

RESEARCH REPORT

One (Rating) From Many (Observations): Factors Affecting the Individual Assessment of Voice Behavior in Groups

Nathan P. Podsakoff
University of Arizona

Timothy D. Maynes
University at Buffalo, State University of New York

Steven W. Whiting
University of Central Florida

Philip M. Podsakoff
University of Florida

This article reports an investigation into how individuals form perceptions of overall voice behavior in group contexts. More specifically, the authors examine the effect of the proportion of group members exhibiting voice behavior in the group, the frequency of voice events in the group, and the measurement item referent (group vs. individual) on an individual's ratings of group voice behavior. In addition, the authors examine the effect that measurement item referent has on the magnitude of the relationship observed between an individual's ratings of group voice behavior and perceptions of group performance. Consistent with hypotheses, the results from 1 field study ($N = 220$) and 1 laboratory experiment ($N = 366$) indicate that: (a) When group referents were used, raters relied on the frequency of voice events (and not the proportion of group members exhibiting voice) to inform their ratings of voice behavior, whereas the opposite was true when individual-referent items were used, and (b) the magnitude of the relationship between observers' ratings of group voice behavior and their perceptions of group performance was higher when raters used group-referent, as opposed to an individual-referent, items. The authors discuss the implications of their findings for scholars interested in studying behavioral phenomena occurring in teams, groups, and work units in organizational behavior research.

Keywords: teams, challenge-oriented citizenship behavior, voice behavior, referent shift, group performance

E pluribus unum

—from the Great Seal of the United States of America

The use of groups and teams has become increasingly common in modern organizations. Indeed, it has been reported that 50%–75% of all U.S.-based organizations now use teams in some form to accomplish organizational objectives (Devine, Clayton, Philips, Dunford, & Melner, 1999). There has also been a corresponding increase in the amount of research conducted on collective entities (groups/teams) in organizations (e.g., Cohen & Bailey, 1997;

Gully, Incalcaterra, Joshi, & Beauen, 2000; Mathieu, Maynard, Rapp, & Gilson, 2008; Stewart, 2006). As a consequence, our understanding of issues related to such collective entities has increased dramatically in recent years. For instance, we now have a much greater understanding of how team cohesion is formed and how it affects group effectiveness (Beal, Cohen, Burke, & McLendon, 2003), how leadership in teams and groups differs from leadership in traditional contexts (Carson, Tesluk, & Marrone, 2007; Kirkman & Rosen, 1999), how group processes affect group functioning (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008; Mathieu et al., 2008), and how aggregate-levels of behaviors such as organizational citizenship influence unit-level outcomes (N. P. Podsakoff, Whiting, Podsakoff, & Blume, 2009). Most of these studies have asked survey respondents to report their perceptions of the collective phenomenon being examined.

Although the advancements noted here have produced valuable knowledge, a fundamental question at the heart of this area of study that has yet to be addressed is this: How do individual people form perceptions of behavioral phenomena occurring at a team or group level in the first place? It would seem that forming perceptions of such entities would be very complex given an observer's need to integrate his or her perceptions of many individuals into a single perception of the collective. Are

This article was published Online First December 8, 2014.

Nathan P. Podsakoff, Department of Management and Organizations, Eller College of Management, University of Arizona; Timothy D. Maynes, Department of Organization and Human Resources, School of Management, University at Buffalo, State University of New York; Steven W. Whiting, Management Department, University of Central Florida; Philip M. Podsakoff, Department of Marketing, Warrington College of Business Administration, University of Florida.

Correspondence concerning this article should be addressed to Nathan P. Podsakoff, Department of Management and Organizations, Eller College of Management, University of Arizona, P.O. Box 201080, Tucson, AZ 85720-1080. E-mail: podsakof@email.arizona.edu

such perceptions based on the *frequency* with which behaviors or actions occur in a group, on the *proportion* of employees in the group who perform a behavior or take a specified action, or a combination of these factors? Despite the fundamental nature of these questions, answers do not currently exist in the literature.

Methodological considerations add complexity to this issue given that it is unclear how shifting from individual- to group-referent measures affects a respondent's ratings when the unit-level attributes of interest comprise individual-level phenomena (e.g., Chan, 1998; Ehrhart & Naumann, 2004; Kozlowski & Klein, 2000). For instance, when a rater evaluates the behavior of individuals within a group, and the referent is not shifted, it is clear that the rater is evaluating individual behavior. However, when the referent is shifted to the group, the criterion for forming a judgment becomes less clear. Would a rater's assessment of a group in which four separate employees perform a specific behavior one time each be the same as it would for a group in which a single employee engages in the behavior four times, or would the ratings be different? Again, we don't have answers to such questions.

Addressing these issues is critical for several reasons. First, perception is at the heart of nearly every topic that we study in organizational behavior, and the study of collective entities (e.g., teams, groups, work units) is no exception. Furthermore, perception plays a foundational role in our methods of assessment and measurement (e.g., responses to survey items). Thus, exploring how a single person forms perceptions of behavioral phenomena occurring at the group level is vital to understanding the very nature of group entities themselves. Second, forming perceptions of collectives (groups of people) is significantly more complex than forming perceptions of individual people (Hamilton & Sherman, 1996) given that evaluating a group's behavior requires a perceiver to consider variation in that behavior *within and across* the members of the group. This fact has implications for the encoding, storage, and retrieval stages of social information processing and, as a consequence, the impressions that people form and hold of collective entities. Third, we have very little knowledge regarding how measurement item referent affects relationships between constructs. For instance, we do not know whether group-referent measures relate in the same way to perceived group performance as individual-referent measures do. Exploring this issue is particularly important because one of the primary factors driving interest in aggregate levels of individual phenomena is the potential impact of these factors on unit functioning and effectiveness (Gully et al., 2002; Mathieu et al., 2008; N. P. Podsakoff et al., 2009; Stewart, 2006).

We addressed these issues in two studies (one field study and one experiment) in which we examined the effects of the frequency of employee voice behavior and the proportion of group members exhibiting voice on a rater's perceptions of group voice. We also explored whether scale item referent (individual vs. group) affects these relationships. Finally, we addressed the question of whether scale referent influences the relationship between ratings of group voice behavior and perceived group performance.

This article makes several contributions to the literature. First, exploring how a person forms perceptions of collectives is important given the implications of such perceptions for construct meaning, measurement, and substantive relationships. Second, given the theoretical importance of scale referent, a better empirical under-

standing of referent effects on construct meaning and measurement is valuable. Finally, examining the impact of scale referent on the relationship between perceptions of a work unit and perceptions of unit performance is important given the centrality of unit performance to this area of study.

Voice Behavior

The primary purpose of this study was to examine how a person forms perceptions of collective levels of individual behavior.¹ Thus, it was important that the behavior selected for examination (a) be an individual- rather than unit-level phenomenon; (b) have implications for unit-level functioning; and (c) have been examined at the unit level in prior research. For these reasons, we examined employee voice behavior. *Employee voice* is a proactive form of organizational citizenship behavior (OCB) that is defined as "speaking out and challenging the status quo with the intent of improving the situation" (LePine & Van Dyne, 1998, p. 853). This form of challenge-oriented OCB is distinct from affiliation-oriented dimensions such as helping, interpersonal facilitation, courtesy, and cheerleading (LePine & Van Dyne, 1998; Van Dyne, Cummings, & Parks, 1995) as well as from task or in-role performance (Van Dyne & LePine, 1998; Whiting, Podsakoff, & Pierce, 2008).

