Team Cognition During a Simulated Close Air Support Exercise: Results From a New Behavioral Rating Instrument

Jerzy Jarmasz[1], Richard Zobarich[2], Lora Bruyn-Martin[2], and Tab Lamoureux[2]

[1]Defence Research & Development Canada, Toronto, Canada

[2]HumanSystems Inc., Guelph, Canada

ABSTRACT

The so-called low-intensity conflicts in which Coalition forces are currently engaged in such as the one in Afghanistan have forced departures in Close Air Support (CAS) practice from accepted US and NATO doctrine. That is, CAS in these environments relies much more on supporting ground troops with non-kinetic effects, and depends much more on effective combat identification (combat ID) in producing these effects as well as the kinetic effects (i.e., firepower) typically associated with CAS doctrine. As combat ID in CAS is mostly a product of team cognition processes, effective training and assessment methods are critical for ensuring effective combat ID in CAS missions. Here we discuss results of a behaviorally anchored rating scale (BARS)-based rating instrument we designed for the purpose of assessing the quality of team processes, based on behavioral markers for team cognition breakdowns (Wilson, Salas, Priest & Andrews, 2007), in distributed simulation-based CAS exercises for the Canadian Forces. The BARS instrument was applied for the first time during Exercise Northern Goshawk, a distributed Close Air Support (CAS) simulation exercise that involved participants from the United States, the United Kingdom and Canada in August 2007. Despite a number of challenges encountered in applying the instrument and analyzing its results, it captured a number of

noteworthy patterns in team cognition during the exercise, including a number with implications for combat ID effectiveness. Thus, this initial application of the BARS instrument shows that it has strong potential for assessing the collective cognitive processes that underlie effective CAS performance in general, and effective combat ID in CAS in particular, during simulated CAS missions. We feel that further research could make this a valuable tool for improving CAS training in simulated environments for the Canadian Forces and the armed forces of allied nations.

INTRODUCTION

The nature of recent wars such as the one in Afghanistan has caused Coalition forces in those theatres of operation to depart from or even completely re-write doctrine in a number of areas. Such is the case with Close Air Support (CAS). United States Joint doctrine defines CAS to be "air action by fixed- and rotary-wing aircraft against hostile targets that are in close proximity to friendly forces and that require detailed integration of each air mission with the fire and movement of those forces" (United States Department of Defense [DOD], 2005, p. I-1). North Atlantic Treaty Organization (NATO) doctrine also defines CAS in very similar terms (NATO, 2005). Integration between air assets and the supported ground forces is typically performed by an air controller attached to ground forces, called a Joint Terminal Attack Controller (JTAC) in US doctrine and often called a Forward Air Controller (FAC) in other NATO countries.

The US and NATO CAS doctrine publications, which cover CAS doctrine for the majority of forces involved Afghanistan conflict (as well as the similar conflict in Iraq), provide guidance for CAS mostly with respect to cold-war-type scenarios, where the purpose of CAS is primarily to deliver kinetic effects (i.e., firepower) onto well-defined, mechanized enemy targets in an environment requiring the establishment of air superiority, defeat of enemy air defenses, and close integration of munitions delivery by air assets with other joint fires (e.g., land or naval artillery; see also Barber et al., 1991, for a human factors analysis of CAS supporting this view). However, as Haun (2006) argues, the realities of current conflicts have forced a shift in CAS away from the

doctrinal focus on delivering kinetic support to ground troops in a non-permissive air environment to a focus on delivering non-kinetic support (e.g., intelligence, surveillance and reconnaissance, shows of force, convoy escort) in an environment where air superiority and the absence of air defenses can be taken for granted. More importantly for our purposes, perhaps, Haun points out that in this new operational environment, CAS "targets" are more often than not individuals (dismounted or in civilian vehicles) or small structures, rather than formations of mechanized infantry or large military headquarters. These targets exceedingly difficult to distinguish from people or buildings belonging to the population that coalition troops are supposed to be supporting. Haun calls this type of CAS "Low intensity conflict CAS" or "LIC CAS" to distinguish it from doctrinal CAS. Haun's comments highlight the crucial role that the ability to effectively and quickly tell friend from foe from bystander, in order to prevent fratricide and civilian casualties ("collateral damage"), plays in contemporary CAS operations. This ability is what commonly known as combat identification or "combat ID."

Fratricide is certainly not ignored in the doctrine manuals; the term itself appears many times in the manuals, and the DOD and NATO manuals devote whole sections to the issue. However, fratricide is a growing concern, with Wilson, Salas, Priest and Andrews (2007) reporting that close to 20% of casualties in recent conflicts have been estimated to be the result of friendly fire. The Afghanistan conflict alone has produced some very high-profile fratricide incidents involving Canadian soldiers (Department of National Defence of Canada [DND], 2006) and American soldiers (Wilson et al., 2007) among others. Furthermore, the issue of civilian casualties stemming from CAS missions

is a common theme in news coverage of the Afghanistan and Iraq wars, as a cursory survey of even mainstream news media will show.

