stag strategies with different quorums (Frey & Goldstone, 2010). Groups in this experiment showed a strong preference for coordination at scale-2 and, to a lesser extent, scale 3, despite many structural factors favoring coordination at higher scales. Our results were consistent with the hypothesis that persistent dyad formation interfered with higher-scale coordination, but the experiment was too complex to establish a causal relation.

Methods

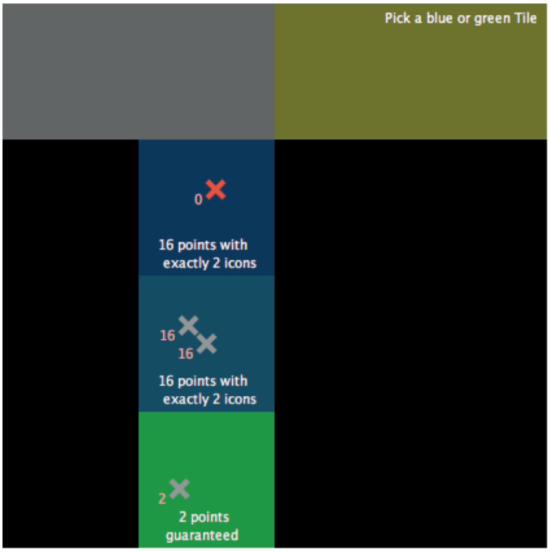

Figure 1. Screenshot of Modular Stag Hunt game interface.
Participants may select either of three tiles. The bottom tile will always be worth 2 points per round. The two upper tiles provide greater earnings of 16 points, but only when they are selected by exactly *n* players. The value of *n* changes by condition between 2, 3, and 4. When the results of the previous round are shared, participants are represented to each other by grey X's. In this screenshot, one player has selected the secure tile, two have successfully coordinated on a profitable tile, and a fourth has tried unsuccessfully to coordinate on the other. The small number next to each icon signals each player's accumulated earnings.

After describing the properties of the classic Stag Hunt and Modular Stag Hunt as single-shot games, we will introduce a design in which the Modular Stag Hunt is repeated for a series of "training" and "test" rounds.

The Modular Stag Hunt

The classic Stag Hunt was first posed to abstract the question "Am I willing to sacrifice the certainty of working alone for the potential economies of scale of a larger but uncommitted group?" In the allegorical Stag Hunt, two neighbors

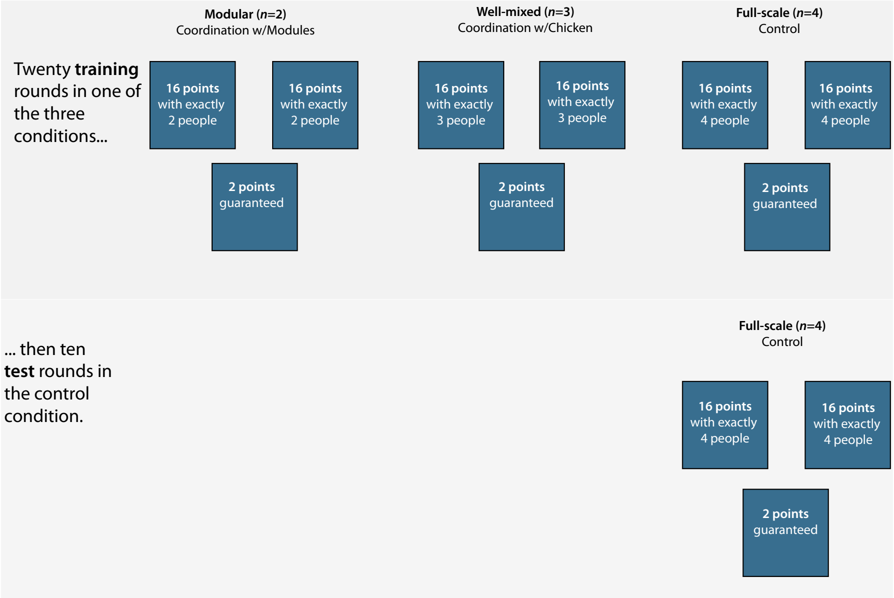

Figure 2. Experiment design, with game structure.
Groups of four picked from three choices over thirty rounds of our Stag Hunt game, moving from twenty training rounds to ten identical test rounds. They played in one of three conditions, Modular, Well-mixed, or Full-scale, each parameterized with a different quorum necessary for the rewards of successful coordination. The purpose of the experiment was to establish which training experience led to the best test performance.

