Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory: A Simulation Study of Group Medical Decision Making

Franziska Tschan, Norbert K. Semmer, Andrea Gurtner, Lara Bizzari, Martin Spychiger, Marc Breuer and Stephan U. Marsch

Small Group Research 2009; 40; 271 originally published online Mar 23, 2009;
DOI: 10.1177/1046496409332928

Small Group Research
Volume 40 Number 3
June 2009 271-300

10.1177/1046496409332928
http://sgr.sagepub.com
hosted at
http://online.sagepub.com

Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory

A Simulation Study of Group Medical Decision Making

Franziska Tschan
University of Neuchâtel
Norbert K. Semmer
University of Bern
Andrea Gurtner
Berne University of Applied Sciences
Lara Bizzari
University of Neuchâtel
Martin Spychiger
Marc Breuer
Stephan U. Marsch
University Hospital, Basel

Teamwork is important in medicine, and this includes team-based diagnoses. The influence of communication on diagnostic accuracy in an ambiguous situation was investigated in an emergency medical simulation. The situation was ambiguous in that some of the patient's symptoms suggested a wrong diagnosis. Of 20 groups of physicians, 6 diagnosed the patient, 8 diagnosed with help, and 6 missed the diagnosis. Based on models of decision making, we hypothesized that accurate diagnosis is more likely if groups (a) consider more information, (b) display more explicit reasoning, and (c) talk to the room. The latter two hypotheses were supported. Additional analyses revealed that physicians often failed to report pivotal information after reading in the patient chart. This behavior suggested to the group that the chart contained no critical information. Corresponding to a transactive memory process, this process results in what we call illusory transactive memory. The plausible but incorrect diagnosis implied that the two lungs should sound differently. Despite objectively identical sounds, some physicians did hear a difference, indicating confirmation bias. Training physicians in explicit reasoning could enhance diagnostic accuracy.

Keywords: *medical emergency driven groups; group decision making; group process analysis*

A correct diagnosis is the basis of good patient care. Finding the accurate diagnosis however, is often difficult. The frequency of diagnostic errors in medicine, estimated from postmortem studies where the diagnoses on the patient chart are compared with what is found in an autopsy, range from 12% to 45% (Graber, 2005). A meta-analysis of 45 autopsy-based studies found a mean of 23.5% diagnostic inaccuracies related to the primary cause of death of the patient. In 5% to 10% of the cases, treatment of the correct disease would have led to prolonged patient survival (Graber, 2005; Shojania, Burton, McDonald, & Goldman, 2003). In a recent analysis of an emergency room, one fourth of early deaths were judged preventable (Lu et al., 2006). Not surprisingly, when hospital physicians were asked to describe their most significant professional mistake, one third named a diagnostic error (Wu, Folkman, McPhee, & Lo, 2003). In addition to the potential harm to the patient, diagnostic errors are costly for the health care system, leading to additional hospital days for patients (Vincent, Neale, & Woloshynowych, 2001). Many diagnostic errors are never detected (Croskerry & Sinclair, 2001), and studies investigating their causes and possible prevention measures are still rare (Graber, 2005; Kuhn, 2002; Norman, 2005). Given the importance of wrong diagnoses and given that more than 70% of them have been classified as preventable (Leape et al., 1991), it seems important to investigate how diagnoses are made.

Teamwork is important in medicine and this includes team-based diagnoses. Often, several physicians are involved in diagnosing a patient (Kee, Owen, & Leathem, 2004), and in an acute emergency situation, ad hoc groups of physicians may have to diagnose and treat a patient under time pressure (Marsch et al., 2003; Tschan et al., 2006; Williams, Rose, & Simon, 1999). Communication issues in groups are likely to contribute to adverse events in medicine (Gaba, 1992; Marsch et al., 2004), including problems of misdiagnosis. Nevertheless, very few studies have examined diagnostic performance of medical teams (Patel, Kaufman, & Arocha, 2002).

In this article, we study the diagnostic process of groups of physicians who are confronted with an ambiguous diagnostic situation in a simulated emergency. We investigate how aspects of team communication and individual

---

Authors' Note: This essay is based on the authors' presentation at the 2008 conference of the Interdisciplinary Network for Group Research (INGRoup), Kansas City, and is part of the special issue (Small Group Research, Volume 40, Issue 3). Please address correspondence to Franziska Tschan, Rue de la Maladière 23, Neuchâtel 2000, Switzerland; e-mail: franziska.tschan@unine.ch.

perception biases are related to the diagnostic accuracy of such groups. Based on theory and research in the areas of individual decision making in medicine and decision making in groups, we suggest that (a) groups that show more extended data gathering and more explicit reasoning will more likely avoid diagnostic errors, and (b) we expect that in groups that first consider a wrong diagnosis, a confirmation bias will be operating, and that there are group processes that will support individual bias. Among these group processes are high communication thresholds and mechanisms we call *illusory transactive memory*, which refers to the erroneous assumption of group members that another person holds important information and would communicate it if it were relevant to the problem at hand. Before we present the study, we briefly review literature on medical decision making.

Medical Diagnostic Process and Diagnostic Shortcomings

In the medical literature (e.g., Bowen, 2006; Flin, O'Connor, & Chrichton, 2008; Kuhn, 2002; Patel et al., 2002; Swartz, 2006), the ideal diagnostic process is described as containing three steps: (a) it starts with data collection, where elements of the patient history, physical exams, and other information are gathered; (b) this yields a first representation of the problem and the generation of a hypothesis, which is matched with an illness script (a schematic, often narrative representation of diagnostic elements and treatment options; Bowen, 2006); and (c) the process terminates with testing the hypothesis and the exclusion of alternate hypotheses.

Comparison of novice and expert physicians shows that experts often adopt forward reasoning strategies. Based on their knowledge, they rapidly formulate a first hypothesis, and additional information is used to confirm, refine, or reject this hypothesis (Kuhn, 2002; Patel et al., 2002). The quality of decision making depends on the expert's pertinent knowledge. Novices more often engage in backward reasoning. They start with a list of plausible hypotheses and compare each of the diagnostic information available with each of the hypotheses (Patel et al., 2002). Here, the quality of decision making depends on accurately weighting and combining information. If the situation is ambiguous even experienced physicians use a mixed forward–backward strategy (Patel & Arocha, 2001). These concepts of medical decision making include rapid and automatic processing (e.g., pattern recognition,) also described in theories of naturalistic decision making (Beach, Chi, Klein, Smith, & Vicente, 1997; Bogner, 1997). They also include aspects of traditional normative decision making, where the conscious combination and weighting of information influences the final choice (Brooks, Norman, & Allen, 1991; Kahnemann, Slovic, & Tversky, 1982). Case analyses of diagnostic errors show that multiple factors influence a misdiagnosis (Graber, 2005). They range from situational aspects such as time pressure (Croskerry & Sinclair, 2001), over *personal characteristics of physicians*, such as overconfidence (Graber, 2005) or experience (Croskerry, 2005) to *patient factors not directly related to the disease* (McKinlay, Potter, & Feldman, 1996).

Some studies have investigated what aspects of the *reasoning process* itself are related to diagnostic accuracy. In accordance with general theories on decision making (Simon, 1955), physicians often show satisficing tendencies, as indicated by omitting information, by faulty, limited, or even absent deliberate reasoning processes, and by premature closure. Kuhn (2002) reports that only 68% of the available and relevant patient information is considered before calling the diagnosis, and in an analysis of 100 diagnostic errors, Graber et al (2005) found 45 instances of faulty or incomplete data gathering. Furthermore, he found that physicians sometimes have problems combining the information available, which can lead to misinterpreting changes in the patient situation (Graber et al., 2005). This finding is corroborated by several studies that found at least partial absence of reasoning processes (Elstein & Schwarz, 2002). Based on think aloud protocols, Denig, Witteman, and Schouten (2002) found that in 48% of the cases, no, or only one, patient-related aspect was considered before a treatment was decided.

