Crew Situation Awareness in High-Tech Settings: Tactics for Research Into an Ill-Defined Phenomenon

Sidney W. A. Dekker
Centre for Human Factors in Aviation
Department of Mechanical Engineering
Linköping Institute of Technology

In this article I discuss research tactics that can gain empirical access to crew situation awareness in high-tech settings, that is, in settings where multiple crew members have to coordinate their activities with those of an automated system in the pursuit of operational goals. Although deemed an important ingredient for safe and efficient operations, crew—or joint—situation awareness remains ill defined, and results regarding its demonstration or manipulation are often unverifiable and inconclusive. In this article I define the problem of crew situation awareness as it occurs in 2-crew automated cockpits and reinterprets a typical case study of crew situation awareness in light of this definition. The remainder of the article is a methodological contribution that reviews a selection of converging or complementary tactics (both field observations and various forms of simulator studies) that can gain empirical access to crew situation awareness and discusses the trade-offs these tactics represent in terms of experimental validity and reliability.

The introduction of powerful automation in a variety of transport applications has increased the emphasis on human cognitive work. Today human operators on, for example, ship bridges or aircraft flight decks spend much time integrating data, planning activities, and managing a suite of machine resources in the conduct of their tasks (Dekker & Hollnagel, 1999; Sarter & Woods, 1995). This shift has contributed to the need for a concept such as situation awareness (e.g., Endsley, 1999), so that one can understand the extent to which an operator is in tune with relevant process data and can form a mental picture of the system and its progress over time. Most high-tech settings are actually not characterized by a single human in interaction with a machine. In almost all cases, multiple people—crews, or teams of operators—jointly interact with the automated system in pursuit of operational objectives. These crews or teams have to coordinate their activities with those of the system (Sarter & Woods, 1997) to achieve common goals.

Despite the weight that crews repeatedly place on having a shared understanding of their system state and on problems to be solved (Orasanu, 1995; Orasanu & Connolly, 1993), however, scientific consensus on a concept of crew situation awareness seems far off (Flach, 1995; Pew, 1995; Sarter & Woods, 1991). It appears that various labels are used interchangeably to refer to the same basic phenomenon—for example, group situation awareness (Wellens, 1993), shared problem models (Orasanu, 1994), team situation awareness (Jentsch, Barnett, Bowers, & Salas, 1994).
1999; Robertson & Endsley, 1995), mutual knowledge (Krauss & Fussell, 1990), shared mental models (Cannon-Bowers, Salas, & Converse, 1993; Stout, Cannon-Bowers, Salas, & Milanovich, 1999), joint situation awareness (Orasanu, 1995), and shared understanding (Sarter & Woods, 1997; Watts, Woods, & Patterson, 1996). At the same time, results about what constitutes the phenomenon are fragmented (Salas, Prince, Baker, & Shrestha, 1995), and ideas on how to measure it remain divided (Pew, 1995), ranging from modified measures of practitioner expertise (Stout et al., 1999) to questionnaires interjected into suddenly frozen simulation scenarios (Endsley, 1988), to embedding implicit probes into unfolding simulations of natural task behavior (Harwood, Barnett, & Wickens, 1988; Orasanu, 1995; Sarter & Woods, 1997). In addition, a common definition or model of crew situation awareness remains elusive. For example, research that claims to identify links between crew situation awareness and other parameters (e.g., planning and coordinated performance [Stout et al., 1999] or crew member roles [Jentsch et al., 1999]) does not define the phenomenon, which renders empirical demonstrations of it unverifiable and inconclusive.

The lack of definition may be related to a tendency in human sciences to rely on folk models to explain what goes on behind complex behavioral sequences (Hollnagel, 1998). Large labels that correspond roughly to mental phenomena we know from daily life are deemed sufficient; they need no further explanation. This is often accepted practice for psychological phenomena because, as humans, we all have “privileged knowledge” about how the mind works (given that we all have one). However, a verifiable and detailed mapping between the context-specific (and measurable) particulars of a behavioral sequence on the one hand and a pertinent concept-dependent model on the other is not achieved—the jump from context specifics to concept dependence immunizes itself against critique or verification (see also Xiao & Vincente, in press).

