EXTENDING THE MULTILEVEL THEORY OF TEAM DECISION MAKING: EFFECTS OF FEEDBACK AND EXPERIENCE IN HIERARCHICAL TEAMS

JOHN R. HOLLENBECK
DANIEL R. ILGEN
JEFFREY A. LePINE
JASON A. COLQUITT
Michigan State University

JENNIFER HEDLUND
Yale University

This study extends previous research on the multilevel theory of team decision making and shows that team-level decision-making accuracy is affected by factors originating at lower levels of analysis. The results from 380 individuals arrayed into 95 four-person teams working on a simulated naval command and control task indicated that the constructs specified by this theory accounted for over half of the variance in team performance and that teams given feedback on those constructs performed better than control teams. We elaborate on the relationships among the core constructs and further elucidate the process by which they affect overall team decision-making accuracy.

For a variety of reasons, the use of teams in business, military, and medical contexts is on the rise. Although there is a great deal of research on group decision making in the behavioral and social science literature, Ilgen. Major, Hollenbeck, and Sego (1993, 1995) showed that much of this research deals with people who are exposed to a common information base and who have to reach consensus on a single decision for which there is no objectively verifiable answer; juries are an example.

However, the teams emerging in contemporary organizations are often characterized by members who differ in their areas of specialization and status (leader versus member), and these teams often make decisions that can be evaluated as right or wrong. For example, concurrent engineering processes that have been adopted by large numbers of organizations employ small teams with differentiated staff members (from marketing, production, and personnel functions, for instance) and a project leader to make decisions about new or existing products. Other examples of these types of teams include managerial staffs, hospital emergency room teams, military command and control teams, and academic research teams.

A group structured this way is best characterized as a team rather than as a set of independent decision makers, for several reasons. First, these individuals are highly interdependent. Each depends on the others for important information related to the team’s success. Second, the members have a common goal and a common fate. The team’s success or failure directly affects the individuals’ own outcomes. Third, members of the team influence each other in the course of making a decision.

Other features of such “leader-staff” teams make them a special case of the more generic concept of team. First, the primary task confronting such a team is a decision-making task; these are not teams that produce a tangible product. Second, there is an established hierarchy in the team—the leader has the final say, and the team does not vote or have to reach consensus. Finally, each member has specialized knowledge, and thus the members are not redundant or easily interchangeable.

There is far less programmatic research on leader-staff teams than on jurylike teams. Several researchers have identified a need to develop conceptual approaches and an empirical literature dealing with these types of teams. Sniezek and Buckley (1995) referred to these teams as judge-advisor systems; Gruenfeld, Mannix, Williams, and Neale (1996) referred to them as partial information groups; Hollenbeck, Ilgen, Sego, Hedlund, Major,
and Phillips (1995) referred to them as hierarchical teams with distributed expertise; and Brehmer and Hagaforss (1986) referred to them as leader-staff arrangements.

In their original study of this type of team, Brehmer and Hagaforss noted this: "One possible reason why there has been so little psychological research on staff decision making may be that there has been no theory to guide research in this area, nor even a pretheoretical framework. Indeed, there has not even been an experimental paradigm for the study of staff decision making" (1986: 182). Hollenbeck and colleagues (1995) developed the multilevel theory of team decision making to address this gap in the literature. This theory is a narrow one in the sense that it focuses on a specific type of team in a specific context. That is, this theory addresses conditions leading to accurate decision making in teams with formal leaders and specialized staffs who work on a recurring set of structured decisions in a situation in which their decisions are evaluated as being right or wrong.

The purpose of the present research was to extend the literature testing this theory both methodologically and conceptually. Methodologically, the study reported here tested whether direct feedback expressed in terms of the theory's core constructs enhanced the performance of decision-making teams. Conceptually, the study focused on the relationships among the core constructs and how they were affected by feedback and experience.

Overview of the Multilevel Theory

The literature on group decision making is vast, and a tremendous number of different theories and variables are discussed in this literature (Bettnerhausen, 1991; Guzzo & Shea, 1992; Levine & Moreland, 1990; Salas, 1995). In fact, Ilgen and colleagues (1993) concluded that one of the biggest problems with this literature was the sheer number of constructs that could be related to team decision making. Many of these are at different levels of analysis, and there has been little in the way of integration across models. Thus, the central emphasis in building a theory of decision making for hierarchical teams with distributed expertise needs to be on parsimonious integration of the existing literature.

To accomplish this goal, Ilgen and colleagues (1995) used a Brunswick Lens model framework (cf. Brehmer & Hagaforss, 1986) to guide the identification of important outcomes at each level of analysis below the team level—the decision level, the individual level, and the dyadic level—that needs to be achieved for such a team to make accurate decisions. In the multilevel theory they then developed around this conceptual frame, Ilgen and colleagues attempted to identify a small core set of variables that primarily drive team decision-making accuracy (Hollenbeck et al., 1995).

A detailed description of the multilevel theory of team decision making is provided in Hollenbeck and colleagues (1995: 293-300). Briefly, according to this theory, team decision-making accuracy is determined by constructs that occur at one of four levels: team, dyad, individual, and decision. The theory identifies the most critical variable at each of the three lower levels and then aggregates these variables at the team level. The goal of the theory is explanation of performance differences within and between teams in terms of decision-making accuracy.

