Confidence Boost in Dyadic Online Teamwork: An Individual-Focused Perspective

Liye Fu[*1], **Andrew Z. Wang**[*†2] **and Cristian Danescu-Niculescu-Mizil**[1]
[1]Cornell University, [2]Stanford University
liye@cs.cornell.edu, anwang@cs.stanford.edu, cristian@cs.cornell.edu.

Abstract

Individuals are often more confident in their solutions when working in teams than when working on their own. This confidence boost is observed even when it is not accompanied by a corresponding gain in performance, raising the question of what other factors might be responsible. We address this question by developing a large-scale experimental setting in the form of a two-player online game that allows us to track the confidence of individuals in naturally-occurring online collaborative tasks. This setting enables us to disentangle and compare the effects of different components of the collaborative process on the confidence of each team member. We show that confidence evaluations are subject to social influence: a low-confidence individual receives a confidence boost as a direct consequence of interacting with their teammate, and the extent of the increase depends more on the confidence, rather than on the competence, of the teammate. The resulting framework can enhance our understanding of confidence boost as an often overlooked byproduct of online teamwork and has implications for designing better online collaboration platforms to meet diverse collaborative objectives.

1 Introduction

One often overlooked byproduct of teamwork is the effect it has on the confidence of those involved. We know that teams are likely to express higher levels of confidence compared to individuals in problem-solving tasks (Sniezek and Henry 1989; Allwood and Björhag 1990; Sniezek 1992), save for a few contrary results (Tindale 1989). Notably, this boost in confidence is not necessarily justified by a corresponding improvement in performance (Stephenson and Wagner 1989; Meudell, Hitch, and Kirby 1992; Heath and Gonzalez 1995; Schuldt et al. 2017), raising the question of what might account for this phenomenon.

In this work, we take a finer-grained perspective and analyze the effect of team interaction on the confidence of *individual* team members. To this end, we develop a large-scale experimental platform that tracks individual confidence over the course of the interaction, allowing us to disentangle and compare factors mediating individual-level confidence boost in team collaborations.

[*]Equal contributions.

[†]Work done at Cornell University.

We first establish through a randomized experiment that low-confidence individuals can get a boost in confidence *as a direct result* of discussing with their teammates. By underlining the role of team interactions in shaping confidence, this finding motivates an investigation into interaction-based factors that are predictive of confidence boost.

We start from the observation that confidence estimations can be thought of as second-order opinions: they are an individual's opinion about the proposed solution. Knowing that individual opinion formation is subject to social influence—whether normative (Asch 1951; 1955) or informative (Turner et al. 1987; Turner, Wetherell, and Hogg 1989)—we may expect an individual's confidence to also be influenceable. In particular, following results from non-collaborative settings (Moussaïd et al. 2013), we hypothesize that low-confidence individuals might be influenced by the confidence of the teammates with whom they interact. We provide support for this hypothesis through a regression analysis, revealing that the confidence of one's teammate is an important predictor of an individual's confidence in the final group decision.

Importantly, we find the teammate's competence—how much the teammate actually knows about the task—and the quality of the team solution—the very subject of the confidence evaluation—to be far less predictive factors. This suggests that individuals are not as much influenced by their teammate's knowledge or by the progress the team makes, as they are by the confidence their teammates exhibit.

While this demonstrates that confidence evaluations are subject to social influence in online collaborative settings, we find that not all individuals are equally influenceable. In particular, low-confidence individuals are more susceptible than high-confidence ones. This uneven effect of confidence influence can serve as an explanation of the overall team-level confidence boost observed in prior literature.

This more nuanced understanding of confidence boost can have implications for the design of platforms for online teamwork to meet diverse collaboration needs. For example, when the quality of the eventual group decision is crucial, platforms might help collaborators to avoid unjustifiably high confidence in poor solutions. In contrast, in scenarios where team-bonding is the priority, platforms could instead aim to facilitate higher participant confidence in the outcome of the collaboration.

In summary, in this work, we:

- Design an online experimental setting that allows us to track changes in confidence exhibited by *individuals* over the course of online teamwork.
- Establish that individuals can undergo confidence boost as a direct result of interacting with their teammates.
- Show that a teammate's confidence (rather than their competence) is an important factor mediating this boost.

2 Background

Performance and Confidence in Teamwork

Performance of task-oriented teams. People often work in teams with the hope that by exchanging information with others, they might eventually produce higher-quality solutions. It is thus not surprising that substantial attention has been devoted to studying team performance. As a result, we have an increasing understanding of mechanisms that are conducive to better-quality solutions in task-oriented teams, both offline (Laughlin and Adamopoulos 1980; Williams and Sternberg 1988; Laughlin et al. 2006; Bahrami et al. 2010; Woolley et al. 2010) and online (Coetzee et al. 2015; Niculae and Danescu-Niculescu-Mizil 2016, inter alia). Some of these insights have already led to the design of platforms to promote more constructive collaborations in online teams. For instance, Salehi et al. (2017) have proposed a system, Huddler, to assemble ad hoc teams while optimizing for fit between team members by harnessing the benefit of member familiarity.

Dyadic confidence expression. While less studied than team performance, confidence expressions in teamwork have been investigated in a number of studies in offline, face-to-face settings. The most common tasks studied include general knowledge questions (Allwood and Björhag 1990; Allwood and Granhag 1996; Schuldt et al. 2017), risky shifts choices (Stoner 1961), and perhaps more consequentially, memory recall tasks (Stephenson et al. 1986; Stephenson and Wagner 1989), which have implications for assessing validity of collaborative testimonies.

It is generally observed that teams tend to express higher levels of confidence compared to individuals working on the same task. This trend holds in both experiments applying between-subject designs (Stephenson and Wagner 1989; Allwood and Björhag 1990) and those using within-subject designs (Allwood and Granhag 1996; Schuldt et al. 2017), although contrary results exist (Stoner 1961; Tindale 1989).

Given that online communication offers different affordances and has characteristics distinct from offline settings, we are drawn to the question of how much this observed trend of confidence boost would generalize to an online environment, and whether we can quantify such confidence boost at an *individual* level.

Role of individual confidence in teamwork. The role of an individual's initial confidence in teamwork has been studied in various contexts. For instance, group members' initial confidences are shown to affect the participation levels (Sniezek and Henry 1989) and the criterion for consensus (Boje and Murnighan 1982) in offline settings. Initial confidence is also shown to be related to an individual's influence over eventual team decisions (Stephenson et al. 1986).

While these studies highlight the role of an individual's initial confidence, in many real-life scenarios, their *final* confidence in their team solution is also important. This is especially true when the actual quality of the team solution remains elusive, and individuals need to rely on their subjective perceptions of the team solution to inform their subsequent actions. These intuitions motivate our study and platform design, which tracks fine-grained individual-level confidence changes over the course of team collaborations.

These considerations lead to our preliminary question: Can we observe and quantify *individual-level* confidence boost in online teamwork?

Confidence and Related Concepts

Confidence has been studied broadly and has branched out to include several related concepts, such as *overconfidence*, which includes, notably, the Dunning-Kruger effect (Kruger and Dunning 1999), and confidence realism (Adams and Adams 1961). In this section, we review a few concepts most relevant to our study.