decide whether to hunt small hare individually or to try working together to hunt stag, which is significantly more rewarding to all. Even though both individuals benefit by coordinating for the larger quarry, they may decide that the costs of coordination, and their uncertainty as to each other's actions', are too great. In its simplest game-theoretic form, the Stag Hunt is a "normal-form" (simultaneous-choice) coordination game with two Nash equilibria. The Nash equilibrium solution concept formalizes a sense of the stability of the different outcomes of strategic situations. Player behavior is at Nash if no individual can increase his or her earnings with a unilateral change of behavior. The concept should not be interpreted to imply a sense of the "best" action (e.g. most prosocial, profitable, moral, mutually beneficial, or even rational), but it is nevertheless surprisingly useful for predicting the behavior of some motivated, experienced participants in coordination games (Camerer, 2003; Camerer & Weber, 2007). In the Stag Hunt, the general incentives and pressures behind coordination are formalized in terms of a conflict between its equilibria. The properties of agents in the game are well-documented, both theoretically and empirically (Camerer, 2003; R. Cooper, DeJong, & Forsythe, 1990; Harsanyi & Selten, 1988), and Skyrms (2004) extended it to models of group formation.

In the Modular Stag Hunt, four (instead of two) players have three (instead of two) choices: they may select a Hare strategy for a small but sure payoff (2 points in our parameterization) or they may attempt to coordinate between two identical Stag strategies. A player who selects Stag earns 16 points if it was selected by *n* players in total, otherwise, it yields 0 points.[1] Changes in the value of parameter *n*, which may be equal to 2, 3, or 4, change the properties of the game dramatically. To simplify analysis and interpretation, we defined the game such that exceeding *n* gives the same null result as falling short of it.

At $n = 4$, which we call the Full-scale version of the Modular Stag Hunt, players who select one of the Stag strategies receive the large payoff only if all four players select it (Figures 1 and 2). In this version there are three pure-strategy Nash equilibria, corresponding to the decision of all four players to converge upon one of the three strategies. Previous work suggests that, as the game is iterated, players will settle into one of these three equilibria (Camerer & Weber, 2007).

For $n = 3$, the Well-mixed version, exactly 3 players must play Stag in order to receive a payoff. If all four players choose the same Stag strategy, then they each receive 0 points. At $n = 3$, the Modular Stag Hunt resembles the game of Chicken, in which the most profitable strategy for individual players gives the lowest payoff if all players select it. Both theory and experiment suggest that coordination successes will be unstable or intermittent in this condition of the game (Bornstein, Budescu, & Zamir, 1997). There are nine pure-strategy Nash equilibria in the Well-mixed condition but it is the only condition of the three in which there is no equilibrium in which all four players always receive the largest payoff. All of the pure-strategy equilibria involving Stag choices require one player to commit to the much less profitable Hare.

In all three of its parameterizations, the Modular Stag Hunt has many mixed-strategy equilibria, but, for simplicity, we focus on the more psychologically salient pure-strategy equilibria. Mixed strategies will not receive much attention in these analyses, except to say that randomizing uniformly between the Stag, a Nash strategy in all three conditions, is a more profitable strategy in the Well-mixed than in the other two conditions. For all of these reasons, stable internal structures should be unlikely to develop in the Well-mixed version.

---

[1] In all three conditions, "Pareto dominance" fails to select between the two Stag strategies, so in addition to the traditional problem of selecting between Hare and Stag there is an equilibrium selection problem between Stag. The complexity of introducing two Stag strategies was necessary to create the three nearly identical experimental conditions.

The Modular version, with $n = 2$, should support what we define as internal group structure. In this condition it is possible to split four players evenly between the two Stag strategies such that all four players receive the most profitable payoff. Players earn nothing when more or fewer than two players select a Stag. The condition is called Modular because groups in it will benefit from pairing off into two "subgroups." We predicted that experiences of profitable behavior in the Modular condition would promote the development of an internal group structure, and that groups in this condition would naturally distribute between the two Stag over time.

Experimental design

We tested the transfer of experience from each of the three versions of the Modular Stag Hunt to the $n = 4$ setting that demands full-scale coordination. In our design, participants repeated the Modular Stag Hunt for 20 training rounds and 10 test rounds. Groups of four were randomly assigned to one of three conditions, and all manipulations were between-subject. In the "Well-mixed" condition, groups trained in the $n = 3$ version before playing the ten $n = 4$ test rounds. In the "Modular" condition, groups played the training rounds at $n = 2$. And in the "Full-scale" condition, which acted as a control, both the training and test rounds were played with $n = 4$. Excepting these differences, the experimental interface was identical in all three conditions (Figure 1).