Thus, effective combat ID is a key factor, perhaps even *the* key factor, in successful CAS in low-intensity environments, whether the mission involves dropping ordnance, performing surveillance and reconnaissance, or providing overwatch for a convoy. Furthermore, as task analyses of CAS (e.g., Zobarich, Lamoureux, & Bruyn-Martin, 2007) show, combat ID, and maintaining situation awareness (SA) in general, are highly distributed collective tasks, where a number of players must communicate effectively to construct a joint awareness of friend, foe and bystander. This is all the more so in coalition missions (DOD, 2005). Thus, as Wilson et al. (2007) point out, ensuring effective team cognition (team communications, team situation awareness, etc.) is imperative for effective combat ID and for minimizing the risk of fratricide and civilian casualties in CAS in the current operational environments.

IMPROVING TEAM COGNITION: RATING SCALES FOR DISTRIBUTED SIMULATION EXERCISES

Distributed Mission Operations (DMO, also called Distributed Mission Training, and "UK Mission training through Distributed Simulation" in the United Kingdom) is a simulation-based training technique with recognized potential as a method for collective training for geographically dispersed, joint and/or multinational operations. DMO could thus be a useful tool for improving CAS performance in general and combat ID performance in CAS in particular. In DMO, trainees located in disparate geographic

locations interact in a common virtual space to train team skills, using simulators linked up via high-speed networks. The US Air Force (USAF) has been researching DMO for a number of years and has developed a number of tools for assessing individual and team performance in DMO settings (Schreiber & Bennett, 2006). Because the trainees are not co-located in physical space, and are often not co-located in virtual space either (e.g., pilots flying different aircraft), assessment of team performance can present special challenges.

The Canadian Forces (CF) Air Warfare Centre is currently standing up its Distributed Mission Operations Centre (DMOC) and developing its DMO expertise. As part of this effort, we were tasked with developing measures of team effectiveness in the context of Exercise Northern Goshawk, a simulated coalition distributed air operations exercise that was facilitated under the auspices of the Technical Cooperation Program (TTCP) Coalition Mission Training Research (CMTR) Project Arrangement. The exercise was designed as a CAS, Time Sensitive Targeting (TST) and Troops in Contact (TIC) training event involving participants and researchers at simulation sites in Canada, the US and the UK on 6-10 August 2007. We focused on the Canadian component of the exercise, which simulated a CAS mission involving a Canadian FAC and coalition (US and UK) pilots. The FAC was supporting a CF Commander (not a member of the primary training audience) who was commanding a small convoy transiting through sparse terrain populated with a few simple structures (small dwellings and a religious building). The scenarios were meant to represent situations typical of the low-intensity conflict

situations, including TIC events, that the CF currently encounter on operations in Afghanistan.

In this paper we discuss our efforts to capture team cognition processes during the Canadian portion of Exercise Northern Goshawk with a Behaviorally Anchored Rating Scale (BARS)-based instrument we designed for this purpose. We start by summarizing the development process for the BARS instrument, which we discuss in more detail in Jarmasz, Zobarich, Bruyn-Martin and Lamoureux (2008) and Zobarich et al. (2007). We then discuss the results we obtained with the instrument, which was applied for the first time at Exercise Northern Goshawk. Finally we discuss challenges and issues which we encountered during this first trial of the instrument, many of which are also treated in more detail in Jarmasz et al. (2008) and Zobarich et al. (2007).

DEVELOPMENT OF THE RATING INSTRUMENT

General Considerations

A BARS instrument was chosen for assessing team cognition largely because it is an unobtrusive method that does not require self-reports by the trainees (time constraints did not allow for a self-report type of assessment during this exercise). Furthermore, BARS have shown to be effective for assessing team effectiveness in a number of settings (e.g., military command-and-control, see Murphy, Grynovicki & Kysor, 2003; medical emergency department training, see Morey, Simon, Jay, Wears, et al., 2002). However, because CAS involves the coordination of different but interrelated tasks by a distributed team, we first needed to identify the aspects of CAS missions that involved team interaction, especially between the FAC, the pilot and the supported commander.

This was accomplished by performing a Hierarchical Task Analysis (HTA) of CAS missions. Having identified the phases of CAS where team behaviors would be most evident, we set about developing anchors for the BARS based on the behavioral markers of team cognition breakdown proposed by Wilson et al. (2007). This was one way of ensuring the instrument focused on the desired construct, namely team cognition.

Task Analyses

The HTA method was chosen because of its origins in systems theory (Annett & Cunningham, 2000; Shepherd, 1998), thus making it suitable for capturing system-level (i.e., team) behavior. The HTA for the overall CAS mission was conducted and validated based on extensive interviews with SMEs knowledgeable in CF procedures in CAS (active FACs and FAC instructors from the CF as well as a CF-18 pilot with CAS experience). This team HTA identified the different members of the broad CAS team (including the Pilots, the FAC, Fire Support Coordination Centre, Air Support Operations Centre, Forward Observation Officer, Signals Officer, and the Convoy Commander), but only developed the FAC and Pilot branches in detail. Each team member's task breakdown included a number of tasks that fed into, or received input from, or required an explicit appreciation of, the tasks of other team members. Those tasks were explicitly represented in both team members' branches. In practice, we used the tasks in the FAC branch which fed into or received from the Pilot task to form the 'skeleton' of the Pilot's branch. This allowed us to identify the critical team coordination points that occur generically across CAS missions. We could then develop sets of BARS to assess the quality of team cognition at each of these coordination points.