Confidence operationalizations. In prior work, confidence has been most commonly studied under the following three operationalizations:[1]

1. Confidence as an estimation of *quality* of performance. Examples of this operationalization include asking students to self-rate their grades,[2] and prompting individuals to estimate task-completion times (Buehler, Griffin, and Ross 1994).
2. Confidence as an individual's *certainty* in the answer. For instance, participants may be asked to rate their confidence with the options such as "certain," "fairly certain" and "doubtful" (Stephenson and Wagner 1989).
3. Confidence as an estimation of one's relative *placement* within a population. With this operationalization, an individual may be asked to rank their performance relative to others (Kruger and Dunning 1999).

In this work, we employ the first confidence operationalization, due to its unambiguity in interpretation. We further explain the rationale behind this choice in Section 3 (Operationalizations and Concepts).

Factors affecting confidence. Several studies have attempted to analyze potential sources behind people's confidence estimates. A correlational analysis suggests that confidence is dependent on the amount and the strength of supporting (but not contradicting) evidence (Koriat, Lichtenstein, and Fischhoff 1980). Tversky and Kahneman propose that internal consistency is a key factor determining individual confidence judgment, which often results in unwarranted confidence from what they term "the illusion of valid-

[1] These operationalizations naturally apply to the study of overconfidence, for which Moore and Healy (2008) has provided a detailed discussion.

[2] See Boud and Falchikov (1989) for a review on the operationalization on student self-assessment studies.

ity" (Tversky and Kahneman 1974). In interactive decision-making settings, experiments show that confidence boost may come from the process of rationale construction (Heath and Gonzalez 1995), instead of other possible theories of "perceived information gain" or "inference certification."

In the context of team collaboration, a study by Stephenson and Wagner (1989) highlights that confidence boost might simply be an effect of the perception that the decision is joint, raising doubts on whether the interaction process in itself has additional value in boosting confidence in team members. This triggers the following question:

RQ1: Does team interaction have a causal role in boosting the confidence of the participants?

Social influence. Broadly speaking, many studies reported social influence in an individual's (first-order) opinion formation, especially in group settings. There are several explanations for such social influence. Individuals may conform as a result of their desire to fit in or adhere to the norm (Asch 1951; 1955; 1956). They may also conform by taking others' opinions as stronger evidence than what they know, in the hope that adopting to others' opinions would lead to better answers (Sherif 1935). As confidence can be considered as a second-order opinion—the opinion about the solution—we expect it to be susceptible to social influence as well.

Perhaps most relevant to our work is the study by Moussaïd et al. (2013). They find that in *non-collaborative* settings, an individual may adjust their opinion and confidence in their opinion upon feedback in the form of another person's opinion and corresponding confidence, especially when the feedback is of higher confidence.

Viewing the dyadic interaction process as a way for both parties to either explicitly or implicitly exchange their respective opinions and their confidence in those opinions leads naturally to the following research question:

RQ2: How does the teammate's confidence and/or knowledge impact an individual's confidence?

Related concepts. Our operationalization of confidence measures how an individual perceives one particular aspect (i.e., the quality of the outcome) of team collaboration. It is thus related, but should not be confused with, other concepts that measure individual perceptions of other aspects of team collaboration, some of which are illustrated in Figure 1.

One important related concept is *self-efficacy*, introduced by Bandura (1982), which describes an individual's self-perception of their capability to handle a prospective situation. This concept naturally generalizes to groups or teams (collective-efficacy), leading to the finer-grained concepts of *team process efficacy* and *team outcome efficacy* (Collins and Parker 2010). The former captures a team's confidence in its capability to work together, while the latter centers on the team's confidence in its capability to achieve team goals. When the evaluation about the team's capabilities span across multiple domains, it is then referred to as *team potency* (Guzzo et al. 1993).

Studies have analyzed how these concepts may interrelate and how they relate to team performance (Gully et al. 2002; Stajkovic, Lee, and Nyberg 2009). Perhaps more relevant to our work is the finding that, in the domain of sports, there is a contagion effect in which a team leader's evaluation of team-efficacy can affect the other team members' perceptions (Fransen et al. 2015; 2016).[3]

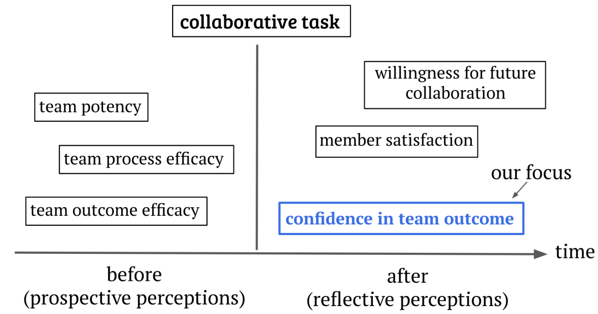

Figure 1: Related concepts that reflect an individual's perception of different aspects of team collaboration. Notably, our measure of confidence can be seen as reflective in nature, focusing on perceptions about *past* decisions.

Although these concepts are related to the concept of confidence that we aim to measure, they focus on a very different aspect of individual perceptions of team collaboration. Specifically, while they attempt to capture an individual or a team's perception in its capability in *prospective* events or processes,[4] we aim to analyze confidence about decisions that an individual or a team has already committed to. As shown in Figure 1, in contrast to the future-looking concept of efficacy, our concept of confidence in the final decision is reflective, which makes it not directly comparable to efficacy-related concepts. We choose to focus on this particular conceptualization as we can more objectively assess the quality of decisions that are already made, compared to assessing capabilities of individuals or teams for future events.

We tackle RQ1 with a randomized experiment and address RQ2 via observational studies. We further describe the details of these studies in Section 3 (Experimental Design).

3 Methods

Large-Scale Experimental Setting

To observe and experimentally intervene on goal-oriented team collaborations at scale, we design a two-player[5] online game in which individuals form ad hoc teams to solve geographical puzzles. The game is a two-player version of the popular single-player game *GeoGuessr*.[6] In our game, players navigate a first-person view of some place in the world. Their task is to identify the exact location of that place with the help of their teammates.

We choose this game design because it allows us to generate a virtually unlimited number of puzzles with so-

[3]Note that the concept of team outcome efficacy may also be referred to as *team outcome confidence* in this domain.

[4]As Bandura (1997) himself puts it, it measures "an affirmation of a *capability* level" and the "strength of that belief."

[5]Our platform allows up to five players to be matched in a team to facilitate future work. However, we focus on the dyadic games to control for the effect of group size in this study.

[6]https://geoguessr.com/.