Our main interest was in the relationship between experiences of Modular and Full-scale training. We identified opposing predictions for how experience in the Modular condition would transfer to the Full-scale condition. If successful coordination transfers easily across scales then $n = 2$ should provide the best training for $n = 4$, but if different scales of coordination somehow compete then $n = 2$ should provide the worst training for $n = 4$. Success at one type of coordination has been shown to transfer to other types (Berninghaus, Ehrhart, & Keser, 2002; Devetag, 2005; Rankin, Van Huyck, & Battalio, 2000; Rick, Weber, & Camerer, 2007; Stahl, 2003). On the other hand, satisfactory performance at a small scale of coordination may be sufficient to disincentivize or otherwise impair performance at other scales (Frey & Goldstone, 2010; Weber & Camerer, 2003).

Subjects and procedure

52 psychology undergraduates in 13 groups of four played the game for cash. There were five groups in the $n = 4$ control condition and four each at $n = 2$ and $n = 3$. To account for the small number of sessions, we set a conservative significance threshold in our statistical analyses. Sessions were conducted over networked

computers contained within separate curtained cubicles. Participants were directed to be quiet, and each cubicle had a small noisy fan to prevent the transmission of explicit verbal signals or implicit verbal cues. Complete instructions were publicly read aloud and participants were then given an opportunity to review them individually on their terminals. The instructions were as follows:

> You will be playing 30 rounds of a game with three other players. This game will last about ten minutes.
>
> Each player's goal is to earn points by choosing tiles with their icon. You will see your icon as a red X. After the game you will be paid $0.01 per point.
> Your earnings will depend on your tile choices and the choices of others.
>
> Green tiles are guaranteed to pay a smaller number of points. Blue tiles pay many more points, but they need exactly the right number of icons to pay anything. If you take too long to choose, you will move to a green tile by default.

After the first round, participants saw their group's previous rounds' choices before the next round began. Experimental sessions lasted just under five minutes on average, with about 10 seconds per round. Participants were paid a small bonus of 1¢ per point and mean earnings were $1.25 (equivalent to $15.00 per hour). Because of its brief duration, the experiment was run in the free time after other collective behavior experiments. However, the behavior we report here is comparable to what we observed in similar experiments that we "crowdsourced" via Amazon's Mechanical Turk, increasing our confidence that our findings are not driven by the experience of previous games, and that they are not peculiar to psychology undergraduates.

Measures

Dependent variable

The main performance variable in this experiment was the number of test rounds, out of ten, in which groups settled on one of the pure-strategy equilibria containing the Stag strategy. If internal group structure can improve higher-level coordination, then Modular groups (those that trained in Modular versions of the game) will coordinate more successfully on $n = 4$ test trials than will Well-mixed or Full-scale groups.

Supporting dependent variables

To help us analyze competing explanations for the experiment's main results, we defined dependent variables for *payoff, fixation, risk aversion*, and *group internal structure*. Though each is defined here at the individual level, all measures were aggregated into per-group, per-round averages for inclusion in our statistical models. The variable payoff provides the mean number of points per individual per round. Fixation gives the observed probability of a participant's highest-probability choice, as measured by the phenomenon of choosing the same strategy repeatedly. By this measure, pure-strategy play (repeatedly selecting only one of the three strategies) registers the maximum fixation of 1.0, and uniformly random mixed-strategy play registers the minimum value of 0.33 (because there were three choices). Risk aversion, defined here as a type of fixation, is the percentage of choices, out of thirty, in which a participant selected Hare. It gave a measure of individuals' susceptibility to strategic uncertainty as to the actions of others in their group. Group internal structure is the standard deviation of the "distances" between participants. In a round, two participants shared distance 0 if they made the same choice, and 1 otherwise. Summed over all rounds, two participants have distance 30 if they never select the same choice and distance 0 if they always select the same choice. By definition, participants' net distances from themselves is always 0. The purpose of the internal structure measure was to detect whether participants coordinated consistently with specific others. A significant increase in the variation between a group's values of distance will be a general symptom of the emergence of the selective pairwise coordination preferences that should emerge only during Modular versions of the game.