In sum, reasoning processes seem to play an important role for diagnostic performance. Although pattern recognition has been empirically established as being valuable, many of the diagnostic shortcomings seem to be rooted in an over reliance on automatic and intuitive processes and the underuse of explicit reasoning (Elstein, 2002; Eva, 2005).

Decision Making and Diagnosing in Groups

As already mentioned, diagnosing a patient sometimes occurs in teams and often under considerable time pressure—especially in emergency situations. If a team diagnoses a patient, it is the group that has to go through the reasoning process described above. Based on theories of groups as information-processing systems (Cranach von, Ochsenbein, & Valach, 1986; Hinsz, Tindale, & Vollrath, 1997; Larson & Christensen, 1993; McGrath & Tschan, 2004), we assume that at least part of the reasoning process will be communicated in the group. The main question is whether groups are likely to display mechanisms similar to those described above.

Information collection. Although one of the most often cited advantages of working in groups is the possibility to integrate expert information, a great body of research has shown that groups often fail to successfully pool the information held by different members (Stasser & Stewart, 1992). Thus, they often act on incomplete information, which may lead to suboptimal decisions. The importance of collecting more information has been corroborated in applied settings, for example in nuclear power plant teams (Waller, Gupta, & Giambatista, 2004) and in teams of physicians (Christensen et al., 2000; Franz & Larson, 2003; Larson, Christensen, Franz, & Abott, 1998). Thus, findings from individual as well as from group decision making support the importance of information collection. We therefore state the following hypothesis:

Hypothesis 1: Groups that consider a greater number of diagnostic information are more likely to find the correct diagnosis.

Decision-making process. Much of the research on group decision making in the field of small groups is of limited applicability for the situation we studied, as it has mainly concentrated on the influence of individual preferences, different tasks, or decision rules on group decision outcomes (Tindale, Kameda, & Hinsz, 2003). However, research investigating whether judgmental biases that are known to influence individual decision makers are attenuated, or exacerbated, in groups is pertinent for this study. As a general summary of this research, one can conclude that decision making in groups does not shield against such biases. On the contrary, some biases seem to be operating in groups even more strongly than in individual decision making (Houghton, Simon, Aquino, & Goldberg, 2000; Kerr & Tindale, 2004; Tindale, 1993).

As stated above, one of the problems that results in individual misdiagnoses is the lack of deliberate reasoning. Applying this thought to medical teams, we expect that communication that reflects explicit reasoning may improve the accuracy of a diagnosis. Team research thus far has stressed the importance of explicit communication for developing a shared representation (Cannon-Bowers, Salas, & Converse, 1993; Endsley, 2000; Gurtner, Tschan, Semmer, & Nägele, 2006) as well as for performance in complex situations (Waller et al., 2004). Salas, Burke, and Samman (2001) suggest that explicit communication is primordial in teams dealing with complex tasks under time pressure because it helps groups to better develop a shared representation and facilitates mutual performance monitoring (Salas, Sims, & Burke, 2005).

In a deliberate reasoning process, pieces of information are related to one another, for instance, by using causal conjunctions such as *because*, *therefore*, *if-then*, and others. By explicitly connecting different aspects, such reasoning makes it easier to detect faults because inconsistencies in reasoning become apparent.

Postulating that findings about the advantages of individual reasoning apply to groups as well, we formulate the following hypothesis:

> *Hypothesis 2:* Groups that show more explicit reasoning during the diagnostic phase are more likely to find the correct diagnosis.

A final mechanism concerns a special type of explicit communication, called *talking to the room*, which may also support building a shared representation of a problem. Talking to the room has been studied in emergency response teams (cf. Artman & Waern, 1999; Waller & Uitdewilligen, 2009). Members of those teams used undirected talk that did not address a specific person by speaking in a louder voice and stating their assessment of the situation, as it were, to the room (Artman & Waern, 1999; Waller & Uitdewilligen, 2009). Talking to the room is a way of communicating that invites other group members to participate in a mutual diagnostic process, as it may increase the chance that the group as a whole pays attention to what is said, detects problems more easily, and feels invited to come up with additional ideas (Waller & Uitdewilligen, 2009). Talking to the room may therefore have a positive influence on diagnostic performance. However, thus far evidence is limited to case studies which demonstrate that talking to the room helps groups rapidly build a common representation of the situation and update everyone if the situation is changing. We formulate the following hypothesis:

> *Hypothesis 3:* Groups that engage in talking to the room are more likely to find the correct diagnosis.

Our hypotheses allow quantitative analyses and hypothesis testing. However, we were interested in exploring more deeply those patterns and events in the interaction of the groups that may be related to diagnostic performance. We were particularly interested in problems of collecting and transmitting information, including indications of confirmation bias, and the exchange of dissenting information. Therefore we conducted additional analyses, both qualitative and quantitative, in an attempt to better understand the processes involved.

Method

Participants

Participants were 53 experienced physicians (38 males), working in groups of three (13 groups) or two (7 groups). Physicians in the same group had never worked together before. Assignment to groups was based on scheduling preferences. The physicians participated for training purposes and received feedback after completing the scenario. All group members gave their written consent to allow the use of their data in a scientific study. The duration of the scenario depended on the time it took a group to start treating the patient and ranged from 7 to 19 minutes (M = 13.4, SD = 3.1).

Simulator

We used a high-fidelity Meti (Medical Education Technologies, Inc., Sarasota, Florida) human patient simulator, which is a full-sized manikin with highly realistic features. The manikin talks (through an intercom) and blinks, and its pupils react to light. Physical signs such as chest wall motion and breathing can be observed. Pulse can be palpated at several places, and heart and lung sounds can be auscultated (the sounds of internal organs heard with a stethoscope). The manikin reacts to interventions (e.g., medication) in real time, but the skin color cannot change. For the current study, the patient was connected to a monitor that displayed blood pressure, oxygen saturation (the concentration of oxygen in the blood), and an electrocardiogram (displaying the rhythm heartbeat). The simulator was located in a hospital room with standard equipment. Two cameras recorded all activities from two different angles in the room.

Procedure and Scenario

On arrival, participants were familiarized with the simulation room and instructed in the use of the instruments. They were given the opportunity to auscultate the patient in order to familiarize themselves with normal and pathological breath and heart sounds of the manikin. They were introduced to the nurse who was present during the scenario (see below for a description of the task of the confederate nurse). The participants left the room for a short break during which the manikin was prepared for the scenario. After reentering the room, participants were greeted by a confederate emergency physician who handed the patient over to them. He told them that the patient had earlier been admitted to the emergency room and was now transferred to the ward. The handover of the patient was done according to a standardized procedure that contains giving (a) basic patient information (demographics), (b) the admittance diagnosis (left sided pneumonia), (c) some of the patient's history, and (d) the medical treatment already started (intravenous penicillin). During the handover the confederate physician's phone rang, and he told the group that he had to leave immediately for an incoming emergency. He handed the patient chart to a member of the group, remarking that all other information was to be found in the chart. Just before leaving the room, the confederate physician explained to the group that he had failed to insert a subclavian catheter (a vein access below the collarbone) on the left side, but was able to insert a catheter elsewhere, and he repeated that he had just started an infusion with penicillin. At this point, the patient made a remark about the pain involved in the failed attempt to insert the catheter. Once alone, the group's task was the surveillance of the patient and the diagnosis and treatment of any occurring problems.

