

Team Dating Leads to Better Online Ad Hoc Collaborations

Ioanna Lykourantzou
Luxembourg Institute of
Science and Technology
Esch-sur-Alzette, Luxembourg
ioanna.lykourantzou@list.lu

Robert E. Kraut
Carnegie Mellon University
Pittsburgh PA, USA
robert.kraut@cmu.edu

Steven P. Dow
University of California,
San Diego
San Diego CA, USA
spdown@ucsd.edu

ABSTRACT

Forming work teams involves matching people with complementary skills and personalities, but requires obtaining such data a priori. We introduce *team dating*, where people interact on brief tasks before working with a dedicated partner for longer, more complex tasks. We studied team dating through two online experiments. In Experiment 1, workers from a crowd platform independently wrote an ad slogan, discussed it with three consecutive people and evaluated their team date interactions. They then selected preferred teammates from a list showing average ratings for people they had dated and not dated. Results show that participants evaluated their dates based on evidence beyond externally judged slogan quality, and relied heavily on their dyad-specific judgments in selecting teammates. In Experiment 2, we replicated the individual and team dating tasks, and formed teams, either i) by honoring pairwise team dating preferences, ii) randomly from their pool of dates, or iii) randomly from those not dated. Results show that teams formed from preferred dates performed better on a final creative task compared to random dates or non-dates. Team dating provides a dynamic technique for forming ad hoc teams accounting for interpersonal dynamics. The initial interactions provided information that helped people select and work with an appropriate teammate.

Author Keywords

Thin slicing; team formation; distributed decision-making

ACM Classification Keywords

H.1.2 User/Machine Systems: Human factors, Human information processing; H.5.3 Group and Organization Interfaces: Collaborative computing, Computer-supported cooperative work.

INTRODUCTION

From multi-national corporations, to global consulting firms [28,29], online games [21,55], and crowd-sourcing [47,60], many organizations form short-term, ad hoc teams

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CSCW '17, February 25-March 01, 2017, Portland, OR, USA

© 2017 ACM. ISBN 978-1-4503-4335-0/17/03...\$15.00

DOI: <http://dx.doi.org/10.1145/2998181.2998322>

that range in size, from dyads to larger groups, to work on collaborative tasks. They need strategies for forming teams of people who have not previously worked together. This is a complex, multi-dimensional problem that requires managers to balance task needs, individual skills, interpersonal relationships, work styles, motivators and other factors [25]. The success of teams depends on the ability of their members to integrate and perform [2,71]. These challenges are amplified in ad hoc work situations where team members must quickly learn to work together.

Current team formation strategies focus primarily on staffing based on individual characteristics, such as skills [68], personality [7], or past worker performance. This approach has two limitations. First, it assumes that the skills required for a particular task *can be* defined a priori and that this information *already exists* in organizational records. However many tasks—like knowledge synthesis and creative problem solving—do not have well-defined skill requirements [30] and thus are not suited for skilled-based team formation strategies. Second, matching people based on individual characteristics ignores interpersonal compatibility, i.e. how well specific individuals work together. This information only becomes apparent *after* the group has formed and interacted. Poor interpersonal compatibility can result in conflicts and low team performance, even when the individual teammates are highly skilled [25,66].

In this work, we introduce *team dating*, a dynamic, task-independent technique for building ad hoc teams. This method allows potential members of a team to briefly work together to assess each other's skills and interpersonal dynamics. From peer ratings, a composition algorithm constructs teams for a more complex task. The key insight behind team dating is the concept of 'thin-slicing', which indicates that people can make accurate inferences about others based on brief exposure [4,17,38]; for example video analysis techniques have been used to predict divorce rates in couples [10]. Curşeu et al. applied the thin slice insight as a team formation strategy in a classroom, providing preliminary evidence for the efficacy of team dating in a face-to-face setting [18]. Our work extends this technique as a team formation strategy for online, ad hoc collaborations, like those used in crowdsourcing complex creative work (e.g. in the context of open innovation [50] or grand challenges [51]). As currently practiced, crowdsourcing is a challenging domain for ad hoc team formation, because those teams often know little about workers' skills and

because crowd teams are composed of strangers who must perform well together in a short time frame.

Through two experiments, our research explores whether crowd workers can extract useful information from team dating and whether using this information in the team formation process leads to better team outcomes.

Experiment 1 (N=33) examined whether working with someone yields information about a person that cannot be extracted through a simple external review of that person's work. Workers created a short writing sample (an ad slogan) and were consecutively paired with three co-workers for 3-minute interactions. After each date, the pairs rated each other on skill and compatibility. After three team dating rounds, workers selected three preferred teammates for a future task from a list that showed the average ratings for their dates as well as for the rest of the workers they had not dated. We also gathered external assessments of each worker's individual writing sample. Results indicate that workers evaluated their co-workers differently than external evaluators, and they prioritized information obtained during their team date interactions when selecting a teammate.

Experiment 2 (N=92) explored whether teams formed based on team dating led to better team outcomes. As in the previous study, workers created individual slogans, then interacted with and evaluated three different co-workers. We then formed final teams in one of three ways: i) *preferred dates*, comprising pairs of workers who gave each other high evaluations after their team date, ii) *random dates*, comprising random pairs of people who had worked together during team dating and iii) *random non-dates* comprising random pairs of people who had not worked together during team dating. Final teams created a second slogan which was also rated by external evaluators. Results show that working with preferred dates improved team performance compared to both random dates and random non-dates. A mediation analysis demonstrated that team dating influenced team outcomes by forming more compatible teams rather than by merely giving teammates more experience working together. There was no significant performance difference between teams comprising random dates or random non-dates.

Overall, our studies show that team dating improved team formation and led to better outcomes, thus providing different organizations an ad hoc technique for team formation that better accounts for interpersonal team dynamics.

RELATED LITERATURE

Selection criteria for team formation

Building effective teams is often defined as “*helping a work group become more effective in accomplishing its tasks and satisfying the needs of group members*” [16], p.792. Prior research has explored what constitutes a successful team [15], how teams develop [43], and how different selection criteria and competencies might lead a team to excel [23][44].

Gilley et al. [25] developed a comprehensive review and integrated theoretical model of selection criteria one must consider for deciding which workers to place in a group, with findings supported by multiple other studies. The researchers highlight pre-existing knowledge and skills—like problem solving, communication competencies, decision-making, goal-setting, performance and workload management capacity [33,34,68]—as the most important selection criteria. Other criteria include: the diversity or similarity of worker skills [62], workers' attitudes, personalities [7,8] and emotional intelligence [37,49]. They also highlight the importance of interpersonal dynamics among the team members, for example conflict, negotiation, strategies for reaching common ground, or trust [48]. In traditional corporate settings, some of these criteria are elicited through individual tests, while others, especially related to interpersonal skills, are determined through interviews in which managers, peers or members of the human resources department infer these attributes [32].

Fully automated team formation algorithms have also recently emerged to place people together in social networked environments. Wang et al.'s [75] comprehensive review of state-of-the-art team formation algorithms list the selection criteria used most often. These include: the set of experts and their task-specific skills, the social network connecting the experts, workload, personnel and communication costs. Such algorithms typically exclude interpersonal compatibility, perhaps because this information requires human assessment and is difficult to obtain at scale.