The multilevel theory: Core constructs and propositions. At the decision level, the most critical variable is decision infomity, which is defined as the degree to which each team member has all the information necessary to perform his or her role in the decision-making process. This is a decision-level variable because a person might be well informed on one decision but poorly informed on another; that is, decisions are nested under individuals. Further, a characteristic of a given decision's context (for instance, time pressure) might affect a team's overall performance because of its effect on this variable. Decision infomity can be aggregated to form a team-level construct referred to as team infomity, which captures how well-informed a team is, on average, across all the decisions it makes. Even if one holds the decision context constant, across a large number of decisions, some teams may end up being better informed than others, and this difference is captured by the construct of team infomity.

At the individual level, the most critical variable according to this theory is individual validity, which is the degree to which any one staff member can make recommendations to the leader that are predictive of the correct decision for the team. Because the staff members are making recommendations and not final decisions, they need not make accurate decisions to be effective team members. Thus, staff members may be biased (off by a constant) and still be valuable in a predictive sense, in that their predictions show a correlation of 1.0 with the true score. As long as the variance in their recommendations is related to variance in the true decision, the team leader (who makes the final decision) can make effective use of their inputs by adjusting for the bias. Thus, a staff member who might make poor decisions if he or she were the team leader (because of bias) may still be an effec-
tive staff member. In teams with multiple staff members, this variable can be aggregated to form a team-level construct referred to as \textit{staff validity}. Operationally, staff validity has been defined as the average level of individual validity.

At the dyadic level, the most important variable according to this theory is \textit{dyadic sensitivity}. This variable reflects the degree to which a team leader correctly weights each staff member’s recommendation to arrive at a team’s decision. For example, for any set of recommendations generated by a group of staff members, there is an ideal or optimal set of weights that might be applied to these recommendations to predict the correct decision. Through policy-capturing techniques, the weights that the leader places on these staff recommendations when arriving at his or her final decision can be identified. The difference between the weight that the leader actually assigns to a staff member’s recommendation and the ideal weight for that staff member’s recommendation has been the traditional operational definition of dyadic sensitivity (a large difference implies low sensitivity). Again, in teams with multiple staff members, this variable can be aggregated to form a team-level variable called \textit{hierarchical sensitivity}, which captures the overall optimality of a leader’s use of his or her staff.

These three concepts—decision informity, individual validity, and dyadic sensitivity—along with their team-level analogues, constitute the core constructs of the multilevel theory. All constructs other than the six listed above are labeled “noncore constructs.” Included in the latter set are constructs that occur at lower levels (the decision, individual, and dyadic level) as well as at the team level. To describe the domain of noncore variables, Hollenbeck and colleagues (1995) adapted a framework developed by McGrath (1976) that divides the influences on groups into six categories (roles, persons, tasks, physical environment, behavior settings, and social environment). According to the multilevel theory of team decision making, the core constructs mediate the relationship of noncore constructs to team decision-making accuracy. Thus, the noncore constructs have a peripheral influence on team decision-making accuracy that is attributable to their effects on the core variables (see Hollenbeck et al. [1995: 296–299, especially Figure 4]).

The multilevel theory: Empirical findings. Hollenbeck and colleagues (1995) introduced the multilevel theory and reported two studies that showed that between 25 and 50 percent of the variance in team decision-making accuracy could be explained by the three core constructs. Noncore variables such as group cohesiveness, familiarity, experience, attrition, job knowledge, and role redundancy explained a smaller percentage of the variance, and almost all of the effects for these variables were mediated by the three core constructs.

The value of decomposing overall team decision-making accuracy into the core constructs could be seen when the impact of some noncore constructs had multiple, but counteracting, influences on different core constructs. For example, familiarity among team members led to both advantages and disadvantages in terms of the core constructs. On the positive side, familiar teams (those composed of friends) were better coordinated and therefore showed higher levels of team informity and staff validity in the initial stages of task performance. On the negative side, team member attrition was much more debilitating to familiar teams than to unfamiliar teams (those composed of strangers) and, over time, the staff validity of unfamiliar teams surpassed that of the familiar teams. When results were taken together, familiarity had no overall effect on team performance, yet one would be mistaken to conclude that familiarity did not matter in terms of important team processes.

Feedback and the Multilevel Theory

Feedback has both informational value, in terms of promoting learning, and motivational value, in terms of promoting effort (Ilgen, Fisher, & Taylor, 1979; Taylor, Fisher, & Ilgen, 1984). Teams have been found to actively engage in feedback-seeking behavior (Ancona & Caldwell, 1992) and to respond to this feedback with changes in behaviors and outcomes. Positive responses to feedback are especially likely when the feedback focuses attention on goal-performance discrepancies (Matsui, Kakayuma, & Onglato, 1987; Mesch, Farh, & Podsakoff, 1994) and promotes learning (Kluger & DeNisi, 1996).

Earley, Northcraft, Lee, and Lituchy (1990) demonstrated that, at the individual level, performance information is especially useful when outcome-type feedback (expressed in terms of bottom-line results) is complemented by process feedback (expressed in terms of intervening steps that lead to bottom-line results). An increase in the effectiveness of outcome feedback when it is paired with process feedback has also been found on numerous occasions in team contexts (Anderson, Crowell, Doman, & Howard, 1988; Flynn, Sakakibara, & Schroeder, 1995).

The results from the initial research testing the multilevel theory suggest that the core constructs capture the key processes that teams need to manage if their overall goal is to generate accurate decisions (Hollenbeck et al., 1995). Thus, this type of
feedback should enhance team performance because of its influence on the core constructs. Thus, our first set of hypotheses is as follows:

**Hypothesis 1a.** Teams provided with process feedback on the core constructs of the multilevel theory of team decision making will perform better than teams that are only provided outcome feedback.