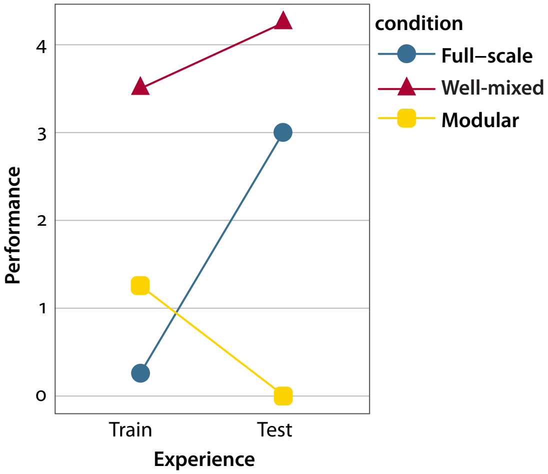

Figure 3. During test, participants in the Modular condition performed much worse than those in the other conditions.
During training, groups in the Well-mixed condition performed significantly better than those in the Full-scale condition. During test, groups in the Well-mixed condition performed significantly better than those in the Modular condition. This figure represents the results of the first row of Table 1.

Results

We used separate ANOVAs for train and test trials to establish differences between the three conditions. These analyses encoded three dummy variables, one for each condition. Additionally, to capture dynamics, multiple regressions tested the effects of increasing rounds of experience. To maintain the independence of observations, these analyses were conducted at the group level. Means and significance values are summarized in Table 1. Because of the small number of sessions and the many comparisons, we set a conservative significance threshold of $p = 0.01$ and established the internal consistency of our findings across the ANOVAs and the regressions.

Descriptives

Looking first at the 20 training trials of the Full-scale condition, it is clear that scale-4 coordination is difficult. On average, participants successfully coordinated on Stag in less than one of the twenty training rounds (0.25 rounds for every 10; first row of Table 1, which Figure 3 plots). Groups performed much better during Well-mixed training, coordinating successfully on a Stag in 7 of the 20 training rounds (and significantly more often than did groups in the Full-scale condition; $F(1,21) = 24$, $p < 0.01$). During training in the Modular condition, participants settled upon the stable configuration (two subgroups of two) in only 2.5 of the 20 training rounds. However, partial solutions were more common, with 21 instances (out of 40), of exactly two subjects successfully coordinating on one of the Stag.

Main test

During the final ten test rounds, performance in the Full-scale condition increased (non-significantly) from 0.5/20 to 2.75/10.[2] Variance in performance was much higher in this condition than in the other two; Full-scale groups tended either to converge stably on coordination success (10/10 test trials) or coordination failure (0/10 test trials), with little behavior between these extremes. Groups that trained in Well-mixed trials showed improvement upon advancing to the test trials ($t(32) = 3.16$, $p < 0.01$). Mean performance on test was 4.25 out of 10 rounds. This was not significantly different from the test performance of Full-scale

---

[2] An alternative comparison is to only the first ten rounds of the Full-scale condition, corresponding to an entirely untrained experience of the game. Across all sessions of the Full-scale condition, no groups successfully coordinated on Stag in the first ten rounds of training. This alternative comparison is similarly uninformative for the other two conditions.

groups. However, a post hoc *t*-test showed that after training in the Well-mixed condition, groups exhibited better test performance than in the Modular condition ($t(3) = -5.67$, $p = 0.010$). Across all test rounds of all groups in the Modular condition, no group successfully coordinated on Stag during test (0/20). Groups in the Modular condition did not show significantly different performance between training and test trials.

Supporting analyses

Group internal structure, payoff, fixation, and risk aversion were modeled in the same manner as performance, with a linear model to establish the effects of increasing experience in the game, and separate ANOVAs for train and test trials to establish differences between the three conditions. These dependent measures were also modeled at the group level. Means and significance values are summarized in Table 1.

Within Modular groups, internal structure did not change significantly with experience in the game (number of rounds played), but during test trials Modular groups had more internal structure than Full-scale groups ($F(1,33) = 6.76$, $p < 0.01$). In the Well-mixed and Full-scale conditions, experience predicted decreased internal structure ($t(32) = -3.95$; $t(32) = -5.95$, respectively; both $p < 0.01$).

Payoffs in the training rounds were significantly higher in both the Modular and Well-mixed conditions than in the Full-scale condition ($F(1,21) = 14.9$, $p < 0.01$; $F(1,21) = 20.5$, $p < 0.01$). These differences did not persist among the test trials for these groups, and the change between train and test trials was not significant in the regression.