Indications for the efficacy of team dating

Thin slicing – quickly extracting useful information

Thin slicing is a term used in psychology to describe the ability of people to find patterns about events based on ‘thin slices’, i.e. narrow windows of experience. Thin slicing was pioneered by Ambady and Rosenthal [4], who showed that short (as low as half a minute) observations of people's behavioral expression allow accurate judgments on a number of outcomes like teacher effectiveness, interviewer trustworthiness or voting behavior; the accuracy of said judgments were not influenced by the type of communication (verbal or non-verbal) or the length of exposure.

Researchers have validated thin slicing—this innate ability to extract information about individual-level traits of others—across multiple traits and settings, such as marital discord [10], work skills [5], socioeconomic status [42], psychological adjustment of recently separated adults [53], general perceptions of psychological well-being [22], deception detection [1], and some group-level characteristics like cohesiveness [69] or affective dynamics [39].

While most thin slicing research investigates face-to-face settings, Ambady et al. [3] (p.203) argue that “*thin slices can be sampled from any available channel of communication, including the face, the body, speech, voice, transcripts or combinations of the above*”, provided that the selected

channel allows the collection of “*brief excerpt[s] of expressive behavior*”, which “*retain much if not most of the information encoded via dynamic, fluid behavior*”. Chat rooms, like the ones used in this study, belong to this category, since they allow the encoding of a dynamic, fluid and expressive behavior.

Thin slicing has proved effective in even more perceptually impoverished settings. For instance, Luborsky found that short written transcripts (30-50 words) of patients' psychotherapy sessions can predict psychological and somatic symptoms, and phobic behavior [45]. Single-word utterances were adequate to accurately assess creative cognition [58]. In an online context, short blog texts have been found to enable participants to accurately rate emotions [24], and online profiles enabled the formation of consistent impressions about a target person [67].

Thin slicing can be used to achieve a predefined goal. For instance, Waller et al. [74] used brief exposure to film excerpts to train students recognize and respond to team dynamics behaviors in situ. Although the thin-slice paradigm has shed light on how people process information and how they make judgments about others, it has rarely been applied to team formation. One exception, which provided inspiration for our current work, is the study by Curşeu and colleagues [18] that used a team dating approach to create student teams in a face-to-face educational setting. Our work builds on this prior work by investigating the efficacy of team dating in an online setting and by offering an efficient algorithm for ad hoc team formation.

Interpersonal compatibility affects team performance

Prior research indicates that team performance is not only mediated by the individual skills of team members, but also by preferences for co-workers and interpersonal compatibility. High preference positively influences long-term team performance, because it signals team members that they will function harmoniously within the boundaries of that group, and increases their cooperative behaviors [19]. Factors such as interpersonal reciprocity and perceived proximity to others—both results of communication and identification processes—have also been found to reduce the ambiguity of working at a distance [78], improve team growth and well-being, and to remain stable over time compared to the early acquaintance period [40]. Our work is based on these findings to create a team formation method that can capture and utilize interpersonal compatibility.

Learning to work together affects team performance

The experience of working together is also known to improve team performance. Transactive memory refers to the idea that people in a continuing close relationship develop shared and efficient mechanisms to encode, store and recall knowledge [76]. Just like individual metamemory allows individuals to know which information is available to them for retrieval, transactive memory allows teammates to know who knows what within the team and how to use this information to improve working together [59]. Learning to

work together, as a result of developing transactive memory and group norms, helps build group performance, improve decision quality and decrease group response time, with proven benefits for organizational knowledge assets [6]. The team dating method proposed in this work also provides people with a space to learn to work together and potentially build common memory systems.

Team Dating Hypotheses

In this work we introduce team dating, a team formation method that brings together people online to interact with different candidate teammates for brief slices of time, and then uses their dyad-specific judgments to form teams for longer, more complex tasks. Based on indications from thin slicing literature we hypothesize that:

H1: Team dating will provide workers with unique information about their candidate teammates and it will affect teammate selection.

Based on indications that interpersonal compatibility and learning to work together can increase team performance we hypothesize that:

H2: Forming teams based on team dating will give better outcomes, and this improvement will be due to interpersonal worker preference, or learning to work together, or both.

To examine how workers behave as a result of team dating, and the effects of the technique on team formation, we conducted two experiments. Experiment 1 (Figure 1), linked to Hypothesis 1, is designed to examine whether team-dating based evaluations differ from pure skill-based external evaluations of the same person, as well as how team dates use the additional information from team dating to choose teammates. Experiment 2 (Figure 2), linked to Hypothesis 2, builds on these findings by examining the impact of team dating on team performance, as well as the factors responsible for this impact.

EXPERIMENT 1 – (HOW) DOES TEAM DATING AFFECT WORKER EVALUATION AND TEAMMATE SELECTION?

This first study examined whether people use information from brief experiences working with others to judge potential teammates, and whether these judgments differ from judgments obtained by external reviewers who evaluated only production outcomes without team dating.

Participants performed an individual creative task and then sequentially briefly interacted in dyads with three randomly chosen co-workers. We call these interactions “dates”. Each date took place in a dedicated text-based chat room. Participants chose a nickname (pseudonym) which identified them to other participants throughout the experiment. After each interaction round, dates evaluated each other in terms of skill and interpersonal compatibility. Finally, each worker was given a list of all workers (both those they had and had not previously worked with), along with their average skill and compatibility ratings from three rounds of team dating,

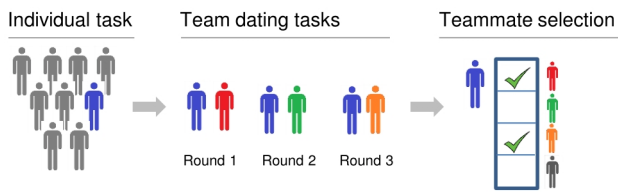


Figure 1. Experiment 1: participants performed individual work and shared it with three team dates. Participants then selected preferred teammates from all dates and non-dates.

and was asked to select three preferred partners. Independent judges—not involved in the team dating process—also evaluated the quality of workers’ individual slogans.

These data (worker evaluations after team dating and external slogan evaluations) were used as dependent measures in a correlation analysis and a logistic regression model that predicts workers’ selections of final partners.

To investigate which information workers relied on when selecting teammates we used logistic regression to predict teammate selection from the following factors: i) dyad-specific competence (i.e., the average skill and compatibility rating the selector gave a specific co-worker), ii) general teamwork competence (the average skill and compatibility ratings a worker received from everyone who worked with him or her) and iii) prior experience working together (a binary variable indicating whether a pair had worked together during the team dating rounds).

Participants

We hired 33 CrowdFlower Level-2 workers and provided them with a link to our web application that implements the team dating process. Gender balance was 33% female and 67% male, the majority (58%) of the participants were between 25-34 years old (33% was 18-24, 6% 35-44, and 3% 45-54 years old), and the majority (46%) had a college degree (30% advanced degree, 24% high school graduates). Participants earned \$4 for 20 minutes of time.

Procedure

Participants arrived at a “preparation” page that explained the overall process, and asked them to select a unique nickname which would help their co-workers identify them throughout the process. This page also explained how the application automatically transitions from one page to the next using a timer, to ensure that each worker will have a team dating partner for each round. Participants then had three minutes to write a short ad slogan for a fictitious product. We chose an ad design task because it fulfilled several key criteria (from [20]):

- Participants could exhibit individual creativity, but it would also benefit from collaboration.
- It is open-ended, complex and accepts different viewpoints, thus it is likely to be affected by interpersonal dynamics.
- It could be completed in a short duration.
- It did not require previous knowledge.

- External judges could rate the quality of the work.

