Social Cognition, Vol. 32, No. 1, 2014, pp. 71–82

PREDICTING GROUP OUTCOMES FROM BRIEF EXPOSURES

Paul E. Stillman
The Ohio State University

Thomas Gilovich
Cornell University

Kentaro Fujita
The Ohio State University

Research on thin slice judgment, or people's ability to make accurate judgments about a target based on very brief exposure, has largely focused on the detection of individual-level traits. The present studies extend this work to group-level characteristics, such as teamwork and cohesiveness, and demonstrate that these inferences can predict behavioral performance outcomes. In Study 1, judgments based on 10-s performance videos of rock bands predicted view-counts of the full performance videos posted on the Internet. In Study 2, judgments of Ultimate Frisbee teams based on 10-s warm-up videos predicted the teams' winning percentages. In Study 3, thin slice judgments based on pictures of boards of directors predicted the companies' success. The authors conclude that judgments of emergent group-level characteristics based on very brief exposures can predict important real-world outcomes.

Sports fans frequently say they "know a good team when they see one." This phrase suggests that at some level people believe that simply by observing a team (or, more generally, a group), they can quickly grasp something of its dynamics and likely future performance. In this article, we examine this assertion empirically. Specifically, we ask whether judgments of group-level characteristics of teamwork and cohesiveness based on brief observations of a group can reliably predict real-world performance outcomes.

This research was supported by a National Science Foundation graduate research fellowship (DGE-0822215) to the first author. We thank Christina Carino, Jun Fukukura, Erik Helzer, Chris Jones, David Pizarro, Vivian Zayas, and our anonymous reviewers for helpful comments on earlier drafts.

Correspondence concerning this article should be addressed to Paul Stillman, Lazenby 107, 1827 Neil Avenue, Columbus OH 43210; E-mail: paul.e.stillman@gmail.com.

THIN SLICE JUDGMENT

People can make reliable and accurate judgments of others based on very limited samples of behavior, a phenomenon known as "thin slice judgment" (e.g., Ambady & Rosenthal, 1993; Ambady & Skowronski, 2008; Todorov, Mandisodza, Goren, & Hall, 2005). In an early investigation, participants' ratings of teachers based only on short clips of classroom performance correlated highly with students' end-of-semester evaluations of the teachers (Ambady & Rosenthal, 1993). Recently, people have been shown to be capable of judging sexual orientation when briefly presented with photographs or short videos of target individuals (Ambady, Hallahan, & Conner, 1999; Johnson, Gill, Reichman, & Tassinary, 2007), even when stimuli are exposed for as little as 50 ms (Rule & Ambady, 2008a; Tabak & Zayas, 2012). People's snap judgments of the competency of political candidates also accurately predict election outcomes (Ballew & Todorov, 2007; Todorov et al., 2005). People thus appear to be highly adept at detecting important characteristics of others on the basis of minimal exposure.

Nearly all of the research on thin slice judgments has focused on the individual as the target of judgment, involving the detection of individual-level characteristics such as sexual orientation (Ambady et al., 1999), personality traits (Carney, Colvin, & Hall, 2007), and psychological states (e.g., distress; Mason, Sbarra, & Mehl, 2010). Less attention has been focused on the perception of groups and emergent group-level characteristics such as teamwork and cohesiveness—characteristics that are not evident in the actions of a single individual, but emerge in the interaction of multiple individuals. To our knowledge, research has yet to examine whether people can reliably and accurately judge such characteristics.

One group-level characteristic that has received considerable attention is cohesiveness. Festinger (1950) described cohesiveness as "the total field of forces that act on members to remain in the group" (p. 274). The construct has been refined over the years (see Friedkin, 2004), and here we use Carron's (1982) definition of cohesiveness as "a dynamic process which is reflected in the tendency for a group to remain united in the pursuit of its goals and objectives" (p. 124).

Of principal interest to small-group researchers is the link between cohesiveness and group performance. Intuitively and anecdotally, more cohesive groups perform better. Although several moderators of the strength of the cohesiveness-performance relationship have been identified (e.g., group size; Mullen & Copper, 1994), group cohesiveness and group performance are positively related across a wide variety of groups (Beal, Cohen, Burke, & McLendon, 2003; Chiocchio & Essiembre, 2009; Evans & Dion, 1991; Mullen & Copper, 1994).