Although voice behavior is exhibited by individuals and has most often been studied at this level of analysis (see Morrison, 2011, for a review of this literature), several recent studies have examined voice behavior as a unit-level phenomenon (e.g., Detert, Burris, Harrison, & Martin, 2013; Frazier & Bowler, in press; Grant, Gino, & Hofmann, 2011; Lam & Mayer, 2014; MacKenzie, Podsakoff, & Podsakoff, 2011; McClean, Burris, & Detert, 2013; Walumbwa, Morrison, & Christensen, 2012). This may not be surprising given that Organ's (1988) original discussion of OCBs suggested that these behaviors have their strongest impact on organizational effectiveness when considered "in the aggregate" and because the unit or organization is the intended beneficiary of individual voice behavior. Therefore, voice behavior is a good example of an individual behavior that has important meaning at the collective level for both organizations and for scholars.

Roles of Frequency and Proportion in the Formation of Perceptions of Collective Behavior

An essential element of research at levels of analysis higher than the individual level (i.e., teams, groups, units, or organizations) is that multiple individuals are nested within the collective entity. In this type of research, scholars often rely on data from individuals to establish the higher level constructs. For example, a supervisor of a group could be asked to evaluate the voice behavior of his or her group. When performing such a task, it becomes necessary for

¹ It is important to note that the focus of our research was the perceptions of the behavior of individuals in a group, not perceptions of the group's environment (e.g., the group's culture or climate) or the characteristics of the group's members (e.g., group member demographics or personalities). The reason this is important is that research on the perceptions of the behavior of group members often uses ratings provided by other people (e.g., peers, supervisors), whereas research on perceptions of the group's environment or group members' characteristics often depends on self-reported measures.

the rater (supervisor) to consider the behavior of each person in the group (individually or in aggregate) rather than focusing on the behavior of a single group member. Thus, raters likely consider variability *within* individuals and *between* group members. This additional source of variance produces complexity for a rater.

In this study, we sought to enhance understanding of this issue by examining the impact of the frequency of voice occurring in groups and the proportion of group members who engage in voice on ratings of group voice behavior. We define *frequency* as the total number of voice events occurring in a team and *proportion* as the number of group members (out of all possible group members) exhibiting voice at least once. We selected these factors for two reasons. First, Ehrhart and Naumann (2004) proposed that frequency of behavioral occurrence and the proportion of group members exhibiting a behavior play an important role in the formation of perceptions of behaviors occurring at the group level. These authors argued that the frequency with which OCB is exhibited in a group and the proportion of group members exhibiting OCB influence perceptions of group levels of OCB because these factors increase the likelihood that each group member will observe incidents of OCB. Second, there are also good theoretical reasons to believe that proportion and frequency will influence an individual's perception of group-level phenomena. Specifically, on the basis of the work of Hamilton and Sherman (1996), we argue that information about proportion and frequency are stored in separate locations in memory and that this information combines with measurement item referent to influence individual perceptions of collective constructs. We elaborate on this argument in the sections that follow.

Frequency, Proportion, and Scale Referent

Hamilton and Sherman (1996) have argued that information and attributes related to individuals are stored in a separate location in memory than information and attributes related to collectives. This suggests that one source of complexity that a person may encounter when rating collective levels of behavior stems from the likelihood that information regarding frequency is stored separately from and has a different structure in a rater's memory than information regarding proportion. More specifically, the frequency (number of times) with which a behavior occurs in a collective entity is likely stored in memory as an attribute of the collective entity rather than as an attribute of the members of the entity, given that frequency is a factor that is separable from individual members (i.e., can be determined without identifying specific members of the unit). In contrast, the proportion of a collective entity that engages in a behavior is information that resides primarily at the individual level and must be combined to determine proportion, because such a determination requires identification of the specific members who did or did not engage in the behavior. These arguments suggest that information about frequency is associated with the collective entity and is, therefore, housed in a single location, whereas information about proportion is stored in separate, distributed locations because it is associated with the multiple, individual members of the entity.

Empirical evidence supports this perspective. Research has demonstrated that people tend to process and store information in memory about individuals differently than groups and that this has implications for how people think about these different types of

entities (see Hamilton & Sherman, 1996). For instance, scholars have found that individuating information about target others is generally stored in a separate location in memory for each person (i.e., people act as organizing nodes in memory; Sedikides & Ostrom, 1988). Further, empirical evidence indicates that groups or other entities also act as organizing nodes in memory and that entity-level characteristics are often stored separately from individuating information about people (see Brewer, Weber, & Carini, 1995; Lickel, Hamilton, Wierzchowska, Lewis, Sherman, & Uhles, 2000).

On the surface, the fact that frequency is an attribute associated with the collective level of an entity, whereas proportion is an attribute that is aggregated from individual-level data, may not appear to be an important distinction. However, this difference has implications for measurement as it relates to scale referent. Generally speaking, scholars have two options when selecting a scale referent for a group phenomenon: They can choose either an individual or a group referent (Chan, 1998; Klein, Conn, Smith, & Sorra, 2001). As implied by the label, measurement items with an *individual referent* refer to individual people in the workplace, such as an employee or coworker (e.g., "this employee frequently speaks up with improvement-oriented suggestions"). As a consequence, responding to items with an individual referent requires an assessment of each member of the entity under consideration.

Conversely, *group-referent* items refer to the collective entity as a whole, such as the members of a group, work unit, or team (e.g., "members of this team frequently speak up with improvement-oriented suggestions"). As a consequence, items for which the referent has been shifted to the group level have a different meaning than items measured at the individual level given that the formation of a judgment about the group does not require an assessment of each individual member of the entity (Chan, 1998). Thus, group-referent items produce a measure of perceived levels of a behavior occurring in an entity rather than perceived levels of individual behavior. Our review of the literature indicates that some researchers have used aggregated individual-referent ratings to create a measure of group voice (e.g., Erez, LePine, & Elms, 2002; Lam & Mayer, 2014; McClean et al., 2013), whereas others have reported the use of referent-shifted items in which voice behavior items directly reference the assessed entity (e.g., Frazier & Bowler, in press; MacKenzie et al., 2011; Walumbwa et al., 2012).

Scale referent choice has implications for the impact of frequency and proportion on ratings of collective behavior. Specifying the referent of measurement as a collective entity focuses a rater's search through his or her memory on attributes of the collective entity rather than on attributes of the individual members of the entity. As a consequence, raters seek out information in memory associated with the collective entity's features and characteristics, not the attributes of the entity's individual members. For instance, asking how many points a basketball team scored in a given game focuses a person's search through memory on the team's total score, not the number of points scored by individual players. In contrast, specifying the referent as an individual focuses a rater's memory search on attributes of the individual, which will lead the rater to seek out information associated with the person's traits, behaviors, and characteristics, not necessarily those of the entity to which the person belongs. Continuing the basketball example, asking how many points a particular player

scored focuses the search on player points rather than team points. Given that information about an entity is housed in a single location, and that information about members of the entity is stored in separate locations, it is likely that a search through memory for an entity's attributes, as prompted by a scale referent, returns information associated with the team or group but not necessarily the individual members of the collective entity. In contrast, individual-referent items are likely to prompt a search that returns information associated with the individuals in the collective entity but not necessarily the entity itself. On the basis of these arguments, we hypothesized the following:

Hypothesis 1: Scale referent affects the relationship between the *proportion* of group members engaging in voice and ratings of group voice behavior such that this relationship is positive when the referent is individual members of the group but is not significant when the referent is the group.

Hypothesis 2: Scale referent affects the relationship between the *frequency* of voice behavior occurring in a group and ratings of group voice behavior such that this relationship is positive when the referent is the group but is not significant when the referent is individual members of the group.