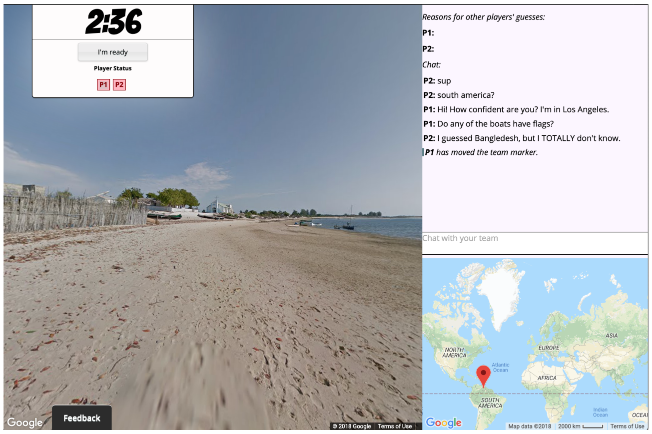

Figure 2: Game interface for the team round. During the team phase, players exchange clues they independently discovered in the solo round, after which no further navigation is possible. They communicate through a chat interface (message pane shown on the top-right corner). The full team chat for this particular game is reproduced in Table 1.

lutions that are known to us but not to the players. The game is public online at https://streetcrowd.us/start and receives hundreds of unique weekly visitors, allowing us to collect a substantial amount of data without the need for financial incentives and designing payment schemes that might otherwise affect subjective judgment (Ipeirotis 2010; Buhrmester, Kwang, and Gosling 2011). Overall, the major advantage of this particular design compared to traditional laboratory setups lies in its naturalness and scalability: the players interact naturally in the game environment and we, as experimenters, can continuously collect data at scale with minimal human efforts and little financial cost. The experiment was approved by Cornell's Internal Review Board.

For each game, two players are randomly matched into a team using a lobby system and assigned to one of the thousands of puzzles available. Each game is divided into two rounds. First, in the *solo round*, players navigate independently around a Google StreetView for a maximum of three minutes to find evidence such as street signs and vegetation that could give insight into which location they were placed in. Once the players are ready, they indicate their current (best) independent guesses of the true location by placing a marker on a world map, constituting their individual solution. Players are then prompted to indicate their confidence in their individual solutions (see Table 2 for the exact phrasing of the prompts).

Next, all team members are placed in a *team round* (Figure 2) in which they are allowed to chat with their teammates via a basic textual chat interface for up to five minutes. (The game concludes either when these five minutes are up or when both team members indicate their agreement with the final solution.) Players may exchange the clues they found in the solo round and deliberate on a final team solution. Each individual in the team is prompted for an independent confidence rating in the team solution.

Table 1: An example team chat for the puzzle shown in Figure 2. In their solo rounds, both players had guessed the wrong continent. **P1** indicated a country-level confidence ($\text{conf}_{\text{ind}} = 3$), and **P2** correctly claimed that it could be anywhere ($\text{conf}_{\text{ind}} = 1$). At the end of the discussion, the players agreed on a solution indicating the Caribbean coast of Venezuela ($\text{qual}_{\text{team}} = 1$, as the correct solution was on a different continent, in Madagascar). While **P1** maintained the same level of confidence, **P2** experienced an unjustified confidence boost as **P2** now estimated the team solution to be in the right region ($\text{conf}_{\text{team}} = 4$).

P1: sup
P1: south america?
P2: Hi! How confident are you? I'm in Los Angeles.
P2: Do any of the boats have flags?
P2: I guessed Bangledesh, but I TOTALLY don't know.
P1: no
P1: hm there is church on the beach
P2: The dude on the sand looks like not African.
He also is wearing a wedding ring.
So probably a former colony of some sort.
P2: There is trash on the beach, too.
P1: i think should be south america
P1: and maybe venesuela?
P2: Oh, and a whitish/Jewish looking guy with a baseball cap.
Definitely a tourist destination.
Probably not Bangledesh.
P2: South America sounds reasonable.
P2: Venezuela?
P1: probably caribean region
P1: yeah my guess is cenezuela
P1: venezuela
P1: i think flag is with blue color like boats :D
P2: The skiffs have large outriggers.
A place definitely historically dependent on fishing.
But that's like half the world.
P2: I'll take your guess.
P2: Venezuela for the win!
P1: ok lets try then
P1 moved the team marker [to the Caribbean coast of Venezuela]

An example team chat session is shown in Table 1. It is important to note that neither the individual solutions nor the individuals' confidence estimates are automatically revealed to the team, making team interaction—through textual messages and movements of the joint marker on the map—the only way to communicate such information.

Operationalizations and Concepts

Confidence operationalization. Our setting allows a natural operationalization of an individual's confidence as their estimation of how precise their solution (guessed location) is. This operationalization has the advantage that all participants are expected to share the same understanding of what each confidence level indicates, even in the absence of explicit explanation and clarification from the experimenters, which makes it fairer to compare between participants.

We prompt each individual to self-report a confidence level on their individual solution (henceforth, conf_{ind}) and on

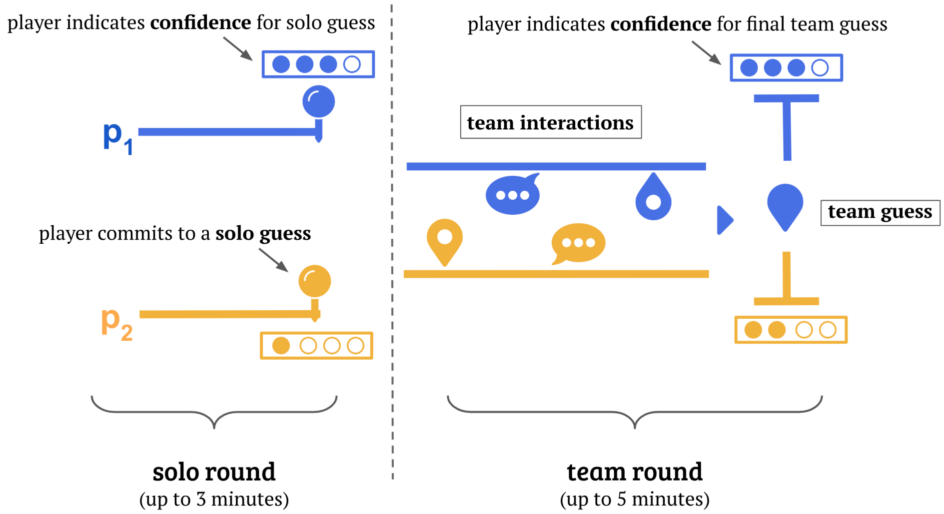

Figure 3: Illustration of the game flow. In the solo round, players p_1 and p_2 explore separately. After they commit to their solo guesses, they are prompted to rate their confidence (conf_{ind}) in this guess. In the team round, players can communicate through chat messages and can move the shared team marker. After both players agree to a final team guess, they are each prompted for an independent confidence evaluation for the team guess ($\text{conf}_{\text{team}}$).