Initial support for our hypothesis that people's snap judgments of group cohesiveness can predict groups' performance outcomes comes from research examining thin slice judgments of dyadic interactions. Bernieri and his colleagues (e.g., Aloni & Bernieri, 2004; Bernieri, Gillis, Davis, & Grahe, 1996; Grahe & Bernieri, 2002) have shown that participants who view brief video clips of a dyad's discussion are able to assess their rapport accurately, as measured by the discussants' subsequent assessments. Following the interaction, the dyad members rated the rapport they felt they had with their partner. The researchers then took a 50-s clip of the interaction and had participants rate how much the individuals liked/enjoyed the interaction. They found that the discussants' self-assessments of rapport

were correlated with the snap judgments of liking and enjoyment (Bernieri et al., 1996). It thus appears that naïve individuals can accurately judge dyadic rapport given a brief observation of the interaction.

Although it is tempting to extrapolate from research on dyads and assume that judgments about groups are similarly accurate, there are several reasons to be cautious about doing so. First, it is unclear whether participants' judgments reflect assessments of the emergent qualities of the dyad or simply judgments of the individual dyad members. Analyses of the cues that participants use to make their judgments of rapport suggest that the most diagnostic cues are those that can be detected by focusing on a single individual within the dyad (e.g., Bernieri et al., 1996; see also Tickle-Degnen & Rosenthal, 1990). Thus, people may be judging the two individuals and not the relationship between them.

A second reason to be cautious is that larger group dynamics may involve emergent phenomena that do not arise in dyads. In a dyad, there is a single relationship between the two individuals. In a group, there are many: the relationship between each member and one another, between a member and various factions or subgroups, and between a member and the group as an entity (Moreland, 2010; Moreland, Hogg, & Hains, 1994). The number and complexity of relationships increases dramatically with each additional member, and it is from the dynamics of these multilayered relationships that group-level characteristics such as cohesiveness emerge. Because judgments of cohesiveness may require evaluation of relationships at multiple levels (member-member, member-subgroup, member-group), past findings on dyadic judgment may not apply.

A final difference between the current work and past research in the thin slice literature on dyads is in how accuracy is assessed. Past research has used self-report among dyad members as the standard of accuracy. Researchers correlated judges' ratings with the subjective assessments of rapport on the part of the two individuals in the target dyad. However, in most studies, there is no objective assessment of the accuracy of these dyadic thin slice judgments (cf. Rule & Ambady, 2008b). Although predicting targets' self-assessments of rapport is useful and informative, it is not the same as predicting actual objective outcomes.

These differences between dyads and groups suggest that it may be premature to extend the results of thin slice research on dyads to groups. Further research is needed to ascertain whether people can make predictive inferences about groups based on short exposure. The goal of the present studies is to demonstrate the predictive utility of thin slice judgments of groups. We investigate groups that have spent considerable time together developing teamwork and cohesiveness (rock bands, Frisbee teams, and boards of directors). We examine whether participants' snap judgments of the cohesiveness of these groups predict objective group performance.

STUDY 1: ROCK BANDS

Rock band performances require careful coordination among band members. If band members are unable to communicate effectively and play off one another, the music they produce will generally be deficient even when the individual band members have superior talent. We predicted that participants will be able to detect

a band's cohesiveness from thin slices of performance, and that these assessments can be used to predict the success of the band over and above participants' personal opinions about the quality of the band's music.

METHOD

Participants. Seventy-seven undergraduates at The Ohio State University (OSU) served as judges for course credit.

Materials and Procedure. We used videos of 10 bands from www.flashrock.com, a website that provides exposure for unsigned bands through posted videos. This allowed us to obtain a sample of bands with which participants were unfamiliar. In each video, lasting 15–18 min, the band played three songs, with video angle and lighting held constant across bands. Participants watched three 10-s clips of each band (30 clips total). The clips were the first, middle, and last 10 s of the second song that each band played. Clips were randomly presented. After watching each clip, participants rated on 7-point scales how much the band members trusted each other, how much conflict there was among them (reverse-coded), how well they worked together, how well they communicated, and how "in synch" they were with one another. To control for participants' personal opinions about the quality of the music, each participant also rated the quality of the band on a 7-point scale. Each participant made a total of 180 ratings (six ratings for each of the three clips of each of the 10 bands).

View-Counts. Flashrock.com makes all full-length videos of each band publicly available on YouTube.com, a popular website used for sharing videos. YouTube.com records and presents the number of times a video has been watched. We retrieved these view-counts at the conclusion of data collection (April 30, 2009) as the measure of band success, reasoning that more popular and successful bands will have higher view-counts.