Diagnostic ambiguity of the scenario. The scenario was designed to produce an ambiguous diagnostic situation. The patient was programmed to be allergic to the penicillin administered, and during the scenario he rapidly developed symptoms of a severe anaphylactic (allergic) shock. The symptoms included increasing difficulty in breathing, and a gradual increase in heart and respiratory rate, a gradual decrease in blood pressure and blood oxygen saturation, and eventually loss of consciousness. When listening with a stethoscope, breath sounds were obstructive (indicating obstruction of bronchi in the lung); the breath sounds were clearly audible and they were identical on both sides. In real patients, a severe anaphylactic shock as modeled in the scenario can lead to death within a short time.

However, some of the symptoms described above (difficulties breathing, low oxygen saturation, increased heart rate), are also likely symptoms of a tension pneumothorax (collapsed lung), a possible—albeit rare—complication from a failed attempt to insert a subclavian catheter (Mansfield, Hohn, Fornage, Gregurich, & Ota, 1994). Given the saliency of the failed subclavian catheter attempt and the gradual development of the symptoms, the situation is ambiguous in the sense that both diagnoses are possible and should be considered. On the other hand, information exists that clearly rule out the tension pneumothorax. First, a very typical symptom of a pneumothorax is sharp chest pain. When asked, the patient declined being in pain. Second, in the case of a tension pneumothorax breathing sounds are absent on the side of the collapsed lung. The manikin was programmed to display identical breath sounds on both sides of the thorax. Therefore, correct auscultation should lead to the exclusion of the pneumothorax diagnosis. Finally, the patient chart contained the patient's history, which indicated, as the seventh of eight previous diagnoses listed, a penicillin allergy. During the handover, the confederate physician mentioned that penicillin treatment had been started and the name of the drug was marked on the infusion bag.

Hints of the confederate nurse for finding the correct diagnosis. A confederate registered nurse was present in the room throughout the whole scenario. His role was to execute demands of the physicians and to help with unfamiliar material. In addition, in groups that incorrectly diagnosed a tension pneumothorax the confederate nurse was instructed to draw attention to information that pointed to the correct diagnosis. For example, the nurse would mention that the patient's skin was getting red (compensating for a shortcoming of the patient manikin in modeling an allergy). Within the context of the group's actions, he also stated that the breathing was bilateral. If the group ignored the patient chart, he drew the group's attention to it by putting it on the bed. He also mentioned that penicillin was administered. The confederate nurse's behavior was scripted in advance. However, to some degree, the confederate's behavior had to be adjusted to the situation.

For groups that did not find the correct diagnosis despite the help of the confederate nurse, the phone rang immediately after the occurrence of a short arrhythmia that indicated the patient was about to die. The group was then told that the patient's wife had called and mentioned the penicillin allergy. This final, blatant indicator was given for ethical considerations. We felt that it was not appropriate to expose a group to a patient who died because they failed to find the correct diagnosis. Furthermore, we wanted to give all groups the opportunity to treat an anaphylactic shock. The ethics committee of the Basel University Hospital approved the study protocol.

Measures

All measures are based on behavioral observation of videotapes of the simulations. Prior to coding, all communication was transcribed word for word. Communication was segmented into utterances according to a proposition made by Moesch (1990), adapted by Tschan (1995), for unitizing communication during overt action. This procedure basically distinguishes units as spoken sentences. The time of the start of each utterance was coded. In addition, selected behaviors (e.g., stopping intravenous medication; reading the patient chart) were coded based on observation. Details are provided below.

Diagnostic performance. Three levels of diagnostic performance were distinguished: (a) *correct diagnosis without help* was attributed to groups that declared the situation being an anaphylactic shock without receiving any hints from the confederate; (b) *correct diagnosis with help* was attributed to groups declaring the anaphylactic shock after having received a hint from the confederate nurse; and (c) *missed diagnosis* was attributed to groups that did not declare an anaphylactic shock, even after receiving hints by the confederate, until the allergy was communicated by phone. Two independent coders coded all 20 groups. They initially agreed on the categorization of 19 groups; the remaining disagreement was resolved by discussion.

Diagnostic information. All communication related to the diagnosis was identified; this includes asking the patient questions related to symptoms, mentioning symptoms, and mentioning aspects of the diagnosis. Again, two raters coded all groups. Initial agreement between the coders was satisfactory (kappas all > .75); remaining disagreements were resolved by discussion among coders. *Diagnoses* were based on the communication of the group; each time a group member called a diagnosis, its content (anaphylactic shock, pneumothorax, or other) was coded. The *time the first diagnosis* called was noted. *Amount of diagnostic information discussed before the first diagnosis* was calculated as the number of different diagnostic information (omitting repetitions of the same information) before the first diagnosis was called. The number of utterances indicating explicit *reasoning* was coded for each group, using a simplified version of a method developed for individual decision makers by Patel and colleagues (Kushniruk, Patel, & Fleiszer, 1995; Patel & Groen, 1986). Each time two or more diagnostic pieces of information were linked using causal conjunctions (e.g., because, therefore, if-then), this was coded as an explicit reasoning unit. For each group, the sum of reasoning units was calculated from the start of the simulation until the diagnosis was called.

Behavioral coding. For the behavioral codings described below, all groups were double-coded. Initial agreement was satisfactory (Cohen's kappa for all categories > .75). Disagreements were resolved through discussion. *Talking to the room* was coded if a physician appeared to address a communication not to a specific group member, but to the whole group before the diagnosis was called. Indications of talking to the room were absence of eye contact with other members and speaking relatively loudly. This was coded as a dummy variable (*at least once* = 1; *never* = 0). *Reading the patient chart*, or information collected from the patient chart, was coded as follows: If a group member opened the patient chart and communicated the allergy diagnosis based on the chart, this was coded as "diagnosis found in the patient chart". If a group member opened the patient chart and visibly read in it but did not communicate the allergy, this was coded as "opened patient chart, allergy not communicated." The final category refers to not opening the patient chart. *Bilateral breathing sounds auscultated* were identified for each group. For each auscultation, it was coded whether the physician commented the result, and specifically, if he or she indicated hearing the breathing sound bilaterally, or hearing a side difference. A dummy variable for each group was then calculated indicating that a side difference in auscultation was mentioned *at least once* (1) or *never* (0). In our scenario, administering epinephrine is an effective way to treat an anaphylactic shock. Therefore, the order to administer epinephrine was taken as an indicator of awareness of the correct treatment of an anaphylactic shock (*treatment applied)*, and the time of this order was noted.

Analyses

Two groups found the correct diagnosis in the patient chart very early in the process. Thus, for these two groups there is no ambiguity about the correct diagnosis, and ways of trying to resolve the ambiguity (e.g., through reasoning) could not be studied. Therefore, these two groups were eliminated from all analyses that dealt with the diagnostic process. We tested whether group size was related to diagnostic accuracy, which was not the case ($p = .43$).

Results

Six of the 20 groups (30%) correctly diagnosed the condition without any help; 8 groups (40%) diagnosed anaphylactic shock after receiving hints from the confederate, and 6 groups (30%) missed the correct diagnosis until told.

Explicit Reasoning and Diagnostic Performance

Hypothesis 1 stated that groups that considered more diagnostic information before calling a first diagnosis are more likely to find the correct diagnosis. Analysis of variance was used to compare the number of diagnostic items mentioned for the three performance groups (correct diagnosis without help, correct diagnosis with help, incorrect diagnosis). There was no significant

Table 1
Influences of Reasoning on Diagnostic Accuracy

Indicator	Correct Diagnosis (C), *M* (*SD*)	Diagnosis Found With Help (H), *M* (*SD*)	Missed Diagnosis (M), *M* (*SD*)	F/χ^2
Diagnostic information considered (#)	2.0 (0.89)	2.4 (1.0)	1.8 (0.91)	$F(2, 17) = 1.092$, nonsignificant
Explicit reasoning: # of linked utterances	4.0 (2.9)	1.13 (1.0)	1.0 (0.6)	$F(2, 15)^a = 5.750$; $p = .014$; C > H > M[b]
Talking to the room	3 of 4 groups	1 of 8 groups	0 of 6 groups	$\chi^2 = 8.598^a$; $df = 2$; $p = .007$
Combined index (reasoning and talking to the room)	1.21 (1.1)	−0.27 (0.48)	−0.45 (0.16)	$F(2, 15)^a = 11.190$; $p = .001$; C > (H = M)[b]

a. Excluding two groups that diagnosed correctly early on, based on reading the patient chart.
b. Group comparisons, Bonferroni corrected.

effect, $F(2, 17) = 1.092$, nonsignificant. Groups that found the diagnosis did not discuss more information ($M = 2.0$, $SD = 0.89$) than groups that relied on help ($M = 2.4$, $SD = 1.0$) or groups that missed the diagnosis ($M = 1.8$, $SD = 0.91$; see Table 1). The first hypothesis is not supported.