Table 2: When self-reporting confidence, players need to click on one of four buttons (first column), corresponding to the four levels of confidence (second column). The average zoom level in players' map interfaces increases with higher confidence (third column), validating that the reported confidence corresponds to perceived levels of guess precision. This discretization is designed to correspond directly with levels of solution quality (last two columns).

reported confidence	confidence level	zoom level	solution quality	quality level
"Could be anywhere"	1	4.35	wrong continent	1
"I know the continent!"	2	4.55	correct continent, wrong country	2
"I know the country!!"	3	5.35	correct country, wrong region	3
"I know the region!!!"	4	7.16	correct region	4

the team solution (henceforth, $\text{conf}_{\text{team}}$). Figure 3 provides a detailed illustration of the game flow, which highlights when such confidence estimations are obtained, and Table 2 lists the choices users are presented with when reporting confidence, as well as the quality levels that directly correspond to these confidence levels.[7]

(Un)justified confidence boost. To measure the degree of change in terms of an individual's confidence, we subtract an individual's initial perception from their final confidence ($\text{conf}_{\text{team}}$- conf_{ind}). An individual undergoes a confidence boost if this value is strictly positive.

Our discretization of confidence is designed to correspond directly to levels of solution quality. For each guess (either by an individual or by a team), we are able to obtain the ground truth location through reverse geocoding, and thus establish an unambiguous mapping between confidence levels and solution quality levels (Table 2). This correspondence allows us to objectively assess whether a confidence boost is *justified* or not by an improvement in quality.

We will refer to the quality of an individual's solo solution as qual_{ind}, the quality of the team's solution as $\text{qual}_{\text{team}}$, and an individual's change in quality as the difference between the two values (i.e., $\text{qual}_{\text{team}} - \text{qual}_{\text{ind}}$). A confidence boost is regarded as *unjustified* if we only observe an increase in confidence but not quality: $\text{conf}_{\text{team}} > \text{conf}_{\text{ind}}$, yet $\text{qual}_{\text{team}} \leq \text{qual}_{\text{ind}}$, as is the case in the example in Table 1.

Experimental Design

Observational Study As a preliminary check, we track and compare confidence data collected from individuals before and after team collaboration to find general patterns of confidence boost. The large-scale setting further offers us the ability to analyze other observed covariates (while maintaining considerable sample size) and identify factors contributing to such changes. Specifically, we focus on analyzing the relation between an individual's eventual confidence

[7]We note that despite being seemingly categorical, our reported confidence categories are designed to fit on a quantitative scale. In fact, the quantized confidence level from 1 to 4 tracks very strongly ($R^2 = 0.975$) with the *average* log of the distance from individuals' initial guesses to their respective correct locations (a log scale is a natural choice for these types of estimation problems (Lorenz et al. 2011)). Hence, the amount of improvement from level 1 to 2 is similar to the amount of improvement from level 2 to 3 and from 3 to 4, making it reasonable to compare between numerical levels and to run standard significance tests on our confidence data.

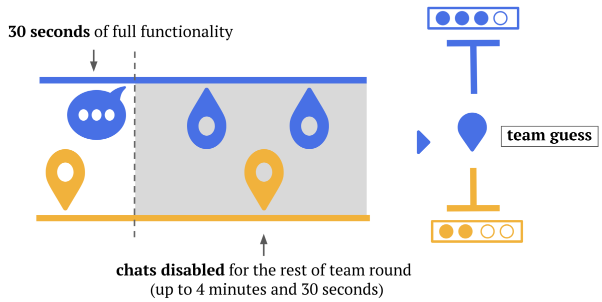

Figure 4: Under the treatment condition, textual communications are disabled after the first 30 seconds of the team round. For the rest of the team round, players can only revise (and see each other's revisions of) the team marker position, without any other types of interaction.

in the team solution and the confidence and competence of their teammate (RQ2).

We collected a total of 1,288 two-player team games involving 2,576 solo rounds, from 1,172 unique players.[8]

Validating the confidence operationalization. To validate that the self-reported confidence values are meaningful, we confirm that players who indicate higher confidence levels are indeed searching for more precise locations on the map. Intuitively, since players with higher confidences are expected to have more precise locations in mind, we would expect them to zoom in more on the map to locate their solution.[9] Indeed, we find that more confident players do have higher zoom levels (Spearman's $\rho = 0.23$, average zoom levels for each confidence level is shown in Table 2, Column 3), suggesting that they are indeed making guesses that they perceive as being more precise.

Randomized Experiment To disentangle the effects of team discussion from non-interactional effects of joint decision-making (Stephenson and Wagner 1989) (RQ1), we conduct a randomized experiment in which the treatment heavily impairs the discussion.

Under the treatment condition, teams are assigned to a "lightning" version of the game, where they only have 30 seconds to discuss. For the rest of the time, they experience the full functionality of the team phase, save for the discussion (see Figure 4 for more details). While the control group experiences the combined effects of team discussion and joint decision-making, the treatment group effectively only reaps the benefits of the latter. Since teams are randomly assigned to one of the two conditions and are identical in expectation in all other aspects, the differences between these groups are thus indicative of the effect of team discussion; this also allows us to compare the relative importance of these two factors.[10]

For a period of two months, we randomly and independently assign teams to the "lightning" treatment before the start of the team round with 50% probability, collecting a total of 392 treatment games and 388 control games. If a dyad is selected for the lightning treatment, we inform each member immediately after completion of the solo phase via a pop-up box with the message "Lightning round! You only have 30 seconds to discuss!" After both players dismiss the box, the 30-second timer begins. Players are notified about the 30-second time limit once again in the chat box, and they receive another warning when only 15 seconds remain. After the 30 seconds run out, the chat box is grayed out to indicate that chatting is disabled, and both players receive the following notification in the chat box: "The discussion period has ended. You can now adjust the map marker to finalize your guess." The players can then use the remainder of their 5-minute team phase to finalize their guess and indicate their respective confidences in this guess.[11] Individuals assigned to the control group experience no additional notifications or changes to the game interface compared to games before this experimental period.

Manipulation check. To verify our randomized assignment, we confirm that the average solo round solution quality and confidence levels are not significantly different between groups. Furthermore, we did not find the response rate to be an issue, with 74% of individuals in the lightning group finishing the game and reporting confidence at the end, compared to 60% for the control group, before filtering to ensure teams have complete information ($n_1 = 776$, $n_2 = 784$). [12]

We validate that our experimental design is effective in limiting discussions by comparing the length of discussions from treatment groups to that from control groups. We use the length of the discussion in terms of number of words uttered as our proxy, and find that lightning-round team discussions tend to be much shorter, at 4.3 words on average per individual versus 17.7 words on average for regular games, after filtering out games with incomplete information ($n_1 = 414$, $n_2 = 520$, Cohen's $d = -1.01$, mixed effects model $p < 0.001$), confirming the treatment effect.[13]

[8]We exclude games that involved any of the authors, and we only consider games for which we received complete confidence feedback and could obtain complete information about solution quality. The average individual played 2.5 games, and the most games played by a single player was 63.

[9]With zoom level 1, the map shows roughly the entire world (as in Figure 2); zoom level 5 corresponds continent-level detail, and one might expect to observe city-level detail at zoom level 10.

[10]We choose this operationalization over disallowing chats completely (which would better isolate the effect of joint decision-making) to keep our setting natural: players may feel that eliminating team discussion removes a crucial mechanic, whereas lightning rounds are a common variation of other popular games.

[11]We preserve the 5-minute time limit used in the regular game mode in order to avoid introducing time stress into the players, and to make the formats of the two team rounds as similar as possible.

