HIERARCHICAL TEAM DECISION MAKING

Stephen E. Humphrey, John R. Hollenbeck, Christopher J. Meyer and Daniel R. Ilgen

ABSTRACT

In this paper, we review the literature on hierarchical team decision making – teams in which a formal leader makes decisions based upon the input from a staff or subordinates or other informed parties. We structure our review around the Multilevel Theory of team decision making (Hollenbeck et al., 1995), integrating the disparate works within this literature. We then provide recommendations to practitioners interested in building, maintaining, and maximizing the effectiveness of hierarchical teams. Finally, we conclude by addressing weaknesses of the literature to date and avenues for future research.

INTRODUCTION

In her now famous August 15, 2001 memo to CEO Kenneth Lay, Enron Vice-President for Corporate Development Sherron Watkins noted that, "I realize that we have a lot of smart people looking at this and a lot of our accountants have blessed the accounting treatment, but none of that will protect Enron if these transactions are ever disclosed to the light of day" (Zellner, Anderson & Cohn, 2002, p. 34). Watkins' advice to suspend the practice of engaging in misleading accounting practices was not heeded by Lay, and within five months,

Research in Personnel and Human Resources Management, Volume 21, pages 175–213.

ISBN: 0-7623-0973-3

the seventh largest corporation in the United States filed for bankruptcy, amid charges of fraud and deceit.

On May 19, 1999, Firestone CEO, Masatoshi Ono received a letter from John Hall, the president of a civil engineering firm in Florida who told him that "all four of the Firestone tires on my Explorer have failed due to tread separation problems and the last one nearly resulted in a serious accident. I address this to you because I fear that my experience cannot be unique, and as president of my own company, I would want to know (Healy, 2000)." Indeed, Firestone's own Claims Department "knew we had a very unusual amount of claims for the ATX," but no one at Firestone ever solicited advice regarding tire performance from the Claims Department, and no one in the claims department ever volunteered this information on their own. Congressional investigations later attributed 119 deaths to the ATX tire, and a series of class action suits against Firestone totaling close to $50 billion threatened its very existence.

On April 14, 1994, an AWACS crew overseeing the no-fly zone in Iraq became aware of a two helicopters that were operating in that area. Based upon their familiarity with Army routines, the AWACS crew presumed it was a pair of Blackhawks that were ferrying people from place to place, and assigned friendly blue "H" symbols to radar return that represented that helicopter. A pair of Air Force F-15's who were responsible for clearing the no-fly zone also detected the helicopters, but based upon a visual identification came to the conclusion that they were Iraqui Hinds that were violating the no-fly zone. Only after shooting down both helicopters and killing 26 people did it become clear that, in fact, they were U.S. Blackhawks carrying a United Nations delegation (Snook, 2000). Many wondered why the AWACS operator, who originally classified the helicopters as friendly, did not intervene and stop the engagement. When asked what his reaction was when the F-15s identification of the helicopters differed from his own, the Mission Control Commander stated, "My initial reaction was – Wow, this guy is good – he knows his aircraft, because not only did he say Hip, but very shortly thereafter corrected it to Hind helicopters and that meant to me – Well my initial ID may have been a mistake; now I've got them" (Andrus, 1994).

Going as far back as Adam and Eve in the Garden of Eden, human decision makers have recurrently received advice from others regarding what course of action they should pursue. Sometimes this advice is heeded, and in other occasions the recommendations are ignored. In some cases, this advice is requested, and in other cases, it arrives unsolicited. Sometimes people in a position to offer good advice say nothing, while those with less valid recommendations confidently sway the decision maker toward disaster.

In today's "information age," the increased number and complexity of choices that have to be made makes the need for "expert advice" even more critical. As Daniel Kadlec noted in a recent cover story for *Time Magazine*, "we are now responsible for so many decisions requiring so much homework that many of us feel helpless and paralyzed. The risks of inaction or unwise action are rising, even as many of the professionals on whom we would like to rely for guidance are proving untrustworthy and even corrupt" (Kadlec, 2002).

Given this state of affairs, it is clear that we need to know how individuals integrate advice and recommendations to arrive at decisions, as well as determine how to select, train, and develop decision making support staff in order to make effective choices. Most research on human decision making groups, however, has focused on how groups arrive at consensus (Ilgen, Major, Hollenbeck & Sego, 1995). For example, studies on jury decision making tends to focus on groups where people are selected for their representativeness, and then, without any formal leader, work to reach a single decision where they rarely learn whether they were right or wrong. This is important research and the jury decision making paradigm has generated a great deal of applied knowledge about choices in this context. Indeed, many people in the justice community now fear that researchers "know too much" about jury selection, in the sense that juries can be "rigged."

Few decisions in organizational contexts, however, are structured like juries. The need in organizational contexts for accountability and speed generally means that hierarchical authorities make decisions, typically after receiving input from a staff or subordinates or other informed parties. As evidence for this, it is instructive to examine one popular normative model of decision making, the Vroom-Yetton Model (1973). This decision tree identifies seven possible styles that a leader can choose to make a decision. Of these seven decision styles, three are hierarchical forms of decision making (AII, CI, and CII). When the situation has a quality consideration and the leader has insufficient information to make a decision alone, the model recommends some form of hierarchical team decision making in all possible contexts

Moreover, unlike juries, in organizational contexts these hierarchical decision making groups make a number of decisions, and these are typically evaluated as being "right or wrong" in terms of producing the desired organizational effect. Thus, most hierarchical teams have a temporal dimension, and successes and failures experienced in the past can dynamically work to influence future decision making processes and outcomes.

Some of this dynamic influence manifests itself so that past errors work forward to increase the probability of errors in the future. For example, after the 1987 U.S.S. Stark incident, in which 37 servicemen died on a vessel that failed to defend itself against a threatening aircraft, there was pressure to change the standard rules of engagement for ships operating in the Gulf. The Secretary of Defense at the time, Casper Weinberger, argued that ships should be operating "under a hair trigger alert, prepared to fire on any plane that approaches in a hostile manner" (Lamar, 1987, p. 13). Weinberger was able to convince his leader, then President Ronald Reagan, who stated that "from now on, if aircraft approach any of our ships in a way that appears hostile, there is one order of battle – defend yourselves, defend American lives" (Jacoby, 1987, p. 17). Soon after this statement was made, the U.S.S. Vincennes mistakenly shot down a passenger plane that was misjudged to represent a threat. Few military experts believe that that the Vincennes incident could have ever occurred had the Stark incident not preceded it.

In addition to creating errors, the temporal and dynamic nature of most hierarchical decision making teams also has a tremendous impact on social relations and team cohesiveness. For example, Cyrus Vance was one of the few Secretaries of State to ever resign his post. Vance resigned from the Carter Administration because the President at that time, Jimmy Carter, was increasingly rejecting his advice in favor of that of national security advisor Zbigniew Brzezinski. The culmination of this process came when Carter rejected Vance's pleas to abort the hostage rescue attempt in 1980. Vance doubted that the elaborate plan would succeed, and feared that it would undermine diplomatic efforts to obtain the release of the hostages. The rescue attempt failed, and eight servicemen died when one of the helicopters involved in the mission crashed into a transport plane in the Iranian desert. The Carter Administration never did free the hostages, and Vance referred to the day of the failed rescue attempt as "one of the most painful days of my life." After that day, he found it impossible to work with Carter, and with a mix of "sadness and frustration" he resigned his post (Berger, 2002).

There has been far less research conducted on hierarchical decision making groups, relative to consensus decision making groups, and unlike researchers in the jury decision making literature, we have little fear of being accused of "knowing too much" when it comes to the operation of this latter type of group. The purpose of this paper is to both review the body of research dealing with this topic, and based upon this existing knowledge base, make recommendations for future research and practice in this area. In this paper we use the Multilevel Theory of hierarchical team decision making (Hollenbeck et al., 1995) to organize the literature in an effort to provide parsimony. We begin by describing the Multilevel Theory of team decision making. Following this, we examine three different streams of literature on hierarchical teams, focusing on their contribution to our understanding of hierarchical team decision making. We conclude by examining some of the prescriptions for practice as well as directions for future research.

THEORY ON HIERARCHICAL TEAMS

Consider the following hierarchical teams:

(1) The President of the United States is confronted with information regarding a possible terrorist attack involving the small pox disease originating from a foreign country. The President is considering making a pre-emptive strike on the foreign country to thwart the attack. In his deliberations, he calls in his National Security Advisor, the Secretary of State, the Secretary of Defense, and the head of the Central Intelligence Agency. Each member of the staff is presented with the same information and asked to make a recommendation regarding the appropriate response by the government. The president must then make a final decision based on these recommendations.

(2) A position has opened in the management department at a university. In an effort to decide which job candidate should be hired, a team is constructed from departmental faculty. This team is led by the department chair who has the final decision making authority, but this person seeks advice from a three-person committee that includes the top researcher, the top teacher, and as well as an affirmative action officer. Each of these three staff members are charged with rating the likelihood that the candidate will make enough of a substantive contribution to the department and university mission to get tenure at this university in six years.

(3) A journal editor must decide whether to accept or reject a manuscript. The study reported in the manuscript tests a controversial theory, and the editor solicits three recommendations in an effort to determine whether the paper in question will be an influential and well-cited article, or ignored and considered trivial by the research community. One of the reviewers is a firm proponent of the theory being tested, the second is a well-known critic of this theory, and the third is a trusted and long-time editorial board member who is not really an expert in the area, but has no stake one way or the other regarding the theory.

In each of the teams described above, a specific person has individual responsibility for making a decision. Moreover, at a later point in time, this decision will be evaluated in terms of some criterion. If the President fails to strike, and thousands die from a terrorist attack, he will go down in history as having made an error. If the department head hires a person who never publishes an article, gets poor teacher evaluations and fails to contribute to the diversity of the university, he or she will be perceived as having made a mistake. If the editor publishes a manuscript that is generally ignored by the research community, the citation rate for the journal will suffer, and he or she will be blamed for diminishing the prestige of the journal.

Although the leader in all these cases has responsibility for the decision, this person does not have all the relevant knowledge, and therefore seeks advice from a set of advisors or staff members. The staff does not have the authority to make the decision, but has relevant information to bring to bear, and will be affected by eventual decision rendered. That is, although the leader will be seen as the primary culprit if there is an error, the culpability of the staff will not be ignored.

