Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, W. B. (2001). Predicting misuse and disuse of combat identification systems. ***Military Psychology***, 13(3): 147-164. (July 2001) Published by Taylor & Francis (ISSN: 1532-7876).

Predicting Misuse and Disuse of Combat Identification Systems

Mary T. Dzindolet, Linda G. Pierce, Hall P. Beck, Lloyd A. Dawe, and B. Wayne Anderson

ABSTRACT

Two combat identification systems have been designed to reduce fratricide by providing soldiers with the ability to "interrogate" a potential target by sending a microwave or laser signal that, if returned, identifies the target as a "friend." Ideally, gunners will appropriately rely on these automated aids, which will reduce fratricide rates. However, past research has found that human operators underutilize (disuse) and overly rely on (misuse) automated systems (cf. Parasuraman & Riley, 1997). The purpose of this laboratory study was to simultaneously examine misuse and disuse of an automated decision-making aid at varying levels of reliability. With or without the aid of an automated system that is correct about 90%, 75%, or 60% of the time, 91 college students viewed 226 slides of Fort Sill terrain and indicated the presence or absence of camouflaged soldiers. Regardless of the reliability of the automated aid, misuse was more prevalent than disuse.

Two combat identification systems have been designed to reduce fratricide by providing soldiers with the ability to "interrogate" a potential target by sending a microwave or laser signal that, if returned, identifies the target as a "friend." Ideally, gunners will appropriately rely on these automated aids, which will reduce fratricide rates. However, past research has found that human operators underutilize (disuse) and overly rely on (misuse) automated systems (cf. Parasuraman & Riley, 1997). The purpose of this laboratory study was to simultaneously examine misuse and disuse of an automated decision-making aid at varying levels of reliability. With or without the aid of an automated system that is correct about 90%, 75%, or 60% of the time, 91 college students viewed 226 slides of Fort Sill terrain and indicated the presence or absence of camouflaged soldiers. Regardless of the reliability of the automated aid, misuse was more prevalent than disuse, $F(1,65) = 31.43$, $p < .01$; $p = .27$ for misuse, $p = .13$ for disuse. Results are interpreted within a general framework of automation use (Dzindolet, Beck, Pierce, & Dawe, 2001).

The physical environment of the battlefield, the high stakes of combat, and the high stress of engagement contribute to a situation that increases the likelihood of fratricide, that is, soldiers erroneously identifying, shooting, and killing friendly troops. The high fratricide rates experienced during Operation Desert Storm led the U.S. Army to search for solutions to this problem (Doton, 1996). Two technological solutions, the Battlefield Combat Identification System (BCIS) and the Individual Combat Identification System (ICIDS; formerly the Combat Identification for the Dismounted Soldier) are currently under development. Both of these systems are being designed as decision-making aids for the identification of friendly troops--the BCIS for armor gunners and the ICIDS for the individual soldier. These systems provide the soldier the ability to "interrogate" a potential target by sending a microwave or laser signal that, if returned, identifies the target as a "friend." Unanswered signals produce an "unknown" response.

Ideally, soldiers will use reliable combat identification systems appropriately, leading the aided soldier (the soldier--system team) to identify friendly targets more accurately and to avoid fratricide more effectively than the soldier could without the system. Preliminary data collected during the BCIS Limited User Test (LUT) indicate this is possible. Knight and Spencer (1996) reported that trained Abrams tank and Bradley Fighting Vehicle gunners provided with the BCIS decision-making aids were able to identify targets with a high degree of accuracy. However, these results do not conclusively support the utility of the BCIS in combat situations. Confounding variables and the artificiality of the LUT environment limit their generalizability. Furthermore, these results are not supported by past research with human--automated "teams." This research has found that human operators do not rely on automated decision-making aids appropriately. Depending on the situation, human operators underutilize (disuse) or overly rely on (misuse) these aids (cf. Dzindolet, Pierce, Beck, & Dawe, 1999; Parasuraman & Riley, 1997). Because of the severity of the consequences of the soldier's decision, additional research is required to understand the likelihood of disuse and misuse of combat identification systems and to define the conditions for appropriate reliance.

If disuse of the combat identification systems is widespread, the systems will be ineffective, and have no impact on fratricide, and the resources expended on system acquisition will be wasted.