[12]Since a difference in response rate could raise problems with the analysis, we only consider dyads in which both individuals completed the game and post-game confidence survey.

[13]To check whether such differences are significant while accounting for dependencies in our data (which make t-test inappropriate), we consider a linear mixed effects model, with a random intercept conditioned on each of the two team members, a random intercept for the team game, as well as a random intercept for the puzzle, and report the p-values for the coefficient of the variable of interest (De Vaan, Schreuder, and Baayen 2007;

4 Results

By prompting team members to reflect on their respective confidence levels before and after team interaction, we are able to track how individual confidence changes through the course of the interaction. This allows us to revisit the phenomenon of confidence boost at the individual level and at a larger scale than in previous work, which has mostly surveyed confidence of whole teams rather than individuals (Stephenson and Wagner 1989; Schuldt et al. 2017). In this section, we establish that the overall confidence boost previously observed in teams is also expressed at the individual level. Then, we zoom into different subpopulations to understand which groups of individuals are more prone to confidence boost.

Previous work has shown that group collaboration generally makes teams as a whole have higher collective confidence (Allwood and Björhag 1990; Sniezek 1992; Schuldt et al. 2017). Echoing these findings on the individual level, we find that post-interaction individual confidences are on average significantly higher than values reported prior to interaction (Cohen's $d = 0.30$, mixed effects model $p < 0.01$). Furthermore, out of the individuals that do report a change in confidence (38%), a large majority (74%) report an increase.

Considering that an individual's degree of confidence change is naturally constrained by the range of confidence values they are allowed to report,[14] it is not surprising that subgroups with different initial confidences exhibit different patterns of change (Figure 5), highlighting the need to control for one's initial confidence level in further analysis.

Recognizing that accurately correcting for ceiling effects is hard, we choose to focus on the subgroup of individuals who start with the lowest possible confidence level as the main subjects of our study. By considering a subgroup with fixed initial confidence, we ensure that all subjects in our analysis have the same amount of room for improvement, thus removing any ceiling effects. Additionally, this subgroup is particularly interesting as it accounts for more cases of confidence boost than all the other confidence levels combined (53%).

It is also important to note that the confidence boost we observe may not be justified: only 30.4% of individuals who report an increase in confidence actually have team solutions that are of higher quality (compared to their pre-interaction individual ones). In addition, out of the 230 individuals who end up with *worse* team solutions than their initial solutions, there are still 18.6% of them who report higher confidence,[15] suggesting that there are quality-independent factors that contribute to this increase.

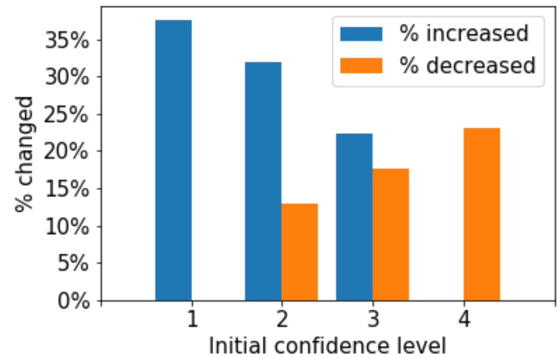

Figure 5: Individuals who start at the lowest confidence level are more likely to experience confidence boost, potentially due to larger room for improvement. We also note that the asymmetrical nature of the two types of confidence changes (compare blue bars with orange bars) further suggests that the population on average shows an increase in confidence after team interaction.

RQ1: Effect of Team Interaction

Noting that confidence boost happens in online teamwork even when it is unjustified by a corresponding increase in solution quality, it is natural to wonder why and when it occurs. Compared to the individual problem-solving process, collaborative decision-making bears two additional components: the process of interacting with other team members to exchange information, and the perception that one is participating in a joint decision, either of which might be responsible for the confidence boost observed. Stephenson and Wagner (1989) attempted to compare the relative effect of these two components in memory recall tasks and concluded that the very perception of making a joint decision might create "an assumption of confidence" in the group.

However, with their experimental design, the separation between interaction and making a joint decision is not complete: the dyads are still able to discuss while in the condition that tries to capture only the act of joint decision-making. Consequently, their design gives no conclusive result on the role of interaction in confidence formation. Our full control over the online interface, on the other hand, provides us the ability to better separate the two components of collaboration, putting us in a better position to evaluate if interaction in itself may have an impact on such confidence boost.

We first note from observational data that the *extent* of interaction is correlated with confidence boost. While our game is designed to encourages conversations, discussions can have varying lengths. At the very extreme, there are teams which complete the task without exchanging any messages (henceforth *silent teams*). If confidence boost is truly independent of discussion, we should expect that the length of discussion should not have any correlation with the extent of confidence boost. Instead, we observe significantly greater confidence boost in teams that have exchanged messages compared to silent teams, with an average increase of

Baayen 2008). In this particular case, the variable of interest is the binary variable indicating whether the team is under the treatment condition or not. Throughout, all mixed effects model p-values are obtained this way.

[14]For instance, we would not be able to observe an increase in confidence from an individual who is already at the highest confidence level. Such ceiling issues are shared by all prior operationalizations of confidence.

[15]25.6% of individuals who stay at the same quality level report higher confidence, while 44.2% of individuals with improved solutions report higher confidence.

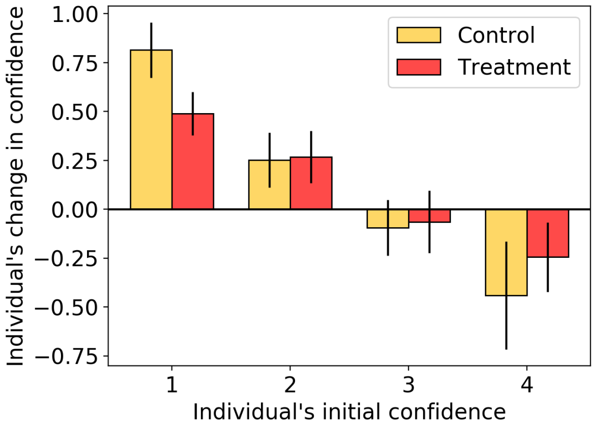

Figure 6: Comparing treatment and control games, we see that individuals with the lowest level of confidence account for most of the overall reduction in confidence boost from the lightning manipulation (Cohen's $d = -0.46$, mixed effects model $p < 0.01$). Error bars represent 95% confidence intervals for the sample mean.

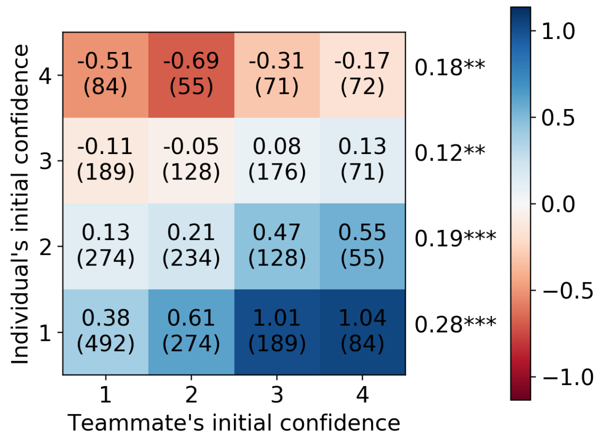

Figure 7: Heat values indicate average change in confidence for individuals in dyads with the given initial confidence composition. We observe that, for the low-confidence individuals ($\text{conf}_{\text{ind}} = 1$), the more confident one's teammate is, the greater one's own confidence tends to become. Trends for high-confidence individuals are less pronounced. Correlation between one's change in confidence and the teammate's initial confidence is computed for each row and shown to the right (*p<0.05; **p<0.01; ***p<0.001).

