Confidence Bias in Situation Awareness

Ketut Sulistyawati and Yoon Ping Chui

Centre for Human Factors and Ergonomics
School of Mechanical and Aerospace Engineering
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
sulis@pmail.ntu.edu.sg, ypchui@unisim.edu.sg

Abstract. In this paper, we explore the concept of confidence bias in Situation Awareness (SA), i.e., the perception of own situational knowledge, a meta-cognition aspect of SA. Two studies were conducted to evaluate the nature of confidence bias across the present and future status, and across individual and team missions, as well as its relation with performance outcome. The results from both studies were consistent. Participants' confidence bias was higher in the future than present status, but did not differ significantly across individual and team missions. Participants who had lower confidence bias were found to have better performance.

Keywords: situation awareness, confidence bias, meta-cognition.

1 Introduction

Loss of Situation Awareness (SA) is often associated as the leading cause of performance errors in high risk, dynamic, and complex environment. SA is commonly viewed as knowing what is going on and projecting what is going to happen in the near future. This definition encompasses time elements of the "present" and "future". Endsley [1] further broke down SA to include three levels, namely perception of elements in the environment (level 1), comprehension of the current situation (level 2), and projection of future status (level 3). Assessing the level of operator's SA can be done objectively by comparing operator's situational knowledge and the actual situation. Various SA assessment tools are derived from this concept, for example, Situation Awareness Global Assessment Technique (SAGAT) [2], Situation-Present Assessment Method (SPAM) [3], Situation Awareness Control Room Inventory (SACRI) [4].

Subjectively, the operators can be asked to provide self appraisal of their SA level. This method requires the operators to direct their attention toward themselves and evaluate the extent to which they are aware about the situation. This self appraisal can be treated as a confidence level of a person towards his situational knowledge. It can determine the selection of actions, e.g., whether to gather more information or to act immediately. Existing tools to measure self-appraisal of SA include Situation Awareness Rating Technique (SART) [5], Situation Awareness Rating Scales (SARS) [6],

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2009, LNAI 5639, pp. 317–325, 2009.

and Crew Awareness Rating Scale (CARS) [7]. There are evidences that self appraisal is independent of the actual quality of the acquired SA. For instance, in a comparison between SAGAT and SART, no significant correlation [8] or only moderate correlation [9] was found between the two.

Alfredson [10] suggested that the difference between what an operator is aware of (i.e., objective SA) and what he thinks he is aware of (i.e., self appraisal, subjective SA) is an important SA indicator. Similarly, Nofi [11] suggested that SA is a function of the two measures, with high objective and high subjective scores being the best SA, and low objective and high subjective scores being the worst SA. For example, a person who thinks that he is fully aware of the situation but in fact he is not, may decide to act directly and confidently without further assessment of the situation. On the other hand, a person who is aware that he has not gain enough understanding of the situation may decide to further assess the situation before taking actions, or select more conservative actions. In her experiment involving air-to-air combat, Endsley [12] noted that low SA elicited using SAGAT did not necessarily accompany low performance. She suggested that this was because the pilots could modify their behavior and act conservatively when they knew that their own knowledge was incomplete.

Research on people' perception of own knowledge can be traced back to the 1980s. Fischhoff [13] revealed the tendency of people to over-estimate how much they knew about general knowledge. Lichtenstein et al. [14] summarized several studies that investigate overconfidence. It was reported to be higher for difficult items compared to easy items in questions related to general knowledge. Overconfidence was also shown for the calibration of future events, although the calibration for future events was somewhat better than for the general-knowledge items. In the context of SA, Lichacz et al. [15] involved 32 individuals in simulations, and collected the subjects' SAGAT scores and confidence in their responses. The result showed that the participants had less over-confidence bias in level 3 compared to level 1 and 2 SAGAT answers. Unfortunately, no other study in SA context is found, and the results from previous studies are somewhat inconsistent.

This paper aims to explore the nature of confidence bias across the present and future elements of SA and across individual and team missions. We hypothesize that there will be differences in the confidence bias across present and future status, and across individual and team missions. In addition, we also aim to investigate the relationship between absolute confidence bias and performance outcome. Based on the theoretical propositions described above, it is plausible to hypothesize that participants with lower absolute confidence bias would have better performance.