The structures of the three staffs differ in one subtle way, however. In the first case, all of the staff is presented with the same information, and although each has a unique perspective, one might expect to see a positive correlation in their recommendations. In the second case, the three different staff members serving the department head not only bring in different perspectives, but are also considering different kinds of information relevant to their recommendation. In this case, one might expect a near zero correlation between the recommendations. Finally, the journal editor, because of the manner by which reviewers were selected in this instance, may expect to see a negative correlation among the judgments of the three people entrusted with evaluating the controversial manuscript.

How does a leader combine and integrate the different recommendations of diverse staff members to arrive at an overall decision for the team, particularly in situations where the staff members disagree? Does this disagreement mean that one of the members is wrong and one is right, and therefore one should be ignored? Does this agreement mean that both staff members are right, but for different reasons, and therefore some kind of compromise is required? How do the staff members interact with the leader – and with each other – in order to insure their own influence, while at the same time promoting the long-term performance and viability of the group? If the leader directly composes the staff in a manner so that disagreement is expected, does he or she resolve that conflict differently than he or she would if the level of disagreement was unanticipated? How do the leader and the staff manage their relationships knowing full well that in the end, one's advice may be accepted, while the others may be spurned? If a decision turns out to be wrong, how does this affect the decision making process the next time there is disagreement between the staff members?

Brehmer and Hagafors (1986) noted how many important decision making teams are structured hierarchically, and they argued that given the pervasiveness of these kinds of teams, there was far too little theoretical and empirical effort directed towards them. They proposed a model of team decision making that was an adaptation of Brunswik's (1955, 1956) lens model of individual decision making. Hollenbeck et al. (1995) subsequently modified this model to form the Multilevel Theory (MLT) of team decision making. In the next section, we will briefly review Brunswik's model and show the progression from his model to the MLT used to organize the hierarchical team literature within this paper.

Individual Decision Making Model

There are several models that have been used to explain individual decision making over the years (Stevenson, Busemeyer & Naylor, 1990). Brunswik (1955, 1956) developed one model that has garnered a lot of interest, entitled the *lens model*. This model is based on his studies on perceptual constancy (Brunswik, 1940, 1943). The lens model was one of the first models to use a probabilistic approach to decision making, doing so through the use of linear regression. The basic premise of this model is that a finite set of cues can be mapped onto a decision object (Y_d) through a weighting scheme. As shown in the right-hand portion of Fig. 1, the linear weights (r_1 through r_k) that are applied

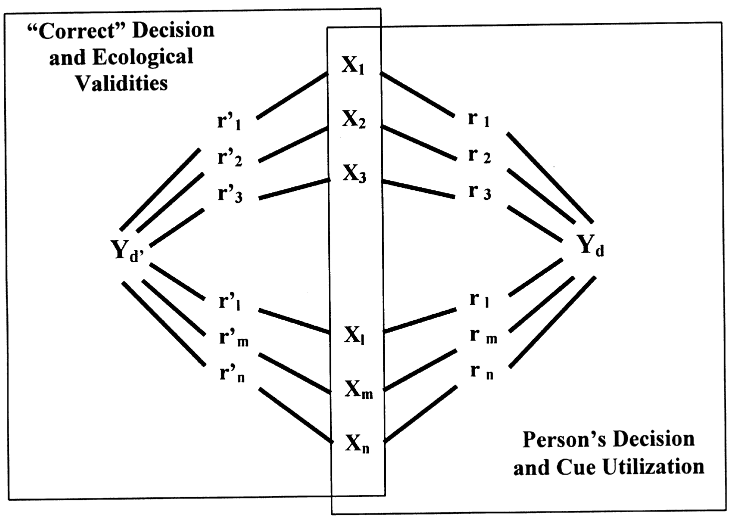

Fig. 1. The Brunswick Lens Model of Decision Making.

Adapted from "Raising an individual decisiom-making model to the team level: A new research model and paradigm," by D. R. Ilgen, D. A. Major, J. R. Hollenbeck & D. J. Sego. In: R. Guzzo & E. Salas (Eds), *Team Effectiveness and Decision Making in Organizations* (p. 126). Copyright 1995 by Jossey Bass. Reprinted with permission of John Wiley & Sons, Inc.

by the *decision maker* to the informational cues (X_1–X_k) that exist in the environment can be compared with an *optimal* weighting scheme (demonstrated in the left-hand portion). The left-hand portion of the figure, often referred to as the *ecological validity*, represents the relationship the individual cues have with the criterion to be predicted. In contrast, the right-hand portion represents how the decision maker of interest has actually utilized these cues.

In perfect conditions, decisions have been shown to adhere to the optimal model. However, decisions tend not to be made under perfect conditions. As March and Simon (1958) argued many years ago, decision makers tend to select satisfactory decisions rather than optimal ones because they cannot identify all relevant cues. Similarly, researchers have identified numerous decision biases from which decision makers suffer (c.f., Arkes, 1991). Because of all of these impediments to optimal decision making, the left-hand side and right-hand side of the figure are often widely different.

Team Lens Model

While the lens model developed by Brunswik (1955, 1956) was intended as a model of the individual decision making process, its framework has been translated to the team level. Brehmer and Hagafors (1986) presented the initial translation of this model to the team level in their study of hierarchical teams, which is demonstrated in Fig. 2. In the team-level version of this model, leaders can reduce the complexity of the decision making process by getting experts to judge a subset of the cues. For example, Fig. 2 shows a situation in which six cues are divided amongst three experts. The experts each make a recommendation based on these cues. The leader then makes a decision based on a combination of the experts' recommendations. When arriving at this decision, the leader needs only to interpret the three experts' recommendations, rather than the total set of cues, thereby reducing the information-processing requirements of the decision.

As with the individual-level lens model, the optimal model (again, the left-hand portion) can be compared with the actual decision (the right-hand portion) to determine where and how the leader deviated from optimality. Referring back to Fig. 2, the optimal decision weights (r'_1–r'_6) can be contrasted with the weights given by the experts (r_1–r_6). This would demonstrate whether the experts made valid judgments based on the cues at hand. In addition, the weights given by the leader (r_7–r_9) can be compared with an optimal aggregation of the experts cues, based on whatever recommendations were made by the experts themselves. This comparison would express the ability of the leader to correctly interpret the accuracy of the experts.

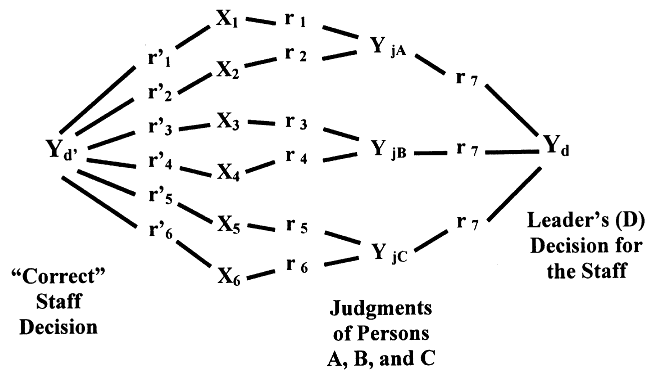

Fig. 2. The Brehmer and Hagafors Model of Staff Decision Making.

Adapted from "Raising an individual decision making model to the team level: A new research model and paradigm." by D. R. Ilgen, D. A. Major, J. R. Hollenbeck & D. J. Sego. In: R. Guzzo & E. Salas (Eds), *Team Effectiveness and Decision Making in Organizations* (p. 126). Copyright 1995 by Jossey Bass. Reprinted with permission of John Wiley & Sons, Inc.

Ilgen et al. (1995) expanded upon the initial model presented by Brehmer and Hagafors (1986) by expressing additional components of the model that those authors did not examine. First, Ilgen et al. identified that the leader may have knowledge of the cues themselves, rather than being completely dependent on the experts' recommendations. As such, the leader must decide whether to make a decision based on the experts' recommendations, the cues themselves, or a combination of both.

Secondly, Ilgen et al. (1995) expressed the communication paths that can exist in the lens model. Figure 3 demonstrates a situation in which advisor B and C have knowledge on an independent set of cues. If cues X_5 and X_6 are relevant cues for B's decision, he is unable to directly learn their values. However, due to the communication channel between them (expressed as the solid black line), advisor B can learn about cues from C directly. In contrast, if advisor A wants to know the levels of X_5 and X_6, she must communicate with D (the leader), who must ask C and then relay it back to A. The longer communication channel has a greater chance of being disrupted by noise, resulting in an inaccurate interpretation of those cues by advisor A. Therefore, an awareness of the communication channel is important in identifying why the leader's decision model deviates from the optimal model.

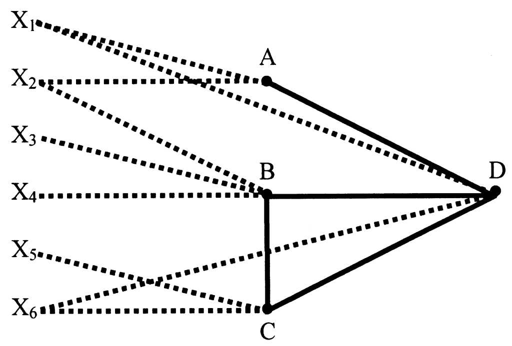

Fig. 3. A Communication Structure for a Four-Person Team.

Adapted from "Raising an individual decision making model to the team level: A new research model and paradigm." by D. R. Ilgen, D. A. Major, J. R. Hollenbeck & D. J. Sego. In: R. Guzzo & E. Salas (Eds), *Team Effectiveness and Decision Making in Organizations* (p. 126). Copyright 1995 by Jossey Bass. Reprinted with permission of John Wiley & Sons, Inc.

The Multilevel Theory of Team Decision Making

Building off of these previous works on decision making, Hollenbeck et al. (1995) developed the MLT of team decision making as a conceptual framework for analyzing decision making in these types of teams. This theory expands upon previous literature by identifying four specific levels of analysis where factors that affect hierarchical team decision making may reside. Then, in an effort to promote theoretical parsimony, the theory identifies the single most critical factor at each level of analysis that determines accuracy.

According to the MLT, the lowest level of analysis that is relevant to hierarchical teams is the *decision level*. That is, decisions are nested under individuals, in the sense that the individuals on the team each make a number of judgments or decisions, and each of these decision opportunities may vary in ways (e.g. time pressure or novelty) that affect the accuracy of the team overall. The next level is the *individual level*, where the focus is on a specific staff member. Staff members are nested within teams, in the sense that each team has multiple staff members, and variance in the characteristics of the staff members (e.g. cognitive ability or agreeableness) will be related to variance in team decision making accuracy.