Another risk is that the definition of crew situation awareness is left to arise from how it is measured indirectly—through a particular study or experimental paradigm. (This often happens with phenomena that are not directly observable [Orasanu, 1995]; cf. the debate on mental workload, e.g., Edwards [1990].) As a result, the distinction between model and measurement becomes blurred (Hollnagel, 1998). Although this can lead to compelling research results based on single measures (e.g., heart rate went up during these or those conditions), such results remain isolated, because the lack of a measurement-neutral definition leaves them untransportable across studies. Constructive debates about the demonstration and manipulation of a not-directly-observable phenomenon become more realistic when model and measurement are decoupled, that is, when a definition is postulated that is not implicitly limited to the measurement used in one paradigm or study (e.g., degree of mental workload is not indicated only by heart rate).

Researchers seem to agree that when a phenomenon is not directly observable, multiple measures or research tactics are necessary to generate converging empirical evidence about its existence and the efficacy of manipulations of it (Woods & Sarter, 1993). Such a complementary, or converging, investigative approach also benefits from a definition in more measurement-neutral terms. On the basis of the idea of converging evidence, Sarter and Woods (1991) proposed several directions for generating empirical data about situation awareness, among them various forms of simulator studies. A decade later, it is appropriate to review the contributions that some of these directions have made or could potentially make. Such a review may be one way to clarify the value of data generated on joint, or crew, situation awareness—that is, to indicate how we can go beyond accepting ad hoc assertions about the demonstration or manipulation of a phenomenon as substitute for more verifiable research results. Next, the problem of crew situation awareness as it occurs in two-crew automated cockpits is introduced and defined. Then a typi-
cal case study of crew situation awareness is reinterpreted in light of this definition. The last part of this article is a methodological contribution that lays out research directions for crew situation awareness. A selection of tactics is reviewed, and their trade-offs in terms of validity and reliability are discussed.

DEFINING JOINT SITUATION AWARENESS IN HIGH-TECH MULTICREW SETTINGS

One way to define a phenomenon as large and comprehensive as (crew) situation awareness is to constrain the examined situation and be clear about what meaning the concept carries in a context like the one studied. In this sense, automated cockpits in today’s commercial airliners represent a good research vehicle (e.g., Orasanu, 1995). Automated cockpits not only form a well-bounded world with a limited number of players (two humans, one powerful machine agent; Sarter & Woods, 1997), but they also have come to the fore as a sponsor of crew situation awareness problems (Federal Aviation Administration [FAA], 1996; Sarter, Woods, & Billings, 1997). At the heart of these problems lies a strong and independent automated system, typically silent about what it is doing, and which can be directed by each human crew member separately through a private interface (the control display unit, of which each pilot has one). The introduction of such automation has in effect put a third crew member (one that is silent and hard to direct) among the two remaining human crew members (Sarter & Woods, 1997). Similar systems are finding their way onto commercial ship bridges as well, creating similar problems (Lindström, 1998).

In these high-tech settings, crew situation awareness can be defined as the extent of convergence between multiple crew members’ continuously evolving assessments of the state and future direction of a process. Such a definition captures various critical aspects, including the importance of multiple subsequent assessments as constituents of awareness and the temporal nature of awareness (Harwood et al., 1988). As circumstances unfold and possible outcomes shift, situation awareness is constantly updated on the basis of recurrent situation assessments (Klein, 1998; Sarter & Woods, 1991). Also, the definition echoes characteristics of situated performance that research keeps reaffirming (e.g., Hutchins, 1995; Rasmussen, 1979; Suchman, 1987). It points to a mental representation (or schema; Neisser, 1976) or cognitive map (Rochlin, 1997) that operators have of the developing situation around them and to how this evolving understanding allows operators to create expectancies for future events and direct their own and others’ attention to interesting changes in the process (Klein, 1998). The formation and updating of a cognitive map of course goes hand in glove with knowing where to look next: Without an internal representation, people would not know what to expect or focus on in the near future. Similarly, directing attention is the formation of a cognitive map: Looking at certain parts of the world at the exclusion of other parts informs, updates, or changes the mental picture of what is going on and where the process is headed.