The specific product we used was a fictitious coffee beverage product, called sCOPA:

sCOPA is a new coffee beverage from Papua New Guinea. The product is based on green and brown coffee beans. Green coffee beans are known for their contribution to a healthier diet. The beverage comes in two types: strong and mild. It can be drunk cold or warm. It will be sold both from super-markets and coffee stores.

The individual task fulfilled several purposes. First, we wanted to familiarize workers with the task specificities and difficulty so that they would be able to evaluate their peers’ skills during team dating. Second, during the team dates, participants could see each other’s individual work as an “ice-breaker” topic. Third, the individual work provided a means to evaluate the “objective skills” of each participant through external, independent ratings.

Next, workers entered the first of three team dating rounds. In each run they were randomly paired with another participant, showed their partner’s nickname, and entered a dedicated chat room for a three-minute conversation. To encourage pairs to have task-related conversations, we showed them each other’s individual slogans with a suggestion to combine them if they wished to.

As soon as the discussion time was up, both participants transitioned to an evaluation page where they were asked to rate each other in terms of skill and compatibility. After this peer evaluation, each participant was randomly paired with another person. The process repeated for three dates.

After three rounds of team dating and evaluations, participants viewed a teammate selection page that listed all potential co-workers (those dated and not dated), along with each person’s average ratings from all of his or her team dates. Workers were asked to select three preferred teammates for a final collaborative task. After their selection, they were informed that there would be no final production task and were paid.

Measures

To assess how participants made their decision to work with each other, we collected three measures:

External ratings of individual work

A new sample of independent CrowdFlower workers rated each individually produced slogan along six dimensions used in prior research [13,31] on a 5-point Likert scale (from “not at all” to “very much”).

- i. Originality: how original and creative is the slogan?
- ii. Customer attraction: how likely to attract customers?
- iii. Simplicity: how simple is the message of the slogan?
- iv. Honesty: how honest is the slogan?
- v. Uniqueness: how well does the slogan highlight the differences between this product and similar products?
- vi. Overall preference: how much did you like the slogan?

Because ratings were highly correlated, we combined them to form an individual slogan quality score ($\alpha=.92$). Each slogan was rated by 10 judges, and their judgments were averaged to calculate the final slogan-quality measure. Because some team dating pairs did not produce a slogan during their 3-minute round, we did not evaluate slogan quality from the team-dating rounds.

Team-date evaluations

After each team dating round, each participant rated the co-worker they had just talked to on a Likert scale (from 1=Not good at all to 5=Excellent) in terms of *skill* (“How good will [worker pseudonym] be in the main task?”) and *compatibility* (“How well do you think you will be able to collaborate with [worker pseudonym] as a team in the main task?”). Participants did not discriminate these two dimensions ($r=0.95$), so we averaged them to create a team date quality score ($\alpha=.97$) in subsequent analyses. This halo effect where judgments of one dimension influence others—even when there is sufficient information to allow for an independent assessment of them—is well-documented in many social judgment settings [57,61,79]. Across multiple team-dates, participants are rated consistently by different teammates (ICC=.61); that is, people rated highly in one team date tended to be rated highly in others.

Teammate selections

After working with all three team dates, each participant chose three teammates, producing a binary value (1=selected, 0=not selected) for every possible rater-ratee pair. This choice represents the dependent variable in the analyses presented below.

Results

Table 1 provides descriptive statistics for the variables used in the analyses. Participants produced relatively short slogans, with an average length of 6.7 words (st.dev=2.5). Slogans ranged from very good “*Green coffee for clean living*” (rated 4.1/5) to poor “*The best!*” (rated 2.7/5).

Team-date evaluations were only modestly correlated with external ratings

Individuals’ decision to select another person to work with in a putative final task was moderately correlated with their own evaluation of the target person, based on their brief team-date experience ($r(76)=.48, p<.001$).

Variable	N	Mean	SD
Slogan quality	52 slogans	3.36	0.41
Target quality rated by all team-dates	33 ratees	3.27	0.96
Target quality rated by final teammate	32 ratees	3.07	1.21
Target quality rated by final non-teammate	32 ratees	3.33	1.04
Worked with (0=no/1=yes)	931 pairs	0.82	0.28
Selected (0=no/1=yes)	931 pairs	0.09	0.28

Table 1. Descriptive statistics for Experiment 1

It was weakly correlated with others’ evaluation of the target based on their team-date experiences ($r(76)=.26, p=.02$) and external judges’ ratings of the target’s individually produced slogan ($r(828)=.10, p=.005$). Finally, the decision was also weakly correlated with whether the chooser had worked on a speed date with the target or not ($r(929)=.17, p<.001$).

We constructed two multivariate, logistic regression analyses to more precisely identify the influences on participants’ selections of partners for the putative final task.

The first analysis predicted 33 participants’ decisions to work with each of the other 25 or 26 other participants in their pool (N=830 choices nested within 33 choosers). Because participants made 25 or 26 choices, their choices were not independent of each other. Therefore, we used hierarchical logistic regression, with the chooser as a random effect, to account for the lack of independent among choices made by a single participant. As Model 1 in Table 2 shows, having worked with the target in a team date increased the odds of selecting that person for the final task (OR=4.18, $z=4.71, p<.001$). Higher external ratings of the target’s slogans also increased selection (OR=2.41, $z=2.53, p=.011$). However, the average rating that the target received from the three team dating partners did not change the odds of selecting him or her (OR=1.17, $z=1.23, p=.15$).

We conducted a second logistic regression predicting an individual’s choices to work with each of the three others with whom they had worked in a team date (N=79 choices nested within 32 choosers). In this analysis, only an individual’s own ratings of the person they worked with predicted selecting them as a partner for the final task (OR=16.15, $z=2.23, p=.03$). Once the chooser’s own evaluations of a target was controlled for, neither external ratings of the quality of the target’s slogans (OR=8.84, $z=1.28, p=.20$) nor evaluations the target received from other team dates (OR=.57, $z=-.83, p=.41$) predicted selecting the target as a teammate for a final team.

Predictor	Model 1: Choosing from all targets in the pool				Model 2: Choosing from targets one has speed dated			
	OR	SE	Z	p	OR	SE	Z	p
Intercept	.00	.00	-5.02	***	.00	.00	-2.07	*
Worked with	4.18	1.27	4.71	***	N/A			
Slogan quality	2.41	.84	2.53	*	8.84	15.05	1.28	
Target quality rated by all team-dates	1.23	.15	1.71		N/A			
Target quality rated by final teammate	N/A				16.15	30.12	2.23	*
N raters	33				32			
N choices	830				79			

Table 2. Logistic regression predicting whether a participant chose a partner as a final teammate

A mediation analysis (not reported in Table 2) shows that while individuals' own ratings of the person they worked had a direct effect on selecting them as a partner for the final task, neither external ratings of the quality of the target's slogans nor evaluations the target received from other team dates had either direct or indirect effects on selecting the target as a teammate.

Discussion

This experiment demonstrates that people use information from their brief work encounters to form judgments about their team dates. Participants did not simply evaluate their team dates based on the quality of their individual work; otherwise we would expect a higher correlation between team-date evaluations and the external slogan ratings.

Merely working with them in the past on a team-date increased wanting to work with them. In addition, one's personal evaluation of a team-date predicted preferring them as a partner in a subsequent task, while others' evaluations of that person based from their team-dating experience did not predict teammate selection and their slogan quality didn't predict the choice once the personal evaluation was accounted for. These results suggest that working with someone provides pair-specific information that influences one's judgments of a partner, and that "second hand judgments" provided by other people do not have the same effects.