Disuse may be mitigated through system design or through training. The systems have been designed to be highly reliable; they can identify a friendly troop at a confidence level above 99% when the hardware on all friendly troops is functioning properly. Providing knowledge of an automated aid's low error rate has been found to decrease disuse (Dzindolet et al., 1999; Moes, Knox, Pierce, & Beck, 1999). Even providing unrealistically low error rates has been found to decrease disuse of alarm systems. For example, Bliss, Dunn, and Fuller (1995) reduced the cry-wolf effect (i.e., the disuse of alarm systems thought to be low in reliability) by having a confederate incorrectly disclose to students that 75% of the alarms that an automated aid would detect were correct. In fact, only 50% of the alarms were correct. Participants led to believe the alarm system was more reliable than it was responded to the alarms more often and more quickly than other participants who did not have this information. Thus, knowledge that the combat identification systems are highly reliable may reduce disuse rates and increase appropriate reliance on the combat identification system.

However, increasing the perceived reliability of the automated systems may lead to overreliance on (or misuse of) the aids. Although the reliability of the combat identification systems under near-perfect conditions is high, on the battlefield there are many variables that may decrease reliability and result in an "unknown" response to queries of friendly targets. For example, friendly troops may be involved in the battle who are not equipped with combat identification. Even if the interrogating system is functioning properly and the friendly target is equipped with combat identification, an "unknown" response will be received if the friendly target does not have a properly working transponder (the hardware that transmits a "friendly" signal back to an interrogating system). An improperly working transponder could be due to a lost antenna or damage during combat. Furthermore, the probability of the combat identification system returning an "unknown" signal after interrogating a friendly soldier changes in unpredictable ways.

Inaccurately informing participants that an aid is unlikely to err, although decreasing disuse, may cause misuse. Expecting the aid's decisions to be more reliable than their own, soldiers may overly trust the combat identification system in battle and may inappropriately rely on its decisions. Lee and Moray (1992,1994) found that individuals who trusted their automated aid's performance more than manual operation were more likely to trust and rely on their automated aid than those whose self-confidence in manual operation was higher than their trust in their aid. Furthermore, Moray, Inagaki, and Itoh (2000) found that, although self-confidence in manual operation was not affected by the automated aid's reliability, participants trusted more reliable automated systems more than they trusted less reliable ones. Therefore, it may be that increasing the perceived reliability of the automated systems, although designed to decrease disuse, may actually increase misuse, especially if soldiers are unsure under what conditions the system is likely to fail. Thus, the likelihood of disuse or misuse may vary with the perceived reliability of the automated system, with disuse being more problematic with less reliable systems, and misuse more problematic with highly reliable systems.

In summary, if soldiers are trained to appropriately rely on combat identification systems developed to reduce fratricide, and if the systems are designed to encourage appropriate reliance, fratricide rates may be reduced. Design and training recommendations depend on whether soldiers are more likely to misuse or disuse automated systems. To date, no study has simultaneously measured misuse and disuse of an automated decision-making aid. Although the propensity toward inappropriate automation use may differ across systems and individuals (Singh, Molloy, & Parasuraman, 1993), simultaneously examining misuse and disuse of a simple automated decision-making aid may help us to understand the likelihood of inappropriate use of the combat identification systems. The purpose of this laboratory study was to examine misuse and disuse of a simple automated decision aid at varying levels of system reliability.

To understand the effect of the reliability of the system (i.e., the accuracy of the system's identification decision) on human operators' likelihood of misuse or disuse, we used a between-subjects design. Due to practical constraints and a desire for a high level of control, we favored a laboratory study over a field study. Some participants worked with an automated aid that was correct about 90% of the time; others worked with an aid that was correct about 75% of the time; and others worked with an aid that was correct only slightly above chance, 60% of the time. A control group was comprised of participants who worked without an automated aid.

If participants appropriately relied on automated aids, then regardless of the aid's (or the participant's) reliability, aided participants should outperform participants who worked without an automated aid. We were particularly interested in examining whether the reliability of the automated system would differentially affect misuse and disuse rates such that misuse would be more prevalent than disuse for participants paired with the highly reliable system (90%) and disuse more prevalent than misuse for those paired with the less reliable system (60%).