Features of the automated flight deck, combined with the dynamic nature of the work that goes on in it, conspire against this convergence of understanding and creation of expectancies. Each pilot can make flight path changes without the other pilot necessarily seeing, knowing, understanding, or agreeing (Segal, 1990; Wiener, 1989; Wiener et al., 1991). At the same time, the automated crew member suppresses cues about what it has been told to do and can keep doing it for long periods without further human intervention. Thus, establishing joint situation awareness on an automated
flight deck is both (a) more critical: Investments in common ground are necessary, because the automation can carry out long sequences of action autonomously while suppressing cues about how and what it is doing and (b) more difficult: Each human crew member has private access to the automation and can make significant changes without the other human knowing or understanding.

Breakdowns in crew situation awareness in automated cockpits can be described as divergences between how the various crew members understand process state and future direction. Divergences are especially likely during high-tempo episodes when automation interaction tasks also typically escalate. Explicit coordination between human crew members about what one is telling the automation to do often falls by the wayside precisely during these high pressure, yet more critical, times (Wiener, 1989).

A CASE STUDY

The events leading up to the crash of an automated B-757 airliner near Cali, Colombia, in 1995 are often held up as an example of situation-awareness problems in high-tech cockpits, yet the official investigation (Aeronautica Civil, 1996) and scientific follow-ups (e.g., Endsley & Strauch, 1997) were limited in their probe of crew, or joint, situation awareness. In those accounts the crew is in effect treated as a homogeneous unit, losing situation awareness as a whole under the pressure of the workload from a new approach and an obstinate, hard-to-direct automated crew member. To the extent that the human crew of the airliner did not share a common understanding of the situation or an idea of where the aircraft was headed, it is implied that they should have, and crew resource management and computer coordination procedures were recited to contrast actual behavior with retroactively invoked standards of proper practice (Aeronautica Civil, 1996).

None of this explains the coordination breakdown behind this accident, however. The real data about crew situation awareness as defined above is in the details of the three interpretations (of the two human crew members and the one automated crew member) and how these diverged. Closer inspection reveals a three-way breakdown among what the automation was doing, what the pilot told the automation to do, and what the copilot understood the pilot had told the automation to do. The explanation below is based on the established facts of this accident (Aeronautica Civil, 1996).

The aircraft in question had flown most of the way from Miami, Florida, to Cali, Colombia, when the approach controller asked it to take a different runway at the arrival airport of Cali. It was already dark by then. Instead of flying around the airport and taking the runway that runs from south to north, the crew was now to fly straight ahead to the reciprocal runway—Runway 19, which runs north to south. The start of an arrival route to the new runway (called the Rozo 1 arrival) lay just to the left (east) of the aircraft’s current flight path (see Figure 1). A navigation beacon named Rozo was located near the top end of Runway 19. The pilot was handling the radio and also instructing the aircraft’s automation where to go. The copilot was the pilot flying; he had never been to Cali before.

In setting up the aircraft for the new approach to Runway 19, the automation was first instructed to take the aircraft to the Rozo beacon. This allowed the crew to concentrate on various other tasks that had to be accomplished (e.g., manual tuning and identification of navigation radios): The aircraft would now fly itself toward Runway 19 (Aeronautica Civil, 1996). Such delegation reflects training for automated aircraft—the automation is used as a cockpit resource to relieve human crew workload. As could be expected under these conditions of sudden change
and high workload (Wiener, 1993), the instruction to the computer was made while the pilot was in a clarifying conversation with the air traffic controller and (as far as the cockpit voice recording can establish) was not verbally coordinated across the cockpit’s crew members (Aeronautica Civil, 1996). There is evidence that the copilot was meanwhile minding the equally important new vertical path of the aircraft (the copilot’s recorded remarks reflect this task; they refer to transition altitudes and other checks relevant during a descent), governed during this time by a lower level automation mode than the lateral path.