0.33 versus 0.13 respectively ($n_1 = 2{,}148$, $n_2 = 846$, Cohen's $d = 0.24$, mixed effects model $p < 0.001$).[16]

These observations can not, however, account for self-selection bias. To exclude this alternative explanation, we conduct a randomized experiment in which we manipulate the degree of interaction (see Section 3, Experimental Design, for further details).

Experiment result. We compare 520 individuals from lightning games with 414 individuals in two-player regular games from the same time frame. We find that 26% of individuals improve in confidence in lightning games, which is significantly lower than the 33% of individuals who improve in regular games (mixed effects model $p < 0.05$, Cohen's $d = -0.15$), confirming that confidence boost comes as a direct consequence of team discussion, rather than simply being a consequence of making a joint team decision.

Notably, most of this difference is accounted for by the individuals who have the lowest level of confidence in their individual solutions ($\text{conf}_{\text{ind}} = 1$), showing that low-confidence individuals are most susceptible to confidence boost through discussion (Figure 6).

RQ2: Effect of Teammate Confidence

Having seen that low-confidence individuals can undergo an (unjustified) boost in confidence as a direct consequence of interacting with their teams, we now seek to explore social factors explaining this increase. Inspired by previous work in non-collaborative settings (Moussaïd et al. 2013), we focus on the effect of the teammate's confidence (RQ2). We hypothesize that confidence is subject to social influence: low-confidence individuals may be influenced by, or even adopt, the confidence of their more confident teammates.

We first explore this hypothesis by comparing the confidence change of individuals who are paired with teammates of varying confidence levels. We find that the confidence boost of low-confidence individuals is more pronounced when they interact with more confident teammates (Figure 7, bottom row): the more confident one's teammate is, the larger the increase in confidence (Spearman correlation $\rho = 0.26$, $p < 0.001$). This teammate effect is less pronounced for more confident individuals (Figure 7, top three rows),[17] suggesting that low-confidence individuals are par-

[16] This overall difference in confidence boost between the two groups persists regardless of whether we consider the subset of individuals in which the solution quality improves ($n_1 = 439$, $n_2 = 125$, Cohen's $d = 0.25$, mixed effects model $p < 0.1$), remains the same ($n_1 = 1{,}516$, $n_2 = 639$, Cohen's $d = 0.19$, mixed effects model $p < 0.01$), or decreases ($n_1 = 193$, $n_2 = 82$, Cohen's $d = 0.40$, mixed effects model $p < 0.05$), suggesting that the confidence boost afforded by discussion is not solely a consequence of an improvement in the quality of the produced solution.

[17] While it might be surprising that the aggregate effect for high-confidence individuals is negative, this phenomenon can be explained by ceiling effects: e.g., for the individuals who initially are at confidence level 4, the only possible observable change is a decrease in confidence. We also note a decreasing trend in the top row as the teammate's initial confidence decreases: while the behavior of high-confidence individuals is beyond our main focus, a potential explanation of this observation is that encountering a lower-confidence individual (who is likely to be more cautious) might make high-confidence individuals realize that the problem is more complex than they initially thought. This observed trend is interesting and merits future exploration as one might as well expect the reverse trend: teammates with lower confidence may have less information to challenge the opinions of high-confidence individuals, thus making the high-confidence individuals believe even more strongly in their solutions, resulting in a boost in confidence.

ticularly susceptible to teammate influence. Importantly, the same pattern holds (with equivalent levels of statistical significance) even if we only consider individuals with unjustified confidence boost.

To quantify this effect, while accounting for other correlates, we fit a linear mixed effects model with the individual's final confidence as the dependent variable (Table 3). We account for each individual, their team, and the particular puzzle as random effects.

As independent variables, we consider the teammate's confidence, the quality of the individual's solo round solution, the quality of their teammate's solo round solution (approximating their respective knowledge on the particular puzzle), the quality of the team's final solution, and the number of words exchanged by each individual along with the number of marker movements made during the discussion (approximating the level of interaction). Finally, in order to disentangle the effect of the teammate's confidence from the confounding secondary effects of both team members' knowledge, we include two-way interaction terms (in italics) between the teammate's confidence and each variable representing solution quality (of each team member's solo solution, and the team's final decision).

We find that for low-confidence individuals ($\text{conf}_{\text{ind}} = 1$), the teammate's confidence is significant in predicting post-interaction confidence (Table 3, column 1). The fact that the teammate's confidence is more predictive than both the actual quality of the team's solution (the very object of the confidence estimation) and the teammate's solution quality is perhaps surprising, discarding the reasonable alternative hypothesis that confidence influence is purely based on team members' knowledge.[18] Instead, we see that confidence boost can occur in systematic, unjustified ways in teamwork, suggesting the presence of non-knowledge-related factors behind confidence boost.

The observation that the teammate's competence—approximated as the quality of their individual solution—appears to be a far less important factor than their confidence is further supported in a predictive setting. In the task of detecting an individual's eventual confidence in the team solution, a logistic regression model that considers only the teammate's initial confidence and the discussion's contents outperforms a model that only has access to information about the quality of the solution (AUC = 0.59 and 0.55, respectively). This phenomenon could be attributed to the difficulty of distinguishing actual knowledge from perceived or expressed competence, especially for the low-competence individuals (Kruger and Dunning 1999; Fu, Lee, and Danescu-Niculescu-Mizil 2017).

To explore this further, we compare the subgroup of low-confidence individuals that show little knowledge in the solo round ($\text{qual}_{\text{ind}} = 1$, Table 3, Column 2) with those that show more knowledge ($\text{qual}_{\text{ind}} = 3$, Table 3, Column 3).[19] We note that the latter group is indeed less susceptible to the influence of the teammate's confidence, presumably demonstrating better discernment of actual knowledge as they interact with their teammate (in spite of their expressed lack of confidence). This observation opens an exciting avenue for future work on identifying under-confident individuals based on their interactional behavior.

[18] In the sense that if one's teammate is more confident, they are likely to be more correct and lead the team towards a more-correct solution: a solution more worthy of high confidence.

[19] We focus on this subgroup instead of individuals at the highest level of competence ($\text{qual}_{\text{ind}} = 4$) for the larger data size.

Table 3: In a linear mixed effects model, the teammate's pre-interaction confidence (rather than their knowledge) is predictive of an individual's confidence in the eventual team solution (Column 1). This factor is even more predictive than the actual quality of the solution. This association is dampened if the individual demonstrates greater knowledge (compare Column 2 with Column 3). Parentheses show standard errors for the respective regression coefficients.