**Hypothesis 1b.** The effect of process feedback on team performance will be mediated by the three core constructs specified by the multilevel theory of team decision making.

**Experience and the Multilevel Theory**

In addition to testing the effects of feedback, this study also examined the role of experience. Few would argue with the general notion that experience with a task and with other team members is likely to improve group effectiveness (Sundstrom, DeMeuse, & Futrell, 1990). Our specific interest, however, was in the degree to which the positive impact of experience on team decision-making accuracy could be explained by the core constructs of the multilevel theory of team decision making. According to this theory, experience is a noncore construct whose effect on team decision accuracy should be mediated by the core constructs. That is, increased experience with the task and other team members should enhance team members’ capacity to share and use information, make accurate recommendations, and weight each person’s unique contribution to the team.

Although we intended to look at main effects for feedback on decision accuracy and the core constructs, we were also interested in exploring the degree to which the provision of process feedback on the core constructs of the multilevel theory could substitute for experience. That is, if one of the major reasons why experience improves performance is that it helps teams learn how to get informed, generate valid recommendations, and appropriately weight different staff members’ inputs, then direct feedback on these core constructs may accelerate the learning process. If this is the case, then direct feedback on the core constructs may substitute for experience in such a way that inexperienced teams performing with feedback perform as well as experienced teams. This idea too has applied implications. All else being equal, one would certainly prefer to have teams composed of people with a great deal of experience working together on a specific task. However, in the real world, this is not always possible. Thus, it is important to explore the possibility that organizations can use process feedback to substitute for experience and speed the team development process. Thus, our second set of hypotheses:

**Hypothesis 2a.** Experienced teams will perform better than inexperienced teams.

**Hypothesis 2b.** The effect of experience on team performance will be mediated by the three core constructs specified by the multilevel theory of team decision making.

**Hypothesis 2c.** The relationship between experience and positive team outcomes (such as decision accuracy and the core constructs) will be stronger in the absence of process feedback, which acts as a substitute for experience.

**Relationships among the Core Constructs of the Multilevel Theory**

The original Hollenbeck et al. (1995) article focused exclusively on the relationships between the core constructs on one hand, and decision accuracy and the traditional variables studied in the group decision-making literature on the other. In particular, the focus of the two studies reported in Hollenbeck and colleagues’ work was on the role of the core constructs as factors that mediated the relationship between the traditional variables studied in the group decision-making literature (i.e., the noncore variables) and decision accuracy.

Although the core constructs of the multilevel theory of team decision making were not the focus of the original research, however, there are also good theoretical reasons to expect relationships among them. Indeed, one of the purposes of this study was to directly study these relationships. For example, in the initial development of the multilevel theory, the core constructs were described as “emanating from lower levels to higher levels.” That is, inferiority at the decision level fed into validity at the individual level, and validity at the individual level fed into hierarchical sensitivity at the next level (see Hollenbeck et al. [1995: 296–296, especially Figure 3]).

Thus, whereas no explicit relationships among the core constructs were specified originally, it is certainly reasonable to expect that the better informed staff members are, the higher their validity is likely to be. Similarly, the better a staff is at rendering recommendations, the better it might be at working with its leader to develop an appropriate weighting system. Indeed, the average correlation between team inferiority and staff validity in the two studies reported by Hollenbeck and colleagues (1995) was .38, and the average correlation

Copyright © 1998. All rights reserved.
between staff validity and hierarchical sensitivity in these initial studies was \(-0.27\).

Indeed, given the tight conceptual link between adjacent core constructs and the past empirical evidence, the prediction of relationships between team informity and staff validity and between staff validity and hierarchical sensitivity is quite straightforward. In addition to these predictions, however, we were interested in exploring the degree to which these relationships were affected by process feedback and experience. For example, the literature on social loafing suggests that accountability is one of the key factors in ensuring that each team member contributes to the best of his or her ability (Miles & Greenberg, 1993; Williams, Nida, Baca, & Latane, 1999). Because the feedback intervention studied here (described below) displayed the value of each staff member’s contribution publicly, we expected that accountability would be higher for staff members working with feedback than for staff members for whom no feedback was available.

However, given the strong relationship between team informity and staff validity, the level of validity a team member can achieve is in many ways constrained by the information he or she has to work with when arriving at final recommendations. Thus, whereas the accountability factor embodied in feedback may make staff members desire to increase their validity, the information they have may limit their ability to translate this desire into results. Thus, feedback may interact with team informity to influence staff validity in such a way that high levels of team informity are more likely to translate into high levels of staff validity when staff members’ contributions are made public via the feedback intervention than when they are less readily apparent.

Whereas feedback is likely to moderate the relationship between team informity and staff validity, one might expect that experience would moderate the relationship between staff validity and hierarchical sensitivity. The complexity of hierarchical sensitivity makes it probably the most difficult of the core constructs for a team to manage and learn. Past research suggests that the general tendency is for leaders to weight staff recommendations too heavily and too uniformly (Brehmer & Hagafora, 1986; Hollenbeck et al., 1995; Snieszek & Buckley, 1995). This basic tendency has to be unlearned, and if this type of learning takes time, the negative effect on hierarchical sensitivity of having a poor staff member would be especially great when the leader and staff lacked experience working together. Indeed, recent research by Gruenfeld and colleagues (1996) showed that one of the major benefits of familiarity among team members is that it makes it easier for a team to effectively deal with conflicting information and opinions. Thus, our final set of hypothesis is as follows:

**Hypothesis 3a.** There will be a positive relationship between team informity and staff validity.

**Hypothesis 3b.** There will be a negative relationship between staff validity and hierarchical sensitivity.

**Hypothesis 3c.** The relationship between team informity and staff validity will be stronger in the presence of process feedback.