Hypothesis 2 stated that groups who show more explicit reasoning during the diagnostic phase more often find the correct diagnosis. Hypothesis 3 stated that more talking to the room should be related to better diagnostic accuracy. Indeed, results show more explicit reasoning in groups that correctly diagnosed the patient on their own ($M = 4.0$, $SD = 2.9$), compared with groups who relied on help ($M = 1.13$, $SD = 1.0$) and groups that missed the diagnosis ($M = 1.0$; $SD = 0.06$). Post hoc tests revealed significant differences between all three performance levels. This supports Hypothesis 2 (explicit reasoning). Talking to the room was generally rare. However, three out of four groups that found the diagnosis on their own did talk to the room, but only one of eight groups that found the diagnosis with help and none of the groups who missed the diagnosis, $\chi^2(2) = 8.598$, $p = .007$. This supports Hypothesis 3. Explicit reasoning and talking to the room correlated quite substantively (r = .51, p = .029). We therefore combined these two variables by performing a z transformation and adding the two values. Comparing the groups with regard to this index (Bonferroni correction) yielded significant differences between the groups who diagnosed the patient on their own and the two other performance levels (see Table 1).

Illusory Transactive Memory and Confirmation Bias

In addition to the analyses related to our hypotheses, we present two additional analyses. These are meant to illustrate phenomena that may be relevant for diagnostic performance and are the basis for suggesting new theoretical directions relevant to this topic. These analyses are both quantitative and qualitative in nature (see Edmondson & McManus, 2007, for a discussion of the use of qualitative data). The first analysis deals with the finding that only 6 of 20 groups read the diagnosis in the patient chart; we will suggest that an illusory transactive memory process may explain this phenomenon. In the second analysis, we describe a relatively frequent auditory bias observed (i.e., hearing unilateral breath sounds), which we feel fits the concept of the confirmation bias.

Recognizing and Communicating Relevant Information in the Patient Chart

The confederate physician who handed the patient over to the group gave the patient chart directly to one physician with the remark that all further information was to be found in the chart. The penicillin allergy was noted as one of eight previous diagnoses in the chart. Thus, the most effective way for the groups to find the allergy was to read the patient chart. However, in only 2 of the 20 groups did one of the physicians find this information early on. In 4 more groups, a group member declared the diagnosis based on the patient chart, but only after the group had received hints from the confederate. This observation leads to two considerations. First, it is important to determine whether physicians ignore the patient chart altogether. Second, if physicians consult the patient chart, is the diagnostic information communicated, and does communicating it have an effect on the group?

The patient chart was never opened in 2 of the 20 groups. Interestingly, both groups rapidly stated the correct diagnosis early on without receiving any help. In the other 18 groups, the chart was consulted. Thus, it seems to be the rule rather than the exception to include a patient chart as potential source of information. However, in 12 of the 18 groups that consulted the chart, the allergy information was either not found or not communicated. To find possible reasons for this lack of information gathering or transmission, we analyzed all instances of chart handling in these 12 groups.

The role of the chart handler. A first observation was that the patient chart was handled by a single person in 10 of the 12 groups throughout the entire simulation. This is obvious if the chart is consulted only once, which was the case in 4 groups. However, in six groups, the chart was consulted repeatedly, but always by the same person. For example, in Group 14 the confederate physician handed the patient chart to a physician who opened it and scanned through the pages, but put it down after 20 seconds. When another group member asked for a specific information, the confederate nurse mentioned the chart. The same physician then reopened the chart and read information to the group; this sequence was repeated again 2 minutes later. In Group 18, the physician who had received the chart opened it, looked at it briefly, and put it away. This same person opened the chart twice later on. In three groups (15, 16, 20) in which only one physician handled the chart, a second physician briefly looked over the shoulder of the chart handler. For example, in Group 16, one of the group members opened the chart briefly at the beginning and later reopened it after a hint from the confederate. He kept the chart in his hands for about 3 minutes, most of the time closed, while he was listening to the breath sounds of the patient. During the time he read the chart, another physician looked over his shoulder. Two different physicians independently opened the patient chart in only two groups (7, 17). Based on these observations, it seems that a role of chart holder seems to develop rapidly, making one physician primarily responsible for the patient chart. Interestingly, in 8 of the 10 groups where only one physician consulted the patient, it was the person who received the chart from the confederate physician.

If one person is the chart holder, this means that the other group members depend on the information from this person. Consequently, successful detection and communication of the relevant information by the chart holder may be particularly important for group performance. We therefore investigated whether, and how, information from the chart was communicated for the groups that did not find the allergy in the chart.

Limited communication about chart information. There were 12 groups in which the chart was consulted, but the consultation did not lead to the correct diagnosis. In 9 of these groups (4, 6, 7, 12, 14, 15, 18, 19, 20) the physician(s) reading in the chart did not successfully communicate any information related to chart content. This even occurred (in Group 19) when a group member explicitly asked about information from the chart. After this question, the chart holder opened the chart and read for about 45 seconds. He did not, however, communicate any chart-related information.

In two groups, there are indications that the chart holder processed information from the chart but did not explicitly communicate it to the group as a whole. After reading in the chart, the chart holder in Group 4 checked the patient's eye (an eye surgery was mentioned in the patient's history), then continued her consultation of the chart, but without verbally communicating. In Group 12, the chart holder reopened the chart after a hint from the confederate. He then immediately stepped to the intravenous line and stopped the penicillin drip, acting as if he knew the correct diagnosis. With a very low voice, he muttered to himself that the problem could be an anaphylactic shock. At the same time, however, another physician had the stethoscope in her ears and was speaking with the third physician. Obviously, the chart holder's statement was not noticed. When the group later on found the correct diagnosis and a physician ordered a discontinuation of the penicillin, the chart holder told the others that he had already stopped it, but still did not mention the allergy stated in the chart. In Group 7, the chart holder tried to get the attention of his two colleagues by pointing on the chart, but the others were busy with something else and did not react. The group discussed treating the tension pneumothorax. After a short time, the chart holder mentioned the possibility of an allergy, but did not state this information as coming from the chart.

Finally, in three more groups, the chart holder read information to the others from the chart (Groups 9, 16, 17), but without mentioning the allergy. This omission of the allergy is especially interesting because, in all cases, the information read to the group was noted on the same page as the allergy.

Based on these observations, we conclude that the failure of most groups to find the relevant information in the chart is not due to ignoring the chart as an information source altogether. On the contrary, consulting the chart is the rule rather than the exception. However, it appears that the chart holder did not find the relevant information, did not communicate it, or did not succeed in communicating it. We will argue in the discussion that these phenomena are indicative of what we call an illusory transactive memory.

Confirmation Bias in Interpreting Breathing Sounds

The second additional analysis examines the occurrence of an auditory illusion observed in several groups. Recall that breathing sounds are especially important for distinguishing between tension pneumothorax (sounds are absent on the affected lung) and anaphylactic shock (sounds are obstructive).