METHOD

Although not designed to be an analog of BCIS or ICIDS, the human--automated-system teams created for this experiment were similar in decision-making aspects to a soldier using a ICIDS system for combat identification. The dynamics of this system apply as well to that of the BCIS decision-making process. Like the soldier with a combat identification system, the participants examined a battlefield, were provided with an automated system's yes-no decision regarding the likelihood of a target, and, with this information, made a decision.

Participants

Ninety-one Cameron University students participated in this study; complete data existed for 89 of the students. Most students received extra credit in a course offered in the Department of Psychology and Human Ecology for their participation, and guidelines set forth in the American Psychological Association Guidelines for Ethical Conduct were strictly followed.

Materials

The workstation contained a Hewlett-Packard Vectra PC, 133-MHz central processing unit with 32 Mb of RAM, including an S3, Inc. Trio 64 Plug-n-Play PCI video card. The 17-in. Hewlett-Packard Ultra VGA monitors were set at high color (16-bit) resolution, 800 x 600 pixels. The operating system used was Windows 95, Version 4.00.950. Slides of Fort Sill terrain were presented for about 0.75 sec. The reliability of the aid was manipulated such that, on average, the aid would be correct for about 60%, 75%, or 90% of the 226 slides, depending on the condition.

Procedure

After signing informed consent forms, participants read an instruction page along with the experimenter. They were told they would view 226 slides displaying pictures of Fort Sill terrain on a computer screen. (See Figure 1 for a sample slide.) The instructions indicated that about 24% of the slides contained one soldier ("target") in various levels of camouflage; the remaining 76% of the slides were of terrain only. Participants were told that sometimes the soldier would be rather easy to spot; other times he would be more difficult to find. Each slide would be presented on the computer screen for about 0.75 sec.