**Hypothesis 3d.** The relationship between staff validity and hierarchical sensitivity will be enhanced by experience.

It should be noted that this set of hypotheses specifies a different role for feedback and experience than do the previous hypotheses. Whereas the two previous sets of hypotheses focus on the role of feedback and experience as factors that directly impact the core constructs, this third set of hypotheses focuses on their influence on the relationships between adjacent core constructs.

### METHODS

**Participants**

Research participants were 380 undergraduate students at a large midwestern university who were arrayed into 95 four-person teams. In return for their participation, each received course credit and was also eligible to earn a cash prize based upon performance.

**Task**

Since this was meant to be a replication and extension of previous studies in this program of research, a similar task (i.e., TIDE\(^2\)) and set of procedures were used (Hollenbeck et al., 1995: 301–303). TIDE\(^2\) is a software program for a decision task simulation that presents participants with values on a number of attributes of a problem or object. At a very basic level, this task is a high-tech analog of common multiple-cue probability-learning tasks widely used in the decision-making literature.

In this particular study, TIDE\(^2\) was programmed to simulate a naval command and control scenario with a leader and three staff members. Each team member sat at a workstation in front of a computer that was networked to all other team members. Each team’s task was to monitor the airspace sur-
rounding it. When an aircraft came into this airspace, each team member needed to gather some information about particular attributes of the aircraft, including its speed, direction, angle, range, and size.

This information was accessed from a "pull-down menu" that listed available information on the incoming aircraft. Staff members also had to share information with each other, because no one person had access to all the information needed to render a judgment. This communication was accomplished with other pull-down menus that allowed for queries and transmissions of raw data or open-form text messages. The cues were independent, and thus none of the information could be extrapolated from cues that were not either directly measured or obtained from other team members who could directly measure the cues.

After accessing, sharing, and considering information, each staff member arrived at a judgment regarding the appropriate response to make toward an approaching aircraft. Staff members were specialized in the sense that one dealt with data related to the aircraft's motion (speed, direction, and angle), another dealt with the aircraft's location (altitude, range, and status with respect to a commercial corridor), and the third with the aircraft's category (size, IFF signals, and radar emissions). So, for example, one team member was responsible for the "motion rule," a standard based on the interaction among cues related to speed, direction, and angle of approach. Staff members did not have to be fully informed to register recommendations, and time limits or poor coordination among some team members often meant that recommendations were based on partial information on the part of one or more staff members.

Judgments and decisions were rendered on a seven-point continuum that varied in aggressiveness from "ignore" (the least aggressive response) to "defend" (the most aggressive response). Intermediate anchors on this scale were arranged in increasing levels of aggressiveness.

A team's decision on each trial was rendered by its leader after he or she had received input from the staff. Once made, the leader's response was compared by the program to the correct decision. This correct decision was based on translating the rules into a linear combination of the attributes and applying an equation to the attribute values of the stimulus aircraft. This equation stipulated that the true score was an interactive function of three sets of cues dealing with an incoming aircraft's motion (speed, direction, and angle), location (altitude, range, and status with respect to a commercial corridor), and category (size, IFF signals, and radar emissions).

These cues interacted in such a way that if any one cue took on a value that was "nonthreatening," the value for the entire rule was nonthreatening; however, the three rules themselves were additive. This feature of the decision task made it highly complex, despite the small number of cues monitored by each staff member. For more information on this task, see Hollenbeck, Segov, Ilgen, Major, Hedlund, and Phillips (1977).

Research Design

This study employed a two-by-two design that crossed (1) whether or not group members had previous experience with each other and the task with (2) whether or not the group received the process feedback intervention.

Process feedback intervention. The normal TIDE computer screen presents each participant with an icon that represents the four team members engaged in the simulation. The top of this screen contains a menu that the participants use to measure aircraft attributes, to transmit information or text messages to other team members, to receive information from other team members, and to send recommendations to the leader (or, if the person is the leader, to enter the team's decisions). There is also a clock in the middle of the screen that lets participants know how much time is left in the current trial (as well as the current trial's number).

We added three features to this screen in the present study for research participants who were assigned to the feedback condition. First, an indicator light appeared next to the icon of any team member who had obtained all the information needed to make a decision based upon his or her area of specialization. The indicator light appeared the moment that person became fully informed. If the light was not on, all team members knew that this person needed some additional information. This missing information was usually not accessible from the focal person's workstation, and therefore it cued the team that one of its members needed assistance.

Second, under each of the icons representing the four team members was a red bar that varied in length depending upon the individual validity of that staff member's recommendations (or, in the case of the leader, his or her decisions). The red bar also included a numerical scale ranging from -1.00 to +1.00. The computer simulation generated the red bar by calculating the correlation between that

---

1 IFF means "identification—friend or foe?"
staff member's recommendations and the true scores for all the trials up to that moment. As with the informinity indicator light, all staff members and the leader could see each others' red bars next to their icons, and hence everyone in the team knew whose recommendations were high in validity and whose were low.

Third, just under each icon's red bar was a green bar that varied in length depending upon the weight that the leader put on that staff member's recommendations. This green bar also included a numerical scale ranging from -1.00 to 1.00. The computer simulation generated the green bar by calculating the correlation between that staff member's recommendations and the leader's actual decisions for all the trials up to that moment. Like the informinity indicator light and the validity bar, each member's green bar could be seen by all other team members and the leader, and hence everyone in the team knew whose recommendations were being weighted heavily by the leader and whose recommendations were receiving less weight.