Fixation, the most common strategy selected by a given subject, was selected for an average of 70% of rounds. This value increased with experience in all three conditions, but did not differ significantly across conditions (Full-scale: $t(32) = 6.06$, $p < 0.01$; Well-mixed: $t(32) = 4.83$, $p < 0.01$; Modular: $t(32) = 2.99$, $p < 0.01$).

Risk aversion, the frequency of selecting Hare, showed the same pattern as payoff. Risk aversion during the training of both Modular and Well-mixed groups was significantly below that exhibited by Full-scale groups ($F(1,21) = 13.2$; $F(1,21) = 12.7$, both $p < 0.01$) but these differences did not persist at test. Over all trials and conditions, only 9% of choices were to the Hare strategy, making the general pattern of coordination failures even more striking. Participants in the repeated Modular Stag Hunt tended to continue to attempt to overcome coordination failure for the duration of the experiment.

Table 1. Summary of experiment results

Dependent measures	Grand Mean	Full-scale (*n*=4; control)			Well-mixed (*n*=3)			Modular (*n*=2)		
		Train	change	Test	Train	change	Test	Train	change	Test
Performance	1.92	0.25	insignf.	3.00	**3.5**[*]	**Up**	*4.25*[a]	1.25	insignf.	*0.00*[a]
Group structure	2.82	3.00	**Down**	0.92	2.90	**Down**	2.25	**3.48**[*]	insignf.	**3.41**[*]
Payoff	4.19	1.60	insignf.	5.90	**4.8**[*]	insignf.	7.20	**5.56**[*]	insignf.	0.71
Fixation [0.33, 1]	0.70	0.66	**Up**	0.93	0.71	**Up**	0.79	0.57	**Up**	0.72
Risk aversion[0,1]	0.09	0.15	insignf.	0.14	**0.074**[*]	insignf.	0.05	**0.045**[*]	insignf.	0.09

Performance is number of successful coordinations on Stag out of 10 rounds.

* Values with bold markings report an effect significant at $p < 0.01$. In a "Train" column, comparison is with respect to training in the Full-scale condition. In a "Test" column, comparison is with respect to tests in the Full-scale condition. Within a condition, "Up" and "Down" report the directions of the significant effects of experience from Train to Test ($p < 0.01$). All regressions report statistics distributed on F(5,30).

a. Italics in upper-right Test columns reflect one significant posthoc test between Modular and Well-mixed test performance.

The results in the first row are plotted in Figure 3.

Discussion

Pairwise coordination prevents coordination at higher scales

In the Full-scale version of the Modular Stag Hunt, groups faced a version of the Weakest Link game (Van Huyck, Battalio, & Beil, 1990). Performance on the test trials was in a sensitive region between the ceiling (10) and the floor (0), leaving room for the other conditions to demonstrate either relative increases or decreases in performance.

For groups trained in the Modular condition, there was a theoretically stable and efficient (Nash-equilibrium and Pareto-dominant) outcome for two subgroups of two in which all four players could receive the largest payoff. Evidence supports the prediction that this environment promoted the emergence of stable subgroups. Groups in the Modular condition did in fact exhibit high internal structure consistent with a modular group structure. However, while groups in this Modular condition earned more during training than other groups, they were generally unable to settle consistently upon two groups of two simultaneously, or to use their experience at $n = 2$ to coordinate on a single Stag strategy during the $n = 4$ test rounds.

By contrast, after transferring from their experience in the Well-mixed coordination problem, groups with no experience of stable subgroups performed at least as well as groups trained in the Full-scale condition. This result is noteworthy because in the Full-scale condition, participants played precisely the same version of the game during both training and test rounds. The fact that the Well-mixed condition transfers at least as well as the Full-scale condition exposes the possibility that preventing the formation of internal structure may be as valuable for achieving full coordination as identical train/test conditions. The apparent benefit of these conditions over the Modular condition supports our claim that small-scale coordination actively interferes with coordination at larger scales.

Alternative explanations

Why else might groups with internal structure have tested worse than groups without internal structure? Supporting measures like fixation, risk aversion, and internal structure are helpful for rejecting a few alternative accounts of our results.

Perhaps game similarity is the best predictor of successful transfer. But the Well-mixed version of the game, in which performance was at least as strong, is the least similar of the three: it was the only one without all-Stag pure Nash equilibria or any outcome in which all four participants could earn the maximum reward.