Effect of Scale Referent on the Relationship Between Group Voice and Group Performance

Several scholars (Ehrhart & Naumann, 2004; Koys, 2001; Organ, 1988; P. M. Podsakoff, MacKenzie, Paine, & Bachrach, 2000) have noted that one of the primary reasons why researchers are interested in collective levels of certain phenomena, such as group voice behavior, is because of the potential effect that these phenomena have on important outcomes at higher levels of analysis such as group effectiveness or productivity. Thus, in addition to the potential effect that scale referent has on the relationships between the frequency and proportion of a specific behavior and ratings of that behavior at collective levels, we were also interested in the influence that (individual and group) scale referents have on the relationship between ratings of collective voice behavior and perceptions of group performance.

One theoretical framework that holds some promise in predicting whether the relationship between OCBs (like voice) and perceptions of group performance is likely to be stronger when these entities are measured with a group, as opposed to an individual, referent is Ajzen and Fishbein's (1977) principle of correspondence. This principle is based on the propositions that (a) entities (such as attitudes or behaviors) may vary on four different elements (the *action*, the *target* of the action, the *context* in which the action is performed, and the *time* at which the action is taken), and (b) the greater the correspondence between the measurement of attitudinal and behavioral entities in terms of these elements, the stronger the relationship that should be observed between them. Consistent with these propositions, several studies (Ajzen & Fishbein, 1977; Kim & Hunter, 1993; Kraus, 1995) have reported that when measures of attitudes have a high correspondence with measures of behaviors, the correlation is uniformly high, whereas the opposite is true when measures of attitudinal predictor variables do not correspond well with measures of behavioral criterion variables.

Thus, we expected that when group performance was measured with items that reference the group, the relationship between ratings of group voice behavior and group performance would be stronger when the measure of voice behavior was taken using a group referent than when it was derived from an aggregation of individually referenced items. This should be the case because there is a greater correspondence between the group-referenced measures in terms of the target of the rating task (the group) than there is when the measure of voice behavior is taken using an individual referent. Thus, we hypothesize this:

Hypothesis 3: Scale referent will affect the relationship between group voice behavior and perceptions of group performance (measured with a group referent), such that this relationship will be more strongly positive when voice is measured with a group referent than when it is measured with an individual referent.

General Procedure

To examine our hypotheses, we gathered field data and conducted an experiment. In the field study, respondents reported (a) demographic information about themselves and the members of their work group, (b) the number of people in their work group, and (c) how many times each group member exhibited voice behavior in the past 2 months. They then rated the group's overall voice behavior and overall performance. In the experiment, participants watched a video of a group interacting at work and then rated the group's voice behavior using either an individual- or group-referent measure. Ratings of group-level helping behavior and group performance were also collected.

Participants for both studies were recruited from Clear Voice Surveys (<http://clearvoicesurveys.com>), a professional survey company that maintains a panel of over 500,000 active respondents. Using demographic information provided by panelists, Clear Voice Surveys is able to target specific respondents within the panel when certain sample characteristics are desirable. Given our hypotheses, the ability to target respondents who work in groups and have experience evaluating group-level phenomena was important. Panelists were only allowed to participate in our study if they were currently working in a team or group comprising at least three other members and had been required to evaluate group performance at some point in their career. Clear Voice Surveys was paid \$14 and \$16 per participant in the field study and the experiment, respectively.

Study 1: Field Study

Sample

Our sample consisted of 220 full-time employees working in the United States. Sample characteristics and respondent demographics are reported in Table 1.

Procedures

An e-mail invitation with a link to the online survey was sent to survey panel members who met the study inclusion criteria (outlined earlier). Respondents reported demographic information,

Table 1
Sample Characteristics

Variable	Study 1: Field study	Study 2: Experiment
Most common industries		
Public sector/nonprofit	0.0%	17.4%
Professional services	17.3%	14.8%
Health care	12.3%	11.7%
Technology	11.8%	10.1%
Consumer goods	8.2%	0.0%
Level of position		
Professional staff	56.4%	67.0%
Supervisor/manager	28.1%	26.7%
Executive	8.2%	2.3%
Vice president/director	7.3%	4.0%
Respondent demographics		
Gender (female)	50.0%	56.8%
Ethnicity (Caucasian)	79.5%	77.1%
Four or more years of college	62.2%	52.3%
35 years of age or more	56.4%	50.3%
10 or more years of work experience	69.5%	54.9%
Two or more years with current team	45.9%	86.5%
Sample size	220	366

group characteristics, the proportion of group members exhibiting voice, and the frequency of voice behavior in the group and provided ratings of group voice behavior, helping behavior, and group performance.

Measures

Demographics and group characteristics. Participants reported their age, gender, education level, years of overall work experience, years of experience working in teams, and current team tenure. They were informed that “work groups refer to the set of employees who report to the same supervisor.”

Group voice proportion and frequency. Participants also reported the number of people in their group and how many times each person had engaged in voice in the past 2 months. This time frame was selected to facilitate accurate recall of events and to standardize time across participants. The number of total voice events (voice frequency) was calculated by summing the reported voice behaviors across members of the group. However, to control for differences in group size, this total was divided by the number

of people in the group. The number of voicing employees in the group divided by the number of employees in the group was used to measure the proportion of the group exhibiting voice behavior.

Ratings of group voice. Participants rated team-level voice behavior using five group referent–shifted items adapted from the work of Van Dyne and LePine (1998) on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency reliability of this scale was .87.

Results: Tests of Relationships

The means, standard deviations, correlations, and internal consistency reliabilities of this study’s variables are reported in Table 2. The results of our regression analysis are reported in Table 3. Eight control variables were entered in Step 1. None of these had a significant effect on ratings of group voice behavior (all $ps > .05$). The hypothesized predictors were entered in Step 2. Voice frequency had a positive effect on group voice ratings ($\beta = .41, p < .01$), but the proportion of group members who voiced did not ($\beta = .05, p > .05$). Thus, Hypothesis 2 was supported. Adding the main effect terms resulted in a change in R^2 of .19.

Study 1 Discussion

There are at least two important findings that should be highlighted from this study. First, rater individual differences (including experience working in teams and tenure with their current team) had little effect on the results. Second, voice frequency (total number of voice events) had a strong positive effect on ratings of group voice, but the proportion of group members exhibiting voice had no effect.

Study 2: Experiment

Although Study 1 provided initial support for some of our expectations, we were unable to address two important issues. First, the causal priority of our predictors could not be established given the cross-sectional nature of our data. Second, we were also unable to fully address Hypotheses 1 and 3 because we couldn’t ask respondents to both report the number of voice behaviors exhibited by each member of their team and then rate the level of each team member’s voice behavior (individual referent) without

Table 2
Study 1 (Field Study) Variable means, Standard Deviations and Intercorrelations

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10
1. Age	20.60	13.15	—									
2. Gender	0.50	0.50	.15*	—								
3. Education	2.53	1.00	.20**	-.05	—							
4. Work experience	15.35	8.80	.91**	.15*	.10	—						
5. Team experience	9.70	6.81	.62**	.09	.11	.70**	—					
6. Age of team	3.87	1.27	.15*	-.05	-.11	.20**	.29**	—				
7. Team tenure	3.45	2.87	.32**	-.08	-.04	.36**	.45**	.36**	—			
8. Proportion of voicers	0.94	0.13	.13	.08	.02	.12	.06	-.12	.05	—		
9. Frequency of voice events	5.04	1.86	.12	-.02	.13	.12	.13	-.08	.02	.59**	—	
10. Voice ratings	4.15	0.62	.00	.03	-.02	.03	.12	.05	.05	.28**	.43**	—

Note. Gender was coded as Male = 1; and Female = 0. The Cronbach alpha for the voice rating measures was .87.