Because of a database anomaly, the automation misinterpreted the instruction to take the aircraft to Rozo. The paper approach chart consulted by the pilot listed Rozo as $R$, but Rozo was
actually not stored in the computer database as \( R \)—it was stored as \textit{ROZO}. Colombia already had an \( R \) for one of its navigation beacons in computer databases. This \( R \) was reserved for Romeo, a beacon near the more important city of Bogota. Romeo lay more than 130 miles to the north-east, over the mountains. For the automation, the command to take the aircraft to “\( R \)” meant “take the aircraft to Romeo.” An underspecified, syntax-based human–computer dialogue allowed this mismatch in interpretation to persist: The computer asked the pilot to verify which “\( R \)” he wanted to approach, without mentioning that none of the \( R \)s on its list actually represented Rozo. The \( R \) highest on the list was selected, because pilots know that the top one is always the beacon closest to the aircraft (Aeronautica Civil, 1996).

Living up to Wiener’s (1989) characterization that computers are “dumb and dutiful,” the automation gently banked the aircraft toward the left, flying away from the inbound approach to Cali and heading for Romeo. Aeronautica Civil (1996) listed numerous tasks that the crew would have had to carry out during this time to prepare for their new approach (e.g., locate and review charts; select and verify radio navigation beacon frequencies; recalculate air speeds, altitudes, and airplane configurations; and monitor the new descent path). The ruffling of paper charts on the cockpit voice recording confirms the crew’s engagement in this and likely other necessary work. There is evidence that as a result of this activity the pilot did not notice the turn. Also, to decelerate the aircraft and increase its descent rate, the speedbrakes were extended during the same time that the automation started to bank the aircraft to the left. This may have eradicated proprioceptive cues generated by the turn. From the onset of the turn, it took more than a minute for the moment of surprise to be reached—not an uncommon lapse for such “automation surprises” (Sarter et al., 1997). Having been preoccupied with other tasks, the pilot checked a particular navigation beacon at this time and noticed the heading change. He asked, “What happened here?” (also not an uncommon question in automated cockpits—see Wiener, 1989). The discovered automation behavior was probably a significant departure from what he expected it to be (fly straight ahead toward Rozo).

In the meantime, the copilot likely did notice the turn. When the aircraft had turned through about 60° (two thirds of a full left turn), he was the one who raised the issue. It is interesting that during the first part of the turn the copilot did not mention it. Equally unaware that the automation was taking the aircraft all the way to Romeo (Aeronautica Civil, 1996), the turn to the left may have initially been consistent with the copilot’s understanding of the situation. Because the Rozo 1 arrival lay to the left of their flight path, a course change in that direction to capture this arrival route was to be expected. It is ironic that not mentioning anything may have been a sign of the copilot’s well-known reputation for good crew resource management (Aeronautica Civil, 1996), which includes knowing when not to interrupt a fellow team member. If the automation is doing reasonable things, why point it out to the other pilot? It was not until automation behavior was becoming clearly unreasonable—when the aircraft was turning too far to reasonably capture the Rozo 1 arrival—that the copilot started to raise questions about where they were headed. In the subsequent right turn—back to Cali airport from their position at the east side of the valley—the aircraft impacted terrain.