The full detail of the HTA is given in Zobarich et al. (2007).

Development of BARS

To keep matters relatively simple we decided to develop the individual BARS from the perspective of the FAC's tasks that involved other team members (thus capturing team cognition from the perspective of the person coordinating the teamwork). The development of BARS proceeded along two main lines: (1) Identification of suitable measurement points in the tasks of the FAC, and (2) development of behavioral anchors for these points.

The first line of work focused on those tasks that contributed to the task of another team member, received information from the actions of another team member, or required a significant understanding of the perspective or activities of another team member. The identification of measurement points indicates where measurement should take place; not what should be measured. Typically, the measurement point selected was not the lowest level of decomposition of the HTA. The lowest level of decomposition was often used to inform the scale anchor behavioral descriptions.

The second line started with the team cognition construct (Wilson et al., 2007), which is presented as having three main dimensions: Communication, Coordination and Cooperation. Wilson et al. further break these categories down into generic team behaviors, and propose a number of behavioral markers (worded as questions) for probing the quality of each behavior. These generic behaviors constituted good candidates for the behaviours that needed to be assessed (i.e., for which BARS needed to be developed) at the measurement points identified in the HTA. Since a rating is

essentially an answer to the question, we reworded the generic behaviors in each of the team cognition categories as questions that were relevant to CAS missions based on our HTA. These items, shown in Table 1, could then serve as a basis for developing specific BARS.

Table 1: Dimensions of Team Cognition as Adapted to the instrument

Communication	How effective was information exchange?
	Was information exchange economical?
	Did closed looped communication go as expected?
Coordination	How well were team members' knowledge requirements managed?
	How well did team members monitor each other's performance?
	How effective was back-up behaviour?
	How adaptable were team members to the changing demands of the situation?
Cooperation	To what extent were team members working toward the same ends?
	How effective were FAC/others as a team?

Thus, a total of 10 BARS, capturing 3 main dimensions of team cognition, could potentially be formulated at each measurement point identified in the HTA. In practice not all were found to be applicable to each measurement point. Table 2 presents the measurement points and the BARS selected for development.

Table 2: BARS Selection Table

Measurement Point	Communication			Coordination				Cooperation		
	1	2	3	4	5	6	7	8	9	10
Determine air assets	✓	✓		✓				✓	✓	
Understand blue situation	✓	✓	✓	✓	✓		✓	✓	✓	
Understand red situation	✓	✓	✓	✓	✓		✓	✓	✓	
Understand white situation	✓	✓	✓	✓	✓		✓	✓	✓	
Understand brown situation	✓	✓	✓	✓	✓		✓	✓	✓	
Understand time	✓	✓	✓	✓	✓	✓	✓			
Maintain personal safety	✓				✓	✓				✓
Transmit immediate CAS request	✓	✓	✓	✓	✓	✓	✓	✓		✓
Receive pilot's scheduled check-in	✓	✓	✓	✓	✓		✓			
Deconflict target area and airspace	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transmit CAS brief	✓	✓	✓	✓		✓	✓	✓	✓	
Communicate remarks	✓	✓	✓	✓		✓	✓	✓	✓	
Communicate options with pilot	✓	✓	✓	✓			✓			
Designate target	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Coordinate with FOO	✓	✓	✓	✓			✓	✓	✓	✓
Transmit talk-on	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Perform BDA	✓			✓						
Abort CAS mission	✓	✓		✓	✓					✓

Note. Blue situation = friendly forces (e.g., location, movement, intent, etc.). Red situation = situation & intent of adversary. White situation = civilian situation, including individuals and landmarks (e.g., culturally significant buildings). Brown situation = weather & terrain factors.

Having identified appropriate measurement points, it was then necessary to develop the BARS for each. We settled upon a 5 point scale as giving sufficient sensitivity to changes in perceived performance while not overwhelming raters with choice. To construct the anchors we started with the behavioral markers, originally phrased as questions, that Wilson et al. (2007) developed for each sub-dimension of Table 1 (e.g., Did team members seek information from all available resources? Did team members pass information within a timely manner before being asked?). Using these, reformulated as statements and combined with information from the lower level of description in the HTA (if there was one), we decided upon what would reflect 'perfect' performance (i.e. a '5' on the scale) and what would reflect very poor performance (i.e. a '1' on the scale), and then developed complementary anchors for the intermediate scale points. Even though the measurement points in Table 2 represented different tasks, the behavioral anchors generally followed similar patterns depending on the generic team behaviour they were based on. A complete listing of all the behvioral anchors we developed is provided in Zobarich et al. (2007). Due to space limitations, here we present sample BARS from each team cognition category for a particular task (FAC transmits the CAS brief and communicates remarks to the pilot) in Table 3.

Table 3: Sample BARS for each team cognition category from the "Transmit CAS brief and communicate remarks" task.