Above this is the *dyadic level*, where the focus is on the one-to-one relationships between team members. For example, a four-person team can be thought of as containing six unique dyadic relationships, three of which are vertical (i.e. leader-staff) and three of which are horizontal (i.e. staff-staff).

Dyads are nested under teams in the sense that each team is comprised of multiple dyads, but characteristics of the dyadic relationships (e.g. experience working together or trust) are seen as influencing higher-level team decision making accuracy. Finally, decision making influences in hierarchical teams also occur at the *team level*, which captures variance attributable to factors unique to that level (e.g. cohesiveness or diversity) that cannot be broken down to any lower level.

Given the many different levels where important predictors of team decision making accuracy may reside, the primary problem in theory development with respect to hierarchical teams is creating a parsimonious framework. Conceivably, there are dozens of variables at each level that could be relevant to decision making processes and outcomes. However, a theory that proposed forty variables would violate all scientific norms for parsimony. The MLT addresses this problem by separating predictors into two sets of core and non-core variables. Each of the core variables of the theory is derived from a Brunswick Lens approach (Ilgen et al., 1995), and represents the single most critical factor that affects team decision making accuracy at each level of analysis. All remaining variables that might be hypothetically linked to accuracy are considered non-core variables, and their influence is primarily transmitted through the core characteristics.

Core Characteristics of the Multilevel Theory: Informity.

The lowest level of decision making is the decision level, and any team or staff member may make multiple decisions. According to this theory, the decision object manifests itself in the form of a set of cue values relevant to the staff member. However, the decision object may not provide complete information, in that it generates levels on a subset of the cues, rather than all possible cues. This means that some information presumed to be relevant to the decision making process may not be available for a specific decision object. The amount of information available about the focal decision object is known as *decision informity*.

Empirically, decision informity is the number of cue values known about the object divided by the total number of cues that are relevant for the decision. Each staff member defines what is a relevant cue differently. Returning to our previous example, the staff member on the academic job search team who is a research expert may want to know five things about the candidate including: (a) work habits; (b) theory development capabilities; (c) methodological skills; (d) access to data; and (e) writing ability. On the other hand, the teaching expert may define different relevant cues, and instead be concerned about the candidate's ability: (a) to effectively structure a course; (b) deliver engaging lectures; (c) manage effective class discussion; (d) create effective homework assignments; and (e) construct fair but demanding exams. In one case, the staff member may be fully informed, meaning he or she has access to all the information that he or she feels is relevant (e.g. if the candidate is graduating from the staff member's own alma mater). In another case (e.g. the candidate is from a foreign university), the same staff member may feel uninformed. Thus, whether one is well informed or poorly informed is determined on a decision-by-decision basis, and will vary both within the team and within the staff member over time.

It is important to note that cues determined to be relevant by one staff member are not always orthogonal to the relevant cues of another staff member. For example, work habits may be important to both the research expert and the teaching expert in the example provided above. In addition, each staff member may have a different number of cues that they deem to be relevant, such that the research expert may only require five pieces of information whereas the teaching expert deems ten cues to be relevant.

Whereas decision informity is exhibited at the decision level for each decision object, there is a parallel to decision informity at the team level. Although the level of informity may be different for each decision object, across a large number of decisions, the team as a whole may be more or less well informed. For example, an academic hiring team at one well-networked institution may be better informed about all the candidates relative to an academic hiring team that has a less well-developed network. Thus, the average level of decision level informity is considered a team-level core variable, referred to as *team informity*. Teams that, on average, know a large amount of the relevant information are highly informed (i.e. high team informity). Teams that know very little about the decision object have a low level of team informity.

Core Characteristics of the Multilevel Theory: Validity

As noted, there are a number of cues that each decision maker may find relevant. When a staff member becomes aware of the values on the cues, he or she will then process them. The staff member then makes a judgment based on these cue values. These judgments represent the individual's contribution to the team, in that he or she takes multiple pieces of information and converts these into a single recommendation. Thus, the research expert and the teaching expert, in our running example, convert the ten pieces of raw, unstandardized, and non-comparable data on each candidate into two pieces of processed, standardized, and comparable data (i.e. two general recommendations) that will be shared with the leader. The degree to which a staff member's recommendation is actually predictive of the criterion is referred to as *individual validity*.

Because the staff members are making judgments and not decisions (see Stevenson et al., 1990 for a more in-depth discussion of the distinction between judgments and decisions), there are no external consequences of the recommendation made by an individual staff member. Instead, it is up to the leader to effectively weigh each team member judgment to make an accurate decision. Interestingly, and non-intuitively, this means that even though certain staff members might make poor decisions themselves, they may nonetheless provide valuable judgments that help the leader make the correct decision. Biased judgments made by a staff member (i.e. recommendations that are off by a constant) can still be valuable if the judgments are highly correlated with the true score. Thus, staff members whose judgments are highly *negatively* correlated with the correct decision are equally as valuable as those that are highly positively correlated if they are instrumental to the leader (i.e. the leader simply makes the opposite decision of the staff member's judgment).

Returning to our previous academic example, the individual validity of each expert can be expressed via the correlation between the recommendation and the criterion. For example, let's assume that all the experts need to make a judgment on the candidate's ability to be successful on a scale of 1 (would fail miserably) to 9 (would become tenured faster than the university specified timeframe). Let us also assume that a number of years later, we learn that the three candidates' criterion scores turn out to be 3, 5, and 7. If the research expert provided judgments of 1, 3, and 5 for these three candidates, he or she has achieved an individual validity of 1.0. Thus, even though this person is too harsh in general (i.e. all estimates are two points lower than they should be), the recommendations provided by this person are still perfectly valid. If the staff member on the team that was an expert on teaching provided ratings of 5, 9, and 1, this person would have a validity of less than 1.0, and one could generally state that the first staff member was more valid than the second when it came to predicting the success of the candidates.

Although individual validity may vary among the staff members within the team, similar to decision informity, individual validity can also be aggregated to the team level. Those team level variables are useful for comparisons among teams. According to the MLT, averaging the staff members' individual validity creates *staff validity,* which represents the predictive ability of the team across all staff members. To assess staff validity, the absolute values of the individual validities are averaged. If the three individual validities were 1.0, 0.80 and −0.60, staff validity for this team would be 0.80. This means that, on average, the experts' judgments are correlated 0.80 with the criterion. This staff would be superior to the staff at another university whose validities might turn out to be 0.20, 0.25 and 0.30, representing a staff validity of 0.25, for the same set of candidates.

Core Characteristics of the Multilevel Theory: Sensitivity

The primary role of the staff is to reduce the amount of information processing that the leader has to engage in, and this is achieved by transforming many pieces of raw, unstandardized and unique information, into a standardized set of recommendations presented on a common scale. In our running example, ten pieces of raw data have been converted into two recommendations, one made by the research expert and one by the teaching expert. The affirmative action officer on the team may provide a third recommendation that is based upon five other raw pieces of data, and hence 15 pieces of data have been converted into 3 specific recommendations.

At the next stage of the decision making process, these three recommendations have to be integrated in order to arrive at a single decision (e.g. which candidate will be hired). Although the leader renders the decision, the structure of the situation ensures that there is influence and interdependence among all team members. Moreover, there is a shared team fate in this context, in the sense that everyone on the team will experience the same outcome (i.e. staff members cannot hire their own choice but must live with the ultimate choice rendered by the leader). Thus, although the leader renders the decision, the decision is best conceived of as a team, rather than individual, product.

Unless there is perfect agreement among the staff members in the process of converting the three recommendations into a single decision, the leader, who consults his or her staff, must apply some set of weights to each of their recommendations to arrive at a single judgment. For example, he or she could weigh each staff member equally, and then hire the candidate that has the highest simple average across the three recommendations. The mathematical aggregation literature has shown that this simple average consistently beats the accuracy of a single decision maker (Fischer, 1981; Libby & Blashfield, 1978).

In a review that compared mathematical and intuitive approaches to aggregation of recommendations, Clemen and Winkler (1999) concluded that complex weighting systems consistently outperformed the simple averaging of the recommendations. Thus, in an effort to improve the accuracy of the team, the leader could place a high weight on one person (e.g. the research expert), a smaller weight on another person (e.g. the teaching expert), and no weight at all to the last person (e.g. the affirmative action expert), and then select the candidate who has the highest weighted average.

Indeed, there is an infinite set of weights that could be applied to the recommendations, and in the Brunswick Lens approach, one "policy captures" the leader's strategy by regressing the leader's decision on the set of recommendations. The regression weight obtained from trying to predict the leader's decision from the staff members' recommendation provides an objective indicator of how much influence each staff member had on the ultimate decision. Obtaining an objective indicator is important in this context, because the literature on policy capturing makes it clear that people's qualitative and introspective reports of weighting strategies are generally inaccurate (Stevenson et al., 1990) relative to their actual behavior. For example, the leader may actually believe that he or she is giving equal weight to the affirmative action officer, but the policy-capturing results could indicate something very different.

Just as one can regress the leader's decision on the staff members' recommendations, after some time period, the criterion score can be obtained (i.e. six years later, the candidates success levels will actually be known), and one can regress the same set of recommendations on the criterion score. This process establishes the "ecological validity" of each of the staff members, in the sense that it shows how well one can predict the criterion from the set of recommendations.

Ideally, in an effective team, the "ecological validity equation" which documents the combinatory strategy that should be used in a normative sense, should be identical to the "policy-capturing equation" that documents the actual combinatory strategy the leader employs in a descriptive sense. In the Multilevel Theory, the ability of the leader and staff to arrive at an accurate set of weights is known as *dyadic sensitivity*. Conceptually, dyadic sensitivity can be thought of as the similarity between the weight assigned by the leader to a specific staff member's recommendation, and the ideal weight for that staff member's judgment. A high similarity between the two weights implies high sensitivity, whereas a large discrepancy implies low dyadic sensitivity.

Thus, in our running example, if the leader places a high weight on the research expert's recommendation, the dyadic sensitivity for that specific dyad within the team is high because this particular staff member was high in validity. If the leader is also placing a high weight on the teaching expert's recommendation, however, the dyadic sensitivity for that particular dyad is low, because this staff member's recommendation is low in validity. Thus, dyadic sensitivity is not a characteristic of the leader, but rather a score assigned to each vertical leader-staff member dyad. This is a dyadic construct because the staff member's behavior (e.g. aggressive self-promotion vs. passive acceptance) will have a strong influence on the leader's weighting scheme.

