

Thin Slices of Workgroups

Patricia Satterstrom, New York University

Jeffrey T. Polzer, Harvard Business School

Lisa B. Kwan, Harvard University

Oliver P. Hauser, University of Exeter Business School

Wannawiruch Wiruchnipawan, Charoen Pokphand Group

Marina Burke, SAS

Corresponding author:

Patricia Satterstrom

New York University

295 Lafayette Street, 2nd Floor

New York, NY 10012

Contact email: patricia.satterstrom@nyu.edu

Contact phone: 617-733-6705

Thin Slices of Workgroups

Abstract

In this paper, we explore whether perceivers can accurately assess the effectiveness of groups, the group properties perceivers use in their judgement, and the contextual and individual differences that allow some perceivers to be more accurate. Across seven studies, we present consistent evidence that perceivers can judge workgroup effectiveness in videos of different lengths—60, 30, and 10 seconds—and in 10-second silent videos and 10-second audio clips. We find that perceptions of collective properties of groups, including cohesion, affective trust, and cognitive trust partially mediate perceivers' ability to accurately judge groups. Furthermore, increased attentional focus improves perceivers' ability to judge group effectiveness. Finally, we find that perceivers with higher levels of social sensitivity are more accurate at judging group effectiveness. We discuss the implications of these findings for the groups literature and social perception literature.

Keywords: Group perception, group effectiveness, thin slices, social sensitivity, attentional focus

Introduction

Are we able to detect when groups are doing well or when they spell trouble? Consider, for example, a professor who oversees a section of 60 students divided into 15 groups. After checking in briefly with each group, she asks three groups to meet during office hours because she senses that they are not on the right track. Can she pick out the “problem teams” accurately? Or, take a medical director at a mid-size clinic who manages eight multidisciplinary primary care groups, and his only opportunity to see each group together is during their five to ten minute morning “huddle” (Bodenheimer & Laing, 2007; Shunk et al., 2014). In recent years, accelerators have changed the entrepreneurship landscape as coaches observe startup teams for short periods and immediately offer feedback, while investors make expensive funding decisions after observing short team pitches (Clingingsmith & Shane, 2017; Gust, Global Accelerator Report, 2016). These examples all share an increasingly common phenomenon—evaluating groups after a brief exposure to their group interaction—and beg the question of whether such perceivers are any good at accurately judging group effectiveness.

A large body of research has helped explain why some workgroups are more effective than others (Mathieu, Maynard, Rapp, & Gilson, 2008; Salas, Cooke, & Rosen, 2008; Stewart, 2006). This research has relied on managers’ assessments of their workgroups when objective performance criteria were not available. Yet, group researchers have not tested the assumption that managers’ (or anyone’s) subjective perceptions of workgroup functioning are accurate when evaluated against objective measures of group performance. Even research on team coaches, whose job is to observe groups, is silent about the accuracy of coaches’ judgments of groups. Coach accuracy is important since studies show a strong relationship between coaches’ perceptions of group effectiveness and their choice of interventions (for a review see Burke et

al., 2006). Behavioral models of coaching, for their part, assume coaches must have extensive knowledge of the task so that they can monitor the team's task performance along with the social-emotional interactions of group members (Hackman & Wageman, 2005; Fleishman et al., 1991). From any vantage point, effectively monitoring workgroup processes requires observers to accurately discern how the group is doing during an action phase (Marks, Mathieu, & Zaccaro, 2001), often before the group's output can be measured or is even fully produced.

Theorizing about how individual perceivers judge group interaction requires an understanding of both individual social perception and collective dimensions of group functioning. Our goal is to take some early steps towards piecing together foundational elements of this multi-layered process. We begin by asking, fundamentally, whether perceivers are able to accurately judge the effectiveness of a group. Our results consistently suggest a resounding "yes" to this question. We then turn to exploring the psychological and social facets that pertain to the "how" and "when" of these accurate judgements, and offer an initial framework to make sense of these facets based on (1) the amount of available stimulus information, (2) the features of groups being judged, and (3) aspects of the perceivers' external (e.g. focus of attention) or internal (e.g. individual differences) state that may aid their ability to make such judgements.

First, based on prior literature on thin slices of individual behavior, we hypothesize that even small snippets of information are sufficient for accurate judgements. In addition, we explore to what extent the available information aids judgment when it is auditory, visual or both; we find that combining audio and video leads to additive effects on accuracy. These results are consistent with research showing that even very short exposure to a single perceptual channel is sufficient to make accurate judgements, but also contributes the new insight that combined audio and visual information can lead to more accurate perceptions.

Second, our theoretical perspective also looked at perceived group dynamics. Arguably, one might expect that groups who do well on a task will be perceived differently along a number of internal dynamics that matter for group decision-making, such as cohesion, trust, and conflict. We find some evidence to support these group internal differences: perceived trust and cohesion partially mediated perceptions of effectiveness, but perceived conflict did not. These findings suggest that perceivers' judgments of distinct dimensions of groups are related in ways that align with patterns found in the literature on group functioning (Beal, Cohen, Burke, & McLendon, 2003; De Jong, Dirks, & Gillespie, 2016).

Finally, we consider the decision-maker—that is, the perceiver whose task it is to judge the effectiveness of a group. Some aspects of the perceivers' effectiveness may be external to them, such as what they are drawn to focus on. Drawing on research that shows that perceivers are more accurate when they focus on the collective versus the individuals (Haberman & Whitney, 2009; Sanchez-Burks et al., 2015), we manipulate the focus of perceivers in our studies to a similar extent; yet, we find mixed results when perceivers focus on the collective and greater accuracy when perceivers focus on individuals compared to a control condition. Other aspects pertaining to the perceiver's effectiveness may be internal, such as individual differences. Based on research suggesting that people differ in their level of social sensitivity—the ability to accurately judge the affective state of others (Bernieri, 2001; Pickett, Gardner, & Knowles, 2004; Woolley, Aggarwal, & Malone, 2015)—we hypothesize that perceivers who do well on detecting emotions in others are also better at discerning the effectiveness of group interaction. Our results support this thesis. Meanwhile, an alternate account based on individual differences is not supported by the data: based on the idea that even very thin slices (i.e. little available stimulus information) about groups can lead to accurate judgment, we speculate that perceivers who tend

to trust their intuitive “gut instinct” may be more accurate than those who tend to reflect on their decisions, but this hypothesis was rejected (and if anything, the opposite appears to be true).

In sum, our investigation provides evidence about some important elements of accurately perceiving group decision-making, and a framework for organizing this evidence. Perceivers *do* make accurate judgements of group effectiveness much of the time, they perceive group features differently between effective and ineffective groups, they use and combine multiple perceptual channels, and they exhibit systematic individual differences in their ability to make these judgments. This framework highlights the interplay of individuals, as perceivers, and groups, as the target of their perceptions. As such, we draw from, and contribute to, both the social perception and group performance literatures, to which we now turn.

Shifting from Social Perception of Individuals to Workgroups

Research on social perception has demonstrated that people can judge individuals across a broad range of contexts, including teachers, physicians, occupational therapists, and car salesmen, in ways that systematically correspond to relevant criterion variables (Ambady & Rosenthal, 1993; Ambady et al., 2002; Ambady, Krabbenhoft & Hogan, 2006; Hecht & LaFrance, 1995; Tickle-Degnen, 1998). These initial judgments of individual targets are important because they set the trajectory for the perceiver’s subsequent behavior, interaction, and social outcomes (Harris & Garris, 2008; Word, Zanna, & Cooper, 1974). Our work responds to Phillips, Weisbuch, and Ambady’s (2014) call to shift research attention from “person perception” to “people perception,” since as organizational scholars we need to better understand how people perceive groups; this is particularly important since much of the perceptual apparatus we use to form social perceptions emerged from interactions with small groups (Neuberg, Kenrick, & Schaller, 2010). In this paper, we are particularly interested in perceptions based on

brief observations of groups who are engaged in a work task, much like the ones medical directors, investors, and others often rely on to gauge workgroup effectiveness.

When individuals are engaged in social interaction, they are especially likely to display revealing cues (Goffman, 1959). Tapping into these displays, perceivers in some thin slice studies judge individual targets after observing them interact with another person. For example, Kraus and Keltner (2009) found that perceivers, after watching short episodes of dyadic interaction, were able to accurately judge the socioeconomic status of the individual targets in the dyad. In another study, perceivers watched thin slices of interaction between speed daters and accurately evaluated individuals' romantic interest in their partner (Place, Todd, Penke, & Asendorpf, 2009). In these studies, individual targets remained the perceptual focus, though they were observed as they engaged in real, spontaneous interaction with another person.

Moving beyond perceptions of individual targets are studies of dyads in which perceivers judge the quality of the interaction or relationship between a pair of people, rather than features of the individual targets themselves. Such dyadic research is a step closer to small group interaction. Gottman and colleagues, for example, conducted an extensive program of research on married couples in which they were able to train perceivers to observe short episodes of interaction between spouses, code various affective cues, and use these codes to predict marital outcomes (Gottman & Notarius, 2000; Gottman, Markman, & Notarius, 1977; Gottman & Levenson, 1992). In one study, trained perceivers coded spouses' affective displays as they discussed issues together that had been ongoing sources of disagreement. Affective displays during the first three minutes of such discussions predicted divorce over the next six years, with couples that displayed more negative affect and less positive affect more likely to get divorced (Carrere & Gottman, 1999).

Research on thin slices of dyads has shown that people convey a great deal of information through nonverbal displays (DePaulo, 1992), exhibiting mimicry of vocalizations (Curhan & Pentland, 2007; Romero, Swaab, Uzzi & Galinsky, 2015) and mirroring of nonverbal behaviors (Chartrand & Bargh, 1999). Bernieri, Gillis, Davis, and Grahe (1996) studied perceivers' judgments of dyadic rapport after viewing brief (30-50 second) videos of discussions between two previously unacquainted people. They operationalized rapport as the extent to which dyad partners liked each other and enjoyed the interaction. Perceivers' ratings were significantly associated with dyad partners' self-reported levels of rapport. Because rapport is an inherent property of the dyad, this research provides evidence that perceivers can accurately judge social collectives beyond their individual components. This is consistent with the idea that human perception is adapted to judging the qualities of social interaction, including among members of a small group.

Dyadic interaction is unique in that two participants talk and listen only to each other, making it relatively straightforward to observe interpersonal synchrony or complementarity (Tiedens & Fragale, 2003). Judging a group that is working towards a common goal, however, requires a perceiver to aggregate observations of three or more individuals—who may not be equally engaged in the interaction—into a single judgment of effectiveness at the group level. Moreover, a perceiver is confronted with numerous relationships that increase exponentially with each additional group member. Members of triads or larger groups can also form coalitions or subgroups (Simmel, 1950), further complicating the perceiver's task. Such features of groups make them distinct from dyads, raising the question of whether research on dyadic perception generalizes to small groups.

Collective Properties of Workgroups

There is compelling research showing that perceivers can detect properties of groups. Recent work by Phillips, Slepian, and Hughes (2018) finds that participants who are shown two grids with four faces each can accurately determine which grid has more racial or gender diversity and which grid has more equal or unequal hierarchy. Similarly, research by Sanchez-Burks, Bartel, Rees, and Huy (2015) on collective affect has tested perceivers accuracy—their emotional aperture—in detecting the average emotions of different size groups (e.g., panels of four or eight faces); they found that perceivers who are primed to look at the whole group were more accurate at assessing the groups' emotions than those who were primed to look at part of the group. Does our accuracy in detecting patterns in static faces extend to groups that are interacting? Research on workgroup mood suggests that a group's mood is influenced by the interaction of group members, the task, and the environment, and that trained perceivers using instruments to code for specific mood cues can accurately detect some moods (Barsade, 2012; Bartel & Saavedra, 2000). These studies did not examine judgments of performance, which force perceivers to not just evaluate group characteristics, but to make predictions about how these characteristics affect the group's outcome.

Two papers have looked at the relationship between perceivers' judgments of real groups and objective group outcomes. Stillman et al. (2014) showed perceivers brief videos of rock bands and Ultimate Frisbee teams and then asked them to rate team cohesion. Perceivers' ratings of cohesion were positively associated with team performance, as measured by bands' Youtube view counts and Frisbee teams' win/loss records, respectively. These authors also found that perceivers' judgments of cohesion in boards of directors, based on looking at publicly-available pictures of boards, were associated with company performance. The only paper that has to our knowledge directly tested the accuracy of perceivers' judgments of group performance, rather

than cohesion, is Tsay (2014), who tested whether perceivers could accurately evaluate the performance of musical groups using brief audio and video clips from live performances. Perceivers were able to accurately judge the winners after viewing six-second silent video clips of the groups or their leader, but were not able to do so after listening to audio clips or observing video and audio together.

Along with providing a provocative start to this line of research, these studies raise many interesting questions. For example, Tsay's (2014) results are based on long-standing groups that choreograph their behavior for music competitions. Since musical groups practice extensively, members' nonverbal cues may reflect how they think they are doing during a performance, which may be why perceptions of expert musical groups are driven by visual and not audio cues. Can people also perceive group effectiveness for workgroups that interact spontaneously as they conduct cognitive decision-making tasks? What if the group has not been working together for a long period of time? Groups that do not rehearse their ability to physically perform to the same beat could pose distinctly different challenges for perceivers trying to evaluate their effectiveness.