* $p < .05$. ** $p < .01$.

Table 3
Summary of Regression Results Predicting Group Voice Ratings in Study 1 (Field Study)

Predictors	Step 1	Step 2
Step 1: Controls		
Age	-.15	-.15
Gender	.03	.04
Education	-.01	-.05
Years of work experience	.04	.01
Years of team experience	.17	.13
Age of team	.01	.06
Current team tenure	.01	.01
Step 2: Main effects		
Proportion of voicers		.05
Frequency of voice events		.41**
R^2	.02	.21
ΔR^2		.19

Note. This table reports standardized coefficients; all group voice ratings in this study were gathered using group referents.

** $p < .01$.

introducing biases that might have threatened the study's internal validity. Thus, we conducted an experimental study to address these limitations.

Sample

The sample consisted of 366 full-time employees working in U.S.-based companies. Sample characteristics and respondent demographics are reported in Table 1.

Procedures

Experimental materials were presented via Web-based software. Participants were randomly assigned to experimental conditions in which they watched a video depicting the members of a four-member audit team working together in a fictitious accounting firm. After the video, participants rated the voice behaviors of the group, evaluated the group's performance, and provided demo-

graphic information. Manipulated factors were (a) the proportion of group members exhibiting voice (low: .25; medium: .50; high: .75), (b) the number of voice events exhibited by each group member (i.e., frequency of voice behavior; low: one event; medium: two events; high: three events), and (c) the referent of the voice behavior survey items (i.e., individual vs. group referent). These factors were fully crossed, resulting in a $3 \times 3 \times 2$ fully factorial design with 18 experimental conditions. The conditions in this study are summarized in Table 4.

Video development and validation. The videos were developed and validated in three steps. These steps were writing the scripts, conducting manipulation checks to establish script validity, and filming the videos. A more complete description of this process is provided in Appendix A, and examples of the voice and nonvoice event scripts are provided in Appendix B. Total running time of the video for each condition was about 30 min.

Measurement referent conditions. In the individual-referent condition, participants were instructed to rate each team member's level of voice, the items made specific reference to the ratings of individual voice behavior, and photos of each team member were included next to the items referring to that team member. In the group-referent condition, participants provided an overall rating of voice behavior exhibited by the group, and the survey items referred to the team as a whole.

Measures

Participants rated voice behavior using the same measure as in Study 1. To test Hypothesis 3 and the discriminant validity of our voice measure, participants also rated team helping and team performance. Helping was measured with LePine and Van Dyne's (1998) seven-item measure. Overall team performance was assessed with MacKenzie, Podsakoff, and Fetter's (1991) three-item measure. In the individual-referent condition, the voice and helping items were preceded by this phrase: "This member of the team . . .," and the order in which the team members (i.e., Adam, Ann, Taylor, and Kirk) were rated was randomized to minimize the effects of priming and comparisons. In contrast, the voice and

Table 4
Summary of Conditions and Voice Ratings in Study 2 (Experiment)

Condition number	Sample size	Number and proportion of voicers	Frequency of voice	Group voice profile	Voice rating: Individual referent		Voice rating: Group referent	
					<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1/2	21/20	1 (.25)	1	0/0/1	3.30	0.64	3.90	0.69
3/4	21/20	1 (.25)	2	0/0/2	3.39	0.41	4.30	0.71
5/6	21/18	1 (.25)	3	0/0/3	3.43	0.55	4.30	0.61
7/8	19/23	2 (.50)	2	0/1/1	3.53	0.49	3.94	0.71
9/10	19/18	2 (.50)	4	0/2/2	3.55	0.56	4.28	0.57
11/12	21/19	2 (.50)	6	0/3/3	3.65	0.39	4.29	0.56
13/14	23/19	3 (.75)	3	1/1/1	3.70	0.48	4.19	0.71
15/16	23/19	3 (.75)	6	2/2/2	3.75	0.42	4.60	0.43
17/18	19/23	3 (.75)	9	3/3/3	3.75	0.51	4.58	0.51

Note. Each row contains both an individual referent and group referent measure condition; the numbers in the second column represent the sample size for the individual referent condition (left-hand numeral) and the group referent condition (right-hand numeral). The numbers in the group voice profile column represent the number of times each group member spoke up in each condition: the first number represents Adam's amount of voice, the second number represents Taylor's amount of voice, and the third number represents Kirk's amount of voice; Ann (the fourth group member) did not exhibit voice in any of the conditions, and Nicole (the team leader) did not exhibit voice behavior and was not rated.

Table 5

Factor Loadings and Average Variance Extracted Values From the Confirmatory Factor Analyses for Study 2

Construct and item	Group referent	Individual referent
Voice behavior (average variance extracted)	(.65)	(.73)
1. Frequently make suggestions about how to do things in new or more effective ways at work	.75	.85
2. Often suggest changes to work projects in order to make them better	.85	.88
3. Often speak up with recommendations about how to fix work-related problems	.84	.89
4. Frequently make suggestions about how to improve work methods or practices	.84	.82
5. Regularly propose ideas for new or more effective work methods	.75	.83
Helping behavior (average variance extracted)	(.55)	(.67)
1. Help others who have been absent when they return to work	.69	.76
2. Help teammates who have heavy workloads	.77	.83
3. Help orient new members to the team	.62	.85
4. Willingly help others who have a work-related problem	.84	.85
5. Are always ready to lend a helping hand to each other	.82	.80
6. Are good at resolving unconstructive interpersonal conflicts between members of the team	.69	.76
7. Raise the spirits of teammates who are having problems at work	.74	.86
Group performance (average variance extracted)	(.81)	(.66)
1. All things considered, this team performs at a very high level	.92	.72
2. This audit team is a very high-performing team	.92	.94
3. This team functions at a very high level	.86	.77
Chi-square values, degrees of freedom, and model fit statistics		
χ^2	150.77	186.99
<i>N</i>	179	187
<i>df</i>	87	87
CFI	.98	.95
IFI	.98	.95
SRMR	.05	.07

Note. All factor loadings are completely standardized lambdas and are significant at $p < .01$. Chi-square values are significant at $p < .01$. Factor loadings in the individual-referent column for the voice and helping items are averages from the ratings of the four group members who were rated on these behaviors. CFI = comparative fit index; IFI = incremental fit index; SRMR = standardized root mean square residual.

helping items in the group-referent condition referred to the group as a whole (i.e., "The members of this work group . . ."). Cronbach's alphas were .91 (individual-referent voice), .90 (group-referent voice), .96 (individual-referent helping), .89 (group-referent helping), and .78 (group performance). We also collected the same demographic variables as those measured in Study 1.

Results

Tests of measurement validity. A series of confirmatory factor analyses were conducted in LISREL 8.8 to assess the validity of our measures. A summary of our results is reported in Table 5. The findings support the factor structure, item convergent validity, and discriminant validity of our measures. Further, the hypothesized three-factor model fit the data significantly better than a measurement model in which all of the items assessing the constructs loaded on a single factor, $\Delta\chi^2(3, N = 179) = 581.92$, $p < .01$, for the group-referent model; $\Delta\chi^2(3, N = 187) = 339.86$, $p < .01$, for the individual-referent model. The variable means, standard deviations, correlations, and internal consistency reliabilities are reported in Table 6.