The patient was programmed to start with normal sounds on both lungs, which gradually became more obstructive, but at all times the respiratory sounds were present in both sides.

In all groups, one or several physicians auscultated the patient. The number of auscultations before the final diagnosis ranged from 1 to 11 (M = 4.8, SD = 2.7). Groups who missed the correct diagnosis auscultated somewhat more often, as compared with groups diagnosing the patient correctly, either by themselves or with help. The mean number of auscultations was 2.7 times per minute (SD = 1.8) for groups that missed the diagnosis, 1.2 times per minute (SD = 0.75) for groups who correctly diagnosed the patient, and 1.2 times per minute (SD = 0.72) for groups that diagnosed with help, $F(2, 15) = 3.949$, $p = .076$.

Auscultation Bias and Diagnostic Performance

All communication related to auscultation was coded with regard to the breathing sounds reported to the group. In half of the groups (10/20), at least one physician stated hearing different respiratory sounds from each side of the thorax. Because the perception of different sounds is in accordance with the hypothesis of a tension pneumothorax, we assume that this auditory illusion is due to a confirmation bias. We therefore tested whether this auditory illusion was related to diagnostic performance. It was present in only one of the groups correctly diagnosing the patient (1/6), in four groups (4/8) that diagnosed with help, but in five (5/6) groups that missed the correct diagnosis, $\chi^2(2) = 5.333$, p (one-tailed) = .035.

One of the advantages of teamwork is the possibility of feedback, mutual performance monitoring, and backup behavior among group members (Salas et al., 2005). Thus, if several physicians auscultate the patient, there is the possibility to correct an individual auditory bias. As in 9 of the 10 groups in which the auditory illusion was observed, at least two physicians auscultated the possibility for correction was given. We therefore analyzed if groups did use the potential for correcting individual errors.

Agreement on the auditory illusion. In one group (10), two physicians independently and at the same time reported hearing the same pitch differences.

Implicit disagreement. In three groups, the physician who was the first to auscultate (from now on designated as A) reported pitch differences after the first auscultation. Another physician (B) also ausculated but neither confirmed nor corrected A's interpretation. There are, however, indications that B heard something different than A, but did not contradict A. In Group 8, B initially confirms A's report of a pitch difference. However, 1 minute later, he listens again. While he is doing so, A repeats that she had heard a difference. B does not answer, but he asks if the patient's skin is red, indicating that he considers an allergic reaction. Similarly, in Group 7, B auscultates three times, each time accompanied by A repeating having heard a difference. B neither confirms nor disconfirms. After listening three times, however, B asks the patient about known allergies. In Group 11, A indicates that there is a difference in pitch and hands the stethoscope to B. B also listens, but does not comment on what he hears. However, he repeats auscultating 3 minutes later and this time he reports hearing an obstruction. Shortly afterwards, the group stops the penicillin.

Changing perceptions. In four groups, at least one physician first reports bilateral sounds, then, on a later auscultation, changes the perception to pitch differences. In Group 12, a physician communicates a bilateral sound, but 2 minutes later she auscultates again and communicates that now she hears a much clearer sound on the right side. Similarly, in Group 14, one physician declares (in a very low voice) that he hears both sides, but 2 minutes later declares that "the breathing sound is now clearly decreased on the left side," and later again states "clearly decreased respiratory sounds left." However, in this group, the same physician later on changes to the correct perception (obstructive and symmetric). After this last statement, the group starts to consider an anaphylactic shock. In two groups, one physician changes from symmetric to side differences, but declares lower pitch on the incorrect (right) side. In both groups it is noticed that this is the wrong asymmetry; however, by then quite some time has already passed.

Contagious illusions. In two groups, one physician might have been influenced by a colleague to change his perception. In these groups, one of the physicians first declares symmetric breathing sounds but is confronted with a colleague who perceives differences.

In Group 20, A listens several times, before he tells the others that he thinks there is a side difference. B starts auscultating later and communicates hearing an obstructive breath sound, but less on the left side than the right. C auscultates last and tells the others that he hears respiratory sounds on both sides, but does not get a reaction from his colleagues. One minute later, he auscultates again, and now says that indeed, there was a lower pitch on the left side.

In Group 19, A communicates breathing sounds being obstructive on both sides. B uses the stethoscope and tells the others twice that he does not hear anything. A auscultates again, and now agrees that there is a “massive problem” on one side. In this group, illusory perceptions continue. The group treats the tension pneumothorax by sticking a wide needle into the patient’s chest to release the excess air. In a case of real tension pneumothorax the air release can clearly be heard. Indeed, one of the physicians declares having heard the air release. Furthermore, he auscultates the patient immediately after the needle decompression and declares “breath sounds are back.”

Discussion

A first important result of this study is that only 6 out of 20 groups (30%) actually found the correct diagnosis on their own, and 6 others treated a tension pneumothorax until they were explicitly given the correct diagnosis. Interestingly, this percentage is in the range of diagnostic error rates found in other studies (Croskerry, 2005; Graber et al., 2005; Kuhn, 2002). It is clear that, based on our limited sample no generalization can be made with regard to error rates. However, it shows that variance in performance in such groups is important.

An alternative explanation for our results could be potential knowledge gaps with regard to the anaphylactic shock. We do not have information about the prior knowledge level of the physicians. However, we could test this alternative interpretation indirectly. If groups lacked sufficient knowledge about the anaphylactic shock, they should not be as efficient in treating the patient once the correct diagnosis is known. We therefore compared the treatment of the anaphylactic shock for the different levels of diagnostic performance. Epinephrine administration is a key treatment for an anaphylactic shock, as it helps to stabilize the failing circulation. We thus assumed that ordering to administer epinephrine would indicate the existence of an illness script of the anaphylactic shock that was at least basically correct. We therefore calculated the time that elapsed between the moment the group had the correct diagnosis and the group ordering epinephrine. An analysis of variance showed no significant difference between the groups, $F(2, 17) = .273$, nonsignificant. If anything, groups that missed the diagnosis were faster ($M = 50.83$ seconds, $SD = 118.7$) than groups that found the diagnosis with help ($M = 84.0$, $SD = 118.7$) and those who found the correct diagnosis on their own ($M = 84.6$, $SD = 0.79$). We therefore conclude that knowledge differences with regard to the anaphylactic shock cannot explain our results.

Our findings suggest that it cannot be taken for granted that medical doctors who are confronted with an ambiguous situation will find the correct diagnosis. Furthermore, working in groups does not prevent shortcomings and biases.

The Importance of Explicit Reasoning for Diagnostic Performance

We based our hypotheses on a model of medical decision making that suggests information gathering as first step, followed by elaborating and combining this information to yield a hypothesis, which can then be tested (e.g., Flin et al., 2008). Contrary to our expectations, the amount of diagnostic information considered, before calling a diagnosis, did not influence its accuracy. This seems surprising, especially against the background of research on group decision making, which shows that the amount of information discussed enhances decision accuracy in groups, including medical teams (Christensen et al., 2000; Larson et al., 1998). The number of different pieces of information discussed in our study before calling the first diagnosis was low, with a mean around 2. However, one has to consider that the groups received important information at the beginning, from the physician handing over the patient as well as spontaneously from the patient, and that this information was sufficient to consider both hypotheses. The experienced physicians may thus not have needed more cues (Patel et al., 2002). Thus, this result may not generalize to situations with less initial information or less experienced physicians.