Furthermore, professors, investors, managers, and others interact with groups across a variety of settings, which may affect the channels of information they use as a basis for their judgments. Managers may listen in on groups' calls to keep track of how they are doing, or walk by a windowed conference room to glance in on a group. Professors may visually scan a loud classroom to identify which groups are struggling. Does their accuracy evaluating these workgroups depend on the type of communication channel they use? While Tsay's (2014) results suggest that visual channels outperform audio channels when evaluating musical groups, some studies suggest that the audio channel has enough information for perceivers to accurately judge

targets. For example, a study in health care found that naive perceivers could predict whether a surgeon would be sued for malpractice based on the surgeon's tone of voice (Ambady et al., 2002). In this paper, we compare perceivers' accuracy when watching silent videos or listening to audio recordings to their accuracy when they can both see and hear groups interact.

Individual Differences in Rating Group Effectiveness

Past work on emotional aperture and workgroup mood suggests that people vary in their accuracy at detecting collective affect. We were interested in testing whether people varied in their ability to accurately detect group performance, and if they did, what explained why some people were more accurate than others. Turning to the literature on group process and performance, recent research has revealed a positive association between a group's level of collective intelligence and group members' mean level of social sensitivity, which presumably confers advantages by helping group members recognize one another's emotional expressions (Engel, Woolley, Jing, Chabris, & Malone, 2014; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Past work has shown that social sensitivity is critical for leaders and managers. A study of emotional intelligence by Kerr, Garvin, Heaton, and Boyle (2006) found a strong positive correlation between supervisors' social sensitivity—as measured by how accurately people identify and perceive emotions in faces and pictures—and their employees' ratings of their effectiveness as supervisors. Interestingly they did not find a correlation between supervisors' knowledge of emotions or ability to manage emotions and employee ratings of supervisors' effectiveness. Also in a study of emotional intelligence, Rosete and Ciarrochi (2005) found that senior executives' social sensitivity—again measured by their ability to perceive emotions—was the strongest predictor of leadership effectiveness (i.e., “how well they achieved

business outputs over the financial year” as assessed by their boss and 360 ratings from employees), beating out other measures of emotional intelligence, personality, and IQ.

In addition to social sensitivity, we explored whether people’s tendency to make intuitive versus reflective decisions affected their accuracy in detecting group performance. We reasoned that if people can make accurate judgments of groups based on brief videos with limited information, then perhaps they are relying on intuitive processes to do so. Is decision accuracy associated with intuitive processes, or instead do judgments of groups benefit from a more reflective mindset? Recent work on cognitive reflection and empathic accuracy has found that perceivers’ who use a reflective mode of thought are more accurate at perceiving others’ feelings compared to perceivers who use a more intuitive mode of thought (Ma-Kellams & Lerner, 2016). Extending this research, we examine whether perceivers’ disposition toward reflection or intuition is correlated with their ability to accurately judge workgroup effectiveness. Weaving together these threads, we propose that social sensitivity and a reflective mindset could play key roles in helping perceivers judge the effectiveness of small groups.

Before we could test the accuracy of perceivers’ judgments, we first needed to create videos of small group interactions for perceivers to judge, which we describe next.

Creating Stimulus Materials

We recruited 160 participants through a university study pool to perform group decision-making tasks in the laboratory. In separate lab sessions, we assembled them into 40 four-person groups of previously unacquainted individuals, from which we excluded one group due to technical difficulties. We recruited group members from different schools and staff positions, aiming to assemble group members with some age diversity and differences in experience and perspective (e.g., avoiding groups of four undergraduates from the same university). Aside from

age, we allowed other demographic characteristics to vary freely within each group. When participants arrived at the lab we randomly assigned them to a role for a group task, had them prepare their role material, and gave them a pre-task survey. After the individual preparation period, participants met one another and worked on group decision-making tasks (see Figure 1).

Groups first performed a brainstorming task for seven minutes (Guilford, 1950) and then an adapted estimation task for fifteen minutes (Littlepage, Robison, & Reddington, 1997). These first two tasks allowed group members to gain some familiarity while working together, though they did not receive feedback on their performance. The third task, which we used to create thin slice videos, was a group decision-making task in which each of the four group members had a different role containing a mix of shared and unique information, creating a hidden profile (Stasser & Titus, 2003). We used an adapted version of PB Technologies (Peterson, 2001) in which the roles described four Vice Presidents of a technology firm who were tasked with deciding which of three job candidates to hire as a new Chief Financial Officer for their company. Prior to the group session, members had indicated their individual preference for which job candidate to hire based on the information in their respective role. During the group session, they were asked to discuss the three job candidates with the goal of agreeing on which one to hire. At the end of their discussion, group members each indicated their preferred choice, but they were not forced to agree. Individuals knew that they would receive a \$5 bonus if they chose the best candidate, along with an additional \$5 if all group members chose the best candidate. The groups we selected for the thin slice videos all made unanimous decisions (though not necessarily of the correct candidate). Through a pilot study, we confirmed that people who read the complete set of information about the three candidates chose the top candidate significantly more often (61%) than either of the other two candidates (17% and 22%,

respectively; $X^2(2, N = 36) = 12.67, p = .002$). After completing the group tasks, participants returned to their individual rooms and completed a post-task survey to measure group process dimensions. Participants then received debriefing information and were paid a base payment of \$40 and the additional bonus payments based on task performance.

Creating video clips

We selected a subset of groups for our video stimuli in order to hold constant the profile of group members' individual preferences prior to group discussion. In 12 of the groups, one member entered the group discussion holding an individual preference for the correct candidate while the other three members held preferences for the other candidates. Of these 12 groups, six reached the correct decision (which we label the “effective” groups) while the other six reached an incorrect decision (the “ineffective” groups). This profile of equivalent inputs but varied outcomes made these 12 groups ideal for creating video clips to test our hypotheses. These 12 groups did not differ significantly from the remaining groups in their demographic composition. The 12 groups worked on the hidden profile task for an average of 15.00 minutes ($SD = 4.61$), with a range from 5.85 to 23.90 minutes. The six effective and six ineffective groups did not significantly differ in task duration ($t(10) = -0.42, p = .250$) or demographic composition (see Tables 1 and 2 for overall descriptives), with the exception of age heterogeneity. Specifically, the ineffective groups ($M = 0.41, SD = 0.14$) were more diverse in age than the effective groups ($M = 0.25, SD = 0.09; t(10) = 2.28, p = .046$; measured as coefficient of variation). As we describe below, we ensured that all of our results were robust in models that controlled for mean age and age heterogeneity.

For each group, we created video clips of three different lengths—60, 30, and 10 seconds—consistent with previous thin slice research (Carney, Colvin, & Hall, 2007). The

starting point of each video clip was the midpoint of each group's hidden profile discussion, equivalent to the timing of Ambady and Rosenthal's (1993) videos, and an important juncture for coaches looking to intervene in a team (Hackman & Wageman, 2005).

Study 1: Testing whether perceivers accurately judge workgroups based on brief observations

In this study and the following, we showed individual perceivers video clips of group interaction. We told perceivers "You will now see two short video clips (60 seconds each) of two groups followed by a question asking you which group you think is more effective. Then you will continue to see pairs of videos of two different groups followed by the same question. Please watch the clips carefully. We realize these video clips provide limited information, but we ask that you make your best guess about which group you think is more effective. The groups you will observe are working on multiple tasks together. The videos you will see were taken during a time group members were working together." We did not give them any further information on the group task or the targets they observed. We used objective group performance on the hidden profile task as our criterion of group effectiveness to evaluate the accuracy of perceivers' judgments. We compensated perceivers a flat rate consistent with the norms of the sample source (e.g., in lab versus online) and the expected duration of the survey. In Study 1, we tested whether perceivers could determine which one of two groups was more effective. We started in the lab with 60-second groups. Once we replicated our in-lab findings online (see appendix A1), we tested the 30- and 10-second videos online.

Study 1a: Paired 60-second videos.

We recruited 65 participants through a university study pool to conduct the study in a computer laboratory; see Table 1 for the demographic profiles of this sample and subsequent samples.

Procedure

Participants were assigned unique identification numbers, sat at individual computer carrels equipped with individual audio headsets, and briefed on what they would do during the study. Each participant viewed six video pairs (twelve 60-second group videos in total, with full audio). Each pair included one “effective group” video and one “ineffective group” video. We created three sets of six video pairs to vary the pairings of effective and ineffective groups. For all participants, the effective group appeared before the ineffective group half the time. No participant saw the same group more than once. After each pair of videos, participants responded to the question “Which group do you think is more effective?” using a forced-choice format. After judging all of the video pairs, participants completed a demographic questionnaire. They were then debriefed and paid.

Results

Perceivers judged groups accurately (i.e., they chose the effective group) in 73.85% of the paired video trials (288 / 390 trials), which is significantly greater than a chance rate of 50% based on a one-sample t-test, $t(64) = 10.45, p < .001, 95\% \text{ CI } [69.29, 78.40]$. This result supports our core hypothesis that people can accurately judge newly-formed workgroups. It is worth noting that most perceivers were accurate more than half the time; specifically, 81.5% of perceivers accurately judged four or more of the six video pairs they viewed.

Study 1b: Paired 30-second videos.

To test whether perceivers could accurately judge group effectiveness from only 30 seconds of interaction, we recruited 66 participants from Amazon's Mechanical Turk (MTurk) with the criteria that they lived in the U.S., were fluent in English, were over 18 years of age and were able to view streaming video. Five participants were dropped due to technical difficulties, leaving 61 participants. We used six sets (instead of three sets) of six 30-second video pairs to represent all possible combinations of the best and worst videos.

Results

Perceivers chose the effective group in 74.32% of the video pairs, a level of accuracy significantly greater than chance, $t(60) = 13.14, p < .001, 95\% \text{ CI } [70.61, 78.02]$, with 85.25% of perceivers choosing the correct video in more than half of their trials.

Study 1c: Paired 10-second videos.

For our final test using paired video clips, we reduced the length of the clips to 10 seconds. We recruited 66 participants via MTurk, two of whom were dropped due to technical difficulties. The procedure was identical to the previous study except that we replaced the 30-second videos with 10-second videos.

Results

Perceivers' accuracy level of 62.24% was significantly greater than chance, $t(63) = 5.81, p < .001, 95\% \text{ CI } [58.03, 66.45]$. Most perceivers (65.62%) selected the effective group in over half of the six trials.

Discussion

The evidence from Studies 1a-1c is the first to demonstrate that perceivers can judge small ad hoc groups working on free-form cognitive work tasks with accuracy levels greater than chance. We gathered this evidence, summarized in Table 1, using a forced-choice paradigm in which participants were required to choose one group as most effective out of each pair of

groups. Given that all three studies reported above used a paradigm in which they jointly evaluated two groups, and research on joint evaluation shows that evaluability is easier when making joint decisions (Hsee et al., 1996; 1999), in the next study we test whether perceivers can accurately judge single groups using a Likert scale requiring them to independently evaluate each group on its own merits, without comparison to any other group. We also examine which collective properties might mediate perceivers' judgments.

Study 2: Perceivers rate single groups

When perceivers evaluate a group's effectiveness, they may rely on various collective properties to inform their judgments. Researchers have identified several group constructs that are reliably associated with group performance and could plausibly be useful to perceivers as indicators of whether a group is functioning effectively. We focus on three constructs that have been studied extensively in the small group literature: cohesion, trust, and conflict. Separate meta-analyses have found that group performance is positively related to cohesion (e.g., Beal, Cohen, Burke, & McLendon, 2003) and intrateam trust (De Jong, Dirks, & Gillespie, 2016) and negatively related to conflict (De Dreu & Weingart, 2003). Despite abundant research studying these phenomena directly within groups, few studies have explored how external perceivers view these group properties. Stillman et al. (2014) provided a notable exception, finding that observers' perceptions of group cohesion were positively related to actual group outcomes. We extend this research in three ways. First, whereas Stillman et al. (2014) combined survey items about trust, conflict, communication, and other dimensions into a single index of cohesiveness, we adapt established scales to ask perceivers separately about cohesion, trust, and conflict. Second, we tested whether perceptions of these constructs would act as mediators of the effect of

objective group performance on perceiver ratings of group effectiveness, rather than as a direct predictor of group performance. Third, we test these relationships by asking observers to judge small groups working on a decision-making task, a context that may elicit more varied distinctions across perceptions of cohesion, trust, and conflict compared to other contexts.

Procedure

In this study, participants viewed one group video at a time, using the same video clips from the 12 groups described above. At the start of the study, we told perceivers “You will now see video clips (60 seconds each) of groups. Please watch each one carefully because you will be asked to answer a set of questions after each clip. The groups you will observe are working on multiple tasks together.” We randomized the sequence of effective and ineffective groups and then showed this sequence of the 12 videos to all participants. After viewing each video, participants rated that group’s effectiveness along with trust, cohesion, and conflict before moving to the next group. We used actual group performance—specifically the binary measure of whether a group reached the correct outcome—to predict perceivers’ judgments of group effectiveness. Perceivers rated group effectiveness on a four-item scale adapted from Edmondson (1999), with each item rated on a seven-point Likert scale. In this and all subsequent studies which utilize this four-item measure Cronbach’s alpha always exceeds .90. We adapted widely-used scales to measure cohesion (3 items; $\alpha=.95$; Seashore, 1954), trust (4 items; $\alpha=.95$; McAllister, 1995), and conflict (6 items; $\alpha=.89$; Jehn, 1995; Bendersky & Hays, 2012).