	Dependent variable:		
	post-interaction confidence		
	(1)	(2)	(3)
	conf = 1	conf = 1, qual = 1	conf = 1, qual = 3
individual solution quality	0.069 (0.077)		
teammate confidence	**0.225**** (0.079)	**0.181*** (0.085)	−0.183 (0.265)
teammate solution quality	0.098 (0.088)	0.038 (0.142)	−0.116 (0.196)
team solution quality	0.098 (0.106)	0.113 (0.142)	**0.382** (0.251)
individual chat length	0.004 (0.002)	0.004 (0.003)	0.017* (0.008)
teammate chat length	0.003 (0.002)	0.001 (0.003)	0.005 (0.007)
total marker moves	−0.022 (0.012)	−0.032 (0.018)	0.016 (0.049)
teammate confidence : indiv. soln. quality	−0.049 (0.033)		
teammate confidence : teammate soln. quality	−0.048 (0.047)	−0.063 (0.076)	0.108 (0.116)
teammate confidence : team soln. quality	0.067 (0.052)	0.086 (0.075)	−0.010 (0.134)
Constant	−0.252 (0.181)	−0.107 (0.193)	−0.388 (0.654)
Observations	1,045	473	142
Log Likelihood	-1,301.706	-569.983	-207.395
Akaike Inf. Crit.	2,635.411	1,167.966	442.790
Bayesian Inf. Crit.	2,714.639	1,226.194	484.172
Note:			*p<0.05; **p<0.01; ***p<0.001

5 Discussion

In this work, we design a large-scale online experimental platform to study how teamwork affects the confidence of individual participants. Our full control over the platform, as well as its scale, allows us to track confidence change at a finer-grained level than in prior work (which mostly took a team-level perspective) and to disentangle between mediating factors.

We find that team interaction directly results in a boost in confidence for low-confidence individuals, and that this boost is explainable as an adoption of confidence—but not necessarily of actual knowledge—from the teammate with whom they interact.

Design Implications

Our work has a number of implications for designing interfaces to better facilitate and support online teamwork. We have demonstrated the feasibility of quantifying, tracking, and even predicting confidence boost for individuals working in teams. Such information could be used in the design of team interaction systems that help modulate team members' confidence to fit the needs of differing collaborative settings:

In risk-heavy result-oriented settings, our observation suggests that individuals without sufficient knowledge may be swayed by more confident teammates to become more confident, even if the proposed solutions do not result in a corresponding increase in quality. To avoid such unjustified confidence buildup, it may be beneficial to design and provide affordances that help individuals better calibrate solution quality. This may be achieved by presenting average performance statistics on similar problems to help individuals ground their confidence estimations, and/or by presenting the performance history of collaborators so that individuals may recognize consistently over-confident partners. Additionally, when individuals' (private) opinions differ, the platform could more directly signal such differences, so that it is more likely for different solutions to be presented and discussed, avoiding the situation that opinions from low-confidence individuals are never brought up, which may lead to groupthink (Janis 1972).

In contrast, in settings in which teammate agreement is more important, e.g., when the goal is team-bonding, we may strive to boost team members' confidence irrespective of its relation to quality to ensure that team members leave with positive feelings. In this scenario, our findings suggest a number of intervention strategies. First, before teams are formed, pre-collaboration confidence levels of individuals may inform team-matching procedures that optimize for the greatest potential collective gain in confidence. During the interaction, collaborative platforms could consider a more direct sharing of teammate confidences to facilitate the process of spreading optimism from higher-confidence individuals to lower-confidence individuals. Furthermore, further work may develop techniques for more accurate and robust confidence detection, enabling platforms to incorporate automatic confidence tracking relying only on activity traces. Such platforms may signal to team managers the cases in which individuals lack confidence in team decisions, so that managers could take appropriate actions promptly.

Limitations and Future Work

While our framework and results hint at potential implications for team management, a transfer into real-world applications would require substantial future work. Next, we highlight a few aspects of the current study that limit the interpretation of the results and demand further investigation.

Affordances of the interface. In our setting, participants communicate primarily through text messages.[20] While this design choice allows us to directly analyze discussion contents automatically, it does raise the questions of how people would react and behave if the interface offered a richer set of affordances. Future work may thus look into online collaborative platforms with audio or video capabilities to search for different types of signals and analyze the effect of increased social presence on confidence expressions in teamwork.

Confidence operationalizations. We make the explicit choice of operationalizing confidence as one's perception of solution quality, since it provides automatic calibration across participants and thus an unambiguous interpretation. This type of confidence, however, might be communicated and perceived differently than confidence operationalized as one's certainty in the solution. In particular, we note that while exchanging solutions is likely to reveal (at least in part) one's evaluation of their qualities,[21] it does not necessarily convey one's certainty in the solution. In fact, Stephenson and Wagner (1989) has noticed that dyads in their study center their discussions primarily on facts and tend to neglect to exchange confidences in their opinions, which might in part explain why the effect of interaction seems weaker in their reports.

Task format. Our problem-solving task follows a specific format which, while common to many online collaboration settings, is not necessarily applicable to all of them. For example, it requires participants to make one single decision. This is in contrast to some previous studies in which participants complete a *set* of questions and provide confidence ratings for each question (Allwood and Granhag 1996; Schuldt et al. 2017). Under those settings, individual and dyadic confidences are generally computed as the averages across the entire set. Since our framework operates at an itemized level, it is not immediately transferable to aggregate confidence measures.

Analysis of the interaction process. In this work, we have focused on only the easily quantifiable aspects of the team discussions, using the length of the chats as crude proxies. Future work could include a more in-depth semantic analysis by utilizing the text exchanges we record. For instance, looking for signals in group affective tones (George 1990; Collins et al. 2013) may help better predict confidence trajectories of individuals.

[20] Although marker movements, which are mutually visible, may also be seen as a form of communication.

[21] For instance, in our setting, "I think it is in France" indicates both an individual's guess as well as country-level confidence.

More generally, while care should be taken when generalizing our observations to other settings, we bring attention to an aspect of online collaboration that is often overlooked. Beyond performance, collaboration platforms should consider a more diverse set of individual-level objectives. Confidence is only one example of many subjective dimensions of teamwork, along with, for example, willingness to collaborate in the future (Whiting et al. 2019), satisfaction with one's contribution or perception of fairness in the decision-making process. Even though team performance might still remain the default objective in most cases, there are scenarios in which it may not be the only goal, or not even the most important one. In fact, in many real-world scenarios, the actual quality of the decision might remain unknown to the participants (e.g., in hiring decisions). A better understanding of more subjective aspects of teamwork is thus crucial to the design of online platforms that can be flexible enough to accommodate diverse collaborative needs.

Acknowledgments

We would like to thank Jonathan P. Chang, Lillian Lee, Justine Zhang, as well as the reviewers for their helpful comments. We are also grateful to all StreetCrowd players and to David Garay, Maheer Iqbal, Jinjing Liang, Xin Lin, Julian Londono, Vlad Niculae, Neil Parker, Avery Smith, and Alicia Zhou for participating in the game's development. This research was supported in part by an NSF CAREER award IIS-1750615 and by an NSF Grant IIS-1910147.