RESULTS AND DISCUSSION

Ratings of the five group-dynamics items were highly correlated ($\alpha = .87$) with high intraclass correlation between judges ($ICC = .97$). We averaged these five measures and aggregated this composite measure across the three clips of each band to create an index of cohesiveness for each band for each participant. Since the individual ratings are nested within judge for each band, we analyzed these data using a mixed-level model with band ratings as the Level 1 variable and judge as the Level 2 variable. This analysis statistically controls for any systematic individual-level variability in judges' ratings to produce more stable estimates of the ratings of each band (Fitzmaurice, Laird, & Ware, 2004). As predicted, cohesiveness ratings significantly predicted YouTube view-counts, $\gamma = 306.34$, $F(1, 692) = 4.89$, $p < .001$. Specifically, judges' ratings of the bands' cohesiveness were positively associated with the popularity of the bands' videos. For ease of interpretation, Figure 1 depicts view-counts as a function of the mean ratings of each band aggregated across judges. Thin slice judgments of a group's cohesiveness thus appear to predict outcomes such as the popularity of rock bands.

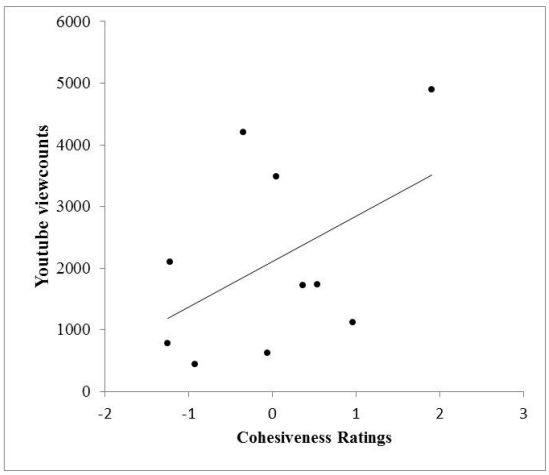

FIGURE 1. Relationship between the judgments of cohesiveness (normalized) and YouTube view-counts. The x-axis represents the mean cohesiveness rating for each band aggregated across judges.

One alternative explanation for this result is that participants' ratings of the bands' cohesiveness were driven by their evaluation of the bands' music. However, cohesiveness still predicted YouTube view-counts after controlling for participants' ratings of band quality, $\gamma = 170.23$, $F(1, 691) = 2.31$, $p = .02$, suggesting that participants' cohesiveness ratings were not driven solely by their personal impressions of the quality of the bands' music.

One potential problem with these data is the relatively small number of bands. Although our *N* is comparable to early work in the thin slice domain (e.g., Ambady & Rosenthal, 1993), having a small sample increases the influence of outliers. To address this possibility, we averaged the mean ratings of each band across judges and conducted an outlier analysis using Cook's distance (Cook & Weisber, 1982). The maximum Cook's distance for our data was .89, under the standard convention of 1.0 suggested by Cook and Weisber (1982).[1] This suggests that our findings are not the product of the undue influence of outliers.

YouTube view-counts, however, are an indirect measure of success. Some videos might be popular because they depict embarrassingly poor performances. To investigate behavioral outcomes more directly, as well as to provide convergent validity for our claims, we turned in Study 2 to the domain of sports.

STUDY 2: ULTIMATE FRISBEE TEAMS

Babe Ruth once said, "The way a team plays as a whole determines its success. You may have the greatest bunch of individual stars in the world, but if they don't

1. Bollen and Jackman (1990) suggest a cutoff of .4, which would exclude one data point. Analyses excluding this data point reveal a pattern that was consistent with our main findings, but did not reach conventional levels of significance ($p = .15$). Note, therefore, that in an independent sample of Cornell University undergraduates ($N = 18$), we replicated the finding that cohesiveness ratings predict view-counts, $\gamma = 668.01$, $F(1, 158) = 5.51$, $p = .02$, even after controlling for quality ratings, $\gamma = 237.87$, $F(1, 157) = 2.66$, $p = .10$. In this sample, all Cook's *d*'s were below the .4 threshold.

play together, the club won't be worth a dime" (Holden Leadership Center, 2009). This is particularly apt for the sport of Ultimate Frisbee, a game played with a Frisbee on a field similar to a football gridiron in which the objective is to score as often as possible by being in possession of the Frisbee inside the end zone. A player is not allowed to move with the Frisbee, but must instead coordinate with team members to pass the Frisbee and advance up the field. Thus, teams that communicate more effectively and anticipate one another's individual movements are more likely to succeed. Team success, moreover, can be assessed objectively by examining a team's win-loss record. We predicted that snap judgments of team cohesiveness would accurately predict team performance.