Given that six of the stimulus groups represented the effective condition and six represented the ineffective condition, we created a dummy variable indicating for each video whether it represented an ineffective (= 0) or effective (= 1) group. We then regressed perceivers’ ratings of effectiveness (as the dependent variable) on the binary indicator of group

performance (the independent variable). We used fixed-effects models to account for differences in the way perceivers used the rating scale, effectively equalizing the mean observation for each perceiver. We also clustered standard errors on the perceiver level to account for interdependence across a single perceiver's ratings of multiple videos, which would otherwise bias the calculations of the standard error and p-value (Bertrand, Duflo, & Mullainathan, 2004; Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013). As noted earlier, we also ran versions of this model controlling for mean age and age heterogeneity of target groups. These controls had no substantive effects on our reported results, with significance levels remaining the same across all studies, and coefficients holding steady or increasing.

Study 2: Single 60-second videos

We recruited 62 participants through a university study pool to conduct the study in a computer laboratory. After viewing and rating the 12 60-second videos, participants completed basic demographic measures.

Results

Perceivers' ratings of group effectiveness showed substantial convergence across the 12 videos ($ICC = .33, p < .001$), demonstrating a significant level of agreement about which groups were most and least effective. But were these judgments accurate? The regression results indicated that the effect of group performance (0 = ineffective, 1 = effective) on perceived effectiveness ratings was positive and significant ($b = 0.89, SE = 0.07, p < .001$), indicating that perceivers were able to judge effectiveness with a substantial degree of accuracy even when viewing single videos.

Next, we tested whether this relationship was mediated by perceptions of cohesiveness, trust, or conflict, using separate regressions to avoid multicollinearity among the mediator

variables. When we added perceived cohesion to the regression, the effect of this mediator was significant ($b = 0.76, SE = 0.03, p < .001$) and the effect of group performance was still significant ($b = 0.31, SE = 0.06, p < .001$), though also significantly reduced (Sobel test: $z = 9.98, p < .001$). This pattern confirms that perceived cohesion partially mediated the effect of group performance on perceived effectiveness. A similar pattern occurred for perceived trust, which was significant ($b = 0.89, SE = 0.03, p < .001$) while reducing the effect of group performance ($b = 0.27, SE = 0.06, p < .001$; Sobel test: $z = 9.48, p < .001$), again indicating partial mediation. The binary group performance predictor had a positive effect on perceivers' ratings of cohesion ($b = 0.76, SE = 0.07, p < .001$) and trust ($b = 0.70, SE = 0.07, p < .001$), a necessary component of a mediated relationship. The indirect effect via ratings of cohesion accounted for 68% of the total effect of group performance; for trust, the indirect effect again accounted for 68% of the total effect. Perceptions of group conflict did not play a mediating role. While having a significant negative direct effect on judgments of group effectiveness ($b = -0.22, SE = 0.07, p < .001$), including perceived conflict in the model did not significantly change the effect of group performance on judgments of effectiveness ($b = 0.90, SE = 0.07, p < .001$) compared to the model that included only group performance ($b = 0.89$, reported above).

Given the mediating role of perceptions of trust, we explored whether the two subcomponents of affective trust and cognitive trust had equivalent mediating effects. When entered into the same model, both subcomponents together partially mediated the effect of group performance ($b = 0.20, SE = 0.06, p < .001$), with cognitive trust ($b = 0.77, SE = 0.03, p < .001$) having a significantly larger coefficient than affective trust ($b = 0.12, SE = 0.04, p < .001$). The indirect effect of these combined subcomponents accounted for 77% of the total effect, with 94% of the indirect effect due to cognitive trust and the remaining 6% due to affective trust.

Supplemental analyses showed that cognitive trust was also a stronger mediator than cohesion, accounting for at least twice as much of the indirect effect as cohesion when they were directly compared in the same model, either with or without affective trust included. Taken together, these results suggest that perceptions of cohesion and trust mediated the effect of binary group performance on judgments of group effectiveness, but that perceptions of conflict did not play a mediating role, which may in part be due to the low levels of conflict in these groups.

Discussion

We found consistent evidence that perceivers were able to accurately judge group effectiveness when independently rating single groups. In the appendix we show that these results are robust to a variety of changes to the experimental design; participants made accurate judgements when observing videos of different lengths, including 60, 30, and 10 seconds, and when watching videos in a controlled laboratory environment or online. We took steps to ensure that the results were not spurious consequences of the order of the videos or the response format of the judgment questions. We also found novel evidence that perceptions of group cohesion and trust partially mediated the effect of group performance on judgments of group effectiveness. These results extend Stillman et al.'s (2014) findings by demonstrating the mediating role of perceptions of group cohesion and group trust, especially cognitive trust.

This robust pattern of results raises interesting questions about the mechanisms through which it occurs (beyond perceptions of cohesion and trust) and the boundary conditions that qualify its generalizability. Thus, we next tested whether focusing on the whole collective or individual members improved accuracy, and in Study 4 we tested whether perceivers can judge effectiveness even when they can only see or hear dynamic interaction, when limited to short audio or visual clips.

Study 3: Manipulating Perceiver Accuracy by Focusing on the Collective or Individual

In this study, we examined whether perceivers' attentional focus affects their accuracy. Research on collective affect suggests that perceivers' attention, and associated judgments, may be guided by either a global focus or an individual focus (Sanchez-Burks et al., 2015). For example, perceivers who were primed to focus on the collective were more accurate at judging a diverse group's emotions than perceivers who were primed to focus on parts of the group. In a similar way, perceivers may be more accurate in discerning group effectiveness when they focus on the whole group rather than an individual group member. We tested this possibility by manipulating perceivers' attentional focus. In one condition, we directed perceivers to focus on evaluating the group as a whole, with the expectation that this would heighten their attention to the combination of group members and the interaction between them. We expected this attentional focus on the whole group to increase perceiver accuracy compared to a control condition with no such instruction. In a second condition, we directed perceivers to focus their attention on evaluating individual group members, expecting this to decrease the accuracy of judgments of group effectiveness, compared to a control. These experimental manipulations allowed us to test whether we could increase perceiver accuracy by shifting the focus of their attention in relatively subtle ways.

Study 3a: Focusing attention on group vs. individuals vs. control

We employed a between-subjects design with three conditions. We manipulated perceivers' attentional focus by inserting a one-sentence instruction into the general instructions for viewing the videos, which were otherwise constant across conditions (i.e., "You will now see video clips (10 seconds each) of groups who are working on multiple tasks together"). Perceivers

read either the group focus instruction (“Please focus on the group as a whole because you will be asked to rate how much you would like to work with this group”), the individual focus instruction (“Please focus on the individual members of the group because you will be asked to choose one group member you’d most like to work with”), or the control instruction (“Please watch each [video clip] carefully because you will be asked to answer a set of questions after each clip”). We randomly assigned each perceiver to one of the three conditions. After each 10-second video, we asked participants first to rate the effectiveness of the group (using an identical question to previous studies) and then asked them an additional rating question, consistent with their condition, to reinforce the manipulation for the remaining videos.

We also used this study to expand our tests beyond the 12 groups we used in previous studies. Recall from the section on “Creating Video Clips” that we selected 12 groups with identical profiles of group member pre-discussion preferences. These 12 groups were also evenly split in reaching the correct or incorrect group decision. For the current study, we included videos from 31 groups that participated in our stimulus creation sessions. Of these groups, 10 reached a correct decision and 21 reached an incorrect decision. Given our interest in judgments of group performance, we continued to use binary decision outcomes as a predictor for testing perceiver accuracy. Each perceiver saw 10-second clips of all 31 videos in random order.

A total of 140 Mturk perceivers participated in the study, with 49 in the group focus condition, 45 in the individual focus condition, and 46 in the control condition. We excluded individuals who did not pass the qualifications test (12), had technical difficulties while playing the videos (10), or started the study but did not finish it (31).

Results

We first confirmed that perceiver ratings of group effectiveness demonstrated a significant level of agreement across the 31 videos ($ICC = .30, p < .001$). We also confirmed that our core result—that perceivers are more accurate than chance—replicated when using all 31 groups ($b = 0.17, SE = 0.05, p < .001$; this and all analyses below control for video order) instead of the 12 we previously used.

Next, we ask whether accuracy differed across the three conditions. To test this, we regressed perceivers' ratings on the binary measure of group performance along with interaction terms to test whether the effect of the group performance measure (indicating accuracy of perceptions) was moderated by dummies for the group focus condition or the individual focus condition (compared to the omitted control condition). As predicted, perceivers in the group focus condition were significantly more accurate than those in the control condition ($b = 0.29, SE = 0.09, p < .001$). Counter to our expectations, however, perceivers in the individual focus condition were also significantly more accurate than the control group ($b = 0.25, SE = 0.08, p < .001$). The group and individual focus conditions were not significantly different from one another ($b = -0.04, SE = 0.09, p < .70$).

These results suggest that subtle shifts in perceivers' attentional focus cause higher levels of accuracy, even for perceptions of videos that are a mere 10 seconds long. The results also establish that perceivers can detect group performance at levels greater than chance across a larger sample of 31 groups. We designed our experimental conditions to focus perceivers' attention on either the group as a whole or the distinct individual group members, expecting these to shift accuracy in different directions, yet both conditions increased perceiver accuracy compared to the control condition. It is possible that our method for focusing perceivers' attention is comprised of two components: the first is to focus on a specific perceptual target

(e.g., whole group versus one individual), and the second is to heighten attention to the overall task. A plausible interpretation of the results is that both treatments heightened overall attention to the task, but the individual focus condition, by directing perceivers to focus on all four group members, did not direct perceivers' focus in a way that was sufficiently different from the group focus condition. We designed our next study to test this possibility by creating a sharper distinction between the individual and group focus conditions.

Study 3b: Focusing attention on group vs. one individual vs. control

Researchers have shown that when perceivers observe a set of stimuli, the information they extract depends on whether they have a wide focus on the entire set or a narrower focus on one or more parts of the set (Alvarez & Olivia, 2008). To test this in the context of perceiving work groups, we took advantage of the fact that the four group members were always arrayed from left to right in the video, allowing us to direct perceivers' focus to one specific member in each group. We used the same design and procedure as the previous study, except that in the individual focus condition we told perceivers: "Please focus on the group member on the far right because you will be asked to rate how much you would like to work with this group member."

A total of 141 Mturk perceivers participated in the study, with 45 in the group focus condition, 47 in the individual focus condition, and 49 in the control condition. We excluded individuals who did not pass the qualifications test (17), had technical difficulties while playing the videos (11), or started the study but did not finish it (23). Each perceiver saw 31 10-second videos in random order.

Results

As in the previous study, perceiver ratings of group effectiveness converged across the 31 videos ($ICC = .26, p < .001$). And again, perceivers overall were more accurate than chance ($b = 0.32, SE = 0.04, p < .001$).

Turning to the experimental conditions, we replicated some, but not all, previous effects. First, perceivers in the individual focus condition were more accurate than the control ($b = 0.18, SE = 0.08, p < .04$), replicating the effect from the previous study. In contrast, perceivers in the group focus condition were not significantly more accurate than those in the control condition ($b = 0.09, SE = 0.09, p = 0.32$), nor did they differ from the individual focus condition ($b = 0.09, SE = 0.09, p = 0.35$). These results are harder to interpret: on the one hand, they did not replicate the earlier result that a group focus increases accuracy relative to a control condition. On the other hand, the group condition was also not significantly different from the individual focus condition, suggesting accuracy falls somewhere in the middle (but without a significant difference compared to either of the other conditions). In other words, the group focus condition appears to have a weaker effect (if any) than in our previous study.

Discussion

Taken together, Studies 3a and 3b make three contributions to our previous studies. First, we expanded the sample of groups beyond the original 12 groups we used for the previous studies. Importantly, our main result remains robust: the thin slice result replicated with this larger sample of groups, even though the additional groups faced different challenges in transforming their profile of individual preferences, which varied considerably, into a group decision. Second, we demonstrated that, on average, when it comes to judging thin slices of group interaction, interventions of focusing perceivers' attention (across our different types of manipulations) can be used to increase their accuracy. Third, we found a pattern of results

suggesting that heightened attention to individuals in the group seems to increase accuracy in judging performance; unlike previous research in which perceivers who focused on the whole set were more accurate than those who focused on parts, we found mixed support for focusing on the group.

We were surprised that our attempt to focus perceivers' attention on a single group member increased their accuracy relative to a control condition. One speculative possibility is that perceivers who were given a specific goal to focus their attention had a stronger accuracy motive (Phillips et al., 2014), which may have increased their attention to all interactions among group members. If so, then a perceiver's initial focus on a particular group member could quickly expand to other group members who actively participate in an interaction episode, whether by speaking, moving, or otherwise exhibiting dynamic cues that draw attention. This may have caused the individual focus manipulation to increase perceiver accuracy. Alternatively, it may be that perceivers judge interacting groups differently than pictures or shapes. The group task may also determine what information perceivers use. For example, Tsay found that perceivers were more accurate at assessing a musical ensemble's performance when looking at a silent video of the leader compared to listening to the entire ensemble.