Although there can be within team variability on dyadic sensitivity (i.e. some leader-staff dyads do better than others), one can still aggregate across the dyads to compose a team-level variable. The team level analog to dyadic sensitivity is called *hierarchical sensitivity.* In this case, hierarchical sensitivity reflects the ability of the team as a whole to arrive at an accurate weighting scheme for all the staff members. Differences at this level imply that, averaging across individual dyads, some teams as a whole are simply better than others when it comes to accurately weighing everyone's contribution.

Note that unlike validity, which considers the staff member in isolation, the sensitivity construct (i.e. especially when operationalized via regression techniques) considers the staff as a unit. Thus, the validity for each staff member, as captured by the correlation between the staff member's judgment and the criterion, may not be the same as the unstandardized regression weight for the ecological validity equation. Moreover, the sum of the individual variance accounted for by each staff may not be the same as the overall amount of variance accounted for in the regression equation employing the three staff members. The critical determinant of how these relate is the correlation among the staff member judgments.

If for example, the three staff members have individual validities of 0.30, 0.30, and 0.30, the unstandardized regression coefficients will only equal 0.30, 0.30, and 0.30 when the three recommendations are orthogonal. If there is a positive correlation among the recommendations, the unstandardized regressions will be lower than 0.30, and if there is a negative correlation between the recommendations, the regression coefficients will be greater than 0.30. A team will probably feel more confident and cohesive when it sees positive correlations among the staff's judgments. However, this confidence is probably unwarranted. Although it may seem non-intuitive, all else equal, a staff that provides recommendations that are *negatively correlated* provides more value than a staff that provides positively correlated recommendations. Thus, in the example that leads off this section, the journal editor that seeks input from both proponents and critics of the theory being tested is specifically structuring the situation in a manner that may lead to negatively correlated recommendations, which, if properly integrated may lead to the best possible outcome.

Non-Core Constructs

The constructs discussed so far (i.e. decision and team informity, individual and staff validity, and dyadic and hierarchical sensitivity) are all termed core constructs within the Multilevel Theory. As indicated in Fig. 4, team decision making accuracy is most proximally affected by the team-level constructs,

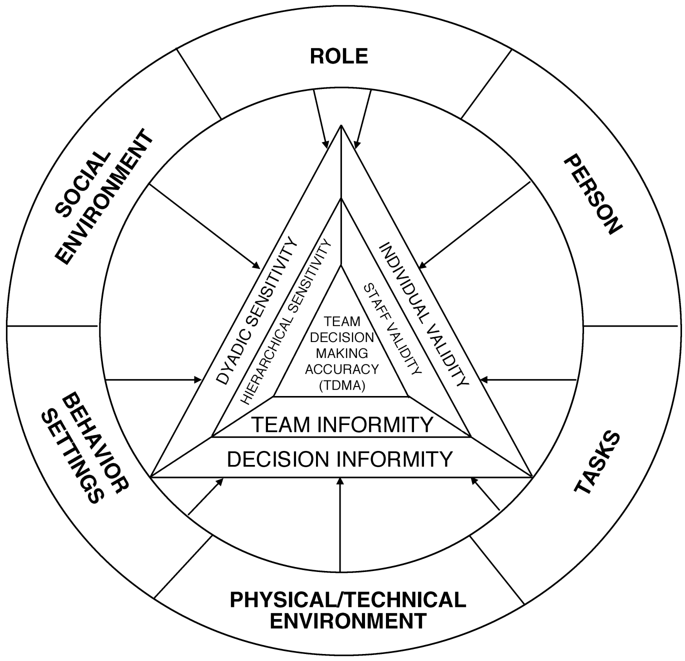

Fig. 4. Overview of the Multilevel Theory of Hierarchical Decision Making.

From "Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise," by J. R. Hollenbeck, D. R. Ilgen, D. J. Sego, J. Hedlund, D. A. Major & J. Phillips, *Journal of Applied Psychology*, *80* (p. 299). Copyright 1995 by the American Psychological Association. Reprinted with permission.

followed by the lower-level core constructs that exist at the dyadic, individual, and decision levels. All other constructs besides the six listed above fall under the category of non-core constructs. These constructs, which have been adapted from McGrath's (1976) framework, influence team decision making accuracy through their effects on the core constructs. Briefly, the categories of non-core constructs are: role, person, tasks, physical/technical environment, behavior settings, and social environment. As Fig. 4 shows, the effects of the non-core constructs on team effectiveness can often be thought of as being mediated by a specific core construct. For example, Hollenbeck et al. (1995) postulated that the characteristics of the person (e.g. cognitive ability, personality, or self-efficacy) are most likely to affect validity, whereas behavior setting (e.g. physical proximity between leader and staff members) is most likely to affect sensitivity and informity. However, whereas the figure indicates categorical links between non-core and core constructs, there may be great variability within categories between specific variables and which core constructs they impact. Because of this, much of the research on the Multilevel Theory has focused on clarifying the linkages between traditional variables studied in the groups literature and the specific core constructs of the theory.

EMPIRICAL RESEARCH ON HIERARCHICAL TEAMS

Investigation of the Core Constructs

The literature examining the aggregation of advisor judgments that we have already reviewed emerged from the mathematical aggregation paradigm (e.g. Ashton, 1986). Brehmer and Hagafors (1986) broadened this literature by examining team decision making through the lens model framework.

Brehmer and Hagafors (1986) were interested in studying hierarchical teams with distributed expertise. They were interested in the weighting process undertaken by team leaders. Specifically, they were interested in determining whether team leaders would reduce their cognitive load in the decision making process by utilizing only the staff members' recommendation, rather than relying on the cues from the environment. Borrowing from social judgment theory (Brehmer, 1986) and Brunswik's (1955) lens model, the authors built a model of hierarchical team decision making, and tested it via a laboratory simulation.

Thirty high school students were paid to act as leaders of a hierarchical team in which three experts analyzed two cues each in making a recommendation to the leader. Similar to many of the other studies presented, the authors simulated the experts rather than use actual people in those roles. Each leader made 90 decisions and was provided with feedback on their accuracy following each trial.

The participants were divided into three different conditions in which the validity of the cues and the validity of the experts varied. In the first condition (i.e. equal cues, equal validity), each cue had the same correlation with the criterion and the experts each provided recommendations that were based on the optimal weights of the cues. In the second condition (i.e. unequal cues, equal validity), the experts still optimally weighted the cues. However, the correlation between the cues and the criterion differed across the three experts. In the final condition (i.e. equal cues, unequal validity), the cues were similar to the first condition, but the experts varied in their utilization of the cues.

This study produced two general findings. First, over repeated decision making cycles, leaders can and do learn how to begin approximating optimal weighting schemes in some situations. Whereas leaders are fairly good at interpreting the equal cue, equal validity situations, they are less successful in the other two conditions. Specifically, in the unequal cue, equal validity condition, the leaders did not learn to ignore the non-valid cues while they simultaneously undervalued the highly valid cues. Similarly, in the equal cue, unequal validity situation, the leader tended to overweight the non-valid expert's recommendations instead of relying more on the cues themselves.

Secondly, the researchers concluded that in situations in which the leader has both the advisor's recommendations and the actual cues, they use both in forming their decisions. This conclusion has several implications. One of the theorized reasons for pursuing a hierarchical team is to reduce information processing demands. However, if the leaders are considering both the advisors' recommendations and the cues, they are actually doing more work then if they had considered the cues themselves. Moreover, decision makers were not able to fully differentiate between the validity of the cues and the validity of the experts, and struggled in the process to weigh both raw input and personalized recommendations simultaneously (Brehmer & Hagafors, 1986).

Brehmer and Hagafors' (1986) results demonstrated the successes and failures that leaders in hierarchical teams can experience. However, due to their use of simulated staff members rather than actual people, they did not capture the richness of the full hierarchical team experience. The next several papers start to fill in that gap, building on the steam of research that Brehmer and Hagafors (1986) began.

Although Brehmer and Hagafors (1986) explored hierarchical teams with distributed expertise in 1986, it was almost 10 years before anyone attempted to develop a formal theory of the leader/staff decision making problem. Hollenbeck et al. (1995) attempted to broaden the understanding of hierarchical team decision making by creating and testing a theory of team decision making. Their paper expanded upon Brehmer and Hagafors (1986) and Ilgen et al.'s (1995) work on the team lens model by proposing the three core constructs of team decision making: informity, validity, and sensitivity.

After building the Multilevel Theory of team decision making (which we reviewed earlier), Hollenbeck et al. (1995) tested it in two laboratory studies. In both studies, the teams participated in a simulation called TIDE,[2] in which each member of a four-person team was trained on a specific expertise. They were then presented with cues from the environment, which had to be interpreted by a particular staff member specializing in a given area. Each staff member was responsible for creating a judgment based on these cues, which the leader used to make a decision. The researchers then compared the decision made by the leader with an optimal decision, which resulted in an accuracy score (calculated in terms of mean absolute error).

In the first study, 84 college students were arranged into 21 four-person teams, with each team making a total of 127 decisions over a four-week period. The authors began by examining the core constructs of the MLT. They found that team informity explained 24% of the variance in team decision making accuracy. In addition, whereas staff validity and hierarchical sensitivity did not have a significant main effect on accuracy, they produced a significant interaction that explained an additional 20% of the variance in accuracy. The plot of the interaction showed that teams high in both of these factors performed better than teams low in either, or both. In total, the core constructs and their interactions explained 64% of the variance in accuracy.

In their second study, Hollenbeck et al. (1995) attempted to replicate and extend the results of the first study. Rather than study a small number of teams over a longer period of time and many decisions, this study examined many teams (i.e. 102 total teams) over a short period of time (i.e. only 3 hours) with only a few decisions (24 per team).

Similar to the first study, the core constructs explained a significant amount of variance in team decision making accuracy ($R^2 = 0.27$). However, as opposed to the first study in which team informity explained nearly all of the variance alone, staff validity explained nearly all of the validity in the second study ($R^2 = 0.18$). Again, there was also a significant interaction between staff validity and hierarchical sensitivity, indicating that the benefits of sensitivity are eroded at low levels of staff member validity.

The Hollenbeck et al. (1995) article demonstrated that the core constructs of the MLT were related to decision making accuracy; however, the lower explained variance in the second study showed that the effect of the core constructs was partially dependent on the reliability of the measures used, which is predominantly a function of how many decision cycles are available for analysis.