References

Adams, J. K., and Adams, P. A. 1961. Realism of confidence judgments. *Psychological Review*.

Allwood, C. M., and Björhag, C.-G. 1990. Are two judges better than one? On the realism in confidence judgements by pairs and individuals. In *Advances in Psychology*.

Allwood, C. M., and Granhag, P. A. 1996. The effects of arguments on realism in confidence judgements. *Acta Psychologica*.

Asch, S. E. 1951. Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership and Men: Research in Human relations*.

Asch, S. E. 1955. Opinions and social pressure. *Scientific American*.

Asch, S. E. 1956. Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.

Bahrami, B.; Olsen, K.; Latham, P. E.; Roepstorff, A.; Rees, G.; and Frith, C. D. 2010. Optimally interacting minds. *Science*.

Bandura, A. 1982. Self-efficacy mechanism in human agency. *American Psychologist*.

Bandura, A. 1997. *Self-Efficacy: The Exercise of Control*. Macmillan.

Boje, D. M., and Murnighan, J. K. 1982. Group confidence pressures in iterative decisions. *Management Science*.

Boud, D., and Falchikov, N. 1989. Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher education*.

Buehler, R.; Griffin, D.; and Ross, M. 1994. Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*.

Buhrmester, M.; Kwang, T.; and Gosling, S. D. 2011. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*.

Coetzee, D.; Lim, S.; Fox, A.; Hartmann, B.; and Hearst, M. A. 2015. Structuring interactions for large-scale synchronous peer learning. In *Proceedings of CSCW*.

Collins, C. G., and Parker, S. K. 2010. Team capability beliefs over time: Distinguishing between team potency, team outcome efficacy, and team process efficacy. *Journal of Occupational and Organizational Psychology*.

Collins, A. L.; Lawrence, S. A.; Troth, A. C.; and Jordan, P. J. 2013. Group affective tone: A review and future research directions. *Journal of Organizational Behavior*.

De Vaan, L.; Schreuder, R.; and Baayen, R. H. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*.

Fransen, K.; Haslam, S. A.; Steffens, N. K.; Vanbeselaere, N.; De Cuyper, B.; and Boen, F. 2015. Believing in us: Exploring leaders capacity to enhance team confidence and performance by building a sense of shared social identity. *Journal of Experimental Psychology: Applied*.

Fransen, K.; Steffens, N. K.; Haslam, S. A.; Vanbeselaere, N.; Vande Broek, G.; and Boen, F. 2016. We will be champions: Leaders' confidence in us inspires team members' team confidence and performance. *Scandinavian Journal of Medicine & Science in Sports*.

Fu, L.; Lee, L.; and Danescu-Niculescu-Mizil, C. 2017. When confidence and competence collide: Effects on online decision-making. In *Proceedings of WWW*.

George, J. M. 1990. Personality, affect, and behavior in groups. *Journal of Applied Psychology*.

Gully, S. M.; Incalcaterra, K. A.; Joshi, A.; and Beaubien, J. M. 2002. A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology*.

Guzzo, R. A.; Yost, P. R.; Campbell, R. J.; and Shea, G. P. 1993. Potency in groups: Articulating a construct. *British Journal of Social Psychology*.

Heath, C., and Gonzalez, R. 1995. Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes*.

Ipeirotis, P. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making*.

Janis, I. L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes.* Houghton Mifflin.

Koriat, A.; Lichtenstein, S.; and Fischhoff, B. 1980. Reasons for confidence. *Journal of Experimental Psychology: Human learning and memory*.

Kruger, J., and Dunning, D. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*.

Laughlin, P. R., and Adamopoulos, J. 1980. Social combination processes and individual learning for six-person cooperative groups on an intellective task. *Journal of Personality and Social Psychology*.

Laughlin, P. R.; Hatch, E. C.; Silver, J. S.; and Boh, L. 2006. Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and social Psychology*.

Lorenz, J.; Rauhut, H.; Schweitzer, F.; and Helbing, D. 2011. How social influence can undermine the wisdom of crowd effect. *In Proceedings of the National Academy of Sciences*.

Meudell, P. R.; Hitch, G. J.; and Kirby, P. 1992. Are two heads better than one? experimental investigations of the social facilitation of memory. *Applied Cognitive Psychology*.

Moore, D. A., and Healy, P. J. 2008. The trouble with overconfidence. *Psychological Review*.

Moussaïd, M.; Kämmer, J. E.; Analytis, P. P.; and Neth, H. 2013. Social Influence and the Collective Dynamics of Opinion Formation. *PLOS One*.

Niculae, V., and Danescu-Niculescu-Mizil, C. 2016. Conversational markers of constructive discussions. In *Proceedings of NAACL*.

Salehi, N.; McCabe, A.; Valentine, M.; and Bernstein, M. 2017. Huddler: Convening stable and familiar crowd teams despite unpredictable availability. In *Proceedings of CSCW*.

Schuldt, J. P.; Chabris, C. F.; Woolley, A. W.; and Hackman, J. R. 2017. Confidence in dyadic decision making: The role of individual differences. *Journal of Behavioral Decision Making*.

Sherif, M. 1935. A study of some social factors in perception. *Archives of Psychology (Columbia University)*.

Sniezek, J. A., and Henry, R. A. 1989. Accuracy and confidence in group judgment. *Organizational behavior and human decision processes*.

Sniezek, J. A. 1992. Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*.

Stajkovic, A. D.; Lee, D.; and Nyberg, A. J. 2009. Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and test of a mediation model. *Journal of Applied Psychology*.

Stephenson, G. M., and Wagner, W. 1989. Origins of the misplaced confidence effect in collaborative recall. *Applied Cognitive Psychology*.

Stephenson, G. M.; Abrams, D.; Wagner, W.; and Wade, G. 1986. Partners in recall: Collaborative order in the recall of a police interrogation. *British Journal of Social Psychology*.

Stoner, J. A. F. 1961. *A Comparison of Individual and Group Decisions Involving Risk*. Ph.D. Dissertation, Massachusetts Institute of Technology.

Tindale, R. S. 1989. Group vs individual information processing: The effects of outcome feedback on decision making. *Organizational Behavior and Human Decision Processes*.

Turner, J. C.; Hogg, M. A.; Oakes, P. J.; Reicher, S. D.; and Wetherell, M. S. 1987. *Rediscovering the Social Group: A Self-Categorization Theory.* Basil Blackwell.

Turner, J. C.; Wetherell, M. S.; and Hogg, M. A. 1989. Referent informational influence and group polarization. *British Journal of Social Psychology*.

Tversky, A., and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science*.

Whiting, M. E.; Blaising, A.; Barreau, C.; Fiuza, L.; Marda, N.; Valentine, M.; and Bernstein, M. S. 2019. Did It Have To End This Way? Understanding The Consistency of Team Fracture. In *Proceedings of CSCW*.

Williams, W. M., and Sternberg, R. J. 1988. Group intelligence: Why some groups are better than others. *Intelligence*.

Woolley, A. W.; Chabris, C. F.; Pentland, A.; Hashmi, N.; and Malone, T. W. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science*.