These results suggest that being able to judge a group's performance is in part driven by contextual factors surrounding the perceiver, such as directing attention to relevant cues. In our final study, we explore another contextual factor—whether video or audio information, or both, is available to perceivers. Furthermore, we investigate the extent to which individual differences among perceivers—specifically their social sensitivity and tendency to engage in cognitive reflection—affect their ability to accurately judge and evaluate the performance of groups. Finally, our next study also disentangles some confounds in our previous study design and

presents the most rigorous and conservative test of our hypothesis, together with a pre-registered analysis.

Study 4: Perceivers rate effective groups higher than ineffective groups in all conditions, and those high in social sensitivity are more accurate

While perceivers' accuracy was quite high in Studies 1a through 3b, there were still perceivers who were less accurate in judging group effectiveness. This is consistent with research on perceiver's accuracy of thin slices of individual targets. To explain this variation across perceivers, researchers have studied individual differences in social sensitivity. Social sensitivity is a person's ability to accurately detect and judge other people's emotions (Bernieri, 2001; Pickett, Gardner, & Knowles, 2004; Woolley, Aggarwal, & Malone, 2015) and is typically measured by a variety of standardized tests, mostly using facial affect. Past research has found that in addition to being able to accurately detect a target's emotions (e.g., anger), there is a strong positive relationship between perceivers' social sensitivity and their ability to accurately detect a target's personal characteristics (e.g., relationship status), social characteristics (e.g., whether a target was watching a funny or scary movie), and personality (for reviews, see Bernieri, 2001 and Hall & Andrzejewski, 2008). Research on collective intelligence has examined group members' aggregated levels of social sensitivity (Engel et al., 2014; Woolley et al., 2015; Woolley, et al., 2010) and found that the higher the average social sensitivity of group members, the better the group performs. Here we extend this research by looking at the social sensitivity of perceivers when judging group decision-making.

While there are many tests of social sensitivity, these studies used the "Reading the Mind in the Eyes" (RME) test in which perceivers are asked to look at series of pictures of targets' eye

area and select which emotion the targets are experiencing (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). In Study 4, we thus used the RME test to examine whether perceivers' social sensitivity is related to their accuracy in judging workgroup effectiveness. Unlike studies of social sensitivity in which perceivers look at individual targets and judge their affect or personality, our criterion variable in this study is objective group performance.

Furthermore, we explore another individual difference that might explain accuracy in judgements. In Studies 1 and 2, we found that people could make accurate judgments of groups after watching brief videos with limited information. To examine whether people who tend to make judgments intuitively are more accurate at judging group effectiveness than those who reflect before making a judgment, in Study 4 we used an adapted version of the Cognitive Reflection Test (CRT). The CRT measures a person's ability or disposition to reflect and resist reporting the response that first comes to mind (Frederick 2005).

Sample size

We recruited 252 participants on Amazon Mechanical Turk. The sample size for this study was determined in advance: we pre-registered the hypothesized main effect—that perceivers will judge groups that completed the task as more effective than those who did not—and we specifically pre-registered that the effect would be detected for both the first decision of each perceiver as well as their average judgements across all decisions (see details of the pre-registration at <https://aspredicted.org/blind.php?x=dv4du5>). To inform our power calculation, we derived an estimated effect size from a small pilot (N=60); for parameter details, see the pre-registration document. Based on the pilot data, our statistical power for the main outcome variable was estimated to be above 95%. In addition, we also specified exploratory analysis to investigate whether social sensitivity or cognitive reflection are predictive of the accuracy of

these group judgements; we had an anticipated estimated power of 95% and 80%, respectively, to detect an effect of these exploratory variables.

Procedure

Participants watched six media clips in a row after reading the instructions (“You will now watch several short clips (10 seconds each) of groups working, followed by a question asking you whether you think the group is effective in reaching a good decision. Then you will continue to see different groups followed by the same question. Please watch the clips carefully. We realize these clips provide limited information, but we ask that you make your best guess about whether you think a group is effective. The groups you will observe are working on multiple tasks together. The video you will see was recorded during a time group members were working together.”). The media clips contained both audio and video, video only, or audio only depending on their randomly assigned condition. After each media clip, they rated the effectiveness of the group on a 1-7 point scale which ranged from 1 = “Not effective at all” to 4 = “Moderately effective” to 7 = “Extremely effective” before moving on to the next media clip. After six media clips, all participants completed the Reading the Mind in the Eyes (RME) test (36 items with $\alpha = .83$; Baron-Cohen et al., 2001), a measure of social sensitivity, and an adapted Cognitive Reflection Test (3 items with $\alpha = .85$; Frederick, 2005), which consists of three questions (e.g., “A textbook and a pen cost \$130 in total. The textbook costs \$100 more than the pen. How much does the pen cost?”).

We manipulated two aspects of the study in a three-by-two factorial design. Along the first dimension, we varied whether participants were assigned to a condition with audio and video combined, video only, or audio only (between-subjects randomization). Along the second dimension, we varied whether the group participants evaluated was an “effective” group or an

“ineffective” group. We randomized the second dimension within-subject, such that all participants saw three effective groups and three ineffective groups in fully randomized order. To avoid spill-over and learning, our main analysis focuses on the first decision only; however, in robustness checks, we also analyze all six decisions for each participant.

For the main analysis which focuses on the first decision each participant made, we used a linear regression model with robust standard errors, predicting the perceived group effectiveness rating with a binary variable indicating whether the group was effective (1 = effective group, 0 = ineffective group). For analyses which include all six decisions a participant made, we use a fixed-effects model across all six decisions, predicting group effectiveness ratings with the same binary group effectiveness indicator, with standard errors clustered at the participant level.

We also report the same analysis for each media-clip condition separately (video and audio versus video only versus audio only) and test the effects of social sensitivity and cognitive reflection on the accuracy of participants’ judgments using random-effects models, clustering standard errors at the participant level. To explore differences between media-clip conditions, we use random effects models with clustered standard errors at the participant level.

Results

We begin by looking at participants’ first decision only (Table 2): participants perceived an effective group as significantly more effective ($M = 4.22$, $SD = 1.24$) than an ineffective group ($M = 3.35$, $SD = 1.27$; coefficient of indicator variable in linear regression = 0.88, 95% CI [0.57, 1.19], $p < .001$). This result supports our main hypothesis that perceivers can distinguish effective groups from ineffective ones. Furthermore, this result holds—and even strengthens in terms of the size of the effect—when we consider all six of the participants’ decisions (effective

group: $M = 4.62$, $SD = 1.34$; ineffective group: $M = 3.39$, $SD = 1.39$; $\text{coeff} = 1.22$, 95% CI [1.09, 1.35], $p < .001$), providing further strong support for our thesis (see Table A2).

Next, we explore whether these effects hold within each of the media-clip conditions independently (see cols. 2-4 in Tables 2 and A2). Participants judged groups that succeeded in solving the task as more effective than those that did not in the video and audio condition (first decision: $\text{coeff} = 1.21$, 95% CI [.68, 1.73], $p < .001$; all decisions: $\text{coeff} = 1.57$, 95% CI [1.34, 1.80], $p < .001$), in the video only condition (first decision: $\text{coeff} = .75$, 95% CI [.17, 1.33], $p = .012$; all decisions: $\text{coeff} = .98$, 95% CI [.78, 1.19], $p < .001$), and in the audio only condition (first decision: $\text{coeff} = .65$, 95% CI [.12, 1.18], $p = .016$; all decisions: $\text{coeff} = 1.08$, 95% CI [.87, 1.28], $p < .001$). Thus, even when participants could only view the video or just listen to the audio of groups interacting, they could still distinguish effective from ineffective groups. However, the effect appears to be strongest in the combined video and audio condition: using the full model with interaction terms, a significant difference emerged for participants in the video and audio condition who rated effective groups higher than participants in the other two conditions (see col. 5 in Table 3: single effect of effective groups in the combined video and audio condition: $\text{coeff} = 1.58$, 95% CI [1.35, 1.81], $p < .001$; coefficient interacted with video only condition is lower: $\text{coeff} = -.59$, 95% CI [-.89, -.29], $p < .001$; similarly, interacted with audio only: $\text{coeff} = -.50$, 95% CI [-.81, -.20], $p = 0.001$; the video only and audio only conditions were not significantly different from each other: $\text{coeff} = 0.08$, 95% CI [-.20, .37], $p = .559$).

We then explored whether individual differences in perceivers' social sensitivity could explain their accuracy in distinguishing effective versus ineffective groups. We find strong support for this proposed mechanism (Table 4): participants who score highly on the RME test were significantly more likely to accurately rate effective and ineffective groups ($\text{coeff} = .13$,

95% CI [.08, .18], $p < .001$). Interestingly, this effect is strongest in the video and audio condition (coeff = .22, 95% CI [.13, .30], $p < .001$) and less pronounced in the video only condition (coeff = .07, 95% CI [-.01, .15], $p = .079$) and in the audio only condition (coeff = .09, 95% CI [.00, .19], $p = .057$). This suggests that social sensitivity plays a role in all three conditions but more so in the video and audio condition (as confirmed in the full interaction model: single effect of RME in the combined video and audio condition is highly significant: coeff = .22, 95% CI [.13, .30], $p < .001$; whereas the coefficient interacted with video only is lower: coeff = -.15, 95% CI [-.26, -.03], $p = .014$; and marginally so in the interaction with audio only: coeff = -.13, 95% CI [-.25, -.00], $p = .049$; see col. 5 in Table 4).

Finally, we looked at whether perceivers were more accurate when they were more intuitive or when they were more reflective as measured by the CRT (Table A3). We found that perceivers who were more reflective were marginally more accurate across all three conditions (coeff = .008, 95% CI [-.00, .02], $p = .051$); however, this relationship did not hold in any of the subanalysis within each condition separately (combined audio and video condition: coeff = 0.01, 95% CI [-.01, .02], $p = .207$; video only: coeff = .003, 95% CI [-.01, .01], $p = .666$; audio only: coeff = .005, 95% CI [-.01, .02], $p = .347$), suggesting the effect is weak at best.

Discussion

This study makes four contributions to our previous studies. First, we test whether our methodology in our previous studies was driving our findings; it is possible that perceivers learned what to expect after the first media clip they saw, shaping their ability to judge subsequent group effectiveness. We powered (and pre-registered) our study so that we could test perceivers' accuracy in judging team effectiveness using only the first clip they saw or heard, and we replicated our previous findings. Second, we found that perceivers were able to judge the

performance of decision-making workgroups whether we presented them with videos with visual and audio, only visual, or only audio information. Though these channels provide substantially different sets of cues, each channel was informative enough by itself to allow perceivers to judge group effectiveness. This evidence suggests that group members engaged in joint decision-making tasks exhibit both visual and audio cues that are associated with their effectiveness. The evidence that perceivers can achieve a significant level of accuracy from viewing silent videos is important because it minimizes the possibility that perceivers are somehow inferring effectiveness only from the substantive, audible content of group discussion (an unlikely possibility when viewing 10 second videos with audio, to be sure, but still logically plausible). This evidence supports that idea that perceivers attend at least in part to nonverbal behaviors when forming judgments of small groups, consistent with research on individual and dyadic targets.

Third, we found a strong relationship between perceivers' social sensitivity and their ability to accurately judge group effectiveness. While this relationship was strongest for the videos with audio and visual information, the marginally significant results in the audio condition suggest that social sensitivity, which we measured using visual stimuli, is not limited to visual judgments but is actually a more generalizable ability. Fourth, we found little evidence that perceivers who generally trust their intuitive judgments were better at judging workgroups. This is intriguing because the amount of information that perceivers receive is very limited (in this case only 10 seconds) and it seems plausible that having a good "gut reaction" based on such short exposure could be a predictor of accurate judgments. However, if anything, our evidence suggests otherwise: we found that higher reflection is marginally associated with more accurate perceptions of a group's effectiveness. One reason why intuition may not have been predictive of

accurate judgments is that trusting your intuition requires having relevant expertise in the task domain (Dane, Rockmann & Pratt, 2012), which perceivers in our tasks did not have (since they did not know about the task the group was working on).

General Discussion

Across a variety of studies, this paper reports the first evidence, to our knowledge, that perceivers can accurately judge the effectiveness of ad hoc decision-making workgroups based on thin slices of group interaction. Perceivers were able to do so when choosing among pairs of videos (Studies 1a-1c) and when independently rating single videos (Study 2); whether in the lab or online; when viewing videos of different lengths, including 60, 30, and 10 seconds; when viewing 10-second video clips without audio, and when listening to 10-second audio clips without video (Study 4). In Study 2, we found that perceptions of collective properties of groups, including cohesion, affective trust, and cognitive trust, partially mediated perceivers' ability to accurately judge groups; conflict, on the other hand, was not a mediator.

We then explored whether individual differences affected perceiver accuracy. In Studies 3a and 3b, we manipulated perceivers' attentional focus by having them focus on evaluating the group as a whole or an individual in the group. We found that increasing perceivers' attention, even with a subtle manipulation, significantly increased their accuracy in the individual focus condition compared to a control condition. However, we did not find a significant difference in perceiver accuracy between the collective focus condition and the individual focus condition. In Study 4 we explored both contextual and individual-level differences. We found that while perceivers were accurate across information channels—they could detect group effectiveness in 10-second silent videos and 10-second audio clips—accuracy was higher when perceivers had

more channels of information available. This suggests that both seeing and hearing a group can improve accuracy over and above just listening to the conversation (e.g. imagine listening in through a conference call) or watching the interaction without audio (e.g. through a window, outside the room). Also in Study 4, we found that perceivers' social sensitivity was positively and significantly related to their ability to judge group effectiveness, and that this relationship was strongest in the video plus audio condition, presumably where they could apply their ability to discern social cues to the largest stream of social information. We also found a marginally significant relationship between perceivers' reflective or more systematic mode of thought and their accuracy in determining group effectiveness.