In the followings, we will present two studies conducted to explore the concept of confidence bias. The first study was conducted in a simulated urban warfare environment, and the second study in a simulated air combat environment. We use the notion of "absolute confidence bias" to represent the difference between self appraisal and what one is actually aware of. A higher bias would be associated with lower SA. Zero confidence bias or a good calibration means good self-awareness, knowing what one knows and does not know.

2 Study 1

2.1 Method

Task. An urban warfare was simulated using a multiplayer first-person shooting game, Counter-Strike™ (CS). There were two opposing teams in CS, namely counter-terrorists and terrorists. The mission's objective was to eliminate the opponent while inflicting the least injury on one's own team. The team members were assigned with the same responsibilities, i.e., to search and kill the opponent players.

Participants. Thirty-two students with age ranging from 18 to 24 years (M = 20.38, SD = 1.70) participated in the study. Their CS experience ranged from 1 to 9 years (M = 3.86, SD = 2.12). They played individual and team missions. Four participants with matched experience levels were involved in an experimental session. Two participants played against each other in the individual missions, and two teams of two participants played against each other in the team mission.

Assessment Tools. A paper-and-pen version of SAGAT was used to objectively assess participant's knowledge of the current and future situation. Examples of SAGAT queries include location, weapon, activity, and health level of each player, identification of opponent that is of a higher threat, whether opponent is within his weapon reach, and prediction of what will happen within the next 10 seconds, e.g., projected location, direction of movement, activity, weapon reach, and health level of all players. The participants were also asked to rate their confidence level (high/low) on each answer given to the SAGAT queries. A score for confidence bias was calculated by the average confidence rating across all items minus the proportion of the same items (SAGAT queries) that were answered correctly [16]. An absolute score of confidence bias indicates the distance from a well-calibrated zero-value of self awareness. The performance outcome was assessed by the health level (survivability) and the number of enemy successfully killed.

Set Up. Five sets of computers (four for participants and one for experimenter) with a network connection were used. The workstations were partitioned so that the participants could not see each other's display. The participants within the same team communicated using headsets through Skype™. The game and communication were recorded using Fraps™. Four small tables were placed behind the participants for them to answer the SA assessment sheets.

Procedure. The participants were first introduced to the study and briefed on the procedure of the experiment. They were given about five minutes to familiarize with the map used in the game. Throughout the missions, the simulation was paused twice to administer the SA assessment. During the pause, the participants were asked to turn around and answer the questions, which were placed on small tables behind them. They took about 1-2 minutes to complete the questions, and then the simulation resumed.

2.2 Result

In this study, the participants were assigned randomly as either terrorists or counter-terrorists. A one-way analysis of variance (ANOVA) was performed to determine whether the SA scores were independent of the side to which each subject was assigned. The analysis showed that none of the scores in individual and team missions was affected by the role assignment. The subsequent analyses will not differentiate between the two roles. The means and standard deviations of participants' absolute confidence bias in individual and team missions are presented in Table 1.

Table 1. Means and Standard Deviations for absolute confidence bias (N = 32)

		Individual mission		Team mission	
		M	*(SD)*	*M*	*(SD)*
Absolute confidence bias	Present	.15	(.11)	.14	(.09)
	Future	.22	(.17)	.28	(.19)

A 2 by 2 within subjects ANOVA was performed to evaluate participant's absolute confidence bias on present versus future status and on individual versus team missions. The result showed a significant difference between the absolute confidence bias for the present and future status, $F(1,31) = 23.00$, $p < .001$, where the participants were significantly more calibrated in how much they know about the present than future prediction (see Figure 1). There was no significant difference between the participant's absolute confidence bias in individual and team missions, $F(1,31) = 1.05$, $p = .31$. The interaction effect was also not significant, $F(1,31) = 2.16$, $p = .15$.