Because the red validity bar sat on top of the green weight bar, the two together also gave the team feedback on dyadic sensitivity. When the red bar and green bar were similar in length, it meant that the leader was weighting that staff member about as heavily as his or her validity warranted (dyadic sensitivity was high). When the red bar exceeded the green bar, the leader was not weighting that staff member's recommendations heavily enough, relative to his or her predictive value. When the red bar was much smaller than the green bar, the leader was making the opposite mistake, giving that person's recommendations too much weight relative to his or her predictive value. Thus, the larger the deviation between the two bars, the lower the dyadic sensitivity.

**Experience manipulation.** Experience was manipulated in terms of the number of sessions a team had. Experienced groups worked together for an initial session that lasted three hours and encompassed roughly 50 decisions (trials). These sessions occurred roughly one week prior to the experimental session. The groups in the no-experience condition came in for just the one three-hour experimental session.

**Measures of Outcomes**

**Team decision-making accuracy.** Accuracy was defined as the degree to which a leader's decision matched the correct decision, which was defined as the mean absolute error of the team's decision. This was a continuous variable that could range from 0 for a perfect match to 6 for the worst match (for instance, the team's decision was to ignore the aircraft [1 = ignore] when they should have shot it down [7 = defend], or vice versa). Following the leader's decision, each person in the team received overall outcome feedback that told the team's performance on the trial. There were 60 aircraft to evaluate (60 decision trials).

**Team informinity.** Team informinity was operationally defined as the total number of staff members who had all the information they needed to make a recommendation in their role. So, for example, if a member responsible for the motion rule had all three pieces of information needed to judge the aircraft according to that rule, he or she was considered fully informed. If any piece of information was missing, the member was considered uninformed. Thus, this measure could take on values ranging from 0 (if no staff member had all the needed information) to 3 (if all three staff members had obtained what they needed).

**Staff validity.** The correlation between each staff member's recommendations and the correct decisions was the index of that member's individual validity. The average of the three validities across staff members was the index of staff validity for the team. This variable could range from 0.00 to 1.00.

**Hierarchical sensitivity.** For each team, regression analysis determined (1) the weights the leader placed on each staff member's recommendations when arriving at the 60 team decisions and (2) the ideal weights that the team leader should have placed on each staff member's recommendations. Hierarchical sensitivity was equal to the average of the absolute difference between the weights actually assigned by the leader to his or her staff's recommendations and the ideal set of weights obtained by regressing the true decision on the staff members' recommendations. A score of .00 indicated that the leader weighted each staff member optimally, whereas a high score indicated a large deviation between the leader's weights and the optimal weights.

**Introducing Variability within Staffs**

To provide a sensitive test of the value of the process feedback intervention, we also varied the relative importance of the different staff positions. Specifically, the potential validity of the role of one of the staff members was weakened by restricting the variance of the cues that made up this person's area of expertise. The lack of range in this person's cues meant there was little relationship between any recommendation coming from this station and the true score, which was instead determined by
TABLE 1
Descriptive Statistics for All Variables Measured at the Team Level\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>s.d.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Team decision-making accuracy</td>
<td>1.20</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Team informity</td>
<td>2.53</td>
<td>0.60</td>
<td>-0.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Staff validity</td>
<td>0.53</td>
<td>0.11</td>
<td>-0.70</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Hierarchical sensitivity</td>
<td>0.17</td>
<td>0.09</td>
<td>0.54</td>
<td>-0.34</td>
<td>-0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Experience</td>
<td>0.50</td>
<td>0.50</td>
<td>-0.53</td>
<td>0.62</td>
<td>0.47</td>
<td>-0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Feedback</td>
<td>0.50</td>
<td>0.50</td>
<td>-0.25</td>
<td>-0.21</td>
<td>-0.11</td>
<td>-0.41</td>
<td>-0.01</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} The value of \( n \) is 95. Decision-making accuracy and hierarchical sensitivity are deviation scores, so low values reflect high standing on the constructs.

\textsuperscript{b} For \( r \)'s greater than .17, \( p < .05 \).
team decision accuracy. The interaction between experience and the feedback intervention did not reach conventional levels of statistical significance. Thus, Hypothesis 2c did not receive support in this analysis—the effect of the feedback intervention on overall performance did not appear to differ for experienced and inexperienced teams.

The Mediating Role of the Core Constructs

The third stage of the analysis assessed the extent to which the effects for the process feedback intervention and experience were mediated by the core constructs (Hypothesis 1b and Hypothesis 2b, respectively). Table 4 shows the results of regressing team decision-making accuracy on experience, the process feedback intervention, and the interaction between these two variables, with the core constructs controlled. After the core constructs were entered, the effect size for experience and the feedback intervention dropped from .33 to .02—a 94 percent reduction in the effect size. Although the effect for experience was still statistically significant (the incremental $R^2$ was .02), the effect of the feedback intervention was driven to 0. Thus, the results of this analysis provided strong support for Hypothesis 1b and moderate support for Hypothesis 2b. The process feedback intervention’s effects on team decision-making accuracy were fully mediated by the core constructs, and the effects of experience were mostly mediated.

To test the nature of these mediating effects, we examined the degree to which the lower-level analogs of team informity, staff validity, and hierarchical sensitivity were predicted by experience and the process feedback intervention. Since variability in the core constructs existed both within and across teams, we used repeated measures regression (Cohen & Cohen, 1983; Hollenbeck, Ilgen, & Sego, 1994) to partition the variance and decompose these effects.