Abundant thin slice research suggests that perceivers can accurately judge individual targets on a wide variety of dimensions (Ambady & Rosenthal, 1992). Yet much of social life occurs in small groups, making it important for people—and especially managers leading groups in organizations—to be able to discern collective properties which can vary widely across groups. A fundamental basis for making such judgments is observing spontaneous social interaction among group members, which we placed front and center in our studies. Our evidence suggests that even short glimpses of such interaction allow perceivers to accurately judge group effectiveness. Just as people are apparently equipped to automatically judge individual characteristics based on facial features (Stirrat & Perrett, 2010), perhaps they are also equipped to quickly evaluate group performance based on glimpses of interaction.

These results are not only relevant for outside observers of groups; they have implications for group members themselves, who are constantly perceiving, judging, and reacting to their own groups. Putting together our results with those from collective intelligence, cognitive reflection studies, and other thin slice research, we suggest that underneath the

deliberative group processes typically studied, including activities like information sharing and tactical negotiations (Mesmer-Magnus & DeChurch, 2009; Jehn, Northcraft & Neale, 1999), there are layers of nonverbal cues (DePaulo, 1992; Ambady & Rosenthal, 1992; Grahe & Bernieri, 1999) that group members are able to perceive and evaluate in collectively meaningful ways. These judgments are likely to be consequential for group functioning, as members continually judge the collective state of their group and adapt their behavior accordingly. It may also be the case that some members are better able to gauge other team members and the group as a whole. Group interaction can be highly dynamic, and the perceptual processes we have studied in this paper may help to explain how individual perceptions and behaviors interlock to become properties of the group.

Limitations

Across several studies, we varied our study procedures to test specific questions under controlled conditions. Our use of similar stimuli—videos of groups performing the same task—naturally limits any claims we can make in generalizing our results to other contexts, tasks, or types of groups. For example, our groups met for short periods of times and experienced low levels of conflict, so it is not surprising that perceptions of conflict did not mediate our findings. In groups that experience high levels of relational, status, or process conflict, perceptions of conflict might mediate the group's poor performance; while perceptions of task conflict might mediate high performance.

In our study, perceivers were unfamiliar with the group members they judged, and group members themselves were previously unacquainted before the study, conditions which may not mimic many day-to-day situations. Nevertheless, using groups of strangers may make the discernable variation that occurred across groups all the more surprising. We also used videos

from the midpoint of group deliberations on the task at hand, which may provide a unique window into group functioning. Using a binary outcome as our criterion for separating effective from ineffective groups may have masked important increments of effectiveness.

We did not focus on demographic and cultural differences among group members and perceivers, though these have been shown to influence perceptions of crowds (e.g., Alt, Goodale, Lick, & Johnson, 2017; Lamer, Sweeny, Dyer, & Weisbuch, 2018). It is possible, for example, that different mechanisms guide perceivers' judgments when observing diverse versus homogeneous groups, which raises another limitation of our study—we did not isolate and manipulate specific cues that may have driven our results. Our intent was to establish a solid foundation of evidence for whether perceivers are able to judge groups consensually and accurately based on small slices of the behavioral stream. In doing so, we hope to open a door to a new direction for future research, a direction that is rich with possibilities.

Future Research

The results of our studies naturally raise questions about the mechanisms through which perceivers judged groups, alongside perceptions of cohesion, affective trust, and cognitive trust. Research can move in a more fine-grained direction to unpack the myriad cues that might be in play, separately and in combination. A key question is how perceivers aggregate numerous cues, both across different channels such as audio and visual, and across different group members. Groups should be more legible when members exhibit similar cues and when cues across different channels align. When inconsistencies occur, do some cues receive the most attention in a stable, systematic way? Do some group members shape perceptions more strongly, or are some behaviors or interaction episodes more potent than others? We can imagine many precise studies designed to tease apart answers to such questions. Such studies will be more powerful to the

extent that they are guided by and inform a broader theory about how cues combine in complex social situations.

It is tempting to focus on affective relations among group members, and the nonverbal cues associated with affective dimensions, as a direction for future research. Some of our results support this inclination, especially the accuracy demonstrated by perceivers who were high in social sensitivity, which itself is related to emotional intelligence. Providing further support for an affective orientation, we found that perceptions of cohesion and affective trust both helped to explain perceiver accuracy. Yet, the story that emerged from our results was not exclusively about affect; in fact, cognitive trust was a stronger mediator of judgment accuracy than its affective counterparts. Moreover, increases in accuracy associated with greater attentional focus could have been associated with perceptions of competence or other task-related dimensions that required more motivated concentration to discern. Ultimately, our results support a theoretical account through which people use all of the dimensions at their disposal to judge group effectiveness, including affective and cognitive dimensions, along with visual and audio channels.

Looking at contextual and individual-level differences could also reveal the abilities that help perceivers achieve more accurate judgments. For example, when does a global focus help in group assessment and when does a local focus improve accuracy? In Study 3a and 3b, we manipulated perceiver focus by having them focus on the group or an individual but did not find a difference between these conditions. Future studies could design stronger manipulations of global and local focus, or measure perceiver's innate tendency toward a local versus global focus to see if perceivers with a global focus are more accurate in their perceptions of groups,

especially perceptions of properties that reside at the group level such as cohesion and effectiveness.

It could also be informative to give perceivers different frames when asking them to judge the same group, testing whether perceivers asked to judge a negotiation, for example, use different cues or interpret cues differently compared to perceivers asked to judge collaborative interaction or teamwork. Framing the task could activate different mental models in perceivers, which in turn influence which cues they look for and notice as indicators of effective group functioning. For example, some perceivers may hold the view that a hallmark of highly effective collaboration is equal participation coupled with lots of back-and-forth exchanges (Woolley et al., 2010), whereas other perceivers might believe that group members with high expertise or who hold a leadership position should be privileged with more airtime. Such mental models could shape perceivers' visual strategies, which could influence accuracy. Perceivers who focus on the content of communication may shift their focus from speaker to speaker as group members banter back and forth, whereas other perceivers might observe how group members react while listening to others. Do observations of listeners allow greater insight into the state of the group, or is it a better strategy to follow the trail of conversation by watching each speaker? Researchers could use eye-tracking technologies to explore where perceivers focus their attention when observing group interaction, and whether this matters for accuracy. This highlights an important advantage of the visual channel—it allows perceivers to witness reactions and nonverbal behavior of those who are *not* talking or participating, potentially providing an important window into group members' engagement.

Studying how people observe groups raises questions about the interplay between perceiving a group and interacting with the group. For new members joining an existing group,

initial judgments of the group could affect the new person's early behavior, in ways both unconscious and strategic. As the group adapts its interaction based on the new member's behavior, group members may update their judgments of how the group is functioning, and attribute any changes to the new person. Shifting from the role of group observer to group participant may also change the motives that guide perceptions. For example, if a member offers an idea to the group, is that member more likely to see cues that confirm the value of the idea, compared to an impartial observer? Beyond new members joining a group, team leaders use judgments of their team to adapt their own approach to leading. Do team leaders who use an explicit, deliberative approach to gauging team engagement—for example by setting aside time in team meetings to debrief recent group decisions—achieve more accuracy than leaders who simply rely on intuitions that are guided by automatic perceptions?

Conclusion

By demonstrating that perceivers can discern group effectiveness from thin slices of group interaction, this research provides a new lens for viewing group functioning, in one direction, and social perception, in the other. A wealth of research on social perception focusing on how one individual perceives another individual has set the stage for studying a distinct class of phenomena in which one individual perceives social interaction between two or more other people. These latter situations have received little attention from researchers, which is surprising given that they describe a broad swath of social perceptions dating far back in time. In this paper we have tried to advance our understanding of these situations to further illuminate the interplay between individuals and groups that is at the heart of social life.

References

- Alt, N. P., Goodale, B., Lick, D. J., & Johnson, K. L. (2017). Threat in the Company of Men: Ensemble Perception and Threat Evaluations of Groups Varying in Sex Ratio. *Social Psychological and Personality Science*. <https://doi.org/10.1177/1948550617731498>
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392-398.
<http://dx.doi.org/10.1111/j.1467-9280.2008.02098.x>
- Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology*, *16*(1), 4-13.
http://dx.doi.org/10.1207/s15327663jcp1601_2
- Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: A clue to malpractice history. *Surgery*, *132*(1), 5-9.
<http://dx.doi.org/10.1067/msy.2002.124733>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256.
<http://dx.doi.org/10.1037/0033-2909.111.2.256>
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431-441. <http://dx.doi.org/10.1037/0022-3514.64.3.431>
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251.

- Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47(4), 644-675.
- Bartel, C. A., & Saavedra, R. (2000). The collective construction of work group moods. *Administrative Science Quarterly*, 45(2), 197-231.
- Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: a meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88(6), 989.
- Bendersky, C. and Hays, N. (2012). Status conflict in groups. *Organization Science*, 23(2), 323-340.
- Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J.A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement*. (pp. 3-20). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, 71(1), 110-129. <http://dx.doi.org/10.1037/0022-3514.71.1.110>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimate? *Quarterly Journal of Economics*, 119, 249-275. <http://dx.doi.org/10.1162/003355304772839588>
- Bodenheimer, T., & Laing, B. Y. (2007). The teamlet model of primary care. *The Annals of Family Medicine*, 5(5), 457-461.
- Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly*, 17(3), 288-307.

- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054-1072.
<http://dx.doi.org/10.1016/j.jrp.2007.01.004>
- Carrere, S., & Gottman, J. M. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process, 38*, 293–301.
<http://dx.doi.org/10.1111/j.1545-5300.1999.00293.x>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.
<http://dx.doi.org/10.1037/0022-3514.76.6.893>
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology, 104*(1), 103-125.
<http://dx.doi.org/10.1037/a0030398>
- Clingingsmith, D., & Shane, S. (2017). Training Aspiring Entrepreneurs to Pitch Experienced Investors: Evidence from a Field Experiment in the United States. *Management Science, Articles in Advance*, 1-16.
- Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology, 92*(3), 802-811. <http://dx.doi.org/10.1037/0021-9010.92.3.802>
- Dane, E., Rockmann, K. W., & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes, 119*(2), 187-194.

- De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of applied Psychology*, 88(4), 741.
- De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, 101(8), 1134.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111(2), 203-243. <http://dx.doi.org/10.1037/0033-2909.111.2.203>
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383. <http://dx.doi.org/10.2307/2666999>
- Engel D, Woolley AW, Jing LX, Chabris CF, Malone TW (2014) Reading the Mind in the Eyes or Reading between the Lines? Theory of Mind Predicts Collective Intelligence Equally Well Online and Face-To-Face. *PLOS ONE* 9(12): e115212. <https://doi.org/10.1371/journal.pone.0115212>
- Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *The Leadership Quarterly*, 2(4), 245-287.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday/Anchor Books.

- Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology, and health. *Journal of Personality and Social Psychology*, *63*, 221–233. <http://dx.doi.org/10.1037/0022-3514.63.2.221>
- Gottman, J. M., Markman, H., & Notarius, C. I. (1977). The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior. *Journal of Marriage and Family*, *39*(3), 461-477. <http://dx.doi.org/10.2307/350902>
- Gottman, J. M., & Notarius, C. I. (2000). Decade review: Observing marital interaction. *Journal of Marriage and Family*, *62*(4), 927-947. <http://dx.doi.org/10.1111/j.1741-3737.2000.00927.x>
- Grahe, J. E., & Bernieri, F. J. (1999). The importance of nonverbal cues in judging rapport. *Journal of Nonverbal Behavior*, *23*(4), 253-269. <http://dx.doi.org/10.1023/A:1021698725361>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454.
- Gust (2016). Global Accelerator Report. Retrieved from: http://gust.com/accelerator_reports/2016/global/.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718-734. <http://dx.doi.org/10.1037/a0013899>
- Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, *30*(2), 269-287.
- Hall, J. A., & Andrzejewski, S. A. (2008). Who draws accurate first impressions? Personal correlates of sensitivity to nonverbal cues. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions* (pp. 87-105). New York, NY: Guilford Publications.