METHOD

Participants. Twenty-six Cornell University undergraduates were paid $5 to serve as judges.

Materials and Procedure. We recorded 10 Ultimate Frisbee teams participating in a mid-level tournament in Autumn 2008. The videos captured each team performing a warm-up drill and coming together for a huddle. The teams' win/loss record in the tournament represented the metric of group performance. Note that, unlike Study 1, in which judgment of group cohesiveness and behavioral performance were based on the very same behavior, assessments of group cohesiveness involved behavior (warm-up and huddle) that was distinct from the behavior constituting group performance itself (i.e., actual game outcomes).

We created three 10-s clips for each team: two involving warm-up drills and one involving the team huddle. These clips depicted only interactions between team members, and no interactions with opponents. The warm-up clips involved 10 s of continual motion without poor throws or dropped passes. The huddle clips involved the 10 s leading up to the convening of the huddle, ending 1 s after the team had convened. Clips were randomly presented. After each clip, participants rated on a 7-point scale how much the team members trusted each other, how much conflict there was on the team (reverse-coded), how well the players worked together, how well they communicated, and how actively engaged they were in supporting their fellow team members.

RESULTS AND DISCUSSION

Ratings of the five group-dynamics items were highly correlated ($\alpha = .94$) , so we averaged across items to create an aggregate measure of perceived cohesiveness and then averaged this aggregate measure across the three clips for each team to get an index of cohesiveness for each team for each judge (*ICC* = .95). We again analyzed the data using a mixed-level model with team rating as the Level 1 variable and judge as the Level 2 variable. As predicted, cohesiveness significantly predicted win percentage, $\gamma = .15$, $F(1, 238) = 37.96$, $p < .001$. Figure 2 depicts win percentage as a function of cohesiveness ratings aggregated across judges. Thus, thin slice judgments of group-level characteristics such as cohesiveness can predict actual group performance.

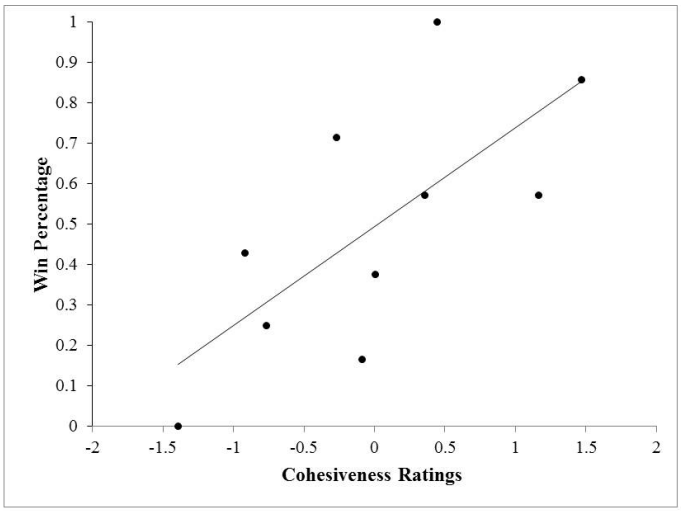

FIGURE 2. Relationship between the judgments of cohesiveness (normalized) and win percentage for a sample of Ultimate Frisbee teams. The x-axis represents the mean cohesiveness rating for each team aggregated across judges.

To address the potential concern that these ratings merely reflect the perceived athleticism of individual team members rather than group cohesiveness, we showed the video clips to 49 additional participants (8 undergraduates from Cornell and 41 from OSU) and asked them to rate each team on athleticism. We averaged their judgments, yielding a single athleticism index for each team. The relationship between cohesiveness and win percentage remained significant after statistically controlling for apparent athleticism, $\gamma = .04$, $F(1, 237) = 2.22$, $p = .03$). Thin slice judgments of group cohesiveness thus predicted wins independently of athleticism.