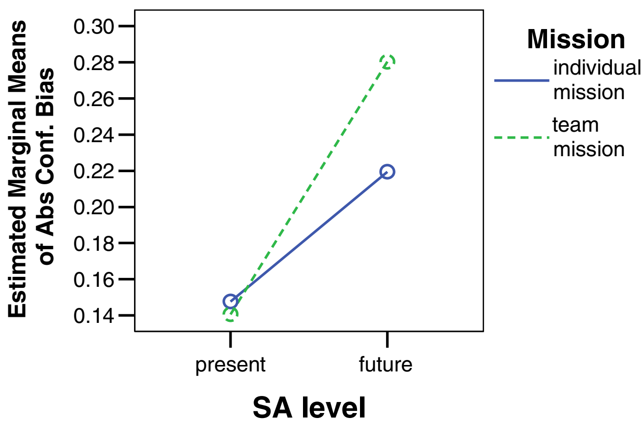

Fig. 1. 2 x 2 ANOVA of absolute confidence bias (Study 1)

In this study, the participants played against each other, and thus the performance score was tied to the opponent participant who played in the same game. The existence of active resistance, i.e., the opponents directly trying to prevent the team from accomplishing its goals, can increase the uncertainty, where the outcome of a team does not only depend on its capability and team processes, but also depends on the capability and processes of the opponent team. As the players' experience levels within one game were balanced, the losing team in that game was not necessarily worse than the winner of the other game. Two observers identified four better and worse performers from the entire pool of subjects. Independent-samples *t* tests were conducted to evaluate whether the better and worse performers differ in their absolute confidence bias. In individual mission, the difference between the better and worse performers were significant, $t(6) = -3.27$, $p < .05$. In team mission, the difference was marginally significant, $t(6) = -.2.26$, $p = .06$. The result suggested that participants with better performance were more calibrated in their SA (lower absolute confidence bias).

3 Study 2

3.1 Method

Task. Air combat environment was simulated using a PC-based simulation game, Falcon 4.0. Two fighter pilots were involved in each experimental session. The mission's objective was to sweep all adversaries along the assigned navigation route. An additional objective to reach a designated checkpoint at specific time was incorporated such that the participants would meet the pre-programmed enemy aircrafts at the targeted location.

Participants. Sixteen military fighter pilots participated in the study. They age between 25 to 37 years ($M = 29.43$, $SD = 3.08$), with flight experience ranging from 390 to 2600 flying hours ($M = 1091$, $SD = 530$). They were assigned into eight teams of two, one as the flight lead and the other as the wingman.

Assessment Tools. Similar to Study 1, a paper-and-pen version of SAGAT was used to assess participant's knowledge of the current and future situation. Examples of SAGAT queries including determining location, altitude, heading, and airspeed of own aircraft; bearing, range, and altitude of enemy aircraft (SA level 1); determining whether the aircraft is within the enemy's weapon envelope (SA level 2); predicting whether they will be in a position to take shots in the next 10 seconds (SA level 3). The participants were also asked to rate their confidence level (high/low) on each answer given to the SAGAT queries. A score of absolute confidence bias was derived as described in Study 1. The performance outcome was measured by the number of times the aircraft was shot down by the enemy (survivability).

Set Up. Two simulator consoles and one control station with a network connection were used. A partition was placed in between the two simulator consoles to prevent the subjects from seeing each other's screen during the team missions. A Thrustmaster® HOTAS Cougar Flight Controller, which is an almost exact replica of the flight control used in F-16 aircrafts, was used to control the simulation. Skype™ was used to facilitate team communication, and Fraps™ to record the simulation.

Procedure. After being introduced to the study, the participants were given adequate time to familiarize with the system. The actual missions were paused several times at specified trigger points for data collection. During the freeze, the participants were asked to turn around and answer the SA assessments placed on tables behind them. The participants took about 1 to 3 minutes to fill up the questions, depending on the number and type of questions asked in each freeze. At the end of the session, the participants were debriefed.

3.2 Result

The means and standard deviations of participants' absolute confidence bias in individual and team missions are presented in Table 2.