**MULTIPLE DIVERGING MIND-SETS**

One explanation can accommodate these data: There was a three-way coordination breakdown among the automation and the two human crew members. In other words, a divergence occurred among (a) what the automation was doing (flying to Romeo), (b) what one crew member thought he had told the automation to do (fly to Rozo), and (c) what the other crew member thought the automation had been instructed to do (capture the Rozo 1 arrival). This divergence was allowed
to grow in part because of how the automation carried out its work (a shallow turn) and in part
because its behavior initially confirmed possible expectations by one of the crew members.
Human and automated crew members each went their own way, unintentionally sponsoring each
other’s misassessments of their individual interpretation of the situation. There are indications
(FAA, 1996) that this kind of scenario carries substantial residual risk not only in aviation but
also in other settings where multiple human users manage and direct a suite of automated
resources in the pursuit of operational goals.

RESEARCH TACTICS FOR GAINING EMPIRICAL ACCESS

To understand breakdowns such as the one that occurred in the previously mentioned airliner
-crash, the experimenter’s challenge is to gain systematic empirical access to a complex behav-
-ioral situation in which multiple ongoing assessments are allowed to diverge. When conducting
research to determine what factors play different roles in sponsoring such breakdowns, how can
one generate internally valid, generalizable, and reliable results? Guides to research in applied
settings exist, of course (e.g., Cook & Campbell, 1979; Woods, 1993), but a focused discussion
related to crew situation awareness as defined in this article seems warranted given the research
results produced to date in regard to this phenomenon. In this section I review different tactics
for creating empirical access to the phenomena in question—each of which involve different
kinds of trade-offs between various sources of experimental validity and reliability. Confirming
the truism that “there is no free lunch,” increasing validity in one part of the experiment often
subtracts it from another. Thus, the experimenter’s challenge is to balance the various sources of
uncertainty (Woods, 1993) and to be explicit about the contributions of the following:

• Internal validity, which derives from the degree to which the experimenter is actually in
control over the variance in the results. The question is this: Can the experimenter create and
control the circumstances that produce the breakdown in flight crew awareness of automation
status and activity and how the various team members interpret what is going on?

• External validity, which is the extent to which these results are actually generalizable to the
situation that the research purports to represent—in this case, a three-way coordination break-
down among automated and human team members.

• Construct validity, which is—in a narrow definition, at least—a special form of external
validity that relates to the experimental operationalization of the concept under investigation. In
this case, how is team situation awareness operationalized? This also determines, or flows from,
how it is measured in the studies. Challenging all attempts at achieving construct validity is the
fact that awareness can be known only indirectly; it has to be inferred from observable behav-
iors related to events in specific contexts.

• Reliability, which relates to the replicability of results across practitioners and situations.
Do the same kinds of circumstances systematically trigger the same kinds of breakdowns in joint
situation awareness?

The research tactics discussed are by no means exclusive—rather, they represent comple-
mentary methods that can generate converging evidence on a phenomenon that is not directly
observable. Table 1 summarizes the selection of tactics discussed here and the trade-offs they
represent with respect to experimental certainty. First, however, I mention some tactics that like-
ly will fail to get such access.
<table>
<thead>
<tr>
<th>Research Tactic</th>
<th>Internal Validity</th>
<th>External Validity</th>
<th>Construct Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field observations</td>
<td>Derives from repeated observations</td>
<td>Generalizability of results depends on translation of observed results into concept-dependent language</td>
<td>Situation awareness (or lack thereof) inferred from relevant process parameters</td>
<td>Replicability of results depends on translation of observed results into concept-dependent language</td>
</tr>
<tr>
<td>Simulator studies with confederate pilot</td>
<td>High through experimental probes and confederate behavior all under experimenter’s control</td>
<td>Results generalizable to multicrow circumstances with 1 crew member out of the loop</td>
<td>Situation awareness (or lack thereof) inferred from relevant process parameters</td>
<td>Studies and results are replicable</td>
</tr>
<tr>
<td>Simulator studies on crew communication</td>
<td>High through experimental probes under experimenter’s control, but both crew members are free to behave naturally</td>
<td>Results generalizable to multicrow settings</td>
<td>Communication reflects investment in crew situation awareness, but relationship is not straightforward; also, crew may be aware of more than they communicate</td>
<td>Studies are replicable, but coding of communication is a reliability challenge</td>
</tr>
<tr>
<td>Simulator studies with eye–brain tracking and communication</td>
<td>High through experimental probes under experimenter’s control, but both crew members are free to behave naturally</td>
<td>Results generalizable to multicrow settings</td>
<td>Crew situation awareness operationalized as joint function of eye and brain behavior and communication between crew members</td>
<td>Studies are replicable, but coding of communication is a reliability challenge</td>
</tr>
</tbody>
</table>
Research Tactics That Do Not Generate Empirical Access