- Harris, M. J., & Garris, C. P. (2008). You never get a second chance to make a first impression: Behavioral consequences of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions*, (pp. 147-168). New York, NY: Guilford Publications.
- Hecht, M. A., & LaFrance, M. (1995). How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *Journal of Applied Social Psychology*, 25(23), 2086-2098. <http://dx.doi.org/10.1111/j.1559-1816.1995.tb02389.x>
- Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational behavior and human decision processes*, 67(3), 247-257.
- Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychological bulletin*, 125(5), 576.
- Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 40(2), 256-282.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, 44(4), 741-763. <http://dx.doi.org/10.2307/2667054>
- Kerr, R., Garvin, J., Heaton, N., & Boyle, E. (2006). Emotional intelligence and leadership effectiveness. *Leadership & Organization Development Journal*, 27(4), 265-279.
- Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science*, 20(1), 99-106. <http://dx.doi.org/10.1111/j.1467-9280.2008.02251.x>

- Lamer, S. A., Sweeny, T. D., Dyer, M. L., & Weisbuch, M. (2018). Rapid visual perception of interracial crowds: Racial category learning from emotional segregation. *Journal of Experimental Psychology: General*, *147*(5), 683-701.
- Littlepage, G., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. *Organizational Behavior and Human Decision Processes*, *69*(2), 133-147.
<http://dx.doi.org/10.1006/obhd.1997.2677>
- Ma-Kellams, C., & Lerner, J. (2016). Trust your gut or think carefully? Examining whether an intuitive, versus a systematic, mode of thought produces greater empathic accuracy. *Journal of personality and social psychology*, *111*(5), 674.
- Marks, M., Mathieu, J., & Zaccaro, S. 2001. A temporally-based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356-376.
- Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, *34*(3), 410-476.
- McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59.
- Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, *94*(2): 535-546.
<http://dx.doi.org/10.1037/a0013773>
- Neuberg, S., Kenrick, D., & Schaller, M. (2010). Evolutionary social psychology. In Fiske S. T., Gilbert D. T., Lindzey G. (Eds.), *Handbook of social psychology* (5th ed., pp.761–796). New York, NY: Wiley. <http://dx.doi.org/10.1002/9780470561119.socpsy002021>

- Peterson, R.S. (2001). *PB Technologies*. Kellogg Dispute Resolution Center Teaching.
- Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology*, *114*(5), 766-785. <http://dx.doi.org/10.1037/pspi0000120>
- Phillips, L. T., Weisbuch, M., & Ambady, N. (2014). People perception: Social vision of groups and consequences for organizing and interacting. *Research in Organizational Behavior*, *34*, 101-127. <http://dx.doi.org/10.1016/j.riob.2014.10.001>
- Pickett, C. L., Gardner, W. L., & Knowles, M. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin*, *30*(9), 1095-1107. DOI: 10.1177/0146167203262085
- Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science*, *20*(1), 22-26. <http://dx.doi.org/10.1111/j.1467-9280.2008.02248.x>
- Romero, D. M., Swaab, R. I., Uzzi, B., & Galinsky, A. D. (2015). Mimicry is presidential: Linguistic Style Matching in Presidential Debates and improved polling numbers. *Personality and Social Psychology Bulletin*, 1-9. DOI: 10.1177/0146167215591168
- Rosete, D., & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadership & Organization Development Journal*, *26*(5), 388-399.
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, *50*(3), 540-547.

- Sanchez-Burks, J., Bartel, C. A., Rees, L., & Huy, Q. (2015). Assessing collective affect recognition via the Emotional Aperture Measure. *Cognition and Emotion*, *30*(1), 117-133.
- Seashore, Stanley E. (1954) *Group Cohesiveness in the Industrial Work Group*. Ann Arbor, MI: University of Michigan Press.
- Shunk, R., Dulay, M., Chou, C. L., Janson, S., & O'Brien, B. C. (2014). Huddle-coaching: a dynamic intervention for trainees and staff to support team-based care. *Academic Medicine*, *89*(2), 244-250.
- Simmel, G. (1950). *The Sociology of Georg Simmel*. Glencoe, IL: Free Press.
- Stasser, G., & Titus, W. (2003). Hidden profiles: A brief history. *Psychological Inquiry*, *14*(3), 304–313. http://dx.doi.org/10.1207/S15327965PLI1403&4_21
- Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, *32*(1), 29-55.
- Stillman, P. E., Gilovich, T., & Fujita, K. (2014). Predicting group outcomes from brief exposures. *Social Cognition*, *32*(1), 71-82.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349-354.
- Tickle-Degnen, L. (1998). Working well with others: The prediction of students' clinical performance. *American Journal of Occupational Therapy*, *52*(2), 133-142. <http://dx.doi.org/10.5014/ajot.52.2.133>
- Tiedens, L., & Fragale, A. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology*, *84*, 558-568. <http://dx.doi.org/10.1037/0022-3514.84.3.558>

- Tsay, C. J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, 124(1), 24-33.
- Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, 24(6), 420-424.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004), 686-688.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10(2), 109-120. [http://dx.doi.org/10.1016/0022-1031\(74\)90059-6](http://dx.doi.org/10.1016/0022-1031(74)90059-6)



Figure 1. Example of group seating arrangement.

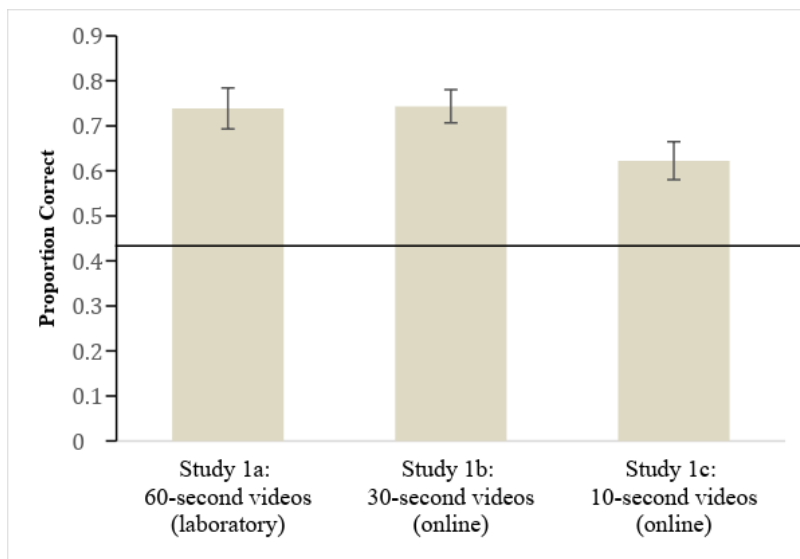


Figure 2. Proportion of correct answers in Studies 1a – 1c as compared to chance (.5). Error bars represent 95% confidence intervals.

Table 1

Summary of Results and Perceiver Demographics for Studies 1, 2, 3, and 4

Study	Sample Source ^a	n ^b	Perceiver Ratings <i>M (SD)</i>	Test statistic ^c	Mean Age	Male	US Citizen	Race					
								African American	Asian	Latino	Multi-racial	Native American	White
1. Paired videos													
1a: 60s videos	Lab	65		$t(64) = 10.45^{***}$	21.9	40%	89%	15%	23%	3%	9%	1%	48%
1b: 30s videos	MTurk	61		$t(60) = 13.14^{***}$	36.1	51%	95%	12%	12%	7%	2%	3%	64%
1c: 10s videos	MTurk	64		$t(63) = 5.81^{***}$	32.6	58%	98%	5%	17%	3%	1%	0%	73%
2. Single videos													
2: 60s videos	Lab	62	4.53 (1.60)	$b = 0.89^{***}$	21.4	43%	87%	18%	21%	12%	7%	0%	43%
3. Manipulating, 10s videos													
3a: Group (G) vs. Individuals (I) vs. Control (C)	MTurk	<u>140</u> 49 G 45 I 46 C		$b = 0.17^{***}$ $b = 0.29^{***}$ $b = 0.25^{***}$	34.0	52%	99%	8%	11%	4%	2%	0%	75%
3b: Overall Group vs. 1 Individual vs. Control	MTurk	<u>141</u> 45 G 47 I 49 C		$b = 0.32^{***}$ $b = 0.09$ $b = 0.18^{**}$	34.4	48%	99%	12%	4%	5%	3%	0%	76%
4. Single 10s clips ^d													
Video & audio	MTurk	89	4.06 (0.80)	$b = 1.57^{***}$	35.4	65%							
Video only	MTurk	82	4.14 (0.88)	$b = 0.98^{***}$	38.3	61%							
Audio only	MTurk	83	3.80 (0.76)	$b = 1.08^{***}$	33.4	54%							
Stimulus materials													
Target Group Members	Lab	48			28.6	63%	83%	4%	21%	8%	8%	0%	58%

^aRN = Research Now

^bThis number does not include participants who were dropped from our sample due to technical difficulties.

^cFor paired studies, we conducted a one-sample *t*-test comparing perceivers' accuracy to chance (.5). For single studies with multiple videos per perceiver, we regressed perceivers' effectiveness ratings onto group performance (effective=1, ineffective=0) with fixed effects and clustered standard errors on the perceiver level. To estimate the coefficients in this table for study 4, we ran the fixed-effects regression for each of the three conditions (video & audio; video only; audio only) separately.

^dIn the pre-registered study 4, we did not collect any demographic variables beyond age and gender since we did not pre-register or plan any subset analysis based on demographic variables. Age and gender were collected because they are the standard demographics recorded by default through the lab account we used.

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

Table 2. Considering only the first decision perceivers make, perceivers accurately judged effective and ineffective groups.

VARIABLES	(1) All Conditions	(2) Video & Audio	(3) Video Only	(4) Audio Only	(5) All Conditions
1 = Effective Group	0.879*** (0.159)	1.206*** (0.264)	0.751* (0.292)	0.650* (0.265)	1.206*** (0.264)
1 = Video Only					0.348 (0.289)
1 = Audio Only					0.063 (0.287)
Effective Group \times Video Only					-0.455 (0.394)
Effective Group \times Audio Only					-0.556 (0.374)
Constant	3.350*** (0.117)	3.220*** (0.205)	3.568*** (0.203)	3.282*** (0.201)	3.220*** (0.205)
Observations	252	88	81	83	252
R-squared	0.110	0.199	0.076	0.070	0.130

Note. OLS model focusing only on the first decision, estimating the perceived effectiveness of the group predicted by an indicator if the group was effective, across and between experimental conditions. Robust standard errors in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3. Using random-effects model across all six decisions of each participant, perceivers accurately judged effective and ineffective groups.

VARIABLES	(1) All Conditions	(2) Video & Audio	(3) Video Only	(4) Audio Only	(5) All Conditions
1 = Effective Group	1.225*** (0.064)	1.581*** (0.116)	0.990*** (0.102)	1.077*** (0.104)	1.579*** (0.115)
1 = Video Only					0.382** (0.138)
1 = Audio Only					-0.015 (0.130)
Effective Group \times Video Only					-0.587*** (0.154)
Effective Group \times Audio Only					-0.503** (0.155)
Constant	3.390*** (0.056)	3.271*** (0.092)	3.654*** (0.104)	3.256*** (0.092)	3.271*** (0.092)
Observations	1,512	529	487	496	1,512
Independent observations	254	89	82	83	254

Note. OLS model focusing only on the first decision, estimating the perceived effectiveness of the group predicted by an indicator if the group was effective, across and between experimental conditions. Robust standard errors in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4. The Reading Mind in the Eyes (RME) test significantly predicts the accuracy of judging the effectiveness of groups across experimental conditions.

VARIABLES	(1) All Conditions	(2) Video & Audio	(3) Video Only	(4) Audio Only	(5) All Conditions
RME Score	0.130*** (0.027)	0.219*** (0.043)	0.073 (0.042)	0.092 (0.048)	0.219*** (0.043)
1 = Video Only					0.059 (0.044)
1 = Audio Only					0.052 (0.046)
RME Score \times Video Only					-0.146* (0.060)
RME Score \times Audio Only					-0.127* (0.064)
Constant	0.568*** (0.019)	0.532*** (0.031)	0.591*** (0.031)	0.584*** (0.033)	0.532*** (0.031)
Observations	1,512	529	487	496	1,512
Independent obs.	254	89	82	83	254

Note. Random-effects model estimating the perceived effectiveness of the group predicted by the score on the Reading Mind in the Eyes (RME) test, across and between experimental conditions. The DV is the extent to which a participant correctly estimated the effectiveness of the group (reversed coded for ineffective groups). Robust standard errors in parentheses are clustered on the perceiver level. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

THIN SLICES OF WORKGROUPS

APPENDIX

DESCRIPTION OF STUDIES A1 TO A5

Below we provide an overview of the thirteen additional studies that we ran but did not include in the main text due to space constraint; we include them here for completeness. All the studies replicated our basic finding that perceivers can accurately judge group effectiveness from short videos. We ran Studies A1 and A2a to ensure that results from online samples were consistent with results from the lab (Study 1a in the main text) before we shifted our studies away from the lab. Study 2 in the paper reported that perceivers could accurately judge single 60-second videos; studies A2b, A2c, and A2d reported below replicate these results with 30-second and 10-second videos. In Studies A3a and A3b, we found perceivers could accurately judge groups with only audio or only video information in separate conditions. However, we did not include Studies A3a-b in the main text because we decided to run a more sophisticated version with 10-second videos instead of 60-second videos, with a control group which included both video and audio, and with measures of social sensitivity and cognitive reflection as potential moderators, which is Study 4 in the main text. In Studies A4a-c, perceivers looked at pictures of two groups and were asked to choose “which group you think is more effective.” Studies A5a-c used pictures of groups but this time perceivers rated group effectiveness individually on a seven-point Likert scale (1 = *very ineffective* to 7 = *very effective*). Perceivers were not consistently accurate across pictures—their accuracy was high for some pictures and low for other pictures of the same group. Given the inconstant results, we chose not to include this study in the main text.