Predictors of mean group voice behavior and scale referent effects. To test our hypotheses, we regressed the ratings of group voice on our manipulated factors. The group- and individual-referent conditions were analyzed separately because the computation of the group voice variable differed across these conditions.² In the individual-referent condition, the variable was created by averaging the voice items for each group member (Ann, Taylor, Adam, and Kirk) and then aggregating these composites across all

four group members. In the group-referent condition, the five voice items were averaged to create the group composite variable.

As indicated in Table 7, all of the control variables were entered in Step 1.³ Years of experience working in teams had a positive effect on group voice ratings ($\beta = .27$, $p < .05$) in the individual-referent condition but not the group-referent condition. None of the other

² Although we believe that our approach was a more appropriate way to analyze these data, we did perform a more traditional moderation analysis using hierarchical multiple regression and a dataset with observations from the individual- and group-referent conditions combined ($N = 366$). To test the (moderating) effects of scale item referent, we created a new variable in which observations using individual-referent ratings were coded as a zero (0), and those using group-referent ratings were coded as a one (1). In this analysis, we entered the predictors in the following steps: (a) control variables; (b) main effects of frequency, proportion, and group referent; and (c) interactions between group referent and voice frequency and between-groups referent and proportion of group speaking up. Results of this analysis were highly consistent with the inferences we made from our study: (a) There was a significant interactive effect ($p < .05$) between group referent and voice frequency, such that the relationship was more positive when the referent was the group than when the referent was individual members of the group, and (b) there was a marginally significant interactive effect ($p = .07$) between group referent and proportion of voicers, such that that the relationship was more positive when the referent was individual members of the group than when the referent was the group.

³ We included these controls in our analyses for two reasons. First, we wanted to make direct comparisons with the results reported from our field study, which measured and controlled these same variables. Second, we found that one control variable (years of team experience) was correlated with our mean voice ratings, and we wanted to control for its effect in our model.

Table 6

Study 2 (Experiment) Variable Means, Standard Deviations, and Intercorrelations

Variable	Individual referent		Group referent		1	2	3	4	5	6	7	8	9	10
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>										
1. Age	37.24	12.84	36.68	12.31	—	.10	-.14	.91**	.78**	-.01	.00	-.05	.08	.17*
2. Gender	0.48	0.50	0.39	0.49	.08	—	-.11	.22**	.13	.04	-.03	-.03	.04	.19*
3. Education	2.46	0.99	2.60	1.01	.16*	.04	—	-.18*	-.12	.09	.01	.03	-.04	-.13
4. Work experience	12.77	9.07	12.46	9.17	.90**	.13	.12	—	.83**	.03	-.01	-.03	.07	.16*
5. Team experience	9.48	6.65	9.47	6.74	.75**	.13	.07	.82**	—	.02	.05	.05	.10	.13
6. Helping behavior	3.43	0.55	4.29	1.03	-.11	-.02	-.15*	-.09	-.07	—	-.12	-.05	.19*	.31**
7. Proportion of voicers	0.50	0.20	0.50	0.20	-.04	-.05	.01	-.03	-.02	.11	—	.69**	.19**	.01
8. Frequency of voice	3.98	2.37	4.05	2.50	-.10	.01	.02	-.11	-.07	.08	.67**	—	.30**	.03
9. Voice rating	3.56	0.51	4.26	0.65	-.11	.01	-.08	-.10	.01	.56**	.29**	.24**	—	.64**
10. Group performance	4.22	0.75	4.34	0.63	.05	-.04	-.19**	.12	.08	.25**	-.01	-.02	.32**	—

Note. Gender was coded as Male = 1; and Female = 0. Correlations from the individual-referent voice rating condition are below the diagonal; correlations from group referent voice rating condition are above the diagonal. For the individual-referent condition, the Cronbach's alphas for helping behavior, voice ratings, and group performance were .90, .91, and .79, respectively; for the group-referent condition, the Cronbach's alphas for helping behavior, voice ratings, and group performance were .81, .90, and .82, respectively. Reliabilities for the individual-referent condition are averages of the reliabilities from the ratings of the four group members who were rated on these behaviors.

* $p < .05$. ** $p < .01$

controls had a significant effect in either condition (all $ps > .05$). The control variables explained 20% and 12% of the variance in group voice ratings in the individual-referent and group-referent conditions, respectively.

In Step 2, proportion of voicers and frequency (number of voice events) were entered into the regression equation. Planned comparisons of the regression coefficients were then conducted for proportion and frequency across the group- and individual-referent conditions to test Hypotheses 1 and 2. We found that the regression coefficient for proportion in the individual-referent condition was positive ($\beta = .25, p < .01$) and significantly larger ($p < .01$) than the estimate in the group-referent condition ($\beta = -.04, p > .05$). Thus, Hypothesis 1 was supported. In contrast, the impact of frequency in the group-referent condition was positive ($\beta = .33, p < .01$) and significantly larger ($p < .01$) than in the individual-

referent condition ($\beta = .06, p > .05$). Thus, Hypothesis 2 was also supported.

Effects of scale referent on relationships between group voice and group performance. The correlation between group-referent ratings of voice behavior and group performance was twice as large as the same correlation for the individual-referent voice measures ($rs = .64$ and $.32$, respectively). A t test indicated that the difference was significant ($p < .01$), providing preliminary support for Hypothesis 3 that there was a stronger relationship between ratings of group voice behavior and perceptions of group performance when the items had the same, rather than different, referents.

Given that prior research has shown an empirical relationship between helping and voice behavior (LePine & Van Dyne, 2001) and that helping behavior influences group performance (N. P. Podsakoff, Podsakoff, MacKenzie, Maynes, & Spoelma, 2014), we conducted a supplementary regression analysis to determine whether ratings of group voice affect group performance beyond group helping behavior. Results of these analyses for both the group- and individual-referent conditions are reported in Table 8. As indicated in the table, group voice behavior had a significant impact on ratings of group performance in both the group-referent ($\beta = .59, p < .01$) and individual-referent ($\beta = .29, p < .01$) conditions, even after controlling for group helping behavior. Further, consistent with our hypothesis, a t test showed that the difference between these regression coefficients was significant ($p < .01$), indicating that the relationship between group voice and group performance was stronger in the group-referent condition than the individual-referent condition.

Study 2 Discussion

Our finding that the mean of the reported level of voice in the group-referent condition was higher than in the individual-referent condition is consistent with the work of Hamilton and Sherman (1996). Indeed, our findings support the view that item referent focuses a rater's memory search on either individual or group characteristics but not both. More specifically, we found that individual-level characteristics (proportion of group members exhibiting voice)

Table 7

Summary of Regression Results Predicting Group Voice Ratings in Study 2 (Experiment)

Variable	Group referent		Individual referent	
	Step 1	Step 2	Step 1	Step 2
Step 1: Controls				
Age	.09	.16	-.10	-.08
Gender	.05	.06	.01	.02
Education	-.03	-.04	-.06	-.06
Years of work experience	-.16	-.16	-.21	-.21
Years of team experience	.16	.09	.27*	.26*
Step 2: Main effects				
Proportion of voicers		-.04		.25**
Frequency of voice events		.33**		.06
R^2	.12	.32	.20	.35
ΔR^2		.20		.15

Note. Group-referent $N = 179$; individual-referent $N = 187$. This table reports standardized coefficients.

* $p < .05$. ** $p < .01$.