Hedlund, Ilgen, and Hollenbeck (1998) applied the Multilevel Theory as an explanatory framework to examine the effect of face-to-face communication vs. computer-mediated communication on team decision making accuracy. Sixty-four teams in a laboratory setting communicated recommendations in this exercise either through face-to-face interaction (FtF) or through computer-mediated interaction (CM). Previous studies have found that the volume and frequency of communication was much higher and different in content in FtF settings compared to CM settings (cf. Hiltz, Johnson & Turoff, 1986; McGuire, Kiesler & Seigel, 1987). Computer-mediated interaction has been associated with more task-oriented messages (Hiltz et al., 1986), lower inhibitions leading to more personal expression (including "flaming") (Dubrovsky, Kiesler & Sethna, 1991), equalization of participation (McGuire et al., 1987), and reduced status differences among members (Dubrovsky et al., 1991; Hiltz et al., 1986). Hedlund et al. (1998) proposed that the effects of medium of communication would be mediated by the three core constructs of the MLT (Hollenbeck et al., 1995) in respect to the relationship on team decision making accuracy.

In this study, the core constructs of the MLT accounted for 43% of the variance in team decision making accuracy, and these in turn were affected by the communication medium. In FtF teams, team informity and staff validity were significantly higher than in CM teams. This is consistent with the fact that FtF teams communicate greater volumes of information because they are not constrained by the technology. Hierarchical sensitivity, on the other hand, was lower in the FtF interaction than in the CM interaction. Hedlund et al. (1998) attributed this to the increased dependence on social cues by leaders in the FtF interaction. In the CM interaction, leaders did not receive social cues; rather, their decisions were based solely on the information communicated over the computer network. Because they were removed from the team, leaders were less apt to make errors of whom to weight more heavily in the decision making process.

Even though the FtF teams suffered from lower hierarchical sensitivity, decision accuracy was still significantly higher for these teams relative to the CM teams (Hedlund et al., 1998).

The implications of this study are important with respect to employing technology to maximizing decision making accuracy in leader-staff situations. On the one hand, whereas the FtF teams had a persistent performance advantage over CM teams in terms of being informed and making valid recommendations, the team did a better job of weighing opinions when they were in the CM condition. This study implies that in practice, team decisions should be made within a sequential structure that changes the communication mode over time. More specifically, in the early stages of the decision making task the staff members should communicate face-to-face, prior to making their recommendations. This would allow greater information flow between the team members. In the second stage, these recommendations should be forwarded to the leader via computer-mediated communication to prevent irrelevant social cues from distracting the weighting process.

While Hollenbeck et al. (1995) and Hedlund et al. (1998) allowed the core constructs to vary naturally, Hollenbeck, Ilgen, LePine, Colquitt and Hedlund (1998) were the first to attempt to directly manipulate the core constructs. Using 95 four-person teams, the authors attempted to replicate the effects of the core constructs on accuracy, as well as examine the role of feedback and experience in hierarchical teams. That is, this study employed a biofeedback-like paradigm, where teams were given direct, visual feedback on the level of team informity, staff validity, and hierarchical sensitivity, to see if they could use this information in a manner that would promote team decision making accuracy.

This study replicated Hollenbeck et al. (1995) and Hedlund et al. (1998), demonstrating that the core constructs of the MLT (plus the interaction between hierarchical sensitivity and staff validity) explained much (63%) of the of the variance in decision making accuracy. More incrementally, this study showed that when outcome feedback (the results themselves) was paired with process feedback (expressed in terms of the core constructs displayed in the form of an on-screen decision aid), teams were more accurate than if they were provided with outcome feedback alone. That is, teams could learn how to become more informed, make more valid recommendations, and develop more optimal weighting schemes when provided with the right feedback. Indeed, this is the first study in the history of this literature to show that teams can arrive at a complex, calibrated, and well-differentiated set of weights that approach optimality. The intervention required to achieve this end had to be precisely tailored to the variables specified by the Multilevel Theory, however, and no team could reach this end state provided outcome feedback alone.

While Hollenbeck and colleagues were examining hierarchical teams using the MLT, Sniezek and colleagues were developing a parallel approach to hierarchical team decision making termed the Judge-Advisor System (JAS) paradigm. This paradigm examined situations in which a single judge (i.e. the leader or formal decision maker) and one or more advisors (staff) provided input into a decision. This literature grew out of Sniezek and colleagues work on confidence in consensus groups (c.f. Sniezek & Henry, 1989; Sniezek & Henry, 1990), but soon expanded beyond that paradigm. However, most of this literature maintained the same focus, in that the research predominately examined how decision makers weight advisors' recommendations (i.e. they examined what impacts hierarchical sensitivity).

The initial work on the JAS conducted by Sniezek and Buckley (1995) focused on the role of confidence in hierarchical teams. In MLT terms, this research examined how staff members' confidence levels impacted hierarchical sensitivity. In this study, team members were provided with cues, which they were then responsible for using to make recommendations. In addition, they gave a measure of their confidence in their judgments. The recommendations, and under some conditions, the confidence ratings, were then passed to the decision maker. There was no appreciable difference in the performance of decision makers that received the confidence information and those who didn't receive the confidence information. In situations where the advisors were in agreement with each other, the judges showed a strong tendency to concur with the advisors. Sniezek and Buckley also found that in situations where the two advisors had conflicting recommendations, the judge most often chose to accept the recommendation of the more confident advisor, even though in reality, confidence was not strongly related to accuracy. Based on these data, the authors concluded that whereas confidence has an impact on influencing the leader (weighting), it does not always have value in promoting accurate decision making. Thus, this study showed that confidence judgments affected hierarchical sensitivity, but not necessarily accuracy.

Building on Sniezek and Buckley's (1995) work, Yaniv (1997) attempted to further clarify the process that decision makers use to assign weights to the various advisors' recommendations when making a decision. According to Yaniv (1997), two methods of advisor recommendation aggregation are *weighting* (i.e. the application of a multiplier to each recommendation before averaging) and *trimming* (i.e. a severe from of weighting where one recommendation is weighted zero, thus in effect, reducing the number of recommendations). In the weighting situation within this study, the judge applied a crude confidence indication as a weight, which was found to be more accurate than the traditional simple average.

In contrast, trimming is removing dissonant data, whether warranted or not. Yaniv (1997) observed that the decision makers engaged in trimming to resolve inconsistencies in the data. The results of this paper showed that decision makers engaged in trimming in situations in which there was outlier data. However, in situations without outlier data, trimming produced results comparable to the results produced by weighting. Based on the data, the author concluded that decision makers do not use the simple averaging method to reach decisions in hierarchical teams. Instead, they use a combination of simple averaging and trimming to produce their final decision. Although this paints a slightly more complex picture of the weighting process, in the end, both unit weighting of all members and zero weighting of some members can still be viewed as quite simple aggregation methods. Certainly, this implies that, without some type of direct process feedback like that employed by Hollenbeck et al., 1998), hierarchical teams are not finely tuned differentiators of the varied inputs that arise within such groups.

Harvey and Fischer (1997) also examined why some advisors are weighted more heavily than others. Leaders were found to be reluctant to reject recommendations, even when those making the recommendation had less information, less training, or less expertise than the leader. This finding was attributed to the desire to spread or diffuse the responsibility for a high-risk decision. Responsibility sharing was dependent not only on the risk of the task, but also on the level of expertise of the leader. That is, the leader was more likely to spread responsibility for that decision to the staff when the leader was low in confidence.

In contrast, Yaniv and Kleinberger (2000) found that decision makers discounted the opinions of others in favor of their own opinions. In this study, the weight placed on the leader's own opinion was significantly higher than the advisors when the recommendations were poor, and nearly equal to the advisors when the advice was good. Even in the instances in which the best advisor was better than the decision maker, the self-weighting of the leader's own opinion was nearly equal to the best advisor, whose weighting should be much higher. The authors suggested that self-inflated opinion bias can be attributed to the fact that an advisor's recommendation is a mere summary of one's cumulative internal knowledge, and is only a small reflection of the advisor's entire store of knowledge. On the other hand, the leader has an awareness of his or her entire internal knowledge base. Yaniv and Kleinberger concluded that knowing the history and collective information that lies behind one's own opinion biases decision makers toward that opinion.

Yaniv and Kleinberger (2000) also found evidence that the reputation of an advisor (i.e. the valuation of past success or failure of an advisor), as well as the formation of that reputation with the leader, can have an effect on the weight placed on that advisor's recommendations. When recommendations declined in quality, reputation was easily lost; however, when the quality of the recommendation improved, the weighting (and reputation) increased very slightly. Thus, it is much easier to lose reputation and trust than to gain or increase reputation and trust, and this type of trust asymmetry (Slovic, 1993) makes the advisor's job a difficult one.

Finally, Harvey, Harries and Fischer (2000) documented additional factors that influence the use of recommendations. Among these are the assessment of the quality of the recommendation (i.e. validity), and the perception of the advisors' expertise. Consistent with past research, the authors found that many leaders could discriminate the quality of staff's recommendations (i.e. the relative correlation between individual staff members judgments and the criterion). However, almost none of these leaders could apply this knowledge to arrive at a finely tuned and effective weighting scheme (i.e. the regression weights to apply to a set of judgments when predicting a criterion). This again points to the need for direct feedback on this aspect of the group decision making process (Hollenbeck et al., 1998).

Thus far, the research we have reviewed has examined the ability of the core constructs to predict decision making accuracy, as well as some of the boundary conditions within which the core constructs operate. The next section describes research that has examined the relationship between the non-core constructs and decision making accuracy.

Examination of Non-Core Constructs

A number of recent studies have examined the operation of the non-core constructs identified earlier. These studies expand previous work on the MLT of team decision making by investigating some of the more distal non-core constructs that have an impact on the decision making process. These studies examine the non-core constructs of social environment (Hollenbeck, Ilgen et al., 1998; LePine, Hollenbeck, Ilgen, Colquitt & Ellis, 2002), role (Hollenbeck, Ilgen et al., 1998), and factors within the person (Colquitt, Hollenbeck, Ilgen, LePine & Sheppard, 2002; Hollenbeck et al., 1995; Hollenbeck, Ilgen et al., 1998; LePine, Hollenbeck, Ilgen & Hedlund, 1997; Phillips, 2001; Phillips, Douthitt & Hyland, 2001; Phillips, 2002).