For decision informity, the unit of analysis was the decision, and hence the number of observations was 5,700 (95 teams each made 60 decisions). For individual validity, the unit of analysis was the staff member, and therefore the number of observations was 285 (95 teams each contained three staff members). For dyadic sensitivity, the unit of analysis was the vertical leader-staff dyad, so the number of observations was also 285 (95 teams each contained three leader-staff dyads).

Table 5 shows the results of this variance partitioning for each of the core constructs. For decision informity, 46 percent of the total variance was due to between-teams variance, whereas 54 percent was attributable to variance across decisions within teams; this variance cannot possibly be explained by between-teams factors such as experience and feedback. Experience and the feedback intervention explained 43 percent of the between-teams variance in informity, a result that is statistically significant ($F = 25.84, df = 3, 91, p < .05$).

The main effects for both experience and the feedback intervention were statistically significant. The nature of these effects indicated, as expected, that greater team experience and the presence of the feedback intervention led to greater levels of informity. In terms of magnitude, the effect size for experience (incremental $R^2 = .39$) was larger than that for the feedback intervention (incremental $R^2$
### TABLE 5
Decomposition of the Variance Explained by Experience and Feedback on Lower-Level Analogs of the Core Constructs

<table>
<thead>
<tr>
<th>Between-Teams Variance</th>
<th><strong>Within-Team Variance</strong></th>
<th>Feedback Experience Residual Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision infornity</td>
<td>54</td>
<td>1.8</td>
</tr>
<tr>
<td>Individual validity</td>
<td>60</td>
<td>1.2</td>
</tr>
<tr>
<td>Dyadic sensitivity</td>
<td>56</td>
<td>7.5</td>
</tr>
</tbody>
</table>

There was also a marginally significant interaction (incremental $R^2 = .02$, $F = 3.02$, $p = .08$) between these two variables. This interaction indicated that the feedback intervention helped inexperienced teams more than experienced ones (feedback substituted for experience on this outcome).

When examining the effects for individual validity, one needs to recall that the validity of one of the staff members was fixed by the experimental design to a level close to zero, and hence it could not vary. Thus, the effects of the noncore constructs on individual validity were only tested on the staff members who could actually alter their validity through increased effort or altered task strategies. Thus, the number of observations for this analysis is 190 (two staff members in each of 95 teams), not 285.

For individual validity, 60 percent of the total variance was due to between-teams variance (i.e., staff validity), whereas 40 percent was attributable to variance across staff members within teams (a finding that cannot possibly be explained by feedback or experience manipulated at the team level). Table 5 shows that experience and the feedback intervention explained 28 percent of the between-teams variability in this core construct, which is statistically significant ($F = 11.79$, $df = 3$, 91, $p < .05$). There were statistically significant main effects in the predicted direction for both experience and the feedback intervention, but no interactions. The magnitude of the effect for experience (incremental $R^2 = .21$) was larger than that for the feedback intervention (incremental $R^2 = .03$).

For dyadic sensitivity, 44 percent of the total variance was due to between-teams variance, whereas 56 percent was attributable to variance across the three leader-staff dyads within teams (a finding that, again, cannot possibly be explained by experience or feedback). Table 5 shows that experience and the feedback intervention explained 20 percent of the between-teams variance in this core construct, which is statistically significant ($F = 7.58$, $df = 3$, 91, $p < .05$). Of the three variables, only the feedback intervention showed a statistically significant main effect. As predicted, the presence of the feedback intervention led to enhanced hierarchical sensitivity, explaining 17 percent of the variance in this outcome.

Thus, overall, the results of these analyses were largely supportive of Hypothesis 1b and Hypothesis 2b. Feedback and experience had positive effects on overall team decision-making accuracy that were almost entirely explained by the core constructs of the multilevel theory of team decision making. The interaction of feedback and experience was not significant, however, for any outcome except team infornity. Thus, for the most part, the results of these analyses did not support the prediction that process feedback expressed in terms of the core constructs could substitute for experience (Hypothesis 2c).

### Relationships among the Core Constructs

Table 1 shows that there were statistically significant relationships among the core constructs supporting Hypothesis 3a and Hypothesis 3b. Team infornity explained 28 percent of the variance in staff validity, and staff validity in turn explained 17 percent of the variance in hierarchical sensitivity.

We tested the possibility that team infornity would interact with feedback to influence staff validity (Hypothesis 3c) with a moderated regression analysis in which staff validity was the dependent variable regressed on team infornity, the feedback intervention, and the interaction between team infornity and feedback. The results supported this prediction, in that the interaction between team infornity and the feedback intervention accounted for a statistically significant increment in $R^2$ over and above that attributable to team infornity (incremental $R^2 = .05$, $F = 6.39$, $p < .05$). When plotted, this interaction supported Hypothesis 3c. High levels of team infornity were more likely to translate into high levels of staff validity when staff members' contributions were made public via the feedback intervention than when they were less readily apparent.

We also tested the possibility that staff validity would interact with experience to influence hierarchical sensitivity (Hypothesis 3d) with an additional moderated regression analysis in which hierarchical sensitivity was the dependent variable regressed on staff validity, experience, and the interaction between these two variables. The results supported this prediction, in that the interaction between experience and staff validity was marginally significant, accounting for an incremental 3
percent of variance explained in this outcome \( F = 3.46, p = .07 \). This interaction was as predicted by Hypothesis 3d in that the deleterious effect of having a poor staff was especially strong when a team lacked experience.