We conducted similar analyses using only ratings based on the clips of the team coming in for the huddle. If ratings based on these clips—which contain no information about Frisbee abilities and instead convey only social interactions—predict their win percentage, it would serve as a stronger test of our hypothesis that participants are detecting cohesiveness. Indeed, cohesiveness predicted performance when using only ratings from the huddle clips, $\gamma = .05$, $F(1, 238) = 10.64$, $p = .001$, even after controlling for athleticism, $\gamma = .02$, $F(1, 237) = 2.26$, $p = .03$.

Again, due to our relatively small sample of stimulus materials, it is possible that our results are the product of undue influence of outliers. We conducted an outlier analysis using Cook's *d* on the mean ratings aggregated across judges. All Cook's *d*s fell below conventional thresholds (*d*s < .25). Although this finding suggests that our results are not the product of outliers, Study 3 sought to replicate Studies 1 and 2 in another domain with a larger sample size and using a different stimulus medium (photographs rather than videos).

STUDY 3: FORTUNE 500 COMPANIES

A company in which key actors do not work in unison is likely to experience infighting, make poor decisions, and lose productivity (Tjosvold, 1984). The cohesiveness of the board of directors may play a particularly influential role in the success of a company, because the directors serve as the governing agency of the company and must make important strategic, financial, and managerial decisions, often under time pressure (Daum & Neff, 2010). We predicted that snap judgments of board-member cohesiveness would be related to a company's success.

METHOD

Participants. Sixty OSU undergraduates served as judges for course credit.

Materials and Procedure. We consulted the 2011 "Fortune 500" list, *Fortune* magazine's annual rankings of the top 500 American companies by revenue. For each company, we searched for publicly available photographs of the board of directors. We found 34 suitable photographs, which we then edited to remove identifying information (stimuli available upon request). Participants viewed each photo and rated them on a 7-point scale how much the board members trusted each other, how much conflict there was between members (reverse-coded), how well the board members worked together, how well they communicated, and how "in synch" they were with their fellow board members.

Following previous work (Rule & Ambady, 2008; Wong, Ormiston, & Haselhuhn, 2011), we obtained multiple measures of company success from CNNMoney.com: revenue, profits, and return-on-assets (profits divided by assets) averaged from the years 2009–2011.

RESULTS AND DISCUSSION

Ratings of the five group-dynamics items were highly correlated ($\alpha = .94$), so we averaged across items to create an aggregate measure of perceived cohesiveness for each company for each participant ($ICC = .84$). We log-transformed the (skewed) company revenues. We analyzed these data using a mixed-level model, with company as the Level 1 variable and judge as the Level 2 variable. As predicted, companies' cohesiveness ratings predicted success, regardless of the metric: revenue: $\gamma = .02$, $F(1, 2000) = 4.974$, $p = .03$; profits: $\gamma = 619.97$, $F(1, 2000) = 26.17$, $p < .001$; return-on-assets: $\gamma = .47$, $F(1, 2000) = 25.26$, $p < .001$, such that greater cohesiveness was associated with greater financial success. Figure 3 depicts average profits from 2009 through 2011 as a function of cohesiveness ratings aggregated across judges. These findings replicate the previous two studies using a larger sample size and different stimulus modality, providing further evidence that participants' snap judgments of groups can predict objective performance outcomes.

One potential concern may be that more successful companies spend more money on image management and skilled photographers, presenting photographs that make their boards of directors appear more cohesive. To address this issue, we asked 21 additional participants to rate the photographs on the following dimen-

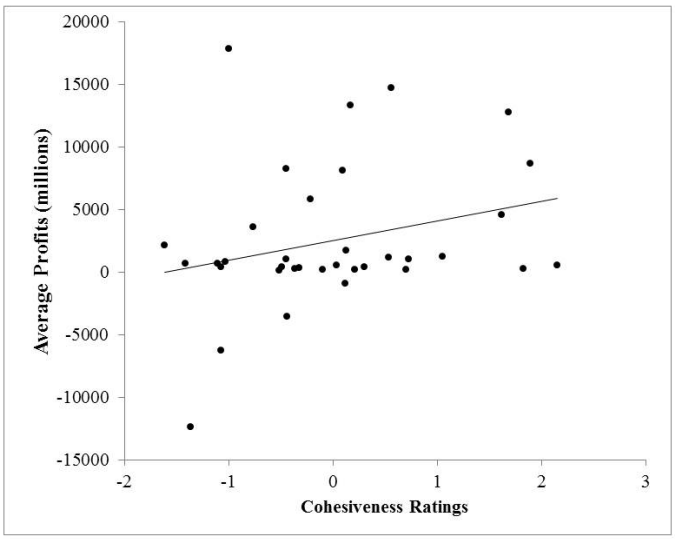

FIGURE 3. Relationship between judgments of cohesiveness (normalized) and company profits in millions on the Fortune 500 list for a sample of 34 boards of directors. The x-axis represents the mean cohesiveness rating for each board aggregated across judges.