This preliminary study raises new questions. Although it shows that team dating influenced choices of a teammate, it provides no evidence that team dating improves the performance of teams formed this way. Moreover, if team dating does help assemble better teams, what are the mechanisms at play? One possibility is selection, in which team dating helps people identify partners with good interpersonal compatibility and complementary skills. The second possibility is learning, in which previously working with another person helps dyads learn to work together. The goal of experiment 2 is to study if teams formed based on preferred teammate pairs perform better than randomly assembled teams, and to tease apart the factors that affect team-date performance differences.

EXPERIMENT 2: HOW DOES TEAM DATING AFFECT TEAM PERFORMANCE?

Experiment 2 extends Experiment 1 by including a final task to measure team performance. Since the workers' personal judgments of the people they worked with during team dating predicted whom they wanted to work with on a final task, we replaced the selection step with an algorithm to form teams based on team-date evaluations. The goal of this algorithm is to optimize team formation by trying to match each worker with the co-worker they mostly preferred working with during team dating, and do so for all final teams. The algorithm can also be parameterized to form teams randomly, ignoring the team dating evaluations, to construct the experiment's control conditions.

Experiment 2 tests the value of forming teams based on team-date evaluations by comparing three team formation methods: 1) *Preferred dates*: using the algorithm to pair workers who mutually rated each other highly in their team-date evaluations, 2) *Random dates*: randomly pairing people who had worked together during team dating without regard to their team-date evaluations, and 3) *Random non-dates*: randomly pairing workers who had not worked together during team dating. Each final team, regardless of condition, was asked to interact for a longer period of time and create a slogan for a new product. At the end workers completed a post-task survey, evaluating their final teams' slogans and final teammates in terms of skill and compatibility. As in experiment 1, we obtained external independent ratings of both the individual and final task slogans.

To investigate whether building teams based on interpersonal compatibility influenced team performance we contrasted the quality of slogans produced by final teams in the *Preferred dates* condition, with the slogans from the *Random dates* condition. To investigate whether the experience of interacting with someone for 3 minutes during team dating by itself influenced team performance we contrasted the team slogan quality in the *Random dates* condition with team slogan quality in the *Random non-dates* condition.

Participants

Ninety-two CrowdFlower, Level-2 workers participated in this experiment in batches of 4 to 12 who entered the experiment simultaneously. Each batch of workers was assigned to one of the three team-formation conditions in a round-robin fashion, resulting in 15 teams formed as preferred dates, 16 as random dates and 16 as non-dates. Participants were paid \$6 for 30 minutes of task time.

Procedure

Figure 2 shows the flow for Experiment 2. As in Experiment 1, each participant working independently first created a slogan for sCOPA coffee. They then participated in three, three-minute long team dating rounds, each with a randomly selected co-worker. Their task was to have an open-ended discussion about how to combine their individually produced slogans. In order to avoid a potential expectation mismatch, people were asked to evaluate their dates but were not explicitly told how these ratings will be used, i.e. for the formation of final teams. Finally, participants were randomly assigned to one of the three team-formation conditions, which were used to create final teams: i) *Preferred dates*, ii) *Random dates* and iii) *Random non dates*.

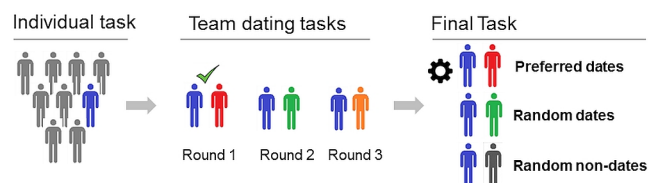


Figure 2. In Experiment 2, participants performed individual work and interacted with three teammates, then were automatically placed into final teams based on three conditions.

Team building algorithm

Figure 3 graphically illustrates the selection process used to form teams in the three experimental conditions. The algorithm first constructs an “affinity” graph for each possible pair of workers. Each edge of the graph has an “affinity value” equal to the average of the skill and compatibility ratings that the two workers gave to each other during team dating. If a pair did not work together in a team date, the value of the edge is zero. Then the algorithm is used to assign participants to one of three experimental conditions as follows:

1. **Preferred dates:** The goal of the algorithm here is to place all workers in a final team with the co-worker that they mostly enjoyed working with during team dating according to their peer evaluations. Computationally this translates to a graph cutting problem where the optimization objective is to cut the “affinity graph” in subsets of exactly two nodes each to maximize sum of the affinity values of the subset edges. Since this graph cutting problem is NP-Hard [11], we used a greedy approximation subroutine that enables the algorithm to run in polynomial time [64]. The subroutine starts by selecting the worker pair with the highest affinity value, reduces the graph and then repeats the process until all workers have been placed in final team pairs.
2. **Random dates:** The goal here is to place workers together in a final team with a random co-worker with whom they have collaborated during team dating. In this case, the algorithm *picks randomly* for each worker node from nodes he or she had previously worked with (i.e., edges between nodes had a value *greater than zero*), repeating the process until all workers have been placed in a final team. That is, pairing of participants is independent of their affinity value.
3. **Random non dates:** The goal here is to place workers together in a final team with a co-worker randomly chosen from among those with whom they have not collaborated during team dating. In this case the algorithm is parameterized to *pick randomly* for each worker node using only the edges that have a value *equal to zero*, repeating the process until all workers have been placed in a final team.

Final task

During the main task, workers are asked to collaboratively create a slogan for a fictitious non-profit organization. Note that the slogan was for an entirely distinct product than the one used in team dating.

EndAIDS is a US nonprofit organization dedicated to mobilizing and inspiring students to fight HIV/AIDS in Africa and promote global health equality. The organization implements two main actions: i) an Educational action, which aims to increase awareness about HIV/AIDS through education programs and actions in schools in the US and Africa, and a ii) a Prevention strategy, which aims to prevent HIV spread to young

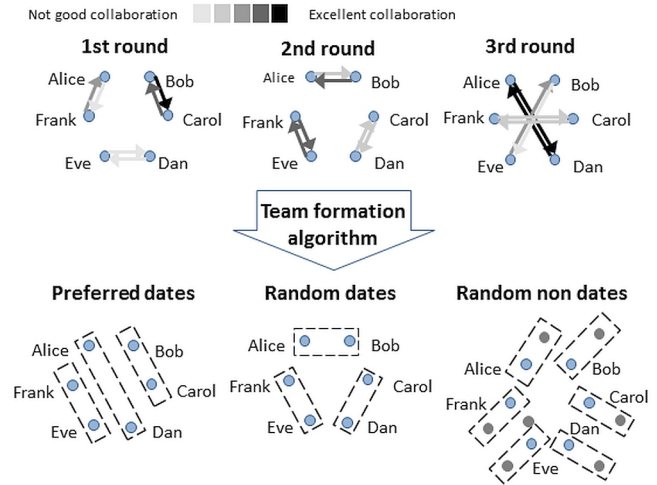


Figure 3. The team formation algorithm receives the ratings provided by the workers during team dating, calculates the affinity values for each possible pair and creates the final teams according to the three conditions.

people through on-site interventions, such as access to rapid HIV tests. These actions are supported by volunteers and a small number of employees. The EndAIDS organizers would like a slogan, to help the organization raise awareness and funds, and engage student volunteers in its actions.

Workers were given eight minutes to develop their ideas and collaborate. Once time was up, participants filled in a short post-task survey, evaluating their final team’s slogan, as well as their final teammate’s skill and compatibility.