Both the Hollenbeck et al. (1995) and the Hollenbeck, Ilgen et al. (1998) studies, which we previously addressed, examined the effects that non-core constructs had upon decision making accuracy. In the Hollenbeck et al. (1995) study, the authors examined three non-core constructs: experience in the task, familiarity with the team members, and team member replacement. These constructs were hypothesized to influence decision making accuracy through their effects on the lower-level core constructs. The results showed that experience led to more accurate decisions, whereas familiarity and attrition of team members did not have a direct relationship with accuracy. Experience was also linked to dyadic sensitivity ($R^2 = 0.03$) and decision informity ($R^2 = 0.26$), whereas the three two-way interactions between the non-core constructs explained 9% of the variance in individual validity. These results implied that the benefits of experience were highest for unfamiliar teams that did not experience attrition. Familiarity and attrition both eroded the benefits of experience, and attrition had especially pronounced negative effects on familiar teams. Finally, the results demonstrated that the experience-accuracy relationship was almost totally mediated by the core constructs.

In the Hollenbeck, Ilgen et al. (1998) study, the authors also examined three additional non-core variables: informational redundancy (the overlap of information between team members), staff member competence, and team cohesiveness. In this study, the non-core constructs were shown to have a significant effect on accuracy ($R^2 = 0.17$), with cohesiveness and redundancy showing particularly strong effects. In general, teams that were high in informational redundancy and cohesiveness performed best, although the effects for these two non-core variables were almost completely mediated by the core constructs.

Whereas the previous two studies examined the effect of several non-core variables on accuracy and tested whether the core constructs mediated their effect, the next several papers do not examine this mediation. Instead, the following studies examined how non-core constructs directly impacted both short and long-term outcomes.

First, LePine et al. (1997) found that in a hierarchical team, it is critical that both the leader and the staff be high in conscientiousness (c) and general cognitive ability (g). High g on the part of the leader or staff was insufficient alone to bring about increased accuracy in the team decision making. That is, a low g or c could neutralize the effect of a good staff (i.e. high in c and g), and that a poor staff (i.e. low in c and g) could also neutralize the effects of a good leader (i.e. high in c and g).

Likewise, Colquitt et al. (2002) found that teams that were more open to experience were more likely to use technology to the benefit of the team in a decision making exercise. Openness to experience was shown to be a moderator of the effects of computer-assisted communication's effectiveness. More specifically, the intellect facet of openness (i.e. ideas and actions) drove this moderating effect, whereas the emotion facet of openness (i.e. feelings, aesthetics, and values) did not significantly moderate this relationship. In addition, open teams were more likely to learn the advantages of computer assisted communication and use those advantages in creative ways to increase decision making accuracy.

Phillips and colleagues (Phillips, 1999; Phillips, 2001; Phillips, Douthitt & Hyland, 2001; Phillips, 2002) have studied the effects of justice perceptions, individual team member differences, and the team leader's confidence in staff on both short-term outcomes (i.e. decision making accuracy) and long-term outcomes (i.e. team viability). Phillips (1999) examined the role of experience with a staff, staff members' past judgment accuracy, and staff members' judgment confidence on both the variance and accuracy of decision weighting by leaders of staff members' recommendations. Drawing from leader-member exchange theory (Schriesheim, Castro & Cogliser, 1999), Phillips (1999) contended that a leader's ability to differentially utilize staff member recommendations is important to team decision making accuracy. Thus, those factors that predict the variance in recommendation weightings (i.e. the range of weightings assigned by the leader), and the accuracy of these weightings, are important components of high-performance hierarchical teams.

In this study, Phillips (1999) found that as experience with a staff increases, the variance in weighting and weighting accuracy increases. Secondly, the author found that the availability of staff members' past judgment accuracy helped increase both the variability and accuracy of recommendation weighting. Third, the availability of staff members' confidence judgments was not related to either the variability or accuracy of recommendation weighting. This result stands in contrast to the results found by Sniezek and Buckley (1995), who found that confidence judgments are related to staff member utilization. Like Sniezek and Buckley, however, this study showed that the confidence of team members was unrelated to their validity, and hence we again see that confidence is a potentially distracting social cue that subverts the team's attempt to arrive at an effective weighting strategy. All too often, members who are high in validity lack confidence, and allow low validity, but highly confident, staff members to dominate the team's decision making process.

Although the Multilevel Theory advocates a cold and rational weighting of staff members based upon their predictive value, one potential negative long-term side effect of this is that a staff member who receives low weight may lose interest in the task and team, and then withdraw. Phillips (2001) was the first to study social outcomes in hierarchical teams with a specific focus on withdrawal intentions and team viability. Phillips found that team viability was positively related to performance regardless of decision influence. That is, if the team was performing well, all the staff members were satisfied, stayed engaged in the task, and were very willing to work together again in the future, regardless of the weight they were being given. Interestingly, in low-performing teams, it was the member who was most influential that was most likely to withdraw when the team performed poorly. This person seemed to internalize the blame for the team's outcomes, and sought to withdraw from the situation.

Phillips et al. (2001) attempted to both integrate the various research findings and expand our understanding of the role of decision influence, leader consideration behaviors, team decision making accuracy, justice, and long-term team outcomes. The results of this study demonstrated that a staff member's influence over a decision, the level of past team decision accuracy, and leader consideration (i.e. the belief that a leader took the staff member's recommendation seriously before making a decision) impacts a team member's perceptions of justice. All three constructs were found to be separate predictors of the justice perceptions of team members, with no interactions among the three.

The justice perceptions affected by the previous constructs lead to two long-term outcomes of hierarchical team decision making: satisfaction with the leader, and team viability. Specifically, as justice perceptions increased, team members were happier with the leader, and were more likely to want to remain with the team. Thus, this series of studies by Phillips and colleagues demonstrates that giving staff members more participation in decision making may lead to higher levels of satisfaction with the leader and team. However, there is no substitute for success when it comes to promoting positive affective reactions and cohesiveness. Clearly, influence in the face of failure is, at best, a two-edged sword.

Although past research has demonstrated that confidence judgments are not related to accuracy (c.f. Phillips, 1999), this research has not discussed why a leader would actually want staff members to provide confidence judgments. Phillips (2002) posited and found that the ability to make confidence judgments led to higher perceptions of procedural justice. Phillips also demonstrated that decision influence was related to procedural justice perceptions, consistent with Phillips et al. (2001) findings. Higher procedural justice perceptions in turn led staff members to feel higher levels of self-efficacy and greater satisfaction with the leader (Phillips, 2002). These two factors combined to reduce task withdrawal by the staff members. Similiarly, Sniezek and Van Swol (2001) showed that the advisor's level of confidence influenced the level of trust between the decision maker and the advisor. Based on these studies, it can be concluded that the ability to express confidence judgments, although perhaps detrimental to the decision making process itself, positively influences long-term outcomes of the team.

Operationalizing Decision Making Accuracy

One of the major foci of the preceding sections has been on the validity of the core constructs surrounding what we consider to be the central output of a decision making team; that is, decision making accuracy. There has been a long history of comparing individual and team accuracy (c.f., Gigone & Hastie, 1997; Hill, 1982). However, the lack of consensus across these studies on how to conceptualize accuracy has limited the development of this literature. Recently, Gigone and Hastie (1997) have provided a compelling argument for studying accuracy using the mean squared error (MSE) over traditional measures such as mean absolute error (MAE; i.e. the absolute difference between decision and true score) and the achievement correlation (r_{xy}; i.e. the linear relationship between the team decision and true score).

In their article, Gigone and Hastie (1997) demonstrated that MSE worked as well or better in many situations. This is attributed to three differences. First, MSE gives more weight to extreme errors than does MAE. Second, it is superior to r_{xy} because it does not ignore the absolute differences between judgments and the true score. Third, MSE contains more information than the other measures alone because it can be decomposed into three components (i.e. mean bias, variability bias, and the achievement correlation) that allow the researcher to pinpoint exactly why a decision is inaccurate. Mean bias can be thought of as being off by a constant in one direction (e.g. harsh or lenient, over-aggressive or under-aggressive) relative to the criterion. Variability bias, in contrast, can be conceptualized as the relative level of range in the decisions compared to the range in the criterion (e.g. the opinions are too extreme in both directions or too restricted in range relative to the criterion). Finally, the achievement correlation within MSE is the same construct as the traditional achievement correlation; that is, the correlation between the decisions and the criterion.

In response to Gigone and Hastie's (1997) article, Hollenbeck, Colquitt, Ilgen, LePine and Hedlund (1998) examined the MLT's ability to explain the three different components of accuracy. The results demonstrated that the MLT is best conceptualized as a theory of the achievement correlation in that teams can achieve ideal levels of both staff validity and hierarchical sensitivity while exhibiting both mean and variability bias. The core constructs explained 52% of the variance in the achievement correlation, while only explaining 10% of the variance in variability bias and a non-significant 2% of the variance in mean bias. Therefore, the MLT does not adequately capture the mean bias and variability bias components of accuracy as delineated by Gigone and Hastie (1997).

However, this finding does not severely limit the use of the MLT to examine hierarchical team decision making. First, Gigone and Hastie (1997) demonstrated that the most popular conceptualization of accuracy (MAE) is highly correlated with MSE ($0.84 < r < 0.95$), implying that they are still not that practically different. Secondly, as demonstrated by LePine et al. (2002), the different components of accuracy can be explained more thoroughly when the MLT is paired with different theories.

LePine et al. (2002) was primarily an examination of the role of sex composition in decision making teams, but in addressing this issue, this article noted that sex composition is likely to affect different aspects of the accuracy (i.e. achievement correlation, mean bias, and variability bias). Sex composition is an important issue because a meta-analysis by Wood (1987) found that all male teams outperformed all other team gender compositions. However, the popular press has argued that gender diversity in teams increases performance. In an effort to resolve this discrepancy, and to integrate an external theory with the MLT, LePine et al. derived hypotheses from social role theory (Eagley, 1987) to predict decision making accuracy in hierarchical teams.

Social role theory (Eagley, 1987) argues that society shapes specific expectations by which men and women are assumed to act. Men are allotted agentic characteristics, meaning that they are expected to be more assertive, controlling, aggressive and competitive than women. If social role theory is correct, a task that appears masculine in nature should lead to these sex differences manifesting themselves in hierarchical team decision making accuracy. Specifically, the aggressive components of male social expectations were expected to influence the mean bias component of accuracy. Supporting these contentions, the authors found that sex composition of the team explained 6% of the variance in mean bias.