Table 8
Impact of Ratings of Group Helping and Group Voice Behavior on Group Performance for Study 2

Variable	Group performance (group-referent condition)		Group performance (individual-referent condition)	
	Step 1	Step 2	Step 1	Step 2
Step 1: Controls				
Age	.24	.23	-.25	-.21
Gender	.18	.15	-.06	-.06
Education	-.10	-.10	-.19**	-.17*
Years of work experience	-.12	-.08	.42*	.48**
Years of team experience	.01	.09	-.06	-.14
Step 2: Main effects				
Rating of group helping		.21**		.07
Rating of group voice		.59**		.29**
R ²	.07	.50	.07	.18
ΔR ²		.43		.11

Note. Group-referent $N = 179$; individual-referent $N = 187$. This table reports standardized coefficients.

* $p < .05$. ** $p < .01$.

positively influenced observer ratings of group voice behavior when the referent was individual employees; but characteristics associated with the group (frequency of voice) did not predict these ratings. In contrast, the reverse was found to be true for group-referent voice measures. That is, voice frequency had a positive effect on group voice ratings, but proportion did not. These findings are consistent with the arguments made by Hamilton and Sherman in that they suggest that information about collective entities is stored separately in memory from information about individuals within the entity.

General Discussion

Our primary objective was to examine how people form perceptions of collective behavioral phenomena. To this end, we examined three separate hypotheses. In this section, we discuss the results of our hypothesis testing and the implications of our findings for future research.

Determinants of Group Voice Ratings and the Moderating Effect of Scale Referent

On the basis of Hamilton and Sherman's (1996) contention that information about groups is stored in a separate location in memory than information about individual members of the group, we hypothesized that the frequency with which voice behavior is exhibited by members of a group and the proportion of group members exhibiting voice behavior would predict ratings of group voice and that these relationships would depend on whether the measurement items reference the group or individual members of the group. Using manipulations and survey measures, we found that both the frequency of voice behavior in the group and the proportion of the group exhibiting voice predicted ratings of group voice behavior. However, consistent with our expectations, these effects were dependent on the referent used to assess group voice items. When an individual referent was used, raters relied on the proportion of group members exhibiting voice to inform their ratings of voice behavior, such that the larger the

proportion of the group exhibiting voice, the higher the mean-level ratings of group voice behavior. In contrast, when provided with group-referent voice items, raters used the frequency of voice behavior occurring in the group to inform their evaluations of group voice behavior, such that the larger the number of voice events in the group, the higher the ratings of composite group voice behavior.

Relationships Between Group Voice and Group Performance

On the basis of Ajzen and Fishbein's (1977) principle of correspondence, our third hypothesis proposed that the relationship between ratings of group voice and perceptions of group performance would be stronger when the measures of group voice and group performance were taken at the same level (group) rather than different levels (voice at an individual level and performance at a group level). Findings from our experimental study supported this view in that ratings of group voice had significantly stronger relationships with group performance when the referent was the group. In addition, in both the lab and field studies, we found that ratings of group voice behavior using a group referent were distinguishable (factor analytically) from group helping behavior.

Although our finding that item referent affects the relationship between group voice behavior and a group's overall performance is important, it should not be used as the basis for choosing one referent over another. Indeed, as noted by Chan (1998) and Kozlowski and Klein (2000), the choice of a scale item referent should be based on a theory of the phenomenon of interest. Consistent with this view, our findings suggest that if a theoretical framework focuses on the frequency of behavior in a group, then scholars should use a group referent because it will lead respondents to recall frequency information from memory. Conversely, if the theory focuses on the proportion of a group exhibiting a behavior, then use of an individual referent would be appropriate because respondents will recall this information.

Implications for Research and Future Directions

This article has several implications for future research. First, our findings indicate that a researcher's choice of group versus individual referent has implications for a variable's conceptual meaning and for the substantive relationships these constructs have with other variables. Indeed, we found that the correlation between group-referenced voice and perceptions of group performance was twice as large as the same correlation for individually referenced voice items. This finding is important given the centrality of performance to the literatures examining collective entities (teams, work groups, etc.).

Second, we examined two substantive factors (frequency of group behavior and proportion of the group exhibiting voice) and one methodological factor (scale referent) that influence how people form and recall perceptions of collective phenomena. This is merely a starting point for research on this topic. It is likely that observer ratings are influenced by additional factors, such as characteristics associated with specific behavioral events. For example, raters may be more likely to attend to, encode, and recall voice events that are highly salient, such as voice behaviors that have a substantial effect on group performance or that produce significant conflict among coworkers. Future research should examine such factors to increase our understanding of how people form perceptions of collective phenomena.

Third, a logical next step for research in this area would be to explore the processes through which individual perceptions of collective phenomena become shared by the members of a group. This is a critical issue because research examining collective phenomena often relies on member perceptions of the phenomena, and agreement among the members of the entity is sometimes a precondition for the phenomena's existence (Chan, 1998). Although there are a number of factors that could impact how perceptions become shared by group members, a few include communication styles, social interaction norms, group member similarity, group member interdependence, and the amount of communication.

Finally, our results show that attributes that are associated with the collective level of an entity (e.g., the frequency of voice behavior in a group) influence ratings when the entity (e.g., the group) is the referent but not when individuals within the group are the referents. We found the opposite to be true for attributes that require the individual identification of the members of the group (e.g., proportion). These findings support propositions derived from previous theory (Hamilton & Sherman, 1996) and add to a growing body of research (Brewer et al., 1995; Lickel et al., 2000) demonstrating that people store information about groups in different locations in memory than they do information about individuals. Although our findings are valuable, we did not examine the cognitive stage (e.g., encoding, retrieval, judgment) at which these differences occur. Therefore, it would be worthwhile to further explore this issue.

Limitations

Although our multimethod approach reduced many of the concerns associated with experimental or field studies by themselves, there are still important limitations to note. One limitation is that in our experimental study, outside observers were asked to evaluate team interactions representing a relatively small portion of the group's life span. It is unknown, therefore, whether the same pattern would hold when people are members of the group. Although this is true, the results of our experiment did converge with those of our field study for group-referent measures, and the field study was not inhibited by these same constraints. Nevertheless, we were unable to test our predictions regarding individual referents in the field study. Therefore, additional research will be needed before a definitive statement can be made regarding the generalizability of individual-referent effects.

Another possible limitation is that the wording of the items used to measure voice behavior tended to refer to the frequency with which these behaviors were exhibited. However, item wording in the group- and individual-referent conditions was nearly identical. Thus, this would not explain why proportion of voicers had a stronger impact on voice ratings in the individual- than in the group-referent condition. Scholars should nevertheless consider using more neutral item wording when examining issues related to frequency or proportion in the future.

A final limitation arises from the fact that participants in our experimental study were not subject matter experts in the field of accounting. Thus, when asked to evaluate the performance of an auditing team, they may not have had a complete understanding of what constitutes effective performance on such a team. Having said this, participants were asked to evaluate the *overall* performance of the team rather than the more technical aspects of performance related to auditing, and all of our participants had significant experience working in teams (10 or more years on average). Thus, it seems

unlikely that this issue could explain our findings. However, researchers should consider such issues when planning and conducting studies in this area in the future.

Conclusion

Notwithstanding the limitations just discussed, the findings of our research are important for several reasons. First, this research is the first to examine the "mental algebra" used by people when evaluating collective phenomena based on individual-level behaviors. Second, our results suggest that the proportion of group members exhibiting a behavior and the frequency of the behavior in the group are both important predictors, although these effects are dependent on the referent used by the rater. However, additional research is needed to better understand how observers develop one rating from many individual observations of a collective phenomenon.