Measures

Team-date evaluations

After each 3-minute team-dating round, the pairs of participants who worked together rated each other’s skill (“How skillful is [worker pseudonym] as a slogan writer?”) and compatibility (“How good will [worker pseudonym] be as a collaborator?”). Although the wording was changed slightly from Experiment 1 to reduce halo effects, the dimensions remained highly correlated ($r=.89$, $df=106$, $p<.001$). Therefore, we combined them into a single team-date evaluation scale ($\alpha=.94$). Participants evaluating a common target agreed moderately ($ICC=.50$), with the target accounting for 50% of the variance in team-date evaluations. For analysis purposes we separated the team-date evaluations into those exchanged among members of a final team (dyadic team-date evaluation) from those given by raters who were not members of the ratee’s final team (average team-date evaluation).

Teammates who worked together in the final 8-minute task also evaluated each other on the same dimensions. Again the items correlated with each other highly ($r=.77$), justifying combining them into a single scale ($\alpha=.77$). These teammate evaluations collected after the final task were moderately correlated with evaluations of these same person collected during the team-dating rounds ($r=.43$).

Slogan evaluations

As in Experiment 1, workers who were not involved in creating slogans evaluated the quality of each slogan written during the individual training task (individual slogan quality), and during the final task (final team slogan quality). Slogans were evaluated on three dimensions: originality, customer attraction and selection likelihood of experiment 1. The judgments were highly correlated and were averaged to construct a highly reliable scale ($\alpha=0.98$).

Because some team dating pairs did not produce a slogan during their 3-minute team dating round, we did not evaluate slogan quality from the team-dating rounds. Members of the final teams also rated the quality of their final slogan, using the same dimensions.

Analysis and Results

Table 3 shows the descriptive statistics per experimental condition of the variables used in the analyses below.

Preferred team dates outperform the other conditions

Our primary analysis examines whether the basis of team formation influenced the quality of the slogans teams produced, controlling for the quality of teammates' individual slogans and the judgments that non-teammates made of them during the team-dating rounds. The unit of analysis is the team nested within batch. Because all workers who participated in a session or batch were assigned to the same experimental manipulation, their performance may not be independent of each other. To control for the non-independence of team nested within batch, we used a random effects regression to analyze the data, with batch as a random effect. The regression also controlled for the quality of teammates' individual slogans and the the team-date evaluations teammates received from people who were not in their final team. Results appear in Table 4 and Figure 4.

Random date versus preferred date. The comparison of the preferred-date versus random-date teams tests whether forming teams based on preferences from initial team dating produced better teams. Results show that it did.

	Preferred date	Random date	Random non-date			
N	15	16	16			
Dyadic team-date evaluation	4.18	.98	3.01	1.03	N/A	
Target evaluation from all team-dates	3.62	.24	3.80	.69	4.15	.64
Target evaluation final teammate	3.73	.27	3.65	.32	3.63	.55
Individual slogan quality	3.25	.24	3.34	.32	3.07	.59
Final team slogan quality	4.06	.47	2.93	.73	3.38	.47

Table 3. Means and SD for variables in Experiment 2

Teams comprising people who had previously worked together in team dates and rated each other highly (preferred dates) produced better slogans than did teams comprising people who had previously worked together but were randomly paired (random team dates: $z=3.52, p<.001$). In addition, teams comprising preferred team dates also performed better than teams comprising non-team dates, i.e. people who had not worked together and then randomly paired (Random non-dates: $z=2.80, p=.038$).

The comparison of Preferred Date teams and Random Date teams tested whether matching potential teammates based on their initial judgments of each other improved final team performance. Mediation analysis (Figure 5) conducted using structural equation modeling shows that teammates' prior evaluations of each other during the team-dating rounds mediated the 28% of the total relationship between the experimental condition and performance outcomes.

Supplementary analyses and control variables included in the regression analysis are reported in Table 2 and rule out the possibility that initial rating of one final teammate by another is a proxy for the rater's generic task and teamwork skills rather pair-specific compatibility. Evaluations made during the team dating rounds by partners who would later be assigned to be teammates and people who were not assigned later to be teammates were not correlated ($r=.16, p=.39$). Second, as shown in Table 3, the quality of slogans teammates produced when working alone, as evaluated by external judges coming from the same population as study participants did not predict the quality of the slogans participants produced with their teammates. Therefore, stable individual skill at writing slogans cannot account for the relationship between team formation methods and final team quality. In addition, as shown in Table 3, team-date evaluations from non-teammates (i.e., those who were not assigned to work together on a final team) did not predict a team's final slogan quality and therefore did not mediate the relationship between experimental condition and team-slogan quality. *Together, these results suggest that during team dating potential partners were judging how well they as a pair would be able to work together, rather than a candidate teammate's generic teamwork ability.*

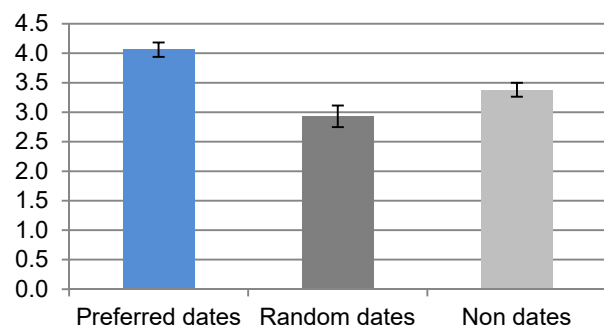


Figure 4. Effects of team formation methods on team's slogan quality (mean and SE).

Independent variable	Coef.	SE	Z	p
Intercept	4.58	0.79	5.78	***
Individual slogan quality	-0.17	0.21	-0.77	
Team-date evaluation from non-teammates	-0.01	0.13	-0.06	
Random date vs preferred date	-1.04	0.29	-3.52	***
Random non-date vs preferred date	-0.65	0.31	-2.80	*

Table 4. Predicting final team slogan quality from team-formation methods.

Random date versus random non-date. This comparison tests whether merely working with someone during the team dating rounds improved the final team performance. Results provide no support for the hypothesis that people who had worked together in the past performed better than people with no work history. Teams whose members were randomly paired with those they had previously worked with, performed no better than teams whose members were randomly assigned to someone with whom they had never worked ($z=1.34$, $p=.18$).

Participants do not change their opinion on a teammate

After completing the final task, teammates independently evaluated the slogan they produced using the same dimensions used by external evaluators, and their teammates on skill and compatibility, as they did during team dating.

The correlation between external evaluations and teammate evaluations of their own slogans was low and non-significant ($r(45)=.13$, $df=2$, $p=.38$), and the experimental condition did not influence participants' evaluation of their team slogan (Wald $\chi^2=.46$, $p=.79$). Nor was experimental condition associated with their final evaluations of their teammates (Wald $\chi^2=.14$, $df=2$, $p=.93$). Although their final evaluations of their teammates was moderately associated with external evaluations of the team slogan ($r(45)=.46$, $p<.001$), this result primarily reflects the high correlation between participants' team dating evaluations and their final evaluations of teammates ($r(31)=.52$, $p=.003$). Among participants who rated each other while team dating and after the final task, the partial correlation between teammates final evaluations of each other and external evaluations of their slogans is non-significant, once the initial team dating evaluation is controlled ($r(28)=.14$, $p=.47$).