As hypothesized, we found that more explicit reasoning (a higher degree of elaboration of information in terms of causal conjunctions) was an advantage. This finding is in accordance with studies on the diagnostic performance of individuals, which have also found a relationship between a more thorough reasoning process and diagnostic performance (Denig, 2002; Graber et al., 2005). However, researchers in the domain of naturalistic decision making show that experts, and teams of experts, often make complex decisions in high time–pressure situations in an automatic and implicit manner, and do not show an elaborate scrutiny of information (Beach et al., 1997). MacMillan, Entin, and Serfaty (2004) even argue that explicit communication and discussions may have too high of a cost in high time–pressure situations. They suggest that in emergencies, groups should strive for more implicit coordination. However, this position is contradicted by other studies. For example, Grote and Zala-Mezö (2004) found that more explicit communication was positively related to performance even for teams with highly standardized procedures. Similarly Kanki, Lozito, and Foushee (1989) demonstrated that more explicit communication was related to better handling of emergency flight situations, even if crew members had experience flying together.

Why is explicit reasoning an advantage? In many of our groups, the diagnostic information was simply stated and not explicitly related to each other. This leaves the burden of drawing conclusions to the other group members. An example of this process can be seen in Group 14 (found diagnosis with help). In this group, physician B said, as the first diagnostic information, "bilateral breath sounds," followed by physician A stating "do we have a recent x-ray?" Both statements imply a hypothesis. Bilateral breath sounds exclude a tension pneumothorax, whereas asking for an x-ray considers tension pneumothorax, as this condition can be seen in an x-ray. However, neither of the implications nor their obvious contradiction were explicitly discussed. One might assume that experienced physicians would infer the implications of each other's utterances more or less automatically. However, they may be focusing their attention on something else. Furthermore, studies on the false consensus effect (Ross, Greene, & House, 1977) suggest that people overestimate the similarity of their own thinking with that of others. The false consensus effect might explain why group members do not fully use the information that other members communicate (they fit it into their own thinking), and why group members may see no need to be more explicit in their communication (they assume that the others are thinking the same way).

Traditionally, studies in the area of group decision making have evaluated whether or not a piece of information was mentioned, but have not investigated how well the pieces of information were related to each other and how deeply they were discussed. Our study suggests that it is valuable to consider the level of explicit reasoning. This is also in accordance with a recent study (van Ginkel & van Knippenberg, 2008) that suggests including information elaboration as a factor influencing decision-making accuracy. Research along these lines is rare, however. We hope that our findings will encourage researchers to assess the level of explicit reasoning in future studies.

The Importance of Talking to the Room for Diagnostic Performance

Talking to the room is likely to invite group members to focus on an issue and to participate in a process of mutual problem solving. Talking in a louder voice signifies that the person talking is not engaging in self-talk but addressing the group, thus making it more likely that other group members will pay attention. This increases the chance that shared mental models will develop (Artman & Waern, 1999), but also that differences in mental models will be noticed. Inconsistencies in reasoning and disregard of important information are more likely to be detected by someone, increasing the chance for mutual correction. In line with the findings by Waller and Uitdewiligen (2009) we assume that talking to the room may be especially important in moments where a group has to discard an earlier hypothesis and consider alternatives.

This impersonal way of addressing the group may invite people to voice potential dissent without, however, creating an obligation for someone to answer, appearing to be lecturing, or putting someone on the spot. Thus, this aspect of talking to the room may have relational advantages.

Barriers to Extended Communication

Both explicit reasoning and talking to the room were shown to be an efficient way of communicating that was related to finding the correct diagnosis. Both variables were highly correlated and combining them into a single index clearly showed the advantage of extended communication. Given their positive effect, the question arises as to why this way of communicating does not occur more frequently. We suspect that very explicit communication may easily be seen as overly explicit, resulting in a perceived violation of important communication rules. Maxims of good communication (Grice, 1975) state that a communication should not be more informative than required. Other group members could interpret engaging in lengthy explanations as lecturing or dominating. The false consensus effect is likely to aggravate this impression, as it induces the assumption that what one is thinking is obvious to the others, at least if other group members are experts themselves. It may therefore be that physicians want to avoid behaving like a teacher towards peers of equal status. Thus, stating information without explaining it may help to avoid taking a relational risk in such a group and could by itself be a relational message (Keyton, 1999). Furthermore, explicit and extended communication carries yet another relational risk. This risk is related to its very advantage in substantive terms, which lies in the greater likelihood of errors in reasoning being detected. Exposing oneself in a way that makes possible errors more transparent carries the risk of losing face. If group members do not know each other, initial psychological safety in the group may be low (Edmondson, 1999, 2003), and group members might be more concerned with face saving than in familiar groups (Gruenfeld, Mannix, Williams, & Neale, 1996). Thus, social aspects may hinder elaborated discussions.

One way to overcome the potentially silencing effect of relational concerns in groups is the standardization of communication. For example, in aviation, some repetitive information exchanges are guided by checklists. One of the functions of these checklists is to ensure that even routine information is explicitly communicated and that all team members are informed (Degani & Wiener, 1993). In medicine, such a highly structured communication may not be feasible, as it is not adapted to the high diversity and high complexity of situations (Helmreich, 2000). However, adopting a somewhat standardized and not personalized way to update information may be useful.

Talking to the room may have some advantages over talking directly with someone with regard to relational concerns. It carries less of a risk of appearing to lecture. No one is explicitly addressed, and therefore no one is singled out as needing to be lectured to. Talking to no one directly also renders the invitation to participate more implicit and less directive than an explicit communication to a specific group member, leaving it up to the other members to decide whether or not to comment. These relational advantages suggest that talking to the room is a particularly effective strategy that can be taught. However, it still carries the risk of losing face by making faulty reasoning apparent, which may be an important reason for why it occurs infrequently. Of course, we did not directly test these considerations in our study. Rather, they represent interpretations of our data. More research is needed.

Illusory Transactive Memory

Theories of transactive memory systems (Austin, 2003; Moreland, 1999; Moreland & Myaskovsky, 2000) assume that groups often have members who specialize in certain aspects of a problem and that the group relies on these experts for specific information. This also applies to medical settings (Patel & Arocha, 2001). A well-established transactive memory system is related to higher performance (Moreland, Argote, & Krishnan, 1996). A positive impact of transactive memory systems requires that the knowledge held by a group member is accessible when needed. Unless other conditions of a well-functioning transactive memory system (trust and good collaboration; Moreland et al., 1996; Moreland & Myaskovsky, 2000; Rau, 2005) are violated, group members will most likely expect that a fellow member who holds important information will volunteer that information, if necessary. Such a behavior would correspond to the norm of sharing information with colleagues if is important for task completion

(Argyle & Henderson, 1985). Consequently, if a group member who holds specific information does not contribute or intervene, the others are likely to assume that there is no important information available. This, however, could be wrong, as our data show. There were many cases where a single physician became the chart holder but failed to communicate important information contained in the chart. At the same time, there were few attempts to independently check the chart by another group member. Thus, the group is acting *as if* it had a well-functioning transactive memory system, even though this is not the case. We call such a malfunction of the transactive memory system an *illusory transactive memory system.* An illusory transactive memory system implies a deceptive sense of security. If our interpretation is correct, the act of reading in the chart by the chart holder may be more damaging to the diagnostic process than if no one had looked at the chart. To the best of our knowledge, there is no research on this phenomenon of an illusory transactive memory system. Because an independent confirmation of our results would have far reaching consequences, we strongly suggest the need to conduct such research.

Illusions and Contagious Illusions

Breathing sounds in the patient dummy are objectively symmetric. Despite this fact, in half of the groups at least one physician stated hearing side differences. We regard such illusory perceptions as a variant of the well-known confirmation bias (Nickerson, 1998; Oswald & Grosjean, 2004). Typically, studies on confirmation bias have concentrated on selecting, interpreting, or remembering information in such a way that one's beliefs are confirmed. This also fits the observation by Patel et al. (2002) for physicians. However, our data show that the effects of a confirmation bias can go beyond attending to, interpreting, and remembering information. Confirmation bias can influence auditory perceptions and even induce auditory illusions.