Perhaps experience in the Modular condition undermines coordination at scale 4 because subjects learned to randomize after many rounds of failure at scale 2. This would be consistent with the moderately efficient "mixed" strategies that exist in this condition. At $n = 2$, randomizing uniformly between the Stag – possibly the most salient mixed strategy, with an expected utility of 6 points per round – is three times more profitable than selecting Hare. However, by this explanation participants in the Well-mixed condition should have performed much worse: the same mixed strategy as above yields an even higher expected payoff of 8 points per round, and it lacks any pure strategy equilibrium that is more efficient. Therefore, the Well-mixed condition should have been even more effective than the Modular condition at eliciting randomized behavior.

Perhaps the poor performance in the test phase of the Modular condition was due to its participants being somehow more likely to "opt out" of coordination and select Hare. However, risk aversion was not significantly higher in the Modular condition than in any other condition during test, and it did not change as participants experienced more rounds.

As reflected in Figure 3, Well-mixed groups were the most successful during training. Perhaps they out-performed Modular groups during test because transfer success is predicted by training success, independent of the details of the game. However, by a more lenient definition of success, groups in the Modular condition were at least as successful as Well-mixed groups. Unlike other groups, Modular groups could experience "partial" coordination success if two of the four players succeeded in coordinating on a Stag strategy. While Modular groups succeeded at full coordination in only 2.5 of 20 training rounds, they succeeded at partial coordination in 21 of 40 training rounds. So there is a sense in which the Modular task was easier in a way that does not seem to have facilitated transfer. Predictions that base transfer success on coordination success will have to articulate the difference between these senses of success.

The measure of group internal structure indicates that participants in the Modular condition implicitly identified each other and formed persistent dyadic subgroups. This experience of stable subgroups seems to have undermined full-scale coordination. This is in contrast to our original intuitions that the structures selected for by the Modular condition would ease the demands of achieving Stag-level coordination in the Full-scale condition.

History dependence

Our arguments have been presented in terms of a structural factor, the competition between different scales of group structure, but this framing obscures the psychological mechanisms behind behavior in these games. In all three conditions, the best predictor of a participant's next choice was their previous choice. Participants tended to persist on one strategy for 60–70% of rounds, and such persistence behavior increased with time. History dependence has been observed in other game theoretic environments (Duffy & Hopkins, 2005), and it is a likely mechanism for the competition we observe between scales. But it cannot be the whole explanation, because simple history dependence cannot explain both the relative success of transfer from the Well-mixed condition and the failure of transfer from the Modular condition. The theory of learning transfer has concepts of positive and negative transfer, in which a past experience either improves or interferes with learning in a later experience. Behavioral game theorists have identified instances of both positive (Devetag, 2005; Rankin et al., 2000; Rick et al., 2007; Stahl, 2003; Woolley, Chabris, Pentland, & Hashmi, 2010) and negative (D. J. Cooper & Kagel, 2005; Grosskopf, 2003) transfer in experimental games. Still, there is no satisfactory account, within the Modular Stag Hunt, for why Modular-to-Full-scale transfer would be negative whereas Well-mixed-to-Full-scale transfer is positive.

Conclusion

The Modular Stag Hunt was motivated by a suspicion that endogenously formed patterns of stable pairwise coordination could be used to "bootstrap" coordination at larger scales. However, our results are consistent only with the opposite claim, that successful coordination between pairs in multi-player games interferes with the formation of larger groups. Experiences of small-scale coordination did not make larger-scale coordination easier to achieve. Rather, we observed competition between the possible scales of stable coordination. Larger-scale coordination in this task may be most successful when pairwise interactions are discouraged, the way a teacher might separate friends during group projects, or a department chair might avoid arranging faculty offices by research focus.

The Modular Stag Hunt is a simple decision environment for testing non-monolithic multi-scale concepts of coordination and organization. The game reveals new kinds of limits on the adaptability of small self-organizing groups. It should continue to help researchers articulate questions about group structure and to organize the many conflicting concepts, across disciplines, that have been proposed to relate group structure, formation, modularity, and adaptability.

Acknowledgements

This manuscript is based on research that was presented at the 2010 and 2012 meetings of the Cognitive Science Society. We wish to thank Tatsuya Kameda, Keigo Inukai, and Tom Wisdom for their suggestions. We also wish to recognize the NSF REESE grant 0910218, the NSF EAPSI grant 1108165, NSF IGERT training grant 0903495 in the Dynamics of Brain-Body-Environment Systems, and the JSPS Summer Program. This research was approved by the Indiana University IRB.