DISCUSSION

In this paper we examined a new technique for online ad hoc team formation, called team dating. Team dating is based on short consecutive trial work sessions, where workers interact with and evaluate candidate teammates. Overall, our results with crowd workers performing a creative task showed that team dating is effective in providing participants with unique information about their co-workers, and that this information can be used to automatically create more effective teams. These results lead to two interesting conclusions:

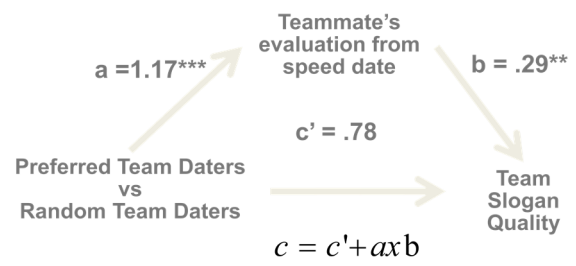


Figure 5. Mediation analysis showing that the relationship between team formation method and team quality is partially mediated by the teammates' prior evaluations of each other. C is the total effect of the experimental manipulation, c' the direct effect and the indirect effect.

First, the effects of team dating stem from dyad-specific teammate preferences, and not from evaluations of candidate teammates' individual skills or general teamwork ability. Dyad-specific preferences were found to be the strongest predictors of teammate selection in Experiment 1 and the only mediator of performance in Experiment 2.

Second, the benefits of team dating do not appear to come from learning to work together. Previous research on transactive memory has demonstrated that the members of a work group perform better if they trained and worked together for longer periods, in part because they learn each teammate's areas of expertise [6,59]. However, in the current studies, simply having worked with a teammate during a speed date did not improve subsequent performance as a team. One explanation may be that the brief work exposure allows people to make a quick intuitive assessment of who will be a good work partner, but may not provide enough time to learn others' strengths, or how to work with them. Another is that learning to work with a partner may not transfer across tasks. Finally, interference caused by rapid switching among speed-dating partners may prevent people from remembering the effective ways of working together.

Implications for Real-World Settings

We examined team dating in a specific environment: an online paid work setting in which two-person, crowdsourcing teams assembled from a relatively small population of workers completed a short creative task. Through this use case we showed that thin slicing information can be incorporated into a team formation algorithm to increase overall final team performance. Our findings are applicable to real-world settings, provided that the latter follow the assumptions of our tested case (online collaboration, one-time creative task etc.). Nevertheless, further research needs to be conducted to examine the extent to which these results apply to different settings. For example, would team dating be valuable in a corporate environment, where the company may have extensive histories about individual workers and the teams are expected to persist for months? Would it produce effective study groups in a large MOOC, where the teams may include 6-12 students working without pay and assembled from a class of thousands of students?

Examining the proposed method in a specific work context is important because different organizational environments impose different performance requirements and have different human resource constraints. These differences eventually impact the design of the team dating process and underlying algorithm. Take the example of a corporate setting where the organization seeks to optimize performance and has the ability to filter out low performing workers. In this case, team dating could be configured for *performance maximization*: the algorithm could construct a list of candidate worker teams ranked in order from highest to lowest performance and then only retain those teams that pass a certain compatibility/performance threshold. In a MOOC setting, where no student can be left behind for the sake of performance, team dating could be configured for *performance balancing*: the algorithm could construct candidate teams with similar compatibility/performance levels to increase the overall student satisfaction.

Additional Information for Team Formation

Our studies explore a method of using individual work samples (individual slogans) and short interactions to create teams. Using work samples is a particularly effective approach used in personnel selection for predicting actual job performance [63]. Moreover, this approach was specifically chosen as it can apply to creative project work where the required skillset may not be known a priori. We chose not to include additional information (CVs, work histories, skill assessments, personality testing, etc.), so as to avoid introducing additional variables into the experimental design, but also so as to not overload the workers with information to read before they could start their team dating interaction. Nevertheless, such information may be available in many real-world work settings and could potentially be integrated into the team formation strategy.

Design Considerations for Team Dating

Scalability issues: population size and number of teams

This paper's notion of team dating consists of two steps. First, a series of short interactions among potential teammates produces a dataset of subjective evaluations. Second, these numeric evaluations serve as input for a team formation algorithm. In our experiment, in step one, each participant had a dyad interaction with three other people (from a pool up to 30 participants) for three minutes each. In step two, the algorithm formed dyads for a team project. Both steps could encounter scalability concerns under a different set of assumptions. For example, the number of "dates" needed to get sufficient information regarding worker-to-worker preferences will vary with the size of the candidate pool. Three dates per worker probably would not suffice for a large worker pool. This relates to the data sparsity problem, often encountered in collaborative recommender systems, where each user has only rated very few items [26]. In team dating, one could address data sparsity using methods like matrix factorization [41] or low-rank matrix approximation [52]. In any case, the number of dates would need to scale relatively to the worker pool size.

The team formation algorithm itself is a greedy graph-cutting implementation that forms worker teams by first constructing a worker-to-worker affinity graph, and then extracting the team with the highest affinity, reducing the graph and repeating the process until all workers have been placed into teams. The algorithm runs in polynomial time and scales computationally [64], so it should still run efficiently with a larger worker pool. For different output choices, such as team size larger than dyads, future research will be necessary to understand the effects of scale. Additional algorithms, such as the minimal cost max network flow algorithm used in [77] could also be examined to optimize team selection.

Length of interactions

The amount of time per team date could have a number of implications. First, in terms of cost, modifying the interaction length changes the overall task time and number of potential partners. Second, while three-minute interactions provided enough information to form effective teams, longer interactions could potentially enable better judgments of skill and compatibility, perhaps allowing workers to distinguish between competent jerks and lovable fools [12].

Communication channels

The communication channel used in this study was synchronous interaction through text-based chat. Participants could see each other's nickname and initial slogan, and the discussion started from that point onwards. From a practical standpoint, chat is more accessible to participants with lightweight internet connections and there are fewer privacy concerns. Future research can examine rich communication channels that allow for non-verbal cues through audio or video, which may help workers make evaluations in shorter periods. This might also help workers remember their co-workers better, allowing us to observe transactive memory effects that were not observed in text chat. Overall, in line with recent studies indicating that rich interaction environments are sometimes but not always appropriate for gathering individual-trait judgments [73], further research is required to determine the most suitable communication channel(s) to deliver team dating with optimal results.

Errors in social assessment

Thin slicing evaluations, like the ones used in team dating, have proven to be accurate predictors of personal traits in various social evaluation contexts. Nevertheless, as with every assessment method, thin slicing is also prone to errors. An error can be either false-positive (a worker overestimating the skill/compatibility of their team dating partner) or false-negative (a worker underestimating their co-worker's skill/compatibility). In case of false-positives, the algorithm may place together people that do not fully match, while in the case of false-negatives it may keep compatible workers apart.

Human judgment error has three possible sources [27]: i) flawed experimental artifacts (e.g. misleading or incorrect instructions to subjects), ii) 'mental short-cuts' (e.g. judging

a person using stereotypes instead of actual facts, to reach a good-enough, fast evaluation [9,65,72]) and iii) error management effects (the brain's adaptive evolutionary bias to favor less costly errors over more costly ones). Subsequently the possible reparation measures also depend on the error source. A consistent experimental design with clear problem formatting and clear content can amend errors due to flawed artifacts. Minimizing external stimuli that contributes to stereotyping [54] (e.g. access to a co-worker's gender, ethnic, social or educational background) can potentially mitigate errors due to mental short cuts. Error due to evolutionary bias may be more difficult and even undesirable to amend, since practically it means that workers may sometimes miss a good collaborator in order to ensure avoiding a non-compatible one [56]. Nevertheless, "as long as immediate impressions are even minimally diagnostic, it may be more advantageous to form these first impressions than to dither and deliberate" [5], which from a task designer's perspective means being aware but tolerating a certain level of innate risk-averse behavior by the workers.