sions: "What is the picture quality of this photo?", "How talented was the photographer?", and "How artistic is the photo?" ($\alpha = .85$). We aggregated these ratings to create an index of photographer's skill (*ICC* = .93). These ratings were significantly related to return-on-assets, $\gamma = 1.36$, $F(1, 1999) = 122.90$, $p < .001$; marginally related to profits, $\gamma = 256.92$, $F(1, 1999) = 2.52$, $p = .11$; and unrelated to revenue, $\gamma = -.007$, $F(1, 1999) = .229$, $p = .63$. Importantly, when these ratings were entered as covariates in our analyses, ratings of cohesiveness still significantly predicted company success across all metrics: revenues: $\gamma = .03$, $F(1, 1999) = 5.19$, $p = .02$; profits: $\gamma = 570.64$, $F(1, 1999) = 20.82$, $p < .001$; return-on-assets: $\gamma = .21$, $F(1, 1999) = 5.11$, $p = .02$. These data thus indicate that investment in image management by more successful companies is not a plausible alternative explanation of our findings.

We again conducted an outlier analysis using Cook's *d* on the mean cohesiveness ratings aggregated across judges. As in the prior studies, all Cook's *d*s fell below conventional thresholds for outlier analyses (all Cook's *d*s < .19), suggesting that our effects are not due to the influence of outliers.

GENERAL DISCUSSION

Our results indicate that snap judgments of emergent group-level characteristics such as cohesiveness can be used to predict a group's performance. Whether they observed short video clips (Studies 1 and 2) or photographs (Study 3), participants' ratings of cohesiveness predicted objective behavioral outcomes in three distinct domains in which cohesiveness is important for success (rock bands, Ultimate Frisbee teams, and Fortune 500 boards of directors). That people are able to detect emergent group-level characteristics such as cohesiveness on the basis of such limited samples of behavior, and that such ratings are predictive of objective out-

comes has several important implications. First, the thin slice literature has generally focused on individual-level rather than group-level judgments. Although a small body of research has investigated thin slice accuracy at the dyadic level (Bernieri et al., 1996), it is unclear to what extent such findings extend to groups. The present studies represent a first step in empirically examining the validity of thin slice assessments of group-level, emergent phenomena, and they confirm that people can accurately judge the quality of the interactions among group members, not simply within dyads, on the basis of very limited stimulus exposure. The surprising accuracy of thin slice judgments thus applies not just to intra-individual variables, such as teaching effectiveness, and dyadic-level variables, such as rapport, but also to fundamentally social, intragroup variables such as cohesiveness.

Second, these data add to the small body of research findings that demonstrate that snap judgments can predict actual behavior, not simply the assessments made by others. Most research has compared snap judgment ratings to target self-ratings, or ratings made by those with more extensive exposure to targets (see Ambady, Bernieri, & Richeson, 2000). Few studies at the individual level (Rule & Ambady, 2008) and no studies at the dyad level have demonstrated the accuracy of thin slice judgments in predicting objective behavioral outcomes.

Finally, the present work has implications for the assessment and evaluation of groups. One obstacle to research on cohesiveness is the difficulty in measuring it effectively and efficiently. Cohesiveness can be measured at the individual level by aggregating group members' ratings of group cohesiveness, at the group level by asking group members to come to a joint consensus on their level of cohesiveness, or by having experts observe and rate group-level activity (Quigley, Tekleab, & Tesluk, 2007). Researchers have recently developed computational algorithms to calculate cohesiveness based on audio-video clips of a group (Hung & Gatica-Perez, 2010). Although these methods have their advantages, they can be costly and time-consuming. Although there has been considerable research on cohesiveness and group performance, to our knowledge no study has demonstrated the predictive validity of snap judgments of cohesiveness or other group-level traits. The present research thus represents a new entry into the domain of group measurement. Although further work must be done to fully substantiate this claim, our findings suggest that thin slice judgments may represent an efficient, cost-effective means of assessing group characteristics.