Various research methods used in the quest for access to crew situation awareness are actually not likely to achieve such access, generally because they are known to inherently miss or destroy the phenomenon of interest. This has been discussed elsewhere (Sarter & Woods, 1991), and such tactics are covered only briefly here. One tactic is the freezing of a mission simulation at random points in time (e.g., Endsley, 1988; Fracker, 1989), when pilots get asked a host of questions on what they remember about the situation. This measures not situation awareness as an unfolding, ongoing phenomenon over time but rather a pilot’s ability to recall data when suddenly deprived of a dynamic flight context in which his or her actions and assessments were once embedded and meaningful. Another tactic is after-the-fact data gathering through questionnaires, interviews, or debriefings (e.g., Stout et al., 1999). This relies on contextually spartan and intentional retrieval of specific information and likely misses aspects of situation awareness that have to do with knowledge that remains tacit until activated by cues in the evolving context in which actual task behavior takes place (Eich, 1984). When it comes to awareness, after-the-fact data gathering is highly selective and probably distorts the nature of knowledge compared to how it was, or would have been, applied by the practitioner in context (Sarter & Woods, 1991).

Field Observations of Crew Awareness of Automation Behavior

Field observations are one way to get empirical data about ongoing human work that takes place in complex environments and, to be carried out, require practitioner expertise in context (Orasanu & Connoly, 1993; Woods, 1993). Repeated observations of crew coordination in high-tech settings can (a) discover and confirm that a particular phenomenon happens; (b) tabulate the factors that seem present in, and perhaps even contributory to, those situations (the independent variables, in classic experimental design language); and (c) begin to sketch patterns for the type of breakdowns that seem to occur over and over.

Field observations do not demonstrate a researcher’s control (and thus cannot prove a researcher’s full understanding) of a phenomenon. They do not allow a researcher to conclude with certainty that the tabulated factors (e.g., not coordinating a computer input) indeed determine a significant portion of the variance in observed behavior. This is an internal-validity problem that frequently makes experimenters reluctant to engage in field studies of any sort (Woods, 1993). A researcher cannot make the phenomenon happen: Many factors that might be relevant or contributory are outside his or her control or even unknown.

Field observation, by means of its inductive and repetitious nature, brings its own sources of experimental control. Facts are gathered and then generalized toward theories that capture underlying psychological mechanisms that may be responsible for observed effects. Field observations are directed at the discovery and description of patterns that recur across similar circumstances, which allows an experimenter to formulate hypotheses—one of the main contributions of exploratory research. Regarding reliability, one problem is often that other researchers must either accept or reject the observer’s interpretation of context-specific data. But field results can support a process of critique, reinterpretation, and follow-ups if they are translated into more concept-dependent levels of description (Hollnagel, Pederson, & Rasmussen, 1981; Woods, 1993; Xiao & Vincente, in press). Such higher level descriptions of the same phenomena also make field results transportable to other settings.
Simulator Studies With a Confederate Pilot

Instead of field observations, an experimenter can make use of full flight simulators, for which specific scenarios are developed and tested on a crew. These scenarios can be designed so that they specifically probe the phenomenon of interest (Orasanu, 1995; Woods & Sarter, 1993); for example, by timing and inserting combinations of changes and events that challenge a crew’s awareness of the operational status of their aircraft. This can deal to a large extent with the internal-validity problem: Observed behavior can be correlated with experimental probes that are under the researcher’s control.