The interpretation of a perceptual confirmation bias fits, at least, the behavior of those who are the first to hear a difference. For those who follow suit an alternative interpretation may lie in conformity effects (Asch, 1956). Someone may have heard the sounds correctly but decided not to communicate this observation because everyone seemed to agree on the pneumothorax diagnosis. The observation that several physicians auscultated several times before they reported hearing differences, and especially the observation that some acted in accordance with a correct perception, without however communicating this deviance, would speak for a conformity effect. This hesitance to communicate a different perception is most likely due to a fear of losing face. If this interpretation is correct, thresholds for communicating deviant perceptions, interpretations, or conclusions in medical groups may constitute a more general problem that deserves more attention (see Redelmeier, 2005, for a case study). Of course, this interpretation is somewhat speculative. However, we feel that our data are suggestive enough to highlight the desirability of a more systematic investigation of such processes. If our explanations should stand a more systematic test, they would imply a refutation of Trope and Liberman's (1996) assertion that a confirmation bias will not occur if the costs of believing in an erroneous hypothesis are high (cf. Oswald & Grosjean, 2004).

Limitations and Strengths

This study has several limitations. First, with 20 groups the sample size is relatively small. Second, all groups were confronted with the same case, which limits the generalizability of the findings to other situations. Third, the scenario could not be fully standardized because the patient's state was also influenced by how the group treated the patient. In addition, the hints provided by the confederate nurse had to be adapted to the situation.

One of the strengths of the study is that we observed decision making by professionals in a very realistic, complex situation that developed and changed in real time. We thus can assume that the communication processes we observed are similar to a real life setting. In addition, motivational problems are not likely to have played a major role in this study, as the participants were aware of the video and were expecting a video debriefing with a specialist after the session.

A special strength of the simulator setting is that it allows control of patient-related information, such as breathing sound. This enables us to unambiguously establish hearing different sounds in the two lungs as being illusory.

Practical Implications

Based on this study, we can make recommendations for groups of physicians who have to diagnose a patient under time pressure. The most important recommendations are to engage in explicit reasoning and to talk to the room. Especially when working in an unfamiliar group, the diagnostic process could profit from physicians stating the links between different observations and explicitly stating the implications of their observations. Talking to the room may have the additional advantage of focusing the group's attention on the diagnostic process and inviting other group members to participate in the process. Explicit reasoning though, may be perceived as odd, because the implied redundancy violates basic communication rules (Grice, 1975). Therefore, such deviances from conversational rules need to be trained explicitly. In fields other than medicine, communication style training is already well-established (Flin, O'Connor, & Mearns, 2002). Our research supports the development of similar training in the medical field (Flin & Maran, 2004). Medical training should also make people aware of common biases in perception and judgment, and suggest considering alternative diagnoses, even in cases that seem to be clear and simple. Finally, medical professionals should be made aware of the necessity to encourage the expression of doubt or dissent, and to establish a psychologically safe climate for doing so.

References

Argyle, M., & Henderson, M. (1985). The rules of relationships. In S. Duck & D. Perlman (Eds.), *Understanding personal relationships* (pp. 63-84). London: Sage.

Artman, H., & Waern, Y. (1999). Distributed cognition in an emergency co-ordination center. *Cognition, Technology & Work*, *4*, 237-246.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, *70* (Whole No. 416).

Austin, J. R. (2003). Transactive memory in organizational groups: The effects of content, consensus, specialization, and accuracy on group performance. *Journal of Applied Psychology*, *88*, 866-878.

Beach, L. R., Chi, M., Klein, G., Smith, P., & Vicente, K. (1997). Naturalistic decision making and related research lines. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 29-35). Mahwah, NJ: Lawrence Erlbaum.

Bogner, M. S. (1997). Naturalistic decision making in health care. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 61-69). Mahwah, NJ: Lawrence Erlbaum.

Bowen, J. L. (2006). Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine*, *355*, 2217-2225.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, *120*, 278-287.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In J. N. John Castellan (Ed.), *Individual and group decision making* (pp. 221-246). Hillsdale, NJ: Lawrence Erlbaum.

Christensen, C., Larson, J. R., Abbot, A., Ardolino, A., Franz, T., & Pfeiffer, C. (2000). Decision making of clinical teams: Communication patterns and diagnostic error. *Medical Decision Making*, *20*, 45-50.

Cranach von, M., Ochsenbein, G., & Valach, L. (1986). The group as a self-active system: Outline of a theory of group action. *European Journal of Social Psychology*, *16*, 193-229.

Croskerry, P. (2005). Diagnostic failure: A cognitive and affective approach. In *Advances in patient safety: From research to implementation* [AHRQ Publications 050021 (1-4)]. Rockville, MD: Agency for Healthcare Research and Quality.

Croskerry, P., & Sinclair, D. (2001). Emergency medicine: A practice prone to error? *Canadian Journal of Emergency Medicine*, *3*, 271-276.
Degani, A., & Wiener, E. (1993). Cockpit checklists: Concept, design and use. *Human Factors*, *35*, 28-43.
Denig, P., Witteman, C. L. M., & Schouten, H. W. (2002). Scope and nature of prescribing decisions made by general practitioners. *Quality and Safety in Health Care*, *11*, 137-143.
Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-383.
Edmondson, A. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, *40*, 1419-1452.
Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*, 1155-1179.
Elstein, A. S., & Schwarz, A. (2002). Evidence base of clinical diagnosis: Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical Journal*, *324*, 729-732.
Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum.
Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education*, *39*, 98-106.
Flin, R., & Maran, N. (2004). Identifying and training non-technical skills for teams in acute medicine. *Quality and Safety in Health Care*, *13*(Suppl. 1), i80-i84.
Flin, R., O'Connor, P., & Chrichton, M. (2008). *Safety at the sharp end. A guide to non-technical skills*. Aldershot, UK: Ashgate.
Flin, R., O'Connor, P., & Mearns, K. (2002). Crew resource management: Improving team work in high reliability industries. *Team Performance Management*, *8*, 68-78.
Franz, T., & Larson, J. R. (2003). The impact of experts on information sharing during group discussion. *Small Group Research*, *33*, 383-411.
Gaba, D. (1992). Dynamic decision-making in anesthesiology: Cognitive models and training approaches. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 123-147). Heidelberg, Germany: Springer.
Graber, M. (2005). Diagnostic errors in medicine: A case of neglect. *Journal on Quality and Patient Safety*, *31*, 106-113.
Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41-58). New York: Academic Press.
Grote, G., & Zala-Mezö, E. (2004). Patterns of adaptive coordination in cockpit crews and anaesthesia teams In R. Tartaglia, S. Baganara, T. Bellandi, & S. Albolino (Eds.), *Healthcare systems ergonomics and patient safety* (pp. 193-196). London: Taylor & Francis.
Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*, 1-15.
Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. (2006). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes*, *102*, 127-142.
Helmreich, R. L. (2000). On error management: Lessons from aviation. *British Medical Journal*, *320*, 781-785.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*, 43-64.

Houghton, S. M., Simon, M., Aquino, K., & Goldberg, C. B. (2000). No safety in numbers: Persistence of biases and their effects on team risk perception and team decision making. *Group Organization Management*, *25*, 325-353.

Kahnemann, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kanki, B. G., Lozito, S., & Foushee, H. C. (1989). Communication indices of crew coordination. *Aviation, Space, and Environmental Medicine*, *60*, 56-60.

Kee, F., Owen, T., & Leathem, R. (2004). Decision making in a multidisciplinary cancer team: Does team discussion result in better quality decisions? *Medical Decision Making*, *24*, 602-613.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, *55*, 623-655.

Keyton, J. (1999). Relational communication in groups. In L. R. Frey, D. S. Gouran, & M. S. Poole (Eds.), *The handbook of group communication, theory, and research* (pp. 192-222). Thousand Oaks, CA: Sage.