Further qualitative studies

Last, in this study we focus on the quantitative side of team dating, in terms of worker selections and measurable team performance. A natural extension is to examine deeper the qualitative processes that take place during and after team dating, in order to discover the mechanisms behind teammate selection and compatibility evaluation. Possible cues to be examined include signals of social or cultural affinity [70], similarity of individual work products, establishment of common ground [14] and trust [35], perceived aptness, equableness of communication and turn-taking style [36], as well as shared mental models [46].

CONCLUSION

In this work we examined team dating, a new technique for online ad hoc team formation. Team dating is based on short consecutive trial work sessions (dates) where workers interact with and evaluate candidate teammates. We studied team dating through two experiments, in a crowdsourcing setting and on the creative task of ad slogan development. The first experiment investigated whether team dating can provide workers with unique information about their candidate teammates and whether this information affects teammate selection. The second experiment examined whether forming teams based on team dating can lead to improved team performance, and the factors behind this. Our results showed that team dating allows workers to extract dyad-specific teammate preferences, which can then be used to automatically create more performant teams.

ACKNOWLEDGMENTS

The authors acknowledge funding support from the Luxembourg National Research Fund (FNR) grant #8734708 and NSF grants #1208382, #1122320, and #1122206.

REFERENCES

- Justin S. Albrechtsen, Christian A. Meissner, and Kyle J. Susa. 2009. Can intuition improve deception detection performance? *Journal of Experimental Social Psychology* 45, 4: 1052–1055.
- Shoshana Altschuller and Raquel Benbunan-Fich. 2010. Trust, Performance, and the Communication Process in Ad Hoc Decision-Making Virtual Teams. *Journal of Computer-Mediated Communication* 16, 1: 27–47.
- Nalini Ambady, Frank J. Bernieri, and Jennifer A. Richeson. 2000. Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in experimental social psychology* 32, 201–271.
- Nalini Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111, 2: 256–274.
- Nalini Ambady and John Joseph Skowronski. 2008. First impressions. *First Impressions* 31, 1: 368. <http://doi.org/10.1016/j.denabs.2007.12.003>
- Linda Argote and Yuqing Ren. 2012. Transactive Memory Systems: A Microfoundation of Dynamic Capabilities. *Journal of Management Studies* 49, 1375–1382.
- Murray R. Barrick, Greg L. Stewart, Mitchell J. Neubert, and Michael K. Mount. 1998. Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology* 83, 3: 377–391.
- Robert Bolton and Dorothy Grover Bolton. 2009. *People styles at work: Making bad relationships good and good relationships better*. AMACOM, New York, NY, USA.
- William P. Bottom. 2004. Heuristics and Biases: The Psychology of Intuitive Judgment. *Academy of Management Review* 29, 4: 695–698.
- Kim T. Buehlman, John M. Gottman, and Lynn F. Katz. 1992. How a couple views their past predicts their future: Predicting divorce from an oral history interview. *Journal of Family Psychology* 5, 3–4: 295–318.
- Thang Nguyen Bui and Curt Jones. 1992. Finding good approximate vertex and edge partitions is NP-hard. *Information Processing Letters* 42, 3: 153–159.
- Tiziana Casciaro and Miguel Sousa Lobo. 2005. Competent Jerks, Lovable Fools, and the Formation of Social Networks. *Harvard Business Review* 83, 6: 92–99.
- Kevin Chai, Vidyasagar Potdar, and Tharam Dillon. 2009. Content quality assessment related frameworks for social media. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and in Bioinformatics)*, 791–805.

14. Susan E. Clark, Herbert H.; Brennan. 1991. Grounding in Communication. *Perspectives on Socially Shared Cognition*: 127–149.
15. D Clutterbuck. 2007. *Coaching the team at work*.
16. Thomas G Cummings and Christopher G Worley. 2008. *Organization Development and Change*.
17. Jared R Curhan and Alex Pentland. 2007. Thin Slices of Negotiation: Predicting Outcomes From Conversational Dynamics Within the First 5 Minutes. *Journal of Applied Psychology* 92, 3: 802–811.
18. Petru L Curşeu, Patrick Kenis, Jörg Raab, et al. 2010. Composing Effective Teams through Team Dating. *Organization Studies* 31, 7: 873–894.
19. Erich C Dierdorff, Suzanne T Bell, and James A Belohlav. 2011. The power of “we”: effects of psychological collectivism on team performance over time. *The Journal of applied psychology* 96, 2: 247–262.
20. Steven P Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel L Schwartz, and Scott R Klemmer. 2011. Prototyping Dynamics : Sharing Multiple Designs Improves Exploration , Group Rapport , and Results. *Acm*: 1–10.
21. Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J Moore. 2007. The Life and Death of Online Gaming Communities : A Look at Guilds in World of Warcraft. *Distribution In Press*,: 839–848.
22. Jacqueline N W Friedman, Thomas F. Oltmanns, and Eric Turkheimer. 2007. Interpersonal perception and personality disorders: Utilization of a thin slice approach. *Journal of Research in Personality* 41, 3: 667–688.
23. James L Gibson, John M Ivancevich, James H Donnelly, and Robert Konopaske. 2012. *Organizations Behavior, Structure, Processes*.
24. Alastair J Gill, Darren Gergle, Robert M French, and Jon Oberlander. 2008. Emotion rating from short blog texts. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*: 1121.
25. Jerry W. Gilley, M. Lane Morris, Alina M. Waite, Tabitha Coates, and Abigail Veliquette. 2010. Integrated Theoretical Model for Building Effective Teams. *Advances in Developing Human Resources* 12: 7–28.
26. Miha Grčar, Dunja Mladenič, Blaž Fortuna, and Marko Grobelnik. 2006. Data Sparsity Issues in the Collaborative Filtering Framework. In *Advances in Web Mining and Web Usage Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 58–76.
27. Martie G Haselton and David C Funder. 2006. The Evolution of Accuracy and Bias in Social Judgment. *Evolution and social psychology*, January: 15–37.
28. Pamela J. Hinds and Mark Mortensen. 2005. Understanding Conflict in Geographically Distributed Teams: The Moderating Effects of Shared Identity, Shared Context, and Spontaneous Communication. *Organization Science* 16, 3: 290–307.
29. Pamela Hinds and Cathleen McGrath. 2006. Structures that work: Social structure, work structure and coordination ease in geographically distributed teams. *20th Anniversary ACM Conference on Computer Supported Cooperative Work, CSCW 2006, November 4, 2006 - November 8, 2006*: 343–352.
30. Eric von Hippel and Georg von Krogh. 2013. Identifying viable “ need-solution pairs”: Problem solving without problem formulation. *von Hippel, Eric von Krogh, Georg*, January 2016: 1–28.
31. Bob Hoffman. 2009. *The ad contrarian*. Fowler Digital Services.
32. John R. Hollenbeck, D. Scott DeRue, and Rick Guzzo. 2004. Bridging the gap between I/O research and HR practice: Improving team composition, team training, and team task design. *Human Resource Management* 43, 353–366.
33. Lawrence Holpp. 1999. *Managing teams*. McGraw-Hill, New York, NY, USA.
34. Stephen E Humphrey, Frederick P Morgeson, and Michael J Mannor. 2009. Developing a theory of the strategic core of teams: a role composition model of team performance. *The Journal of applied psychology* 94, 1: 48–61.
35. Sirkka L. Jarvenpaa and Dorothy E. Leidner. 1999. Communication and Trust in Global Virtual Teams. *Organization Science* 10, 6: 791–815.
36. Brigitte Jordan and Austin Henderson. 1995. interaction analysis: Foundations and practice. *Journal of the Learning Sciences* 4, 1: 39–103.
37. Peter J. Jordan, Neal M. Ashkanasy, Charmine E J Härtel, and Gregory S. Hooper. 2002. Workgroup emotional intelligence. Scale development and relationship to team process effectiveness and goal focus. *Human Resource Management Review* 12, 195–214.
38. Malte Jung, Jan Chong, and Larry Leifer. 2012. Group hedonic balance and pair programming performance: affective interaction dynamics as indicators of performance. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 829–838.