References

- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888–891. <http://dx.doi.org/10.1037/0033-2909.84.5.888>
- 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. <http://dx.doi.org/10.1037/0021-9010.88.6.989>
- Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology*, 69, 29–40. <http://dx.doi.org/10.1037/0022-3514.69.1.29>
- Carson, J. B., Tesluk, P. E., & Marrone, J. A. (2007). Shared leadership in teams: An investigation of antecedent conditions and performance. *Academy of Management Journal*, 50, 1217–1234. <http://dx.doi.org/10.2307/20159921>
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83, 234–246. <http://dx.doi.org/10.1037/0021-9010.83.2.234>
- Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23, 239–290. <http://dx.doi.org/10.1177/014920639702300303>
- Detert, J. R., Burris, E. R., Harrison, D. A., & Martin, S. (2013). Voice flows to and around leaders: Understanding when units are helped or hurt by employee voice. *Administrative Science Quarterly*, 58, 624–668. <http://dx.doi.org/10.1177/0001839213510151>
- Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations prevalence, characteristics, and effectiveness. *Small Group Research*, 30, 678–711. <http://dx.doi.org/10.1177/104649649903000602>
- Ehrhart, M. G., & Naumann, S. E. (2004). Organizational citizenship behavior in work groups: A group norms approach. *Journal of Applied Psychology*, 89, 960–974. <http://dx.doi.org/10.1037/0021-9010.89.6.960>
- Erez, A., LePine, J. A., & Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology*, 55, 929–948. <http://dx.doi.org/10.1111/j.1744-6570.2002.tb00135.x>
- Frazier, M. L., & Bowler, W. M. (in press). Voice climate, supervisor undermining, and work outcomes: A group-level examination. *Journal of Management*.
- Grant, A. M., Gino, F., & Hofmann, D. A. (2011). Reversing the extraverted leadership advantage: The role of employee proactivity. *Academy*

- of *Management Journal*, 54, 528–550. <http://dx.doi.org/10.5465/AMJ.2011.61968043>
- Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beuven, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology*, 87, 819–832. <http://dx.doi.org/10.1037/0021-9010.87.5.819>
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, 103, 336–355. <http://dx.doi.org/10.1037/0033-295X.103.2.336>
- Kim, M.-S., & Hunter, J. E. (1993). Attitude-behavior relations: A meta-analysis of attitudinal relevance and topic. *Journal of Communication*, 43, 101–142. <http://dx.doi.org/10.1111/j.1460-2466.1993.tb01251.x>
- Kirkman, B. L., & Rosen, B. (1999). Beyond self-management: Antecedents and consequences of team empowerment. *Academy of Management Journal*, 42, 58–74. <http://dx.doi.org/10.2307/256874>
- Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology*, 86, 3–16. <http://dx.doi.org/10.1037/0021-9010.86.1.3>
- Koys, D. J. (2001). The effects of employee satisfaction, organizational citizenship behavior, and turnover on organizational effectiveness: A unit-level, longitudinal study. *Personnel Psychology*, 54, 101–114. <http://dx.doi.org/10.1111/j.1744-6570.2001.tb00087.x>
- Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco: Jossey-Bass.
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, 21, 58–75. <http://dx.doi.org/10.1177/0146167295211007>
- Lam, C. F., & Mayer, D. M. (2014). When do employees speak up for their customers? A model of voice in a customer service context. *Personnel Psychology*, 67, 637–666.
- LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61, 273–307. <http://dx.doi.org/10.1111/j.1744-6570.2008.00114.x>
- LePine, J. A., & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of Applied Psychology*, 83, 853–868. <http://dx.doi.org/10.1037/0021-9010.83.6.853>
- LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with Big Five personality characteristics and cognitive ability. *Journal of Applied Psychology*, 86, 326–336. <http://dx.doi.org/10.1037/0021-9010.86.2.326>
- Lickel, B., Hamilton, D. L., Wierzchowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, 78, 223–246. <http://dx.doi.org/10.1037/0022-3514.78.2.223>
- MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salesperson performance. *Organizational Behavior and Human Decision Processes*, 50, 123–150. [http://dx.doi.org/10.1016/0749-5978\(91\)90037-T](http://dx.doi.org/10.1016/0749-5978(91)90037-T)
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Challenge-oriented organizational citizenship behaviors and organizational effectiveness: Do challenge-oriented behaviors really have an impact on the organization's bottom line? *Personnel Psychology*, 64, 559–592.
- 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, 410–476. <http://dx.doi.org/10.1177/0149206308316061>
- McClean, E. J., Burris, E. R., & Detert, J. R. (2013). When does voice lead to exit? It depends on leadership. *Academy of Management Journal*, 56, 525–548. <http://dx.doi.org/10.5465/amj.2011.0041>
- Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management Annals*, 5, 373–412. <http://dx.doi.org/10.1080/19416520.2011.574506>
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., & Klinger, R. L. (2013). Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures. *Journal of Applied Psychology*, 98, 99–113. <http://dx.doi.org/10.1037/a0029570>
- Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., Maynes, T. D., & Spoelma, T. (2014). Consequences of unit-level organizational citizenship behaviors: A review and recommendations for future research. *Journal of Organizational Behavior*, 35, S87–S119. <http://dx.doi.org/10.1002/job.1911>
- Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual- and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, 94, 122–141. <http://dx.doi.org/10.1037/a0013079>
- Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26, 513–536. <http://dx.doi.org/10.1177/014920630002600307>
- Sedikides, C., & Ostrom, T. M. (1988). Are person categories used when organizing information about unfamiliar sets of persons? *Social Cognition*, 6, 252–267. <http://dx.doi.org/10.1521/soco.1988.6.3.252>
- Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, 32, 29–55. <http://dx.doi.org/10.1177/0149206305277792>
- Van Dyne, L., Cummings, L. L., & Parks, J. M. (1995). Extra-role behaviors: In pursuit of construct and definitional clarity (a bridge over muddied waters). *Research in Organizational Behavior*, 17, 215–285.
- Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management Journal*, 41, 108–119. <http://dx.doi.org/10.2307/256902>
- Walumbwa, F. O., Morrison, E. W., & Christensen, A. L. (2012). Ethical leadership and group in-role performance: The mediating effect of group conscientiousness and group voice. *Leadership Quarterly*, 23, 953–964. <http://dx.doi.org/10.1016/j.leaqua.2012.06.004>
- Whiting, S. W., Podsakoff, P. M., & Pierce, J. R. (2008). Effects of task performance, helping, voice, and organizational loyalty on performance appraisal ratings. *Journal of Applied Psychology*, 93, 125–139. <http://dx.doi.org/10.1037/0021-9010.93.1.125>

(Appendices follow)

Appendix A

Summary of Video Development and Validation Procedure For Study 2

The videos were developed and validated in three steps, following the general procedure outlined by Podsakoff, Podsakoff, MacKenzie, and Klinger (2013). First, scripts were written for each video in the study. Scripts depicted scenes in which the members of an auditing team of a fictional accounting firm (GWK) were performing the year-end financial statement audit for one of their clients (HiTech Distributors, Inc.). The team comprised four staff accountants (Adam, Anne, Kirk, and Taylor), led by one supervisor (Nicole). The number of team members engaging in voice behavior and the total number of voice events varied in the scripts such that one, two, or three of the team members (Adam, Kirk, and/or Taylor) each spoke up either zero, one, two, or three times. Other team interactions were also scripted to (a) provide temporal spacing between the voice events, (b) take the place of voice events in conditions in which fewer voice events would be displayed, and (c) enhance the realism of the videos. Voice behaviors in the scripts focused on improving efficiency or work quality.

Second, video manipulations were recorded. Professional actors were hired to play the audit team members, and a professional film crew recorded the videos. All actors were dressed in professional business attire. The actor playing the role of the manager was in her mid-30s, and the actors who played team member roles were in their mid- to late-20s. A total of nine voice events (and nine nonevents) were filmed (i.e., three voicers speaking up three times each).