That people are able to make predictive judgments about complex systems like groups begs the question of how much further the phenomenon extends. Can people accurately judge the harmony or atmosphere within much larger groups, such as army divisions or legislatures? Is the predictive validity we found in judgments of team cohesiveness more a positive result of people detecting exceptional chemistry when they see it or is it more a negative result of people detecting exceptional dysfunction? Are some intragroup processes such as cohesiveness harder or easier to spot in brief exposures than other processes such as respect? Regardless of the outcome of future studies that address such questions, the current work indicates that the ability to make predictive snap judgments is more broad and pervasive than originally thought.

REFERENCES

Aloni, M., & Bernieri, J. (2004). Is love blind? The effects of experience and infatuation on the perception of love. *Journal of Nonverbal Behavior, 28*, 287–295.

Ambady, N., Bernieri, F., & Richeson, J. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32,* 201–271.

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83*, 947–961.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology, 77*, 538–547.

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,* 431–441.

Ambady, N., & Skowronski, J. J. (Eds.). (2008). *First impressions*. New York: Guilford Press.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *USA, 104*, 17948–17953.

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*, 989–1004.

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,* 110–129.

Bollen, K. A., & Jackman, R. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Eds.), *Modern methods of data analysis* (pp. 257–291). Newbury Park, CA: Sage.

Campion, M., Medsker, G., & Higgs, C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology, 46,* 823–850.

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*, 1054–1072.

Carron, A. V. (1982). Cohesiveness in sport groups: Interpretations and considerations. *Journal of Sport Psychology, 4,* 123–138.

Chiocchio, F., & Essiembre, H. (2009). Cohesion and performance: A meta-analytic review of disparities between project teams, production teams, and service teams. *Small Group Research, 40,* 382–420.

Cook, R. D., & Weisber, S. (1982). *Residuals and influence in regression*. New York: Chapman & Hall.

Daum, J. H., & Neff, T. J. (2010). Cornerstone of the board: Lessons on creating or rebuilding a board. *SpencerStuart*.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research*, *22*, 175–186.

Festinger, L. (1950). Informal social communication. *Psychological Review*, *57*, 271–282.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. New York: Wiley.

Friedkin, N. E. (2004). Social cohesion. *Annual Review of Sociology, 30,* 409–425.

Grahe, J. E., & Bernieri, F. J. (2002). Self-awareness of judgment policies of rapport. *Personality and Social Psychology Bulletin, 28,* 1407–1418.

Griffith, J. (1988). Measurement of group cohesion in U.S. army units. *Basic and Applied Social Psychology, 9,* 149–171.

Holden Leadership Center. (2009). *Great leadership quotes*. Retrieved March 7, 2011, from http://leadership.uoregon.edu/resources/quotes

Hung, H., & Gatica-Perez, D. (2010). Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia, 12,* 563–575.

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology, 93*, 321–334.

Mason, A. E., Sbarra, D. A., & Mehl, M. R. (2010). Thin-slicing divorce: Thirty seconds of information predict changes in psychological adjustment over 90 days. *Psychological Science, 21,* 1420–1422.

Moreland, R. L. (2010). Are dyads really groups? *Small Group Research, 41*, 251–267.

Moreland, R. L., Hogg, M. A., & Hains, S. C. (1994). Back to the future: Social psychological research on groups. *Journal of Experimental Social Psychology, 30*, 527–555.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*, 210–227.

Quigley, N. R., Tekleab, A. G., & Tesluk, P.E. (2007). Comparing consensus- and aggregation-based methods of measuring team-level variables: The role of relationship conflict and conflict management processes. *Organizational Research Methods, 10*, 589–608.

Rule, N., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology, 44,* 1100–1105.

Rule, N., & Ambady, N. (2008b). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19,* 109–111.

Tabak, J. A., & Zayas, V. (2012). The roles of featural and configural face processing in snap judgments of sexual orientation. *PLoS ONE 7*(5): e36671.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry, 1,* 285–293.

Tjosvold, D. (1984). Cooperation theory and organizations. *Human Relations*, *37,* 743–767.

Todorov, A., Mandisodza, A., Goren, A., & Hall, C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308,* 1623–1626.

Wong, E. M., Ormiston, M. E., & Haselhuhn, M. P. (2011). A face only an investor could love: CEOs' facial structure predicts their firms' financial performance. *Psychological Science, 22,* 1478–1483.