Table 2. Means and Standard Deviations for absolute confidence bias (N = 16)

		Individual mission		Team mission	
		M	*(SD)*	*M*	*(SD)*
Absolute confidence bias	level 1	.10	.08	.17	.15
	level 2	.11	.07	.15	.17
	level 3	.24	.17	.36	.21

A 2 x 3 within subjects ANOVA was performed on participants' absolute confidence bias. The result showed significant main effect of the SA level, $F(2,28)$ = 17.96, $p < .001$. The absolute confidence bias was significantly higher for level 3 than for level 1 and 2 SA (see Figure 2). The absolute confidence bias did not differ significantly across individual and team missions, $F(1,14)$ = 3.58, p = .08. The interaction effect was also not significant, $F(2,28)$ = .36, p = .70.

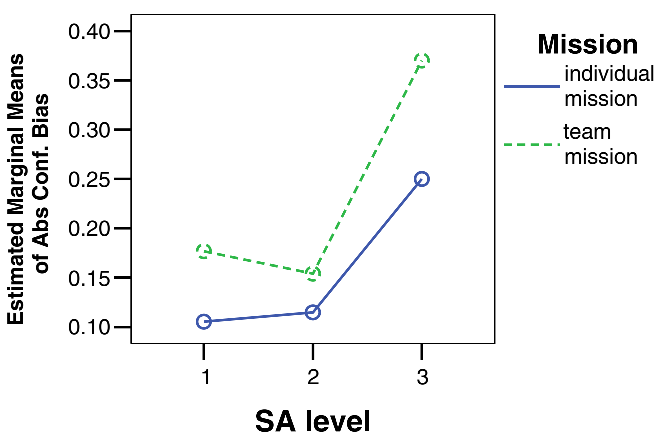

Fig. 2. 2 x 3 ANOVA of absolute confidence bias (Study 2)

Pearson product-moment correlation coefficients (one-tailed) between the absolute confidence bias and performance scores in individual and team missions were calculated. The results were significant in both individual mission, r = .63, $p < .01$, as well as team mission, $r = .68$, $p < .05$, indicating that participants with better SA calibration (lower absolute confidence bias) had better survivability (less damage due to enemy shots).

4 Discussion

The results from the two studies were consistent. The effect of SA levels on confidence bias was significant. The participants in both studies had higher confidence bias in their responses in the future (or level 3 SA) compared to the present situation (or level 1 and 2 SA). In other words, they were more calibrated on what they thought they knew about the present than the future situation. This contradicts to the similar study by Lichacz et al. [15], who reported that the participants had less confidence bias in level 3 compared to level 1 and 2 SAGAT answers. However, this is somewhat inline with Lichtenstein et al. [14], who reported that the confidence bias was higher for difficult compared to easy items. SA probes of the present situation required the participants to report specific information such as current location, flight parameter, and location of enemy. The participants usually knew when they did not report this information accurately and reported a low confidence when they were unsure on the accuracy of their answer. On the other hand, it was harder to gauge whether their prediction of what is going to happen would be correct or not, thus resulting in a poorer judgment on the accuracy of this future knowledge.

The impact of mission type (individual or team) on confidence bias was not significant in both studies. Although the effect was not statistically significant, Figure 2 suggests that we cannot totally dismiss this factor. As things get more complicated in team mission, due to the presence of teammate and more enemies, there might be a chance that the confidence bias in team mission would be higher than in individual mission. Further studies to explore this issue are warranted.

With respect to performance outcome, both studies suggested that better performing individuals and teams had significantly better calibration on what they knew (lower confidence bias), which was consistent to the hypothesis. When people are over-confident, they will prematurely close off the search of evidence, feeling that they "know the truth", and they will less likely to seek additional needed information, and confidently making decision and taking action that are prone to errors. Over-confidence of own ability is dangerous as one would bear to take higher risk, thinking that they are doing better than they really are, and does not hesitate to take aggressive approach when the situation indeed should be handled with greater caution. For instance, as observed in the second study, some of the fighters over-confidently continued with enemy engagement as they thought they knew the situation well, which indeed they did not. This poor decision resulted in low survivability (being shot by the enemy). A less over-confident pilot would rather abort the engagement and planning for retarget. On the other hand, people who are under-confident mainly showed low situational knowledge and that the correctly guessed answers to SAGAT questions were mainly due to luck. Under-confidence can make people to be more cautious and possibly hesitant to engage decisions and actions, resulting in less chance to win the mission.