Communication: How effective was information exchange?	
5	FAC passed complete and accurate brief (following theatre standard), & provided all key remarks info (e.g., weapons effects, attack geometry, ACA measures, number of attempts, level of risk for blue and white forces, danger close initials).
4	FAC passed all items of brief and most key remarks but provided some remarks only when prompted (e.g. danger close initials)
3	FAC passed all items of brief & remarks available to him but had to communicate with others to obtain missing info requested by pilot (e.g. MAXORD, ACA measures)
2	FAC omitted important brief & remarks items that were available to him.
1	FAC failed to provide sufficient brief & remarks for pilot to complete mission.
Coordination: How well were team members' knowledge requirements managed?	
5	FAC/pilot implicitly coordinated in an effective manner (e.g., did not require special coordination or discussion beyond standard turn-taking), and displayed a common understanding of CAS brief SOPs and of the brief/remarks.
4	FAC/pilot coordinated explicitly and effectively (e.g., discussed coordinates formats), and achieved a common understanding of the CAS brief SOPs & brief/remarks with little effort.
3	FAC/pilot made explicit attempts at coordinating knowledge (e.g., asked each other questions), and at great effort (e.g., much time spent in discussion) achieved common understanding of the SOPs & brief/remarks.
2	FAC/pilot made explicit attempts at coordinating knowledge, and achieved an incomplete common understanding of the SOPs & briefs/remarks.
1	FAC/pilot failed to display a common understanding of the SOPs & brief/remarks.
Cooperation: To what extent were FAC/Pilot working towards the same ends?	
5	FAC/Pilot collaborated to ensure CAS brief SOPs (mandatory CAS brief, or that collective SA did not require brief) were adhered to and that all required info was passed & understood for the attack.
4	FAC/Pilot collaborated to understand all relevant CAS brief & remarks info but both had slightly different priorities on brief SOPs, which were easily resolved or accepted.
3	FAC/Pilot collaborated to understand all mandatory CAS brief info, but did not cooperate fully on understanding remarks or had a significant disagreement about brief SOPs; pilot has most but not all info required for talk-on and attack.
2	FAC/Pilot collaborated poorly to achieve joint understanding of brief & remarks, and disagreed significantly about brief SOPs; pilot had only a fraction of the required information for the talk-on and attack.
1	FAC/Pilot could not agree on brief SOPs and did not collaborate to ensure pilot received & understood brief & remarks; pilot did not have any useable information to proceed with talk-on & attack.

Protocol for applying the BARS instrument

Prior to piloting the BARS instrument during Exercise Northern Goshawk, we organized the BARS according to five generic phases of the CAS missions we expected to see during the exercise based on our knowledge of the daily scenarios and the CAS HTA: (1) Pre-check in, (2) Pilot Check-in, (3) generic non-kinetic CAS support tasks covering reconnaissance, shows of force, convoy escort and so on, (4) CAS target prosecution tasks, and (5) generic post-attack tasks (battle damage assessment, re-tasking, egress, etc.). The BARS categories assigned to these tasks and subtasks are shown in Table 4 (see Zobarich et al., 2007 for a complete list).

Table 4: CAS Tasks to be rated during Northern Goshawk, grouped into generic CAS mission phases

CAS mission phase	Tasks
1. Pre-check in	Understand Situation Updates Transmit Immediate CAS Request (as needed)
2. Pilot Check-in	Receive pilot's check-in (includes situation brief)
3. Non-kinetic CAS support	Communicate Options with Pilot Situation Updates
4. Target prosecution	Transmit CAS brief and communicate remarks Transmit talk-on Deconflict Target Area Designate Target Abort Mission (as needed)
5. Post-attack	Perform Battle Damage Assessment Communicate Options with Pilot

Two of the authors (JJ and RZ) were designated to perform the ratings. After a dry run applying the BARS instrument on the first day of the exercise (see exercise schedule below), we determined some basic ground rules for applying the instrument. To the extent possible, ratings would be made every time a task was performed (e.g., every time the FAC talked a pilot onto a target) rather than giving an overall rating to the task for the mission. Also, during the target prosecution phase of a mission, where CAS team members were likely to perform many tasks in parallel and the number of possible simultaneous ratings was expected to overwhelm the abilities of a single rater, one rater (JJ) was designated to rate primarily the CAS brief and target-talk on tasks, and rate the other tasks as resources allowed, while the second rater (RZ) was designated to rate primarily the Deconflict Target Area and Designate Target tasks. In cases where a number of tasks were being performed simultaneously by the team (e.g., designating target while deconflicting air space), each rater focused on a designated subset of the BARS. Minor changes were also made to the BARS after the dry run on the first day (the wording of BARS for different tasks that assessed the same dimension was harmonized, and a few specific BARS were either added or dropped from specific tasks depending on their perceived utility).

RATINGS AND OBSERVATIONS DURING NORTHERN GOSHAWK

Collecting the ratings

Ratings were collected by the raters on each day of the exercise, for the duration of the Canadian portion of the missions scheduled for each day (approximately 2 hours per day for 4 days). As discussed below, Day 1 was used to perform a dry run of the ratings and to finalize the instrument. The raters were able to directly observe the FAC and the Convoy Commander while listening to the radio communications between the FAC and the air radio net (pilots and ASOC).