Thus, in general, the results of these analyses were largely supportive of the third set of hypotheses. There were statistically significant relationships among adjacent core constructs (Hypotheses 3a and 3b), and these relationships were affected by feedback and experience (Hypotheses 3c and 3d).

**DISCUSSION**

**Replicating and Extending Research on the Multilevel Theory**

This study replicated several aspects of the initial research testing the multilevel theory of team decision making (Hollenbeck et al., 1995). The theory's core constructs explained a large percentage of the variance in overall team decision-making accuracy (83%), and the direction of the effects was as predicted. These results also replicated the earlier findings that indicated that the core constructs mediated the effects of other variables on team decision-making accuracy. That is, the core constructs explained almost all of the effect of increased experience with the task and other team members. In terms of the multilevel theory, the reason that experience is so important to these kinds of teams is that it enhances the levels of team informality and staff validity. Experience alone was not sufficient, however, to enhance hierarchical sensitivity.

These results extend previous findings by showing that the multilevel theory can be used as something more than a conceptual vehicle to parsimoniously explain results associated with the traditional variables studied in the group and team decision-making literature. These results show that one can design interventions around the core constructs of this theory and use these interventions to control team performance. In this study, teams that were provided graphic feedback on the core constructs performed better than teams that just received global performance feedback. Indeed, all the variance attributable to the feedback intervention was accounted for by the core constructs.

Although the feedback intervention had effects on all the core constructs, its main virtue was that it promoted hierarchical sensitivity. That is, the presence or absence of process feedback had a major impact on hierarchical sensitivity, which in turn had a major impact on overall team decision-making accuracy. This is an important finding because this was the one core construct that did not appear to simply improve with experience.

This observation suggests that in this type of leader-staff team, developing an optimally differentiated set of weights for staff members is difficult, even for experienced teams. This latter finding is highly consistent with findings from other programs of research on leader-staff teams or judge-advisor systems (Brehm & Hagaors, 1986; Snieszek & Buckley, 1995). In many ways, hierarchical sensitivity is the most complex of the core constructs, and identifying problems on this dimension is probably difficult for many teams. In lieu of any direct feedback, a leader and staff members probably have only a general idea of the relative value of each staff member or the weight being assigned to each staff member during real-time task engagement. This lack of information limits a team's ability to develop a highly accurate and differentiated weighting structure.

For example, as in previous studies, in this study the weights the leaders placed on staff recommendations were too high relative to the ideal. The leaders also tended to show less than ideal variability in their weighting of staff members. Thus, the leaders relied too heavily on the staff and failed to sufficiently differentiate among staff members. Apparently, the red and green bars that were provided in the feedback intervention made this weighting task much easier.

A secondary virtue of the feedback intervention was that it enhanced the teams' ability to get information distributed to the right people. That is, the indicator light that appeared when all staff members had obtained all the information they needed to make recommendations helped to focus teams on efficient information distribution. It was especially useful in highlighting team members who were missing some critical piece of information. With experience, all teams eventually reached relatively high levels of team informality; however, this aspect of the intervention was particularly helpful for teams that lacked experience with each other and the task. Thus, feedback seemed to substitute for experience when it came to team members getting efficiently informed.

The final virtue of the feedback intervention was that it enhanced staff validity. This effect was small relative to the effects on the other core variables, however. Instead, staff validity was affected mainly by experience. That is, experience led to higher staff validity, which in turn led to higher overall team performance. Apparently, whereas the red bar that this intervention used to represent validity levels was useful for identifying that a staff member was having problems, it did not provide much as-
sistance in helping this person learn how to make better judgments. The problems identified by the feedback intervention in the area of hierarchical sensitivity and team informality are probably easier to rectify than validity. A leader can always simply ignore a staff member if it is clear he or she is providing no value. Similarly, a team, through a little effort, can always figure out why one staff member is always short one piece of information and then design a distributional plan that eliminates that problem. However, whereas learning how to make valid recommendations improves over time, apparently knowing where one presently resides on the validity scale does not speed this learning process.

Relationships among the Core Constructs

Although the feedback intervention did not have a large impact on staff validity, another fact made clear by examining these data is that team informality was strongly related to this outcome ($r = .53$). Indeed, results from both this study and the earlier research on the multilevel theory (Hollenbeck et al., 1995) suggest that there is a strong, probably causal, relationship between the two variables. For example, the average correlation between team informality and staff validity in the two studies reported by Hollenbeck and colleagues (1995) was .37. A similar relationship seems to exist between staff validity and hierarchical sensitivity in this study, where the value of the correlation was .41.

Although in the initial development of the theory (Hollenbeck et al., 1995: 296–298), the core constructs are described as “emanating from lower levels to higher levels” (for instance, informality at the decision level fed into validity at the individual level), no explicit relationships among the core constructs were specified originally. The data provided here, however, suggest that propositions regarding the relationships among the core constructs should also be included in this theory.

As we noted earlier, the feedback intervention had only small direct effects on team informality and staff validity. However, when one recognizes the relationship between these latter two variables, another manner in which feedback could affect these two core constructs is that it might moderate their relationship. In line with the literature on social loafing, our results showed that accountability is one of the key factors ensuring that each team member contributes to the best of his or her ability (Miles & Greenberg, 1993; Williams et al., 1989).