In simulation studies there are various approaches to the operationalization of situation awareness, but in general these studies enjoy tight control over the factors that generate and influence the phenomenon of interest. One way of achieving this in multicrew conditions is to employ a confederate practitioner. In this case, the confederate is one of the two pilots in the cockpit. He or she knows about the experiment and its goals and plays a part in conducting the study. In such a setup the other pilot is the only subject, and his or her behavior is of empirical interest. The confederate represents a large source of internal validity, because through him or her the experimenter can carefully control factors that affect the subject pilot’s awareness and conclude with certainty that those factors are responsible for any breakdowns observed.

Tight internal control, however, creates trade-offs elsewhere. The experimenter must be aware of the extent to which the cost of tighter internal control is borne by external validity or, more specifically, construct validity. The issue lies in the operationalization of pilot awareness of automation status and activity with a confederate in the seat next to the subject. In the flight deck automation work done by Sarter and Woods (1997), one of the scenarios that probed pilot awareness of automation activities presented a loss of a navigation signal on approach—a reason for immediate go-around. In another scenario, the aircraft stayed in an unexpected mode after go-around, which could result (and did, in many cases) in overspeeds. Awareness (or the lack thereof) could be inferred from the subject pilot’s lack of reaction, the aircraft’s continued behavior along an unintended path, or both. The nature of the setup does not allow the confederate pilot to contribute to the subject pilot’s awareness of critical changes. What is in effect measured is how long the confederate and the automation can jointly lead a subject pilot into not picking up changes in automation status and behavior. Not only does the automation have to be strong and silent, but the confederate must also be silent. It is ironic that this situation mimics reality very well, and thus the presence of a confederate has only a limited impact on external validity. A survey conducted by Sarter and Woods (1997) pointed out that single-pilot cockpits are in effect often created on automated flight decks. This can occur, for example, when one pilot is instructing the computer and the other is engrossed in more manual tasks, such as reading charts and setting the aircraft’s basic navigation instruments up for an approach (indeed part of the Cali scenario). In these situations the other crew member is as good (or as useless) as the silent confederate in this experimental setup. The results achieved in controlled experimental settings with a silent confederate are externally valid to the extent that they map onto these specific target situations.

Simulator Studies on Communication and Crew Awareness

An obvious operationalization of joint air crew awareness is to measure communication among the crew members in situations where both can play their natural role. Communication is a natural byproduct of multiple people cooperating to work toward particular goals (see Miyake, 1986; Suchman, 1987). Here it is assumed that communication represents the extent to which
team members make individual investments in creating common ground or achieving common understanding (Orasanu, 1995; Watts et al., 1996) and that communication is thus a reflection of crew or joint situation awareness. There are advantages to measuring situation awareness this way: Communication is readily available and can be recorded unobtrusively (Ericsson & Simon, 1980; Orasanu, 1995; Sarter & Woods, 1991; Woods, 1993).

The relation between crew situation awareness and communication is not straightforward, however. Orasanu (1995) explored a series of cockpit simulation studies that presented difficult in-flight problems for crews to solve or find their way around. In these kinds of studies, not all talk is problem-related talk, and not all problem-related talk is directed at investing in a common understanding of the problem. Orasanu found that more effective crews sometimes communicate less, depending on the nature and urgency of the problems afflicting their aircraft (see also Wiener, 1993). What Orasanu found, however, was that a greater proportion of higher performing pilots’ (as compared to lower performing pilots’) total talk was devoted to flight problems (i.e., statements of goals, plans, and information requests). This effect was actually greater during high-workload, abnormal phases of flight. By such coding, crew utterances can be examined for their contribution to crew situation awareness. In Orasanu’s (1995) studies, two types of crew utterances were of particular interest—those related to an individual’s situation assessment and those related to planning a course of action, reflecting the critical ingredients of the definition of crew situation awareness in this article. More effective crews contained pilots who verbalized a greater number of plans (Orasanu, 1995).