A number of challenges were encountered in applying the instrument. These challenges are discussed in detail in Jarmasz et al. (2008). In general terms, the raters found it challenging to perform all the ratings prescribed by the protocol in real-time. This was partly due to the quick pace of events and could be remedied by reviewing the recordings of the exercise. The raters found that their ability to follow events and thus perform more of the ratings increased each day of the exercise, despite the scenarios becoming more complicated each day.

Due to these challenges and to the fact the neither rater produced a complete set of ratings on any day due to the rating protocol described above, we cannot perform inferential statistics on the ratings that were collected. Thus the inter-rater reliability of the instrument and any day-to-day changes in team cognition ratings cannot be statistically assessed at this time. Instead, here we simply describe possible team cognition patterns by presenting summary statistics for each day of the exercise (except Day 1), as well as individual ratings for selected incidents in the exercise that had combat ID implications.

Summary Statistics

For each day of the exercise where ratings were collected with the stable BARS instrument (i.e., Days 2, 3 and 4), we computed average ratings for each team cognition category in each major task of the CAS missions identified in Table 4, by combining ratings from both raters across all BARS in the major team cognition categories (communication, coordination, cooperation) for all separate instances of a given task (i.e., all ratings for the communication category from all talk-ons in a given day were combined into one average). Since the sample size for each average rating is different (ranging from $n = 27$ for one BARS applied to multiple instances of the same task by both raters, to $n = 1$ in a few cases) we do not attempt statistical comparisons of the rating averages. Given the sheer number of values that were obtained even with this procedure, here we report (Table 5) only the averages for tasks in Phases 3 (non-kinetic CAS support) and 4 (Target prosecution) of the CAS missions, which are the ones which directly lead up to attacks on targets in CAS missions, and therefore are likely the ones most relevant to combat ID and fratricide performance.

A visual inspection of Table 5 suggests the rating averages were on the whole relatively high, with most ratings being higher than 4.0 and none being lower than 3.83. A calculation of the "grand average" of all the ratings for each day (including those not included in Table 5) reinforces this impression, with the lowest grand average being 4.22 on Day 3. There are a few noteworthy "large-scale" patterns in evidence in the ratings. The presence of ratings for the Abort mission tasks on Days 2 and 4 indicate aborted attacks on both of those days. One of the aborted attacks was due to technical problem with one of the simulated instrument displays the FAC was using; however the second

aborted mission will be discussed in more detail below. Also, almost all of the rating averages in the cooperation category achieved the highest possible value of 5, suggesting high degrees of cooperation between participants throughout the exercise. Also note that all of the sub-4 ratings for the Phase 3 and 4 tasks occur on Days 2 and 3, mainly for communications ratings and to a lesser extent for coordination ratings, which may be an indication that team communications and coordination during these phases of the CAS missions improved over the duration of the exercise. We note that no fratricide or collateral damage events were observed during the exercise.

Table 5: Summary statistics of ratings for Phase 3 and 4 tasks in Exercise Northern Goshawk

Task	Team cognition category	Day 2	Day 3	Day 4
3.1 Communicate options with pilot	Communication	3.83 (1.17)	3.67 (0.98)	4.44 (0.73)
	Coordination	3.50 (0.84)	4.00 (0.00)	4.33 (0.52)
3.2 Understand situation updates	Communication	3.89 (0.93)	4.33 (1.03)	4.13 (0.52)
	Coordination	4.00 (0.00)	3.17 (0.41)	4.17 (0.39)
	Cooperation	4.67 (0.58)	4.00 (1.15)	4.50 (0.58)
4.1 Transmit CAS brief and communicate remarks	Communication	4.22 (1.20)	4.33 (1.03)	4.50 (0.84)
	Coordination	4.50 (0.76)	4.75 (0.50)	4.50 (0.55)
	Cooperation	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)
4.2 Transmit talk-on	Communication	4.22 (0.97)	3.83 (1.17)	4.29 (0.69)
	Coordination	3.83 (0.83)	4.13 (0.83)	4.08 (0.90)
	Cooperation	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)
4.3 Deconflict target area	Communication	4.33 (1.00)	4.00 (1.00)	4.33 (1.12)
	Coordination	4.13 (0.75)	4.00 (0.00)	4.33 (0.52)
	Cooperation	5.00 (0.00)	4.00 (0.00)	4.00 (0.00)
4.4 Designate target	Communication	4.14 (1.07)	4.33 (0.58)	4.50 (0.84)
	Coordination	NR	4.50 (0.58)	4.43 (0.55)
	Cooperation	NR	NR	NR
4.5 Abort mission	Communication	5.00 (0.00)	NR	5.00 (0.00)
	Coordination	4.67 (0.58)	NR	4.75 (0.50)

	Cooperation	5.00 (NA)	NR	5.00 (0.00)

Note: Values in parentheses represent standard deviations; NA = standard deviation cannot be computed because mean was based on a single sample. NR = no rating was made for this measure during the event.