Third, checks of our manipulations were obtained in a sample of 48 working professionals. Participants were randomly assigned to view either the voice event or the nonvoice event for each of the nine behavioral incidents (i.e., participants viewed a total of nine separate videos). After each video, participants were provided with a definition of voice behavior and then asked to identify whether the scene depicted a person engaging in voice behavior or not. Results indicated that participants distinguished voice from nonvoice events, given that raters correctly categorized the videos 87% of the time.

To assess the effects of other potentially confounding variables, participants also rated (a) the physical attractiveness of the actors who played a voicing role in our videos, (b) the quality of each voice suggestion, (c) the convincingness of the arguments presented in favor of the suggestion, and (d) the perceived magnitude of the effect that the suggestion would have on team performance. A repeated-measures multivariate analysis of variance revealed no significant differences across the nine videos for any of these variables (all F s < 1.28, all p s > .26). Thus, taken together, the evidence obtained from the manipulation checks indicated that the video manipulations tended to be appropriately perceived as either voice or nonvoice behavior and that the observed effects of these manipulations on the outcome measures were not likely the result of differences in perceived voicer attractiveness, convincingness of the arguments, quality of the suggestions, or magnitude of the effect of the suggestion on team functioning.

(Appendices continue)

Appendix B

Examples of Voice Event and Nonvoice Event Manipulations Used in Study 2

Example 1

Introduction Script

Nicole (Team Supervisor): Now, as you know, the work of the audit is divided into sections based on the different areas of the balance sheet. So, please turn to the first page of the financials, which is where the balance sheet is, and let's review the assignments I've made for this year. Starting at the top of the balance sheet, Taylor, you'll be auditing cash and accounts receivable; Kirk, you'll be auditing inventory and fixed assets; Adam, you'll be working on accounts payable and unearned revenue; and Ann, you'll be auditing stockholder's equity and long-term debt. Finally, the first person finished with their sections will work on calculating the deferred tax asset or liability. Does that sound okay?

Voice Event Script

Kirk: Overall, the assignments look great Nicole, but I'd like to make one suggestion. You have Taylor working on cash and accounts receivable, and I think it would be a good idea for whoever is auditing cash to also audit accounts payable because the audit tests in both of those sections require that a person look at the bank statements. So, it would be more efficient for one person to audit both the cash and accounts payable sections together. It would only require a simple change to make this happen. You assigned Adam to work on accounts payable and Taylor to work on accounts receivable. You could just switch those two assignments. That way Taylor would work on accounts payable and cash and Adam would work on accounts receivable and unearned revenue. This change also makes sense for Adam because the unearned revenue and accounts receivable sections are highly related to each other as well. Anyway, this is just a thought. Feel free to do whatever you think is best.

Nonvoice Event Script

Kirk: Overall, the assignments look great Nicole. From my end, I don't foresee that I'll have any trouble completing my sections on time. As I look at the financial statements though, I do have a couple of questions that you might be able to answer since you've met with HiTech's CFO a couple of times since last year's audit. First, I noticed that the balance in the accounts payable account is significantly higher than it has been in prior years. Second, HiTech is also carrying much more inventory than they usually do. Do you know why this is the case?

Nicole: Good questions. I also noticed these things, and asked the CFO about them. Apparently, at the end of the year one of HiTech's suppliers, in an effort to boost year-end sales, offered a significant discount on several of HiTech's most popular items. So, HiTech purchased many more items from this supplier than they typically do at the end of the year, and that is why the balances in both accounts payable and inventory are higher this year.

Example 2

Introduction Script

Adam, Ann, & Kirk (to Nicole and Taylor as they exit the room together): See you guys next week.

Voice Event Script

Taylor: Hey Nicole, do you have a minute?

Nicole: Sure, what's up?

Taylor: Well, I've been working on the section in which we test the balance in the deferred tax asset account, and I'm running into some trouble. The problem is that anytime someone makes an adjustment to any income statement account, the adjustment flows through to the deferred tax asset account, and I have to verify the balance in the account again. I've already done this several times, and I'm sure that I'll have to do it several more times throughout the course of the audit. I know that in the past we've done this section at the same time that we've been working on the other areas of the audit, and I recognize that you may disagree with what I'm about to say, but I recommend that we hold off and test the deferred tax asset account after all of the other work has been done. That way, we won't have to keep replicating our work. It seems to me that this would be a more efficient and effective way to get this section done.

Nonvoice Event Script

Nicole: So, how are things coming with your sections of the audit?

Taylor: Things have been coming along fairly well. I finished auditing the cash section, and I've been working on the deferred tax asset account, which as you know, can be tough because nearly every adjustment that gets made to the numbers in the other sections ends up impacting the deferred tax asset calculation. Other than that, it's been smooth sailing though. How are we doing with respect to meeting our deadlines?

Nicole: We're doing well. We should be finished with fieldwork in about a week. After that, we'll just have a few items to complete before we're done and can move on to the next audit.

(Appendices continue)

Appendix B (*continued*)**Example 3***Introduction Script*

Nicole: Well, I got pulled away on another client with a crisis situation. The client is going public next year. So, the workload is significantly greater than a traditional audit. Plus, there's a learning curve because the client is new to us. Anyway, things on the audit weren't going well, the audit team wasn't going to make its deadlines, and the client wasn't happy. So, the firm pulled in a few extra people for two days to help the team get back on track, and I was one of those people. All I can say is that I'm glad to be back working with you guys.

Voice Event Script

Adam (walks over and sits in the chair next to Nicole): Nicole I've been working on these unearned revenue accounts and I think that our audit procedures in this area are not as efficient as they could be, and I've got a suggestion regarding how we can improve them. As has been true in the past, HiTech's contracts with their customers are structured in such a way that revenue is not earned until products that have been shipped are received by the customer. This makes it difficult to calculate the appropriate amount of unearned revenue that should be recorded at the end of the year because we don't know whether the customer has received the shipment or not. To simplify the process, HiTech came up with a method to estimate the amount of unearned revenue that should be recorded based on the date the items were shipped and the destination of the shipment. The problem is that the method is not very reliable and has resulted in incorrect numbers being recorded in the accounting records. Furthermore, it's very difficult to verify whether the numbers are correct. However, I think there is a simple fix to this problem. HiTech uses an independent shipping company to deliver orders to customers, and the company has an online package tracking system. We could get a printout of this and see what shipments were still in transit as of the end of the year. In this way, we could determine the precise amount of revenue that had not yet been earned at year end.

Nonvoice Event Script

Adam (walks over and sits in the chair next to Nicole): Hey Nicole, I've been working on these unearned revenue accounts and I just want to make sure I understand what's going on before I complete the work. It looks like HiTech's contracts with their customers are structured in such a way that revenue is not earned until products that have been shipped are received by the customer. Is that right?

Nicole: Yes. HiTech's contracts are generally structured such that they are responsible to ensure that the product is delivered safely and in an acceptable condition to the customer. If a shipment arrives with damaged products, the customer can return the shipment without cost and without an obligation to pay for the goods. If the contracts specified that as soon as the products were shipped, the customer assumed all liability for the products, then HiTech could record the revenue the moment the items were shipped, but that just isn't the case here. The key thing to think about is when does liability of the products transfer from HiTech to the customer. Does that make sense?

Adam: Yeah, your explanation is very helpful. I think I've got it now. If I have any other trouble, I'll let you know. Thanks.

Received December 30, 2013

Revision received October 30, 2014

Accepted November 3, 2014 ■