Providing feedback on this form of bias helps improve performance, as mean bias feedback explained 16% of the variance (LePine et al., 2002). More interesting though is the interaction of sex composition and mean bias that accounted for 30% of the variability in team decision making accuracy. The interaction between feedback and sex composition showed that whereas all male teams are overly aggressive in their evaluations of decision criteria, this effect is eliminated when they are provided direct feedback on mean bias in the form of an on screen decision aid. Thus, while sex composition does affect performance, its effects can be overcome.

In addition to these results, LePine et al. (2002) found that the core constructs of the MLT explained significant variance in decision accuracy above and beyond social role theory. Specifically, the authors found that although social role theory addresses the mean bias component of accuracy, the MLT addresses the linear consistency component. Thus, when combined, MLT and social role theory can explain two components of MSE; that is, mean bias and the achievement correlation.

RECOMMENDATIONS FOR PRACTICE AND FUTURE RESEARCH

Practical Implications

Clearly, a great deal of decision making in organizations takes place within hierarchical teams, and thus, it is important for managers involved in these teams to have guidelines for optimizing their experience. The following section documents a list of general recommendations that managers may wish to consider about hierarchical teams, and some guidelines we have extracted from the literature (i.e. from theoretical deduction or empirical results). Of course, the irony of offering advice on how and when to use advice is not lost on us, nor are we unaware of the small and nascent nature of this literature. Still, we offer these points in the spirit of moving this literature forward. We invite future empirical work that might contradict these statements, and in the process, further our ability to make more valid recommendations.

First, research indicates that decision makers begin to derive benefits from a staff when there are as few as five cues that need to be monitored, and that this benefit increases as the number of cues increases. Because many complex decisions involve many more than five relevant variables, this may explain the widespread ubiquity of hierarchical teams in organizational contexts.

Second, in terms of team composition, the traditional positive effects found for conscientiousness (c) and general cognitive ability (g) across diverse contexts are also in evidence in hierarchical teams, with one critical caveat. Specifically, c and g need to be present at both the staff level and the leader level, because research shows that a poor leader can totally neutralize the effects of a good staff and vice versa. Also, all male teams tend to show a bias toward aggressiveness and risk taking that can be mitigated by feedback or by diversifying the sex composition of the team.

Third, the three sub-criteria that need to be accomplished in order to insure accurate decision making in hierarchical teams include: (a) team informity (i.e. making sure that each staff members has access to all the relevant information); (b) staff validity (i.e. the ability of each of the staff members to convert unstandardized raw data on the cues for which he or she is responsible into standardized recommendations that predict the criterion); and (c) hierarchical sensitivity (i.e. arriving at an optimal manner to weigh each staff member's input when formulating the team's decision). These three factors have explained between 25% and 65% of the variance in team decision making accuracy, and the effects for other variables such as experience, ability, cohesiveness, familiarity, attrition, informational redundancy, communication media, and feedback have all been shown to be largely mediated by these three core variables.

Fourth, cognitive ability, experience, familiarity, informational redundancy, cohesiveness, and face-to-face communication media have all been found to reliably improve team informity and staff validity. Attrition in the team's composition, however, is more disruptive to familiar and cohesive teams than teams of strangers who, with sufficient experience, seem to be able to derive more robust patterns for exchanging and converting raw information into valid recommendations.

Fifth, left to their own devices, most teams fail to arrive at an optimal scheme for integrating the diverse opinions of varied staff members, even when provided with timely and accurate outcome feedback. There are a number of reasons why teams struggle with hierarchical sensitivity. First, teams tend to use a simple, intuitive averaging approach to all of the staff members, although they will sometimes augment that by trimming outliers (i.e. giving them a weight of zero). Teams rarely give enough weight to the best member, and almost always give too much weight to the worst member.

Sixth, when variance in weights is in evidence, it tends to be based upon the staff member's confidence, rather than his or her past performance, and confidence turns out to be a poor surrogate for ability in the contexts studied to date. More weight is also given to trusted staff members, but the trust seems to be hard won, and easily lost. Indeed, the asymmetrical nature of trust may promote conservatism among the staff members, who, in order to protect themselves, generate ambiguous, two-handed (i.e. "on the one hand this, but on the other hand that") recommendations. These two-handed recommendations are of little a priori value to the leader, although they may be of great post hoc value to the staff member who can evade responsibility for bad outcomes. Variance in weights also seems to be based on status, in the sense that teams seem to treat the leader's opinion as if he or she is the best performing member, even when this is not necessarily the case.

Seventh, teams that are provided with direct feedback in the form of on screen decision aids regarding their levels of hierarchical sensitivity can dramatically improve the ability to arrive at an optimal weighting pattern. This feedback needs to include, at the very least, an objective indicator that shows how much weight is being assigned to each staff member, as well as an objective indicator that documents the validity of each staff member. This type of feedback also can be used to reduce mean bias errors that have been documented in all-male decision making teams. Communication media (e.g. computer mediated) that reduces distracting social cues and portrayals of confidence, and instead focuses the leader and staff more narrowly on the precise recommendation being offered, also can be used to promote hierarchical sensitivity.

Eighth, although a cold, rational approach to staff member weighting strategies may demand placing very low weights on some team member's judgments, this has been found to have some negative social side effects. Staff members who receive lower relative weights tend to perceive this as an injustice, reducing their satisfaction with the leader, and threatening team viability. Some of this can be mitigated by success, in the sense that if the team is viewed as being highly successful, the negative response of the disenfranchised staff member seems to be muted. Interestingly, in terms of withdrawal, the person most likely to leave the team is the individual who was weighted the most on a team that failed. Staff members on failing teams that were accorded less weight in the past, were more likely to stay, perhaps, hoping that the failure experience may reverse their fortunes in terms of being influential in the future.

Suggestions for Future Research

Relative to what is known about juries and consensus decision making groups, the short list of practical implications listed above stands as evidence that we know far less about hierarchical decision making teams. One of the areas in which we need more research is on situations where the staff members are expected to show negative correlations among their recommendations. To date, most research has focused on situations where the staff is likely to generate positively correlated opinions, or opinions that at worst, are unrelated to each other (due to distributed expertise). All else equal, however, a staff that provides valid but negatively related opinions has the best opportunity to display the type of synergy that is often hoped for in this type of context (i.e. an overall regression equation that explains more variance in the criterion relative to the simple sum of their squared correlations). We suspect that few decision making teams will be able to recognize and take advantage of this type of "statistical suppressor effect," and will instead see this as frustrating and intractable disagreement among staff members.

Second, although the descriptive literature is clear that confidence promotes influence, even when it is unfounded, to date, there has been little in the literature that would help leaders overcome this bias. The near zero correlation between confidence and performance documented in this literature implies that in some cases confidence is warranted, and in other cases it is not. What are the behavioral factors that leaders can use to discriminate false bravado from well-grounded assuredness? What are the interpersonal behaviors that highly valid but poorly weighted members engage in that limit their impact?

Third, the current research base tends to confound time with the number of decision cycles. Research that lasts longer in duration tends to have the teams go through more decision cycles; however, these two variables are not necessarily tightly linked in all environments. One team might make and receive feedback on ten decisions in one year, whereas another team might make and receive feedback on fifty decisions in the same time period. Yet another team might make and receive feedback on fifty decisions in one month. In terms of developing into an effective decision making team, in what ways do the number of cycles and time interact? More specifically, can one speed the team's development by substituting cycles for time? Training programs that force teams through a large number of decision cycles, even if in a simulated environment, might promote team development in a way that would take years in a real decision environment.

Fourth, research to date has focused on leaders who do not have any strong link to their functional staff members. However, in many real world contexts, the leader could be an ex-staff member who has been promoted to the leader's position (e.g. a Dean of a School of Business who was formerly the chair of a specific department like Management of Finance, or a U.S. President that used to be the head of the CIA). How does this person's functional knowledge affect his relationship to the staff member assigned this functional duty? Does their redundancy weaken the influence of the staff member, as the leader bypasses him or her and goes straight to the raw data, or does this common interest and frame of reference increase the amount of weight assigned to the staff member with the same functional background? How does having a specific functional background affect how the other staff members interact with the leader? Do they immediately presume they will be part of an "out-group."

Fifth, research on the Brunswick Lens Model has shown that decision makers are poor at detecting configural cues, such as interactions; however, this has not been directly studied in leader-staff contexts. This is important, because in many contexts, a staff member could be more valuable in some situations vs. others. To date, none of the studies that have examined hierarchical teams have created interactions between the staff member's validity and characteristics of the situation. For example, one staff member might be valuable when not faced with any time constraints, but prove worthless when operating under time pressure. One staff member might start off displaying low validity in the beginning, and then increase steadily in validity over time. Can leaders detect this type of moderator effect, and if not, what can be done to increase the sensitivity of the leader to changes in the value of various staff members?

Sixth, what role does agreeableness among staff members play in hierarchical teams? In consensus groups, agreeableness is quite important, because without it, it is impossible to arrive at the total level of agreement required to render a decision. At the same time, in consensus groups, agreeableness has been linked to decision making errors due to pre-mature consensus and "groupthink" type of problems. Because hierarchical teams can proceed without a strict consensus, the dual-sword nature of agreeableness may be avoidable, and one may be able to develop highly effective teams despite relatively low levels of agreeableness among staff members. This would provide a structural alternative to consensus groups when the people involved lack the necessary interpersonal traits or can be expected to disagree for political or functional reasons.

Finally, the last area for future research deals with context, in the sense that most research on hierarchical teams to date has taken place in laboratory contexts. Researchers in this area have gravitated toward the laboratory because to study accuracy, one must have a context where there are many teams facing the same decision, and the criterion manifests itself in a timely manner. For example, a set of mutual fund management teams that are assembling a portfolio will generate a large number of decisions that later can be evaluated in terms of their accuracy. However, it might take years to learn which decisions were good and which ones were bad. Similarly, a team of underwriters for an insurance company will make a large number of decisions, some of which are profitable, and some of which come back to haunt the company. However, once again, it might take years to discover which are which. Graduate student selection committees will accept and reject a large number of applicants, but the true success levels of these students will not be known for a long time.

Thus, even though there are countless "real world" teams that engage in repetitive hierarchical decision making where accuracy can be measured, for the most part, researchers have solved the "criterion problem" by using ad hoc groups in tightly controlled laboratory contexts. In these contexts, researchers can increase the speed and frequency of the decision cycle process, while relying on objective measures of accuracy and influence on the leader. If one could develop construct valid perceptual measures of accuracy and influence, this would help migrate this research from lab to field contexts. Unfortunately, those who know the policy capturing literature well recognize that people are often poor reporters of their own policies. If people lack self-insight into their own decision making processes, this limits the use of perceptual measures that, in turn, makes it difficult to conduct publishable research on this topic outside the laboratory context. Thus, the lack of valid perceptual measures becomes the single biggest hurdle to broadening the scope of research on hierarchical teams.