39. Malte Jung and Larry Leifer. 2011. A Method to Study Affective Dynamics and Performance in Engineering Design Teams. *International Conference on Engineering Design, ICED'11*.
40. David A Kenny and Lawrence La Voie. 1982. Reciprocity of interpersonal attraction: a confirmed hypothesis. *Social Psychology Quarterly* 45, 1: 54–58.
41. Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8: 42–49.
42. Michael W. Kraus and Dacher Keltner. 2009. Signs of socioeconomic status: A thin-slicing approach. *Psychological Science* 20, 1: 99–106.
43. Vincent Lenhardt. 2004. *Coaching for meaning: The culture and practice of coaching and team building*. Palgrave Macmillan, New York, NY, USA.
44. Daniel Levi. *Group dynamic for teams*. Thousand Oaks, CA: Sage.
45. Lester Luborsky. 1996. Onset conditions for psychological and psychosomatic symptoms during psychotherapy: A new theory based on a unique data set. *American Journal of Psychiatry* 153, 7: 11–23.
46. Andy Luse, James C. McElroy, Anthony M. Townsend, and Samuel Demarie. 2013. Personality and cognitive style as predictors of preference for working in virtual teams. *Computers in Human Behavior* 29, 4: 1825–1832.
47. Ioanna Lykourantzou, Angeliki Antoniou, Yannick Naudet, and Steven P. Dow. 2016. Personality Matters : Balancing for Personality Types Leads to Better Outcomes for Crowd Teams. *CSCW 2016*: 260–273.
48. Merce Mach, Simon Dolan, and Shay Tzafrir. 2010. The differential effect of team members' trust on team performance: The mediation role of team cohesion. *Journal of Occupational and Organizational Psychology* 83, 3: 771–794.
49. Gretchen A. Macht, Robert M. Leicht, and David A. Nembhard. 2013. Emotional Intelligence, Communication, and Team Performance. *Proceedings of the 2013 IIE Annual Conference*, 2810–2819.
50. Tammy L. Madsen, Jennifer Woolley, and Kumar Sarangee. 2012. *Using Internet-based collaboration technologies for Innovation: crowdsourcing vs. expertsourcing*. Boston.
51. Thomas W. Malone. 2011. Solving Climate Change with Crowdsourcing. *MIT Sloan Experts Blog*.
52. Ivan Markovsky. 2012. *Low Rank Approximation - Algorithms, Implementation, Applications*.
53. Ashley E Mason, David a Sbarra, and Matthias R Mehl. 2010. Thin-slicing divorce: thirty seconds of information predict changes in psychological adjustment over 90 days. *Psychological science : a journal of the American Psychological Society / APS* 21, 10: 1420–1422.
54. Craig McGarty, Vincent Y. Yzerbyt, and Russell Spears. 2002. Social, cultural and cognitive factors in stereotype formation. *Stereotypes as explanations: the formation of meaningful beliefs about social groups*: 1–15.
55. B Nardi and J Harris. 2010. Strangers and friends: Collaborative play in World of Warcraft. *International Handbook of Internet Research*: 395–410.
56. Randolph M. Nesse. 2001. The Smoke Detector Principle - Natural Selection and the Regulation of Defensive Responses. In *Unity of Knowledge – The Convergence of Natural and Human Science*. 75–85.
57. Richard E Nisbett and Timothy Decamp Wilson. 1977. The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Feisonality and Social Psychology* 35, 4: 250–256.
58. Ranjani Prabhakaran, Adam E Green, and Jeremy R Gray. 2014. Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior research methods* 46, 3: 641–659.
59. Ray Reagans, Linda Argote, and Daria Brooks. 2005. Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together. *Management Science* 51, 6: 869–881.
60. Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, et al. 2014. Expert Crowdsourcing with Flash Teams. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*: 75–85.
61. Frank E. Saal, Ronald G. Downey, and Mary a. Lahey. 1980. Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin* 88, 2: 413–428.
62. Kai Sassenberg, Kai J Jonas, James Y Shah, and Paige C Brazy. 2007. Why some groups just feel better: the regulatory fit of group power. *Journal of personality and social psychology* 92, 2: 249–267.
63. Frank L. Schmidt and John E. Hunter. 1998. The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin* 124, 2: 262–274.
64. Robert Sedgewick and Kevin Wayne. 2011. *Algorithms*.
65. Herbert A. Simon. 1956. Rational choice and the structure of the environment. *Psychological Review* 63, 2: 129–138.

66. Scott A. Snook and Jeffrey T. Polzer. 2004. *The Army Crew Team (Case 9-403-131)*. Cambridge, MA.
67. Kristin Stecher and Scott Counts. 2008. Thin slices of online profile attributes. *proc. ICWSM*: 127–135.
68. Michael J. Stevens and Michael A. Campion. 1999. Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management* 25, 2: 207–228.
69. Paul E Stillman, Thomas Gilovich, and Kentaro Fujita. 2014. Predicting group outcomes from brief exposures. *Social Cognition* 32, 1: 71–82.
70. Fons Trompenaars and Charles Hampden-Turner. 1993. Riding the waves of culture: Understanding cultural diversity in business. *Long Range Planning* 26, 5: 153.
71. Mary Ann C. Tuckman, Bruce W. Jensen. 1977. Stages of small-group development revisited. *Group & Organisation Management* 2, 4: 419–427.
72. Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185, 4157: 1124–1131.
73. Helen J. Wall, Paul J. Taylor, John Dixon, Stacey M. Conchie, and David A. Ellis. 2013. Rich contexts do not always enrich the accuracy of personality judgments. *Journal of Experimental Social Psychology* 49, 6: 1190–1195.
74. Mj Waller, G. Sohrab, and Bw Ma. 2013. Beyond 12 Angry Men Thin-Slicing Film to Illustrate Group Dynamics. *Small Group Research* 44, 4: 446–465.
75. Xinyu Wang, Zhou Zhao, and Wilfred Ng. 2016. USTF: A Unified System of Team Formation. *IEEE Transactions on Big Data* 2, 70–84.
76. Daniel M Wegner, T. Giuliano, and P. T. Hertel. 1985. Cognitive interdependence in close relationships. *Compatible and incompatible relationships*: 253–276.
77. Miaomiao Wen, Keith Maki, Xu Wang, Steven Dow, James D. Herbsleb, and Carolyn Penstein Rosé. 2016. Transactivity as a Predictor of Future Collaborative Knowledge Integration in Team-Based Learning in Online Courses. *Proceedings of the 9th International Conference on Educational Data Mining*, 533–538.
78. J. M. Wilson, M. Boyer O’Leary, a. Metiu, and Q. R. Jett. 2008. Perceived Proximity in Virtual Work: Explaining the Paradox of Far-but-Close. *Organization Studies* 29, 7: 979–1002.
79. David J. Woehr and Allen I. Huffcutt. 1994. Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology* 67: 189–205.