References

Ahn, T., Esarey, J., & Scholz, J. (2009). Reputation and Cooperation in Voluntary Exchanges: Comparing Local and Central Institutions. *The Journal of Politics*, 71(02), 398–413. Retrieved from http://www.journals.uchicago.edu/doi/full/10.1017/S0022381609090355

Ahn, T., Isaac, R., & Salmon, T. (2008). Endogenous group formation. *Journal of Public Economic Theory*, 10(2), 171.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 22(6), 725–730.

Bavelas, A. (1952). Communication patterns in problem-solving groups. *Cybernetics: Circular Causal, and Feedback Mechanisms in Biological and Social Systems*, 1–44.

Berninghaus, S. K., Ehrhart, K. M., & Keser, C. (2002). Conventions and local interaction structures: Experimental evidence. *Games and Economic Behavior*, 39(2), 177–205.

Berninghaus, S. K., Ehrhart, K. M., & Ott, M. (2010). Cooperation and forward-looking behavior in Hawk-Dove games in endogenous networks: experimental evidence. *Mimeo*, Karlsruhe Institute of Technology, Faculty of Economics.

Bornstein, G., Budescu, D., & Zamir, S. (1997). Cooperation in intergroup, N-person, and two-person games of chicken. *Journal of Conflict Resolution*, 41(3), 384.

Camerer, C. F. (2003). Behavioral game theory: Experiments in strategic interaction. Princeton University Press.

Camerer, C. F., & Weber, R. (2007). Experimental Organizational Economics. *In the Handbook of Organizational Economics*, Eds. R. Gibbons and J. Roberts. New Jersey: Princeton University Press.

Camerer, C. F., & Weber, R. (2008). Growing organizational culture in the laboratory. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results*. Amsterdam: Elsevier.

Christie, L., Luce, R., & Macy, J. (1952). Communication and learning in task-oriented groups. *Technical Report*. MIT.

Cooper, D. J., & Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review*, 95(3), 477–509.

Cooper, R., DeJong, D., & Forsythe, R. (1990). Selection criteria in coordination games: Some experimental results. *American Economic Review*, 80(1), 218–233.

Devetag, G. (2005). Precedent transfer in coordination games: An experiment. *Economic Letters*, 89, 227–232.

Duffy, J., & Hopkins, E. (2005). Learning, information, and sorting in market entry games: theory and evidence. *Games and Economic Behavior*, 51(1), 31–62.

Frey, S., & Goldstone, R. L. (2010). Group Stratification and Coordination Failure in a Continuous N-Player Stag Hunt. Presented at the 2010 Proceedings of the Cognitive Science Society.

Fu, F., Hauert, C., Nowak, M. A., & Wang, L. (2008). Reputation-based partner choice promotes cooperation in social networks. *Physical Review E*, 78(2), 026117.

Glanzer, M., & Glaser, R. (1961). Techniques for the study of group structure and behavior II: Empirical studies of the effects of structure in small groups. *Psychological Bulletin*, 58(1), 1–27.

Goldstone, R. L., Roberts, M., Mason, W., & Gureckis, T. (2010). Collective Search in Concrete and Abstract Spaces. In T. Kugler, J. C. Smith, T. Connolly, & Y. Son (Eds.), *Decision Modeling and Behavior in Complex and Uncertain Environments* (pp. 277–308). New York: Springer Verlag.

Goldstone, R. L., Wisdom, T. N., Roberts, M. E., & Frey, S. (2013). Learning Along With Others. In *Psychology of Learning and Motivation* (Vol. 58, pp. 1–45). Elsevier. doi: 10.1016/B978-0-12-407237-4.00001-3

Grosskopf, B. (2003). Reinforcement and directional learning in the ultimatum game with responder competition. *Experimental Economics*, 6(2), 141–158.

Guetzkow, H., & Simon, H. (1955). The impact of certain communication nets upon organization and performance in task-oriented groups. *Management Science*, 1(3/4), 233–250.

Harsanyi, J. C., & Selten, R. (1988). A General Theory of Equilibrium Selection in Games. *MIT Press Books*, 1.