Study A1: Paired 60-second videos

After finding initial support for our hypothesis in a controlled laboratory setting (Study 1a in the main text), we sought to replicate this result using the same video pairs in an online environment, where we had less control over the participant experience and to ensure that results online were comparable to results from the lab. We recruited participants through Research Now, a company that maintains a high-quality panel of online research participants. Participants were eligible if they lived in the U.S., were fluent in English, were over 18 years of age, and had high speed internet access. Before beginning the study, all participants completed a 10-second test to ensure they could stream video and audio. We dropped 16 participants who experienced technical difficulties, leaving 47 perceivers.

Procedure

As in Study 1a, each participant saw six video pairs. Given the online environment, after each video pair participants were asked if they had any technical difficulties viewing the videos. We used six sets (instead of three sets) of six video pairs to represent all possible combinations of the effective and ineffective group videos. After viewing and choosing the most effective group from each pair, participants completed demographic survey items along with additional scales that formed part of a larger data collection effort.

Results

Replicating the results of Study 1a, perceivers chose the effective group in 71.28% of the paired trials, a level of accuracy significantly greater than chance, $t(46) = 7.70, p < .001, 95\% \text{ CI } [65.72, 76.84]$. Again, a majority of perceivers were accurate more than half of the time, with 80.85% of perceivers choosing the correct group in four or more of the six trials.

Studies A2a, A2b, A2c, and A2d: Single 60-second, 30-second, and 10-second videos

Study A2a: Single 60-second videos

Seeking to replicate the effect from Study 2a for single 60-second videos in an online context, we recruited 60 participants from Research Now, using the same criteria reported in Study A1 for this context. We omitted 16 of these participants who experienced technical difficulties during the study, leaving a sample of 44 participants. Replicating the results of Study 2a, perceivers' ratings of group effectiveness converged across the 12 videos ($ICC = .32, p < .001$), with the regression results revealing a significant positive effect of group performance on effectiveness ratings ($b = 0.93, SE = 0.11, p < .001$).

Study A2b: Single 30-second videos

We recruited 62 participants via Research Now, a company that maintains a high-quality panel of online research participants. Participants were eligible if they lived in the U.S., were fluent in English, were over 18 years of age, and had high speed internet access. Before beginning the study, all participants completed a 10-second test to ensure they could stream video and audio. We included the 48 participants who did not experience any technical difficulties. Participants viewed and rated the 12 30-second videos.

Results

After establishing that perceivers exhibited significant agreement in their ratings of group effectiveness ($ICC = .35, p < .001$), we again found that participants systematically rated the effective groups as more effective than the ineffective groups ($b = 0.65, SE = 0.09, p < .001$). The effect of group performance on ratings of group effectiveness was again partially mediated by perceptions of cohesion ($b = 0.83, SE = 0.02, p < .001$; group performance $b = 0.18, SE = 0.05, p < .001$) and trust ($b = 1.00, SE = 0.02, p < .001$; group performance $b = 0.06, SE = 0.03,$

$p < .05$); conflict did not reduce the effect of group performance ($b = 0.67, SE = 0.09, p < .001$) and therefore did not mediate this effect, though it did have an independent negative effect on perceptions of effectiveness ($b = -0.45, SE = 0.12, p < .001$). Regarding the subcomponents of trust, cognitive trust again exhibited a stronger mediating effect than either affective trust or cohesion. This pattern of results for 30-second videos mirrored the pattern from 60-second videos.

Study A2c: Single 10-second videos, independently randomized

We conducted this study of single 10-second videos by recruiting 61 people via MTurk. We excluded one participant who experienced technical difficulties, giving us valid data from 60 participants. In each single-video study reported above, we created a set of twelve videos in a randomized order and presented the same set order to each participant. To ensure that the effect of group performance was not an artifact of a particular sequence of videos, in this study we randomized the sequence of the twelve videos independently for each perceiver.

Results

As in previous studies, group effectiveness ratings were clustered by video ($ICC = .26, p < .001$). Regression analyses revealed that group performance significantly affected perceivers' effectiveness ratings in the predicted direction ($b = 0.61, SE = 0.10, p < .001$). Consistent with judgments of the 60- and 30-second videos, the effect of group performance on ratings of group effectiveness was partially mediated by perceptions of cohesion ($b = 0.75, SE = 0.03, p < .001$; group performance $b = 0.21, SE = 0.04, p < .001$) and trust ($b = 0.94, SE = 0.02, p < .001$; group performance $b = 0.12, SE = 0.03, p < .001$), but not conflict ($b = -0.44, SE = 0.09, p < .001$; group performance $b = 0.58, SE = 0.10, p < .001$). Once again, cognitive trust accounted for a

dominant share (88%) of the mediated indirect effect of trust, with only a small portion of the effect due to affective trust (12%).

Study A2d: Single 10-second videos with short response format

An additional study of single 10-second videos ensured that a specific feature of our experimental paradigm was not influencing the results. In studies 2a-2c, we had perceivers rate the effectiveness of each group on a four-item group effectiveness scale, which itself was embedded within a longer survey of cohesion and related measures. The fact that participants answered a variety of survey questions may have heightened their attention to each subsequent video, potentially activating more deliberate perceptual processes due to a procedural artifact. Since research on collective affect suggests that people are more accurate when they make holistic ratings instead of compiling particular group features, we modified the response format for the current study. Specifically, after each video in study 2d, perceivers answered only the single question “How effective is this group?” on a seven-point Likert scale (1 = *very ineffective*; 7 = *very effective*). Like the previous study, we also randomized the sequence of the 12 video clips independently for each perceiver. With this procedure in place, we recruited 71 participants via Mturk. We excluded nine participants who experienced technical difficulties, leaving 62.

Results

Perceivers’ effectiveness ratings converged across the 12 groups ($ICC = .33, p < .001$). The regression estimate replicated the results from the previous studies ($b = 0.58, SE = 0.08, p < .001$), and it remained unchanged after controlling for the order of videos ($b = 0.58, SE = 0.09, p < .001$). This result increased our confidence that perceivers’ ability to accurately judge group effectiveness based on observing a thin slice of group interaction is robust to longer versus shorter survey response formats.

Studies A3a and A3b: Audio-only and visual-only studies

We explored whether perceivers needed visual, audio, or both channels to accurately judge the effectiveness of decision-making groups.

Study A3a: Paired 60 second audio recordings without video

60 MTurk participants listened to a 60-second audio clip of group interaction (using the same clips as for the 60-second video studies, but omitting the video image), followed by a screen stating that the next group was about to begin (with the statement remaining on the screen for 10 seconds), followed by another 60-second audio clip. Participants were then asked to choose which of the two groups was most effective.

Results

Participants were randomly assigned to one of six sets of pairs of groups, with each pair comprised of one effective and one ineffective group presented in random order. Perceivers chose the effective group in 60.83% of the paired trials, an accuracy level that is significantly greater than chance, $t(59) = 4.18, p < .001, 95\% \text{ CI } [55.65, 66.02]$.

Study A3b: Paired 60-second video recordings without audio

67 MTurk participants were randomly assigned to one of six video sets and watched 6 randomized video pairs. All videos were 60 seconds in length and contained no audio.

Results

Perceivers made accurate choices in 64.68% of the paired trials, a percentage significantly greater than chance $t(66) = 5.88, p < .001, 95\% \text{ CI } [59.70, 69.66]$.

Studies A4a, A4b and A4c: Perceivers rate paired group pictures

Pictures represent a relatively simple set of static cues from a single point in time, containing substantially less information than even a short video. By necessity, pictures only represent the visual channel. Even with such limited information, researchers found that perceivers could determine status hierarchy and dominance in groups of faces using pictures (Haberman & Whitney, 2009; Phillips, Slepian, & Hughes, 2018). Stillman, Gilovich, and Fujita (2014) found that when perceivers viewed pictures of company boards of directors, their ratings of the boards' cohesiveness were associated with company performance. Given this evidence, we tested whether perceivers could accurately judge workgroup effectiveness from pictures.

Procedure

We used still shots from the beginning, middle, and end of the 60-second videos we used in the previous studies. Using one picture from each of the twelve groups, we paired pictures together in a round robin format so that each pair included one ineffective group and one effective group. This resulted in 36 effective-ineffective pairs of pictures. We then segmented these 36 pairs into six different sets of pairings. Each participant was randomly assigned to view one of the six sets of paired pictures. We then asked participants to view each pair of pictures for as long as they wanted, and after each pair, to choose the group they judged to be most effective.

Results

For the first set of pictures, Study A4a, taken from the beginning of the 60-second videos, 62 MTurk perceivers selected the effective group in 51.61% of the paired trials, which was not significantly different from chance (50% accuracy, $t(61) = 0.69, p > .250, 95\% \text{ CI } [46.91, 56.32]$). For the second set of pictures, Study A4b, taken from the middle of the 60-second videos, 61 perceivers recruited from MTurk chose the effective group in 48.91% of the paired trials, which was not significantly different from chance, $t(60) = -0.55, p > .250, 95\% \text{ CI } [44.94, 52.87]$. For the third set of pictures, Study A4c, taken from the end of the 60-second videos, 60

participants from MTurk were significantly more accurate than chance, choosing the effective group in 64.72% of the paired trials ($t(59) = 6.56, p < .001, 95\% \text{ CI } [60.23, 69.21]$).[4]

Discussion

Perceivers reliably judged performance when viewing 60-, 30-, and 10-second videos, yet their judgments were less reliable when viewing still pictures. Regardless of whether we showed them paired pictures or single pictures, perceivers were no better than chance with two sets of pictures (sets 1 and 2), though they were better than chance with a third set of pictures. One reason our findings were less consistent than those of Stillman et al. (2014) is that they used posed pictures of boards of directors and a criterion of past company performance. Our pictures, captured in the midst of performing a decision task and prior to receiving any performance feedback, may have exhibited more mixed cues.

Studies A5a, A5b, and A5c: Single pictures

In these three studies, we asked participants to view a single picture of a group and then rate the effectiveness of that group before moving to the next picture. We conducted separate tests of the three sets of pictures used in Studies A4a-4c.

For Study A5a we recruited 62 participants on MTurk. Participants evaluated the 12 photographs used in Study A4a, presented in random order to each participant. After viewing each picture for as long as they wanted, perceivers answered the single question “How effective is this group?” using a seven-point Likert scale (1= *very ineffective* to 7 = *very effective*). Across the 12 photographs, participants’ ratings of effectiveness converged significantly ($ICC = .18, p < .001$). However, using the same analytic approach as Study 2, and consistent with the results of Study A4a, we found that the effect of group performance on effectiveness ratings was not

significant ($b = 0.14$, $SE = 0.09$, $p = .122$). Continuing with the same procedure, in Study A5b we substituted the second set of pictures that we tested in Study A4b above. We recruited a new set of 62 participants from MTurk, whose ratings again converged ($ICC = .13$, $p < .001$).

Consistent with the null result in Study A4b, the effect of group performance on perceivers' effectiveness ratings was not significant ($b = 0.05$, $SE = 0.08$, $p > .250$).

Study A5c tested the set of pictures taken from the endpoint of the 60-second video clip of group interaction, which elicited a significant difference in perceivers' judgments in Study A4c. A new group of 62 people from MTurk participated in the study, and their ratings converged across photographs ($ICC = .22$, $p < .001$). Furthermore, the regression model revealed a significant positive effect of group performance on perceivers' effectiveness ratings ($b = 0.61$, $SE = 0.09$, $p < .001$), consistent with the significant effect in Study A4c.

Discussion of Picture Results

Perceivers reliably judged performance when viewing 60-, 30-, and 10-second videos, yet their judgments were far less reliable when viewing still pictures. Regardless of whether we showed them paired pictures or single pictures, perceivers were no better than chance with two sets of pictures (A4a/A5a, A5b/A5b), though they were better than chance with a third set of pictures (A4c/A5c). One difference between videos and pictures is that the amount of time perceivers viewed videos was set by the length of the video (e.g., 10 seconds), whereas perceivers in our picture studies could choose how long they viewed each picture. One possibility is that perceivers were less accurate judges of pictures simply because they viewed them for shorter durations than videos. To examine this, we captured the amount of time in seconds perceivers spent on each page, which contained both a picture of a group and the rating

question ($M = 7.95$ seconds, $SD = 18.26$). We clustered judgments by two-second increments and re-ran the analyses on each subsample, but no clear pattern emerged to indicate that perceiver accuracy was a function of time spent viewing the picture. For example, across the three studies, perceivers exhibited the same accuracy (or inaccuracy) when viewing pictures for 0-2 seconds as they tended to exhibit for longer increments of time.

^a RN = Research Now

^b This number does not include participants who were dropped from our sample due to technical difficulties.