Finally, we also noted that the distribution of confidence bias in Study 1 ranged from under-confident to over-confident, reflecting that some participants felt less confident on what they knew while others thought that they knew more that they actually did. Some participants, who were not sure on the answers to SAGAT queries, were allowed to write down their guess and asked to rate "low" in the respective confidence level. As a result, under-confidence bias partly represents the number of guessed answers given by the participants. In Study 2, however, we found that the pilots were rarely under-confident on their responses, which can be associated back to the nature of pilots who tend to be highly confident towards what they know and can do. This finding provided us some insights on how job nature and personality might have an impact on confidence bias. The individual differences might also be the reason for the inconsistent findings between the two studies reported here and the study by Lichacz et al. [15].

In summary, this paper provided a significant contribution to the fundamental theory of SA, specifically on the confidence bias towards own situational knowledge, which has not been much researched so far. With the significant impact of confidence bias on the performance outcome, future studies to better understand this concept are necessary.

References

1. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors 37, 32–64 (1995)
2. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE National Aerospace Electronic Conference, pp. 789–795 (1988)
3. Durso, F.T., Hackworth, C.A., Truitt, T., Crutchfield, J., Manning, C.A.: Situation awareness as a predictor of performance in en route air traffic controllers. Air Traffic Quarterly 6, 1–20 (1998)
4. Hogg, D.N., Folleso, K., Strand-Volden, F., Torralba, B.: Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plan control rooms. Ergonomics 38, 2394–2413 (1995)
5. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design, in Situational Awareness in Aerospace Operation (AGARD-CP-478). NATO-AGARD, Neuilly Sur Seine, France (1990)
6. Waag, W.L., Houck, M.R.: Tools for assessing situational awareness in an operational fighter environment. Aviation, Space, and Environmental Medicine 65, A13–A19 (1994)
7. McGuinness, B., Foy, L.: A subjective measure of SA: The Crew Awareness Rating Scale (CARS). In: Human Performance, Situation Awareness and Automation Conference. Savannah, Georgia (2000)
8. Endsley, M.R., Selcon, S.J., Hardiman, T.D., Croft, D.G.: A comparative analysis of SAGAT and SART for evaluations of situation awareness. In: Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society, pp. 82–86. Human Factors and Ergonomics Society, Santa Monica (1998)
9. Endsley, M.R., Sollenberger, R., Stein, E.: Situation awareness: A comparison of measures. In: Human Performance, Situation Awareness and Automation: User Centered Design for the New Millennium Conference. SA Technologies, Inc., Savannah, GA (2000)
10. Alfredson, J.: Differences in Situational Awareness and How to Manage Them in Development of Complex Systems. Linkoping University, Linkoping (2007)
11. Nofi, A.A.: Defining and Measuring Shared Situation Awareness. Center for Naval Analyses (2000)

12. Endsley, M.R.: Predictive utility of an objective measure of situation awareness. In: Proceedings of the Human Factors Society 34th Annual Meeting, pp. 41–45. Human Factors and Ergonomics Society, Orlando (1990)
13. Fischoff, B.: Perceived informativeness of facts. Journal of Experimental Psychology: Human Perception and Performance 3, 349–358 (1977)
14. Lichtenstein, S., Fischhoff, B., Phillips, L.D.: Calibration of probabilities: the state of the art to 1980. In: Kahneman, D., Slovic, P., Tversky, A. (eds.) Judgement under Uncertainty: Heuristics and Biases, pp. 3–20. Cambridge University Press, New York (1982)
15. Lichacz, F.M.J., Cain, B., Patel, S.: Calibration of confidence in situation awareness queries. In: Proceedings of the 47th Annual Meeting Human Factors and Ergonomics Society, Santa Monica, CA, pp. 222–226 (2003)
16. McGuinness, B.: Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self ratings. In: 9th International Command and Control Research and Technology Symposium: The Power of Information Age Concepts and Technologies, San Diego, California (2004)