Hornsey, M. J. (2008). Social Identity Theory and Self-categorization Theory: A Historical Review. *Social and Personality Psychology Compass*, 2(1), 204–222. doi: 10.1111/j.1751-9004.2007.00066.x

Kearns, M., Suri, S., & Montfort, N. (2006). An experimental study of the coloring problem on human subject networks. *Science*, 313(5788), 824.

Klamt, S., Haus, U.-U., & Theis, F. (2009). Hypergraphs and Cellular Networks. *PLoS Computational Biology*, 5(5), e1000385. doi: 10.1371/journal.pcbi.1000385.g002

Kun, Á., Boza, G., & Scheuring, I. (2010). Cooperators Unite! Assortative linking promotes cooperation particularly for medium sized associations. *BMC Evolutionary Biology*, 10(173). Retrieved from http://www.biomedcentral.com/1471-2148/10/173

Leavitt, H. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 46(1), 38–50.

Maldonado, H., Klemmer, S. R., & Pea, R. D. (2009). When is collaborating with friends a good idea? Insights from design education (Vol. 1, pp. 227–231). Presented at the Proceedings of the 9th international conference on Computer supported collaborative learning, International Society of the Learning Sciences.

Mason, W., Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology*, 137(3), 422–433.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, 12(1), 3–23.

Newman, M. (2004). Detecting community structure in networks. *The European Physical Journal B*, 38, 321–330.

Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2).

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

Ravasz, E., Somera, A., Mongru, D., Oltvai, Z., & Barabási, A. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586), 1551.

Rick, S., Weber, R., & Camerer, C. F. (2007). Knowledge Transfer in Simple Laboratory Firms: The Role of Tacit vs. Explicit Knowledge. *Department of Social and Decision Sciences Working Paper*.

Riedl, A., Rohde, I. M. T., & Strobel, M. (2016). Efficient Coordination in Weakest-Link Games. *Review of Economic Studies*, 83(2), 737–767. doi: 10.1093/restud/rdv040

Salmon, T. C., & Weber, R. A. (2016). Maintaining Efficiency While Integrating Entrants From Lower Performing Groups: An Experimental Study. *The Economic Journal*. doi: 10.1111/ecoj.12308

Skyrms, B. (2004). The stag hunt and the evolution of social structure. *Cambridge Univ Press*.

Skyrms, B., & Pemantle, R. (2000). A dynamic model of social network formation. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), 9340.

Stahl, D. O. (2003). Action reinforcement learning versus rule learning. *Greek Economic Review*, 22, 27–56.

Weber, R. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review*, 96(1), 114–126.

Weber, R., & Camerer, C. F. (2003). Cultural conflict and merger failure: An experimental approach. *Management Science*, 49(4), 400–415.

Wenzel, M., Mummendey, A., & Waldzus, S. (2007). Superordinate identities and intergroup conflict: The ingroup projection model. *European Review of Social Psychology*, 18(1), 331–372. http://doi.org/10.1080/10463280701728302

Woolley, A., Chabris, C., Pentland, A., & Hashmi, N. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*, 330, 686–688. doi: 10.1126/science.1193147

Authors' addresses

Seth Frey
Neukom Institute for Computational Science
& Dept. of Psychological and Brain Sciences
Dartmouth College, Moore Hall
Hanover NH 03755
USA

seth.frey@dartmouth.edu

Robert L. Goldstone
Cognitive Science Program
Indiana University
1101 East Tenth St.
Bloomington, IN 47405
USA

rgoldsto@indiana.edu

Biographical notes

Seth Frey studies human decision behavior in complex social environments, approaching computational social science from the perspectives of behavioral game theory and cognitive science. He is presently a Postdoctoral Fellow at Dartmouth College's interdisciplinary William H. Neukom Institute for Computational Science. He was a postdoctoral researcher at Disney Research, a part of Walt Disney Imagineering. In 2013, he earned a Ph.D. in Cognitive Science and Informatics at Indiana University. He earned a B.A. in Cognitive Science from the University of California at Berkeley in 2004.

Since receiving his Ph.D. in psychology from University of Michigan in 1991, **Robert Goldstone** has been a professor in the Department of Psychological and Brain Sciences at Indiana University. His research interests include concept learning, collective behavior, and computational modeling of human cognition. He was awarded the 2000 APA Distinguished Scientific Award for Early Career Contribution to Psychology in the area of Cognition and Human Learning, and a 2004 Troland research award from the National Academy of Sciences. He served as editor of Cognitive Science from 2001–2005. In 2006 he became the director of Indiana University's Cognitive Science Program.