^c For paired studies, we conducted a one-sample t-test comparing perceivers' accuracy to chance (.5). For single studies, we regressed perceivers' effectiveness ratings onto group

Appendix: Table A1

Summary of Results and Perceiver Demographics Across Studies Not Included in Main Text

Study	Sample Source ^a	n ^b	Perceiver Ratings <i>M (SD)</i>	Test statistic ^c	Mean Age	Male	US Citizen	Race ^d						
								African American	Asian	Latino	Multi-racial	Native American	White	
A1. Paired videos														
60s videos	RN	47		t(46) = 7.70***	50.3	32%	100%	4%	4%	4%	2%	0%	85%	
A2. Single videos														
A2a: 60s videos	RN	44	4.75 (1.55)	b = 0.93***	47.9	29%	97%	7%	7%	0%	0%	0%	86%	
A2b: 30s videos	RN	48	4.68 (1.43)	b = 0.65***	43.4	27%	100%	0%	0%	2%	0%	2%	96%	
A2c: 10s videos	MTurk	60	4.94 (1.41)	b = 0.61***	36.0	53%	98%	12%	0%	7%	5%	2%	75%	
A2d: 10s videos	MTurk	62	4.77 (1.49)	b = 0.58***	35.5	52%	95%	13%	7%	3%	2%	0%	76%	
A3. Paired videos														
A3a: 60s audio	MTurk	60		t(59) = 4.18***	33.6	53%	100%	10%	10%	10%	3%	2%	65%	
A3b: 60s video	MTurk	67		t(66) = 5.88***	31.8	54%	96%	4%	9%	4%	4%	3%	75%	
A4. Paired pictures														
A4a: pictures	MTurk	62		t(61) = 0.69	32.1	53%	98%	3%	11%	3%	2%	0%	81%	
A4b: pictures	MTurk	61		t(60) = -0.55	32.1	47%	98%	5%	13%	8%	5%	0%	67%	
A4c: pictures	MTurk	60		t(59) = 6.56***	31.1	67%	100%	7%	8%	0%	0%	0%	85%	
A5. Single pictures														
A5a: pictures	MTurk	62	4.77 (1.46)	b = 0.14	36.5	42%	98%	13%	3%	5%	5%	0%	74%	
A5b: pictures	MTurk	62	4.85 (1.27)	b = 0.05	33.5	56%	98%	6%	6%	8%	3%	0%	76%	
A5c: pictures	MTurk	62	4.71 (1.42)	b = 0.61***	33.1	58%	100%	5%	3%	5%	5%	0%	82%	

performance (effective=1, ineffective=0) with fixed effects and clustered standard errors for perceivers.

^d In Study 3b, 2% of perceivers identified as "other" regarding race. In Study 5a, 1 person identified as "other."

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

Table A2. Using fixed-effects model across all six decisions of each participant, perceivers accurately judged effective and ineffective groups.

VARIABLES	(1) All Conditions	(2) Video & Audio	(3) Video Only	(4) Audio Only
1 = Effective Group	1.219*** (0.064)	1.572*** (0.116)	0.981*** (0.103)	1.076*** (0.104)
Constant	3.395*** (0.032)	3.276*** (0.058)	3.664*** (0.051)	3.256*** (0.052)
Observations	1,512	529	487	496
Independent observations	254	89	82	83
R-squared	0.235	0.316	0.185	0.200

Note. Fixed-effects model across all six decisions estimating the perceived effectiveness of the group predicted by an indicator if the group was effective, across and between experimental conditions. Robust standard errors in parentheses are clustered on the perceiver level. *** p<0.001, ** p<0.01, * p<0.05

Table A3. A high (reflection) score on the Cognitive Reflection Test (CRT) is not significantly associated with accurately judging the effectiveness of groups.

VARIABLES	(1) All Conditions	(2) Video & Audio	(3) Video Only	(4) Audio Only	(5) All Conditions
Intuitive CRT Score	0.008 (0.004)	0.009 (0.007)	0.003 (0.007)	0.005 (0.006)	0.009 (0.007)
1 = Video Only					-0.030 (0.024)
1 = Audio Only					-0.027 (0.020)
Intuitive CRT Score X Video					-0.006 (0.010)
Intuitive CRT Score X Audio					-0.004 (0.009)
Constant	0.646*** (0.008)	0.667*** (0.017)	0.637*** (0.018)	0.640*** (0.011)	0.667*** (0.017)
Observations	1,512	529	487	496	1,512
Independent observations	254	89	82	83	254

Note. Random-effects model estimating the perceived effectiveness of the group predicted by the score on the Cognitive Reflection Test (CRT), across and between experimental conditions. The DV is the extent to which a participant correctly estimated the effectiveness of the group (reversed coded for ineffective groups). Robust standard errors in parentheses are clustered on the perceiver level. *** p<0.001, ** p<0.01, * p<0.05

References

- Alt, N. P., Goodale, B., Lick, D. J., & Johnson, K. L. (2017). Threat in the Company of Men: Ensemble Perception and Threat Evaluations of Groups Varying in Sex Ratio. *Social Psychological and Personality Science*. <https://doi.org/10.1177/1948550617731498>
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392-398.
<http://dx.doi.org/10.1111/j.1467-9280.2008.02098.x>
- Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology*, *16*(1), 4-13.
http://dx.doi.org/10.1207/s15327663jcp1601_2
- Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: A clue to malpractice history. *Surgery*, *132*(1), 5-9.
<http://dx.doi.org/10.1067/msy.2002.124733>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256.
<http://dx.doi.org/10.1037/0033-2909.111.2.256>
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431-441. <http://dx.doi.org/10.1037/0022-3514.64.3.431>
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251.

- Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47(4), 644-675.
- Bartel, C. A., & Saavedra, R. (2000). The collective construction of work group moods. *Administrative Science Quarterly*, 45(2), 197-231.
- Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: a meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88(6), 989.
- Bendersky, C. and Hays, N. (2012). Status conflict in groups. *Organization Science*, 23(2), 323-340.
- Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J.A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement*. (pp. 3-20). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, 71(1), 110-129. <http://dx.doi.org/10.1037/0022-3514.71.1.110>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimate? *Quarterly Journal of Economics*, 119, 249-275. <http://dx.doi.org/10.1162/003355304772839588>
- Bodenheimer, T., & Laing, B. Y. (2007). The teamlet model of primary care. *The Annals of Family Medicine*, 5(5), 457-461.
- Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly*, 17(3), 288-307.

- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054-1072.
<http://dx.doi.org/10.1016/j.jrp.2007.01.004>
- Carrere, S., & Gottman, J. M. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process, 38*, 293–301.
<http://dx.doi.org/10.1111/j.1545-5300.1999.00293.x>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.
<http://dx.doi.org/10.1037/0022-3514.76.6.893>
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology, 104*(1), 103-125.
<http://dx.doi.org/10.1037/a0030398>
- Clingingsmith, D., & Shane, S. (2017). Training Aspiring Entrepreneurs to Pitch Experienced Investors: Evidence from a Field Experiment in the United States. *Management Science, Articles in Advance*, 1-16.
- Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology, 92*(3), 802-811. <http://dx.doi.org/10.1037/0021-9010.92.3.802>
- Dane, E., Rockmann, K. W., & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes, 119*(2), 187-194.

- De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of applied Psychology*, 88(4), 741.
- De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, 101(8), 1134.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111(2), 203-243. <http://dx.doi.org/10.1037/0033-2909.111.2.203>
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383. <http://dx.doi.org/10.2307/2666999>
- Engel D, Woolley AW, Jing LX, Chabris CF, Malone TW (2014) Reading the Mind in the Eyes or Reading between the Lines? Theory of Mind Predicts Collective Intelligence Equally Well Online and Face-To-Face. *PLOS ONE* 9(12): e115212. <https://doi.org/10.1371/journal.pone.0115212>
- Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *The Leadership Quarterly*, 2(4), 245-287.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday/Anchor Books.

- Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology, and health. *Journal of Personality and Social Psychology*, *63*, 221–233. <http://dx.doi.org/10.1037/0022-3514.63.2.221>
- Gottman, J. M., Markman, H., & Notarius, C. I. (1977). The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior. *Journal of Marriage and Family*, *39*(3), 461-477. <http://dx.doi.org/10.2307/350902>
- Gottman, J. M., & Notarius, C. I. (2000). Decade review: Observing marital interaction. *Journal of Marriage and Family*, *62*(4), 927-947. <http://dx.doi.org/10.1111/j.1741-3737.2000.00927.x>
- Grahe, J. E., & Bernieri, F. J. (1999). The importance of nonverbal cues in judging rapport. *Journal of Nonverbal Behavior*, *23*(4), 253-269. <http://dx.doi.org/10.1023/A:1021698725361>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454.
- Gust (2016). Global Accelerator Report. Retrieved from: http://gust.com/accelerator_reports/2016/global/.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718-734. <http://dx.doi.org/10.1037/a0013899>
- Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, *30*(2), 269-287.
- Hall, J. A., & Andrzejewski, S. A. (2008). Who draws accurate first impressions? Personal correlates of sensitivity to nonverbal cues. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions* (pp. 87-105). New York, NY: Guilford Publications.

- Harris, M. J., & Garris, C. P. (2008). You never get a second chance to make a first impression: Behavioral consequences of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions*, (pp. 147-168). New York, NY: Guilford Publications.
- Hecht, M. A., & LaFrance, M. (1995). How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *Journal of Applied Social Psychology*, 25(23), 2086-2098. <http://dx.doi.org/10.1111/j.1559-1816.1995.tb02389.x>
- Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational behavior and human decision processes*, 67(3), 247-257.
- Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychological bulletin*, 125(5), 576.
- Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 40(2), 256-282.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, 44(4), 741-763. <http://dx.doi.org/10.2307/2667054>
- Kerr, R., Garvin, J., Heaton, N., & Boyle, E. (2006). Emotional intelligence and leadership effectiveness. *Leadership & Organization Development Journal*, 27(4), 265-279.
- Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science*, 20(1), 99-106. <http://dx.doi.org/10.1111/j.1467-9280.2008.02251.x>

- Lamer, S. A., Sweeny, T. D., Dyer, M. L., & Weisbuch, M. (2018). Rapid visual perception of interracial crowds: Racial category learning from emotional segregation. *Journal of Experimental Psychology: General*, *147*(5), 683-701.
- Littlepage, G., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. *Organizational Behavior and Human Decision Processes*, *69*(2), 133-147.
<http://dx.doi.org/10.1006/obhd.1997.2677>
- Ma-Kellams, C., & Lerner, J. (2016). Trust your gut or think carefully? Examining whether an intuitive, versus a systematic, mode of thought produces greater empathic accuracy. *Journal of personality and social psychology*, *111*(5), 674.
- Marks, M., Mathieu, J., & Zaccaro, S. 2001. A temporally-based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356-376.
- Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, *34*(3), 410-476.
- McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59.
- Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, *94*(2): 535-546.
<http://dx.doi.org/10.1037/a0013773>
- Neuberg, S., Kenrick, D., & Schaller, M. (2010). Evolutionary social psychology. In Fiske S. T., Gilbert D. T., Lindzey G. (Eds.), *Handbook of social psychology* (5th ed., pp.761–796). New York, NY: Wiley. <http://dx.doi.org/10.1002/9780470561119.socpsy002021>

- Peterson, R.S. (2001). *PB Technologies*. Kellogg Dispute Resolution Center Teaching.
- Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology*, *114*(5), 766-785. <http://dx.doi.org/10.1037/pspi0000120>
- Phillips, L. T., Weisbuch, M., & Ambady, N. (2014). People perception: Social vision of groups and consequences for organizing and interacting. *Research in Organizational Behavior*, *34*, 101-127. <http://dx.doi.org/10.1016/j.riob.2014.10.001>
- Pickett, C. L., Gardner, W. L., & Knowles, M. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin*, *30*(9), 1095-1107. DOI: 10.1177/0146167203262085
- Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science*, *20*(1), 22-26. <http://dx.doi.org/10.1111/j.1467-9280.2008.02248.x>
- Romero, D. M., Swaab, R. I., Uzzi, B., & Galinsky, A. D. (2015). Mimicry is presidential: Linguistic Style Matching in Presidential Debates and improved polling numbers. *Personality and Social Psychology Bulletin*, 1-9. DOI: 10.1177/0146167215591168
- Rosete, D., & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadership & Organization Development Journal*, *26*(5), 388-399.
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, *50*(3), 540-547.

- Sanchez-Burks, J., Bartel, C. A., Rees, L., & Huy, Q. (2015). Assessing collective affect recognition via the Emotional Aperture Measure. *Cognition and Emotion, 30*(1), 117-133.
- Seashore, Stanley E. (1954) *Group Cohesiveness in the Industrial Work Group*. Ann Arbor, MI: University of Michigan Press.
- Shunk, R., Dulay, M., Chou, C. L., Janson, S., & O'Brien, B. C. (2014). Huddle-coaching: a dynamic intervention for trainees and staff to support team-based care. *Academic Medicine, 89*(2), 244-250.
- Simmel, G. (1950). *The Sociology of Georg Simmel*. Glencoe, IL: Free Press.
- Stasser, G., & Titus, W. (2003). Hidden profiles: A brief history. *Psychological Inquiry, 14*(3), 304–313. http://dx.doi.org/10.1207/S15327965PLI1403&4_21
- Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management, 32*(1), 29-55.
- Stillman, P. E., Gilovich, T., & Fujita, K. (2014). Predicting group outcomes from brief exposures. *Social Cognition, 32*(1), 71-82.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349-354.
- Tickle-Degnen, L. (1998). Working well with others: The prediction of students' clinical performance. *American Journal of Occupational Therapy, 52*(2), 133-142. <http://dx.doi.org/10.5014/ajot.52.2.133>
- Tiedens, L., & Fragale, A. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology, 84*, 558-568. <http://dx.doi.org/10.1037/0022-3514.84.3.558>

- Tsay, C. J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, 124(1), 24-33.
- Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, 24(6), 420-424.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004), 686-688.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10(2), 109-120. [http://dx.doi.org/10.1016/0022-1031\(74\)90059-6](http://dx.doi.org/10.1016/0022-1031(74)90059-6)