Ratings for specific events

The summary statistics above hide a number of events during the exercise that had combat ID implications. We provide three examples here. The first event involved ordnance dropped on target by a pilot without clearance from the FAC, in clear violation of CAS procedure (NATO, 2005; US DOD, 2005). At the end of the talk-on phase, at a point when the FAC was already satisfied that the pilot had visually acquired the correct target, the pilot stated his intention to drop ordnance on target with terminology that was unknown to the FAC. The FAC requested the pilot repeat his transmission a number of times but due to the low quality of the simulated radio channel (see discussion below) the FAC could not make out what the pilot said. The pilot then proceeded to drop the ordnance, which destroyed the target and surprised the FAC. At the time, the rater who was responsible for rating the talk-on tasks (as per the protocol above) rated one of the three communications BARS for the task as a 3, and two of the four coordination BARS as a 3 and a 2 respectively (see Table 6).

Thus ratings for the interaction between the FAC and the pilot on this attack suggest poor performance on some aspects of team communication and team coordination. Subsequent discussion of the incident by the FAC and the pilot, as well as discussion at the mission debrief, established that the pilot had interpreted the FAC's confirmation that

the correct target had been identified as clearance to attach, whereas the FAC was unfamiliar with the terminology the pilot used to signal his intention to expend ordnance. Poor transmission quality on the simulated radio channel also made it difficult for the players to understand each other, even though on the whole they seemed to be trying to cooperate. Thus, the incident seemed to involve a breakdown in communications (non-standard terminology, need for frequent repetition, and failure to ensure complete mutual understanding) as well as coordination (failure to ensure everyone fully agreed on procedures before the attack). The ratings for this incident made during the event (Table 6) are consistent with the interpretation of events in the debrief, suggesting the BARS instrument is able to track such breakdowns. Fortunately the correct target was hit and no collateral damage was incurred, but the situation could have had serious consequences if, for instance, this had been a real attack and the FAC had been concerned that the pilot's angle of attack or chosen ordnance might have effects on nearby friendly forces or civilians.

Table 6: Ratings for talk-on task of Incident 1

Team cognition category	BARS	Rating
Communication	How effective was information exchange?	5
	Was communication economical?	3
	Did closed looped communication go as expected?	4
Coordination	How well were team members' knowledge requirements managed?	3
	How well did team members monitor each other's	2

	performance?	
	How effective was back-up behavior?	5
	How adaptable were team members to the changing demands of the situation?	4
Cooperation	To what extent were team members working toward the same ends?	5
	How effective were FAC/others as a team?	5

The second event we focus on involves the second aborted attack of the exercise. In that incident, the FAC had been prosecuting a cluster of targets (small buildings and vehicles) with multiple aircraft simultaneously. This was a complicated attack with the FAC and the pilots spending a lot of time questioning each other to ensure each pilot had positive visual identification of the correct target. The FAC seemed to lose track of which pilot he had just cleared for an attack (he subsequently explained in an informal debrief that this occurred due to having too many pilots talking to him on the same channel). Rather than risk the wrong target being hit, the FAC aborted one attack, re-assessed the situation with the pilots, resumed the attack and prosecuted the remaining targets successfully, with no fratricide or collateral damage. Thus in this incident, team cognition was heavily taxed by the complexity of the attack, yet the willingness of the players to try to cooperate and coordinate seems to have contributed to avoiding an inappropriate drop. The ratings for the talk-on and abort tasks of this attack, shown in Table 7, appear to be consistent with this interpretation: the communication and coordination categories for the talk-on contain a number of '3' ratings, whereas the cooperation categories, as well as the ratings for the Abort task contain mostly '5' ratings (the one '4' rating for the Abort task

reflects the fact that the FAC and pilots explicitly coordinated by re-discussing the abort codes during the task).

The third incident involved a situation where a technical problem with some simulators (a so-called 'sim-ism') impaired some of the participants' situation awareness and ability to perform combat ID. The scenario involved a target area that included a number of static objects (buildings) and mobile objects (vehicles). The pilots were reporting a failure to see some of the expected objects at the coordinates indicated by the FAC, generating some confusion. After requesting that the pilots describe the general features of the target area, the FAC was satisfied that they were in the right location and could see at least some of the target area objects, and suspected (correctly) that a 'sim-ism' was preventing them from seeing some of the targets. Given the absence of nearby friendly forces or civilians, the FAC was able to talk the pilots onto the right targets by marking the target area and by referring the pilots to the targets they were able to see. The FAC was able to gauge their situation awareness enough to realize there was a problem, and take corrective measures to ensure that all team members then supported each other in successfully prosecuting an attack on a target. As a result, the ratings for the talk-on for this attack, shown in Table 8, are generally high, with only one scale, the "Was communication economical" BARS, obtaining a rating below 3 due to extra discussion needed and some ongoing transmission problems.