Kuhn, G. J. (2002). Diagnostic errors. *Academic Emergency Medicine*, *9*, 740-750.

Kushniruk, A. W., Patel, V. L., & Fleiszer, D. M. (1995). Complex decision making in providing surgical intensive care. In *Proceedings of the seventeenth annual conference of the Cognitive Science Society* (pp. 287-292). Hillsdale, NJ: Lawrence Erlbaum.

Larson, J. R., Christensen, C., Franz, T. M., & Abott, A. S. (1998). Diagnosing groups: The pooling, management, and impact of shared and unshared case information in team-based medical decision making. *Journal of Personality and Social Psychology*, *75*, 93-108.

Larson, J. R. J., & Christensen, C. (1993). Group as problem-solving units: Toward a new meaning of social cognition. *British Journal of Social Psychology*, *32*, 5-30.

Leape, L. L., Brennan, T. A., Laird, N. M., Lawthers, A. G., Localio, A. R., Barnes, B. A., et al. (1991). The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *New England Journal of Medicine*, *324*, 377-384.

Lu, T. C., Tsai, C. L., Lee, C. C., Ko, P. C. I., Yen, Z. S., Yuan, A., et al. (2006). Preventable deaths in patients admitted from emergency department. *Emergency Medical Journal*, *23*, 452-455.

MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Process and performance at the inter- and intra-individual level* (pp. 61-81). Washington, DC: American Psychological Association.

Mansfield, P. F., Hohn, D. C., Fornage, B. D., Gregurich, M. A., & Ota, D. M. (1994). Complications and failures of subclavian-vein catheterization. *New England Journal of Medicine*, *26*, 1735-1738.

Marsch, S. U., Hunziker, P., Spychiger, M., Breuer, N., Semmer, N., & Tschan, F. (2004). *Teambuilding delays crucial measures in simulated cardiac arrests.* Paper presented at the Schweizerische Gesellschaft für Intensivmedizin, Interlaken, Switzerland.

Marsch, S. U., Müller, C., Marquardt, K., Conrad, G., Tschan, F., & Hunziker, P. R. (2003). Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. *Resuscitation*, *60*, 51-56.

McGrath, J. E., & Tschan, F. (2004). Dynamics in groups and teams: Groups as complex action systems. In M. S. Poole & A. H. van de Ven (Eds.), *Handbook of organizational change and development* (pp. 50-73). Oxford, UK: Oxford University Press.

McKinlay, J. B., Potter, D. A., & Feldman, H. A. (1996). Non-medical influences on medical decision-making. *Social Science & Medicine*, *42*, 769-776.

Moesch, K. (1990). *Praedikatzentrierte Einheiten als Basissegmente zur Analyse von Diskursen: ein pragmatischer Loesungsversuch*. [Predicate-centered unitizing for the analyses of discourse. A pragmatic solution]. Unpublished manuscript, Psychologisches Institut, Universität Bern, Switzerland.

Moreland, R. L. (1999). Transactive memory: Learning who knows what in work groups and organizations. In L. L. Thompson, J. M. Levine, & D. M. Messik (Eds.), *Shared cognition in organizations* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum.

Moreland, R. L., Argote, L., & Krishnan, R. (1996). Socially shared cognition at work. Transactive memory and group performance. In J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition?* (pp. 57-84). Thousand Oaks, CA: Sage.

Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior and Human Decision Processes*, *82*, 117-133.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220.

Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education*, *39*, 418-427.

Oswald, M., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook of fallacies and biases in thinking, judgment and memory* (pp. 79-95). New York: Psychology Press.

Patel, V. L., & Arocha, J. F. (2001). The nature of constraints on collaborative decision making in health care settings. In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 383-405). Mahwah, NJ: Lawrence Erlbaum.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, *10*, 91-116.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision making. *Journal of Biomedical Informatics*, *35*, 52-75.

Rau, D. (2005). The influence of relationship conflict and trust on the transactive memory: Performance relation in top management teams. *Small Group Research*, *36*, 746-771.

Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annuals of Internal Medicine*, *142*, 115-120.

Ross, C. E., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279-301.

Salas, E., Burke, C. S., & Samman, S. N. (2001). Understanding command and control teams operating in complex environments. *Information, Knowledge, Systems Management*, *1*, 311-323.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, *36*, 555-599.

Shojania, K. G., Burton, E. C., McDonald, K. M., & Goldman, L. (2003). Changes in rates of autopsy-detected diagnostic errors over time: A systematic review. *Journal of the American Medical Association*, *289*, 2846-2849.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99-118.

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology*, *93*, 426-434.

Swartz, M. H. (2006). *Textbook of physical diagnosis* (5th ed.). Oxford, UK: Elsevier.

Tindale, R. S. (1993). Decision errors made by individuals and groups. In N. J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 109-124). Hillsdale, NJ: Lawrence Erlbaum.

Tindale, R. S., Kameda, T., & Hinsz, V. B. (2003). Group decision making. In M. A. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 381-405). London: Sage.

Trope, Y., & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239-270). New York: Guilford Press.

Tschan, F. (1995). Communication enhances small group performance if it conforms to task requirements: The concept of ideal communication cycles. *Basic and Applied Social Psychology*, *17*, 371-393.

Tschan, F., Semmer, N. K., Gautschi, D., Hunziker, P., Spychiger, M., & Marsch, S. U. (2006). Leading to recovery: Group performance and coordinative activities in medical emergency driven groups. *Human Performance*, *19*, 277-304.

van Ginkel, W. P., & van Knippenberg, D. (2008). Group information elaboration and group decision making: The role of shared task representations. *Organizational Behavior & Human Decision Processes*, *105*, 82-97.

Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: Preliminary retrospective record review. *British Medical Journal*, *322*, 517-519.

Waller, M., & Uitdewilligen, S. (2009). Talking to the room. Collecitve sensemaking during crisis situations. In R. A. Roe, M. J. Waller, & S. R. Clegg (Eds.), *Time in organizational research* (pp. 186-203). London: Routledge.

Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of adaptive behaviors and shared mental model creation on control crew performance. *Management Science*, *50*, 1534-1544.

Williams, K. A., Rose, W. D., & Simon, R. (1999). Teamwork in emergency medical services. *Air Medical Journal*, *18*, 149-153.

Wu, A. W., Folkman, S., McPhee, S. J., & Lo, B. (2003). Do house officers learn from their mistakes? *Journal of the American Medical Association*, *265*, 2089-2094.

Franziska Tschan is a full professor of social psychology at work at the University of Neuchâtel, Switzerland since 1996. She received her PhD in psychology (1990) and her Habilitation (1999) from the University of Bern, Switzerland.

Norbert K. Semmer is a full professor of industrial and organizational psychology at the University of Bern, Switzerland. He received both his PhD (1983) and his Habilitation (1987) in psychology from the Technical University of Berlin.

Andrea Gurtner received her PhD in psychology at the University of Bern, and is a faculty member in the Department of Management and Administration at the Berne University of Applied Sciences and a research associate with the University of Neuchâtel, Switzerland.

Lara Bizzari received her MS in work and organizational psychology from the University of Neuchâtel in 2008.

Martin Spychiger is a registered nurse (1988) and holds a degree as a specialized nurse in intensive care (1992). He is a research assistant in the SNF project Human Factors in Medicine.

Marc Breuer is a registered nurse (1992) and holds a degree as a specialized nurse in intensive care (1999). He is a research assistant in the SNF project Human Factors in Medicine.

Stephan U. Marsch is a professor of medicine at the University Hospital in Basel, Chair of the Simulation Center at the Intensive Care Unit at the University Hospital in Basel, and Chair of the Intensive Care Unit in Basel. He holds an MD from the University of Zurich, Switzerland (1989), and a PhD from the University of Oxford (1995).