SUMMARY AND CONCLUSIONS

On October 7, 2001, after the ruling Taliban regime in Afghanistan refused to hand over leaders of the terrorist group al Qaeda to U.S. officials, a coalition of nations invaded Afghanistan. The mission, code named Operation Enduring Freedom was lead by General Tommy Franks. Both the mission itself, as well as Franks approach to the mission, that relied almost exclusively on heavy bombing and surgical air strikes by the U.S., combined with ground pressure from indigenous anti-Taliban forces, received criticism from a number of sources. Many felt that based upon the Soviet's experience in Afghanistan years earlier, the combination of rugged geography and intrepid warrior culture of the Afghan people would make it impossible to dislodge the Afghan government. Even among those who felt that an invasion could be successful, few thought this could be achieved without amassing a huge U.S. based ground force such as that employed in Operation Desert Storm. The indigenous anti-Taliban forces were viewed as untrustworthy and even cowardly. Even the Secretary of Defense, Donald Rumsfelt was uncomfortable with Franks' plan, stating publicly in early November that it was "too conservative" (Campbell, 2001). Despite advice to change his approach, Franks persisted with his original plan, and in less than 60 days, a new government was in place in Kabul, and Franks was being publicly praised by many of his former detractors.

In business, medical, political and military contexts, a great deal of decision making occurs in contexts where the decision maker is receiving advice from a number of different sources. The individual charged with the authority to make a decision does not have all the relevant knowledge for rendering the judgment, and the outcomes of the eventual decision will have far reaching implications beyond those for the decision maker. In these contexts, there is neither time nor a reasonable expectation that consensus will be developed. Thus, the individual must take all the available advice and attempt to integrate it in order to arrive at a decision that utilizes the unique viewpoints offered by those whose opinions are solicited (or for that matter, unsolicited).

As we showed with the three vignettes that opened this paper, decision makers often make errors in this context because they either refuse to heed the advice of their staff or take the advice of the wrong sources instead of the right ones. Problems also can occur among the staff members, who either fail to offer advice when they have relevant data, or are too quick to retract their advice when counter-advice is offered up by more confident sources. As this last vignette shows, success in some contexts may hinge critically on ignoring all the advice that is being offered.

Research is only beginning to uncover factors that promote decision making effectiveness in these kinds of hierarchical teams. In this paper, we reviewed theory and empirical research on hierarchical teams, and based upon this, generated recommendations for practice and future research. Unfortunately, relative to their ubiquitous use in practice, there is a relative paucity of research on these types of teams, especially when compared to the amount of research directed toward consensus groups such as juries. Hopefully, future research will redress this balance, and provide a much stronger theoretical and empirical base for helping improve the effectiveness of teams engaged in this type of decision making.

ACKNOWLEDGEMENTS

This research was supported in part by Grant N00014-96-1-0983 from the Cognitive and Neural Sciences Division of the Office of Naval Research. Although support for this work is gratefully acknowledged, the ideas expressed herein are those of the authors and not necessarily endorsed by the funding agency.

We would like to thank Jason Colquitt, Janet Sniezek, Jean Phillips, and Jeff LePine for helpful comments on an earlier version of this manuscript.

REFERENCES

Andrus, J. G. (1994). *AFR 110–13 aircraft accident investigation board report of investigation: U.S. army black hawk helicopters 87–26000 and 88–26060*. U.S. Air Force.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*, 486–498.

Ashton, R. H. (1986). Combining the judgments of experts – How many and which ones. *Organizational Behavior and Human Decision Processes*, *38*(3), 405–414.

Berger, M. (2002). Cyrus R. Vance, a confident to presidents, is dead at 84. *The New York Times* (January 13).

Brehmer, B. (1986). The role of judgment in small-group conflict and decision making. In: H. R. Arkes & K. R. Hammond (Eds), *Judgment and Decision Making: An Interdisciplinary Reader* (pp. 293–310). New York, NY: Cambridge University Press.

Brehmer, B., & Hagafors, R. (1986). Use of experts in complex decision making: A paradigm for the study of staff work. *Organizational Behavior and Human Decision Processes*, *38*, 181–195.

Brunswik, E. (1940). Thing constancy as measured by correlation coefficients. *Psychological Review*, *47*, 69–78

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, *50*, 255–272.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA, U.S.: University of California Press.

Campbell, D. (2001). Tommy Franks: He's no Stormin' Normin: Conservative general proves critics wrong [Electronic Version]. *The Guardian* (November 17).

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, *19*, 187–203.

Colquitt, J. A., Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., & Sheppard, L. (2002). Computer-assisted communication and team decision making performance: The moderating effect of openness to experience. *Journal of Applied Psychology*, *87*, 402–410.

Dubrovsky, V. J., Kiesler, S., & Sethna, B. N. (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision making groups. *Human Computer Interaction*, *6*, 119–146.

Eagley, A. H. (1987). *Sex differences in social behavior: A social role interpretation*. Hillsdale, NJ: Lawrence Erlbaum & Associates.

Fischer, G. W. (1981). When oracles fail – A comparison of four procedures for aggregating subjective-probability forecasts. *Organizational Behavior and Human Performance*, *28*, 96–110.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*, 149–167.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes*, *81*, 252–273.

Healy, J. R. (2000). Firestone may have known of trouble in 1994. *USA Today* (October 4).

Hedlund, J., Ilgen, D. R., & Hollenbeck, J. R. (1998). Decision accuracy in computer-mediated vs. face-to-face decision making teams. *Organizational Behavior and Human Decision Processes*, *76*, 30–47.

Hill, G. W. (1982). Group vs. individual performance: Are N+1 heads better than one? *Psychological Bulletin*, *91*, 517–539.

Hiltz, S. R., Johnson, K., & Turoff, M. (1986). Experiments in group decision making: Communication process and outcome in face-to-face vs. computerized conferences. *Human Communication Research*, *13*, 225–252.

Hollenbeck, J. R., Colquitt, J. A., Ilgen, D. R., LePine, J. A., & Hedlund, J. (1998). Accuracy decomposition and team decision making: Testing theoretical boundary conditions. *Journal of Applied Psychology*, *83*, 494–500.

Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Colquitt, J. A., & Hedlund, J. (1998). Extending the multilevel theory of team decision making: Effects of feedback and experience in hierarchical teams. *Academy of Management Journal*, *41*, 269–282.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology*, *80*, 292–316.

Ilgen, D. R., Major, D., Hollenbeck, J. R., & Sego, D. (1995). Raising an individual decision making model to the team level: A new research model and paradigm. In: R. Guzzo & E. Salas (Eds), *Team Effectiveness and decision Making In Organizations* (pp. 113–148). San Francisco: Jossey-Bass.

Jacoby, T. (1987). A tragedy in the gulf. *Newsweek*, *77* (June 1), 16–19.

Kadlec, D. (2002). You're on your own. *Time*, *159* (January 28), 24–28.

Lamar, J. (1987). Rough seas and new names. *Time*, *68* (June 29), 13.

LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., & Hedlund, J. (1997). Effects of individual differences on the performance of hierarchical decision making teams: Much more than g. *Journal of Applied Psychology*, *82*, 803–811.

LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., Colquitt, J. A., & Ellis, A. (2002). Sex composition, situational strength, and team decision making accuracy: A criterion decomposition approach. *Organizational Behavior and Human Decision Processes*, *88*, 445–475.

Libby, R., & Blashfield, R. K. (1978). Performance of a composite as a function of number of judges. *Organizational Behavior and Human Performance*, *21*, 121–129.

March, J. G., & Simon, H. A. (1958). *Organizations*. New York: John Wiley & Sons, Inc.

McGrath, J. E. (1976). Stress and behavior in organizations. In: M. D. Dunnette (Ed.), *Handbook Of Industrial and Organizational Psychology* (pp. 1310–1367). Chicago: Rand McNally.

McGuire, T. W., Kiesler, S., & Siegel, J. (1987). Group and computer-mediated discussion effects in risk decision making. *Journal of Personality and Social Psychology*, *52*, 917–930.

Phillips, J. M. (1999). Antecedents of leader utilization of staff input in decision making teams. *Organizational Behavior and Human Decision Processes*, *77*, 215–242.

Phillips, J. M. (2001). The role of decision influence and team performance in member self-efficacy, withdrawal, satisfaction with the leader, and willingness to return. *Organizational Behavior and Human Decision Processes*, *84*, 122–147.

Phillips, J. M. (2002). Antecedents and consequences of procedural justice perceptions in hierarchical decision making teams. *Small Group Research*, *33*, 32–64.

Phillips, J. M., Douthitt, E. A., & Hyland, M. M. (2001). The role of justice in team member satisfaction with the leader and attachment to the team. *Journal of Applied Psychology*, *86*, 316–325.

Schriesheim, C. A., Castro, S. L., & Cogliser, C. C. (1999). Leader-member exchange (LMX) research: A comprehensive review of theory, measurement, and data-analytic practices. *Leadership Quarterly*, *10*, 63–113.

Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, *13*, 675–682.

Sniezek, J. A., & Henry, R. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, *45*, 1–28.

Sniezek, J. A., & Henry, R. (1990). Revision, weighting, and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes*, *45*, 66–84.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*, 159–174.

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*, 288–307.

Snook, S. A. (2000). *Friendly fire: The accidental shootdown of U.S. black hawks over northern Iraq*. Princeton, NJ: Princeton University Press.

Stevenson, M. K., Busemeyer, J. R., & Naylor, J. C. (1990). Judgment and decision making theory. In: M. D. Dunnette & L. M. Hough (Eds), *Handbook Of Industrial and Organizational Psychology* (2nd ed., Vol. 1, pp. 283–374). Palo Alto, CA: Consulting Psychologists Press, Inc.

Vroom, V. H., & Yetton, P. W. (1973). *Leadership and decision making*. Pittsburgh: University of Pittsburgh Press.

Wood, W. (1987). Meta-analytic review of sex differences in group performance. *Psychological Bulletin*, *102*, 53–71.

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes*, *69*, 237–249.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281.

Zellner, W., Anderson, S. F., & Cohn, L. (2002, January 28). A hero-and a smoking-gun letter. *Business Week*, *3767*, 34–35.