Table 7: Ratings for Talk-on and Abort tasks for Incident 2

Task	Team	BARS	Rating

	cognition category		
Transmit talk-on	Communication	How effective was information exchange?	5
		Was communication economical?	3
		Did closed looped communication go as expected?	5
	Coordination	How well were team members' knowledge requirements managed?	3
		How well did team members monitor each other's performance?	3
		How effective was back-up behavior?	5
		How adaptable were team members to the changing demands of the situation?	3
	Cooperation	To what extent were team members working toward the same ends?	5
		How effective were FAC/others as a team?	5
Abort mission	Communication	How effective was information exchange?	5
		Was communication economical?	5
	Coordination	How well were team members' knowledge requirements managed?	4
		How well did team members monitor each other's performance?	5
	Cooperation	To what extent did team members display mutual trust?	5

Table 8: Ratings for talk-on task for Incident 3

Team cognition category	BARS	Rating
Communication	How effective was information exchange?	5
	Was communication economical?	3
	Did closed looped communication go as expected?	5
Coordination	How well were team members' knowledge requirements managed?	4
	How well did team members monitor each other's performance?	4
	How effective was back-up behaviour?	5
Cooperation	How adaptable were team members to the changing demands of the situation?	4
	To what extent were team members working toward the same ends?	5
	How effective were FAC/others as a team?	5

The three incidents we described above all represented challenging situations for team cognition, with the first two being more challenging than the last one. Accordingly the ratings for the last incident were generally higher than for the first two. More generally, many of the attacks and non-kinetic CAS support tasks in the exercise came off with little or no problems, and the ratings for those events were generally higher than for the three presented above (almost all '4' and '5' ratings). Thus it seems that the BARS instrument is able to capture differences in team cognition processes that seem to correspond to

actual differences in team performance in CAS, some of which, as we discussed, could have significant implications for combat ID effectiveness and the prevention of team fratricide and civilian casualties.

DISCUSSION AND CONCLUSIONS

Congruence of the ratings with exercise events

On the whole, the ratings produced by the BARS instrument we developed to rate team cognition breakdowns seems to have produced results consistent with the subjective observations of observers. The lower average ratings that were obtained for the communications categories discussed above were reflected in comments collected from local observers of the event who seemed to agree that the simulated radio transmissions were at times of poor quality (noisy or otherwise distorted), and that the high number of players on the same channel also adversely affected the ability to understand transmissions at times. The high ratings for almost all cooperation BARS and many coordination BARS was also echoed by the comments of many observers to the effect that participants seemed quite willing to cooperate and "play" together. The seeming improvement for the ratings for the non-kinetic and target prosecution tasks over the last three days was also consistent with the comments of the local participants that the participants seemed to be developing a "rapport" and cooperating better.

The examination of the ratings for a few selected events also seems to suggest that the BARS instrument seems to be congruent with events in the exercise at a "micro" level, in that the ratings seemed to capture different degrees and different types of team

cognition impairment for different events in the exercise. Thus, there is reason to think that the instrument we developed is capable of tracking differences in the quality of team processes at both a "macro" (averaged ratings) and "micro" (ratings for specific events) levels, at least under certain circumstances.

Challenges in applying the BARS instrument

Despite the promising results discussed so far, we have already noted that due to the way the ratings were collected, we were not able to perform any inferential statistics on our ratings. Also, we were able to find parallels between the ratings we collected and the subjective impressions of the observers and participants who were involved in the Canadian portion of the exercise, but we lacked objective measures with which to compare the ratings. Thus, we are not in a position to state with much certainty that our instrument does in fact track real changes in the quality of team cognition during a team task, or that whatever changes are tracked do in fact correspond to objective differences in combat ID performance or the ability to avoid fratricide or unwanted casualties.

These are but two of a number of challenges and issues that were raised in this initial attempt to apply the BARS instrument to a distributed CAS exercise. These challenges and potential solutions are discussed in detail in Jarmasz et al. (2008). We summarize these briefly below.

Convergent and discriminant validity. Some aspects of the scenarios affected the degree to which the properties of the instrument (namely discriminant validity) could be assessed. Many of the participants had extensive CAS experience (the Canadian FAC was in fact a qualified CAS instructor) and all seemed willing to cooperate with each other for

the sake of the exercise. Also, friendly forces and civilian entities were deliberately kept at some distance from CAS targets by exercise organizers, due mainly to technical issues affecting some of the simulators. Thus, a kind of performance ceiling effect limited opportunities to verify the instrument's ability to discriminate between good and bad team performance. This could be remedied by using the instrument on CAS exercises involving wide ranges of trainee experience and task difficulty (especially regarding combat ID issues).

Problems with the instrument itself. One problem was the fact that performing the ratings in real-time during the exercise was difficult for two reasons: the sheer number of BARS to be applied (especially in cases involving multiple simultaneous tasks) and the fact that applying them required good familiarity with CAS exercises. One of the raters was observing a CAS event for the first time during Exercise Northern Goshawk, and often found it challenging to make required ratings during the exercise, especially in the early days of the event. Providing prospective raters with prior experience with CAS events (e.g. via an audio recording of a previous CAS exercise) could ensure raters are more comfortable applying the ratings in real time. The instrument could also be made easier to apply by reducing the number of BARS, or by applying it to after-action reviews or audio recordings of CAS exercises so raters can apply it at their leisure after first observing the event in real time.

The exercise also revealed some deficiencies in specific BARS (some of these were addressed after the dry run of the instrument on Day 1), which could be addressed by refining them by reviewing data from other exercises and more consultation with SMEs.