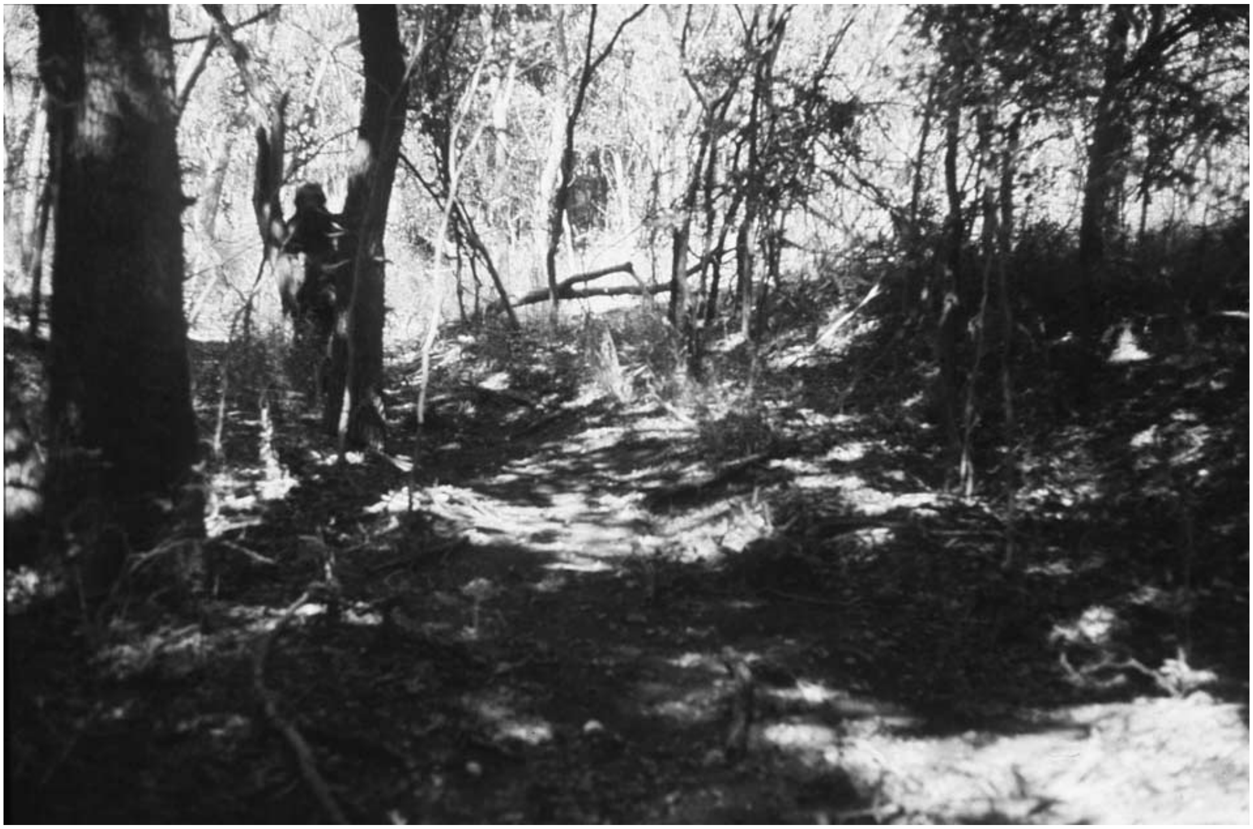

Figure 1 Sample slide shows Fort Sill terrain with a soldier present.

Participants in the aided condition were told that a computer program routine had been written to assist them in performing their task. They were told that the routine rapidly scanned the photograph looking for contrasts that suggested the presence of a human being. If the contrast detector routine determined that the soldier was probably present, the word present and a red circle would appear in the contrast detector box. If the contrast detector routine determined that the soldier was probably absent, the word absent and a green circle would appear in the contrast detector box.

Regardless of the condition, a screen asking the participants to indicate whether they believed the soldier was in the slide would appear next. They were told they had as much time as they needed to make their decision. Finally, they would be asked to indicate the extent to which they were certain their decision was correct. A 5-point scale ranging from 1 (highly confident) to 5 (not at all confident) was provided. Thus, the aid's decision was provided to participants after the slide was presented but before participants indicated their decision and their level of decision confidence.

The instructions explained that there were two possible errors that could be made: (a) indicating that the soldier was present when, in fact, the soldier was not; and (b) indicating that the soldier was not present when, in fact, the soldier was. Participants were told that both errors were equally serious and should be avoided.

Participants who were provided with the decisions of the automated aid were informed that the contrast detector was not perfect. Depending on the condition, they were told that previous research had found the routine to be correct on 60%, 75%, or 90% of the trials and incorrect on 40%, 25%, or 10% of the trials, respectively. Furthermore, they were told that they were not bound in their decision making by the response of the automated aid. Participants were instructed that, in all cases, the ultimate decision as to whether the soldier was present or absent belonged to them.

Participants performed four practice trials, were provided an opportunity to ask questions, and then proceeded to view the 226 slides. When they completed the task, participants completed a brief survey concerning their experience.

RESULTS

Overall Performance

Overall performance was measured by using signal detection rating-scale analysis (Dorfman & Alf, 1969; Dorfman, Beavers, & Saslow, 1973; Macmillan & Creelman, 1991). Response counts were determined for each participant for each confidence rating category contingent on whether a soldier was actually present. An overall cumulative response matrix for each participant then was determined beginning with a highly confident response that a soldier was absent and proceeding through the opposite extreme of high confidence that a soldier was present. The cumulative proportions were then z transformed and plotted. Such plots represent empirically determined receiver operator characteristics (ROCs). (See Table 1 for a sample participant's initial and cumulative responses, cumulative proportions, and z-transformed proportions.) For each participant, the slope of the ROC plotted in standard coordinates was determined through the use of an initial least-squares regression solution that was fed into a Marquardt procedure to find the maximum likelihood of fit for a Gaussian distribution (Marquardt, 1963). (See Figure 2 for a sample participant's ROC.) The resulting slopes then were tested to determine if d' would be an appropriate statistic to serve as a measure of detection. A t test comparing the mean slope of .49 to the null value of 1.0 indicated that the standard deviations of the noise and signal-plus-noise distributions were not equivalent, t(88) = -29.52, SE = .02, p < .01. This result indicated that the use of d' as a general index of detection sensitivity was inappropriate. Instead, the complement of the area under the ROC, A(z), was employed as the dependent measure of detection sensitivity (Simpson & Fitter, 1973). This statistic represents the probability of an error in a two-choice, forced-choice situation and is represented on a ROC plot as the area above the curve. To assess the effects of the various conditions, two other dependent measures were collected: (a) the slope of the ROC, which provides a measure of the ratio of the variances associated with the target plus noise (soldier present) and noise-only (soldier absent) distributions; and (b) c, a nonparametric measure of response bias at the no-yes decision criteria (Macmillan & Creelman, 1990, 1991).