Some problems with the operationalization of joint situation awareness by means of communication remain. Coding, of course, represents a reliability challenge: Would different experimenters code the same utterance the same way? Also, independent of the reliability in coding problem-related talk, construct validity remains an issue: Practitioners may be aware of much more (even jointly) than they communicate explicitly. Communication may fail to reflect moment-to-moment changes in situation awareness (Orasanu, 1995). Presumably it lags behind the event of interest (the experimental probe) for some indeterminate time. Furthermore, verbal communication is just one data trace in tracking coordinative behavior. Other behavior (e.g., pointing, different kinds of body language) may help establish joint awareness as well.

Simulator Studies With Eye Tracking and Other Data Traces

One way out of the experimenter’s dilemma is to back up the verbal protocol of activities in a naturalistic setting (with two nonconfederate crew members) with various other data traces. These traces can help calibrate the extent to which cross-crew coordination is representative of awareness of changes in the automation. For example, by tracking head and eye movements and measuring visual cortical brain activity of individual crew members (see, e.g., Visual Interaction and Human Effectiveness in the Cockpit, 1998), the experimenter can calibrate the lags between annunciation (e.g., of a mode change in the automation), perception, and cross-crew coordination. As in Donders’s (1868/1969) experiments, the lag between annunciation and coordination (taken as one measurement unit in research that uses communication as the operationalization of awareness) could be filled with another marker: the perception of the annunciation by individual crew members. Thus the experimenter could gain insight into the different reasons behind the subsequent phases of the lag between annunciation and coordination (“Did you make it do that?”). The lag between annunciation and perception would likely be a joint function of the individual’s expectations and properties of the automation feedback (Neisser, 1976).
This could help us understand, for example, why the pilot in the Cali scenario did not comment on the turn toward Romeo. There was nothing in his mental model that helped him expect such behavior, and subsequently—busy as he was with other things—he was not looking for evidence of it, either. The second lag, between perception and coordination, would shed light on different processes altogether. This second lag has to do with the observer’s expectations of what the other pilot will have told, or will expect, the automation to do, what the other pilot is busy with at that moment, and any other assumptions or observations about the other pilot. This second lag can help explain the copilot’s silence about the aircraft’s behavior: The turn was consistent with his mental model and assumed to be consistent with the pilot’s mental model as well.

CONCLUSION

Crew situation awareness has been identified as a critical factor in the effectiveness and safety of a team of humans who have to coordinate their activities with highly automated systems in the pursuit of operational goals. Yet scientific consensus on what crew situation awareness is and how to influence it appears to be far off. Many labels are used interchangeably to refer to the same basic phenomenon; results about what constitutes the phenomenon and how to measure it are fragmented, and a common definition or model is rarely given, not even in studies that claim to demonstrate it. In this article crew situation awareness was defined as the extent of convergence between multiple crew members’ continuously evolving assessments of the state and future direction of a process. This definition captures various critical aspects, including the importance of multiple subsequent assessments over time as constituents of situation awareness and how this evolving series of assessments allows operators to create and share expectancies for future events.

In this article I have stressed how during empirical investigations of crew situation awareness

- The phenomenon of interest must be defined, even if it is within the narrow context of study or application domain. We cannot rely on intuitive folk explanations of crew situation awareness, because it leaves conclusions about its demonstration or manipulation unverifiable.
- Multiple research tactics must typically be used to produce converging evidence on the existence of, and ability to influence, a phenomenon that is not directly observable. Complementary research tactics discussed here included field observations and various forms of simulator studies.
- All of these tactics involve different kinds of trade-offs between sources of experimental uncertainty. Being explicit about the selected mix of experimental control can sponsor debate about alternative interpretations of results rather than endowing the knowledge base with ad hoc assertions that the phenomenon under investigation was demonstrated or influenced.

ACKNOWLEDGMENT

This work was supported in large part by a grant from the Swedish Flight Safety Department to the Swedish Centre for Human Factors in Aviation.
REFERENCES