We avoided the critical incident (Flanagan, 1954) and retranslation of expectations methodologies (Smith & Kendall, 1963), which base behavioral ratings on the opinion of SMEs rather than a theoretical construct. However, applying these to the existing instrument could help improve the diagnostic value of some of the BARS, as well as determine whether some of the BARS or even some of the tasks used in the current form of the instrument have less diagnostic value than others. These could then be eliminated, shortening the instrument and making it easier to use.

Inter-rater reliability. Ultimately, the construct validity and the inter-rater reliability of the instrument will have to be assessed and improved by having a number of raters apply it to a number of CAS missions or exercises. Fortunately, Exercise Northern Goshawk was part of a broader international research effort under The Technical Cooperation Program, and other performance measures and digital logs from a number of similar DMT and DMO exercises have been collected by our research partners (see Schreiber & Bennett, 2006). Thus there is an extant body of data which could potentially be used for improving and refining the BARS instrument.

Conclusion

The BARS instrument we developed to assess the quality of team cognition in CAS exercises appears to have captured some team performance patterns which seem to be consistent with subjective appraisals of the participants' collective performance in the exercise. Thus we feel that, once work required to refine the instrument and improve its inter-rater reliability has been performed, the instrument has the potential to be very useful for capturing team cognitive processes that underlie effective combat ID and team

performance in CAS. This is particularly important with respect to avoiding friendly fire incidents, which are a very real and serious issue in CAS as some recent events in Afghanistan and Iraq have shown. Further, we feel it could play a useful role in supporting the development and assessment of distributed simulation exercises for CAS, and by extension CAS mission safety and effectiveness, especially for low-intensity conflicts, for Canada and allied nations. Finally, the general methodology we present (start with an HTA of a team task, generate behavioral anchors for key interaction points using Wilson et al.'s (2007) behavioral markers, validate and refine with SMEs) could be a valuable methodology for assessing team performance and training in other distributed team tasks, in particular those where failure of team cognition could lead to friendly fire or other lethal consequences.

ACKNOWLEDGEMENTS

The authors thank the members of 1 Royal Canadian Horse Artillery, Canadian Forces Base Shilo, especially LCol Williams (then Commanding Officer), WO Cyr (FAC Supervisor), Maj Allen and Sgt Everett, as well as Capt Williams of the Artillery School at Combat Training Centre Gagetown for their support and expertise in the development of the BARS and execution of the exercise.

REFERENCES

Annett, J. & Cunningham, D. (2000). Analysing command team skills. In J.-M. Schraagen, S. Chipman and V. Shalin (Eds), *Cognitive Task Analysis*. Mahwah, NJ: Erlbaum.

Barber, A., Brown, D., Chandler, E., Davis, D., Dye, C, Moyer, J., Perez, W., and Wolcott, J. (1991). Human-centered analysis of the future Close Air Support/Battlefield Air Interdiction (CAS/BAI) Mission (DTIC Technical Report HSD-TR-1991-0021). Cameron Station, Alexandria, VA: Human Systems Division.

Department of National Defence (2006). Board of Inquiry Minutes of Proceedings, A-10 Friendly Fire Incident 4 September 2006 Panjwayi District, Afghanistan. Retrieved from http://www.forces.gc.ca/site/focus/opmedusa/A10_BOI_Report_e.pdf

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-357.

Jarmasz, Zobarich, Bruyn-Martin & Lamoureux (2008). Avoiding Friendly Fire: Constructing Behaviorally-Anchored Rating Scales to Assess Team Cognition in Distributed Mission Training for Close Air Support. Paper to appear in the Proceedings of the Human Factors and Ergonomics Society 52nd annual meeting.

Morey, J. C., Simon, R., Jay, G. D., Wears, R. L., Salisbury, M., Dukes, K. A., & Berns, S. D. (2002). Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams Project. *Health Services Research, 37,* 1553-1581.

Murphy, J., Grynovicki, J. O, & Kysor, K. P. (2003). Case study of a prototype set of Behaviorally Anchored Rating Scales (BARS) for C2 assessment. *Proceedings of the 8th International Command and Control Research and Technology Symposium, 17-19 June 2003, Washington DC.*

North Atlantic Treaty Organization (2005). Tactics, Techniques and Procedures for Close Air Support Operations (ATP-3.3.2.1(A)). Brussels, Belgium: Author.

Schreiber, B. T., & Bennett, W. Jr. (2006). Distributed mission operations within-simulator training effectiveness baseline study: summary report (Vol. 1) (AFRL-HE-AZ-TR-2006-0015-Vol I). Mesa, AZ: Air Force Research Laboratory.

Shepherd, A. (1998). HTA as a framework for task analysis. *Ergonomics, 41,* 1537-1552.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* 149-155.

United States Department of Defence (2005). Joint Tactics, Techniques and Procedures for Close Air Support (JP 3-09.3). Washington, DC: Author.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the heat of battle: taking a closer look at shared cognition breakdowns through teamwork. *Human Factors, 49*, 243-256.

Zobarich, Bruyn-Martin, & Lamoureux (2007). Forward Air Controller: Task Analysis and Development of Team Training Measures for Close Air Support. (DRDC Toronto CR 2007-156). Defence R&D Canada – Toronto.

This page intentionally left blank.