The feedback intervention displayed the value of each staff member’s contribution publicly, creating potentially higher levels of accountability for teams provided process feedback on core constructs. However, given the strong relationship between team informality and staff validity, the level of validity participants could achieve was in many ways constrained by the information they had to work with when arriving at their final recommendations. Thus, whereas the accountability factor embodied in the feedback may have made staff members wish to increase their validity, the nature of the information they had limited their ability to translate this desire into results. That is, feedback led to higher levels of staff validity only when team informality was high.

Whereas the feedback intervention had little in the way of direct effects on team informality and staff validity, the experience manipulation had no direct effect on hierarchical sensitivity. However, given the relationship between staff validity and hierarchical sensitivity, another way in which experience could affect these two core constructs is moderation of their relationship. For example, as we noted earlier, given the complexity of hierarchical sensitivity, it is probably the most difficult of the core constructs for a team to manage and learn. The general tendency is for leaders to weight staff recommendations too heavily and too uniformly. This basic tendency has to be unlearned via experience. Assuming this unlearning would take time, we predicted that the negative effect on hierarchical sensitivity of having a poor staff would be especially great when a team lacked experience.

This prediction was partially confirmed by the regression results predicting hierarchical sensitivity from staff validity, experience, and the interaction between these two variables. Although experience had no main effect on hierarchical sensitivity, the interaction between experience and staff validity was marginally significant and suggested that the negative impact of a poor staff was especially strong when a team lacked experience.

Thus, as a whole, these additional analyses not only reveal strong relationships among the core constructs, but also suggest that the strength of these relationships varies under different conditions. For example, although the feedback intervention had no direct impact on staff validity, it did enhance the relationship between team informality and staff validity. Similarly, whereas experience had no direct impact on hierarchical sensitivity, it did enhance the relationship between staff validity and hierarchical sensitivity.

Limitations and Recommendations for Future Research

The nature of the sample, task, and context studied here limits the degree to which the results of
this study are generalizable to other contexts. Future research in real-world contexts might examine the degree to which other process feedback interventions aimed at highlighting a leader’s sensitivity to the real value of staff member contributions lead to better decision making in applied contexts. Validating this theory in a rigorous scientific sense would be difficult in a real-world context, however, because the time lag for determining which decisions were right and which were wrong may often be longer than what was experienced here. In addition, the unstructured nature of many of the decisions made in applied settings may create difficulties testing the theory in these contexts. However, difficulties associated with testing this theory in applied settings do not necessarily mean the theory is not valid in those contexts.

Possible applied interventions might include making leaders directly accountable for considering and publicly discussing the weights they give to staff members’ opinions. Interventions might also be targeted toward making leaders develop more accurate and objective methods for keeping track of staff recommendations, such as keeping diaries detailing the various opinions raised by their staffs and their final decisions. These records could later be compared to actual outcomes. Similar interventions might be aimed at staff members so that they will develop accurate perceptions about conditions under which they are likely to be high or low in validity, as well as when they are likely to be high or low in influence on the leader.

Of course, these interventions are predicated on the assumption that decision-making accuracy is the primary outcome of interest. As we noted at the outset, the multilevel theory is narrow in the sense that it does not place a great deal of emphasis on group outcomes other than accuracy. One avenue for future research aimed at extending the boundary conditions of this theory would deal with expanding the set of dependent variables emphasized. In some organizational contexts, other potential outcomes, such as understanding decisions, commitment to decisions, implementation of decisions, and satisfaction with a team or leader may be more critical than decision accuracy. Under these conditions, the cold, objective practice of basing the weight given each staff member solely on past validity may be counterproductive.

The fact that not all the variance in experience was explained by the core constructs also suggests the need for additional core constructs. Experience had some positive effects on overall decision-making accuracy that were not attributable to the core constructs. One possible reason for this is that the theory deals with only the vertical aspect of within-team dyads (that is, hierarchical sensitivity is a function of leader-staff dyads). Experience may also have beneficial effects on horizontal dyads (staff-to-staff relations) in terms of enhanced coordination or mutual support, which in turn leads to more accurate decision making for reasons other than the three currently posed in the theory.

Conclusion

This study replicates and extends previous research on the multilevel theory of group decision making. The results of this study indicate that (1) the constructs specified by this theory account for over half the variance in team performance and (2) teams that are provided direct feedback on the theory’s core constructs can manage these variables and use this information to improve overall team performance. This research also elaborated on the relationships among the core constructs and showed that (3) team infomory affects staff validity, especially when teams are provided feedback, and (4) staff validity affects hierarchical sensitivity, especially when teams are experienced.

REFERENCES


Bettenhausen, K. 1991. Five years of group research: What have we learned and what needs to be addressed? Journal of Management, 17: 345–381.


John R. Hollenbeck is a professor in the Management Department at Michigan State University. He received his Ph.D. in management from New York University. His current research interests include team decision-making effectiveness, self-regulation theories of work motivation, and organizational staffing practices.

Daniel R. Ilgen is the John A. Hannah Professor of Psychology and Management at Michigan State University. He received his Ph.D. in psychology from the University of Illinois. His current research interests include team decision-making effectiveness and work motivation.

Jeffrey A. LePine is a doctoral candidate in organizational behavior at Michigan State University. His primary research interests include the effects of individual differences on team performance and adaptability, as well as extra-role behaviors and role development.

Jason A. Colquitt is a doctoral candidate in organizational behavior at Michigan State University. His primary research interests include team composition, dispositional influences in work motivation, and theories of organizational justice.

Jennifer Hedlund is a postdoctoral fellow at Yale University. She received her Ph.D. in industrial and organizational psychology from Michigan State University. Her research interests include team decision-making effectiveness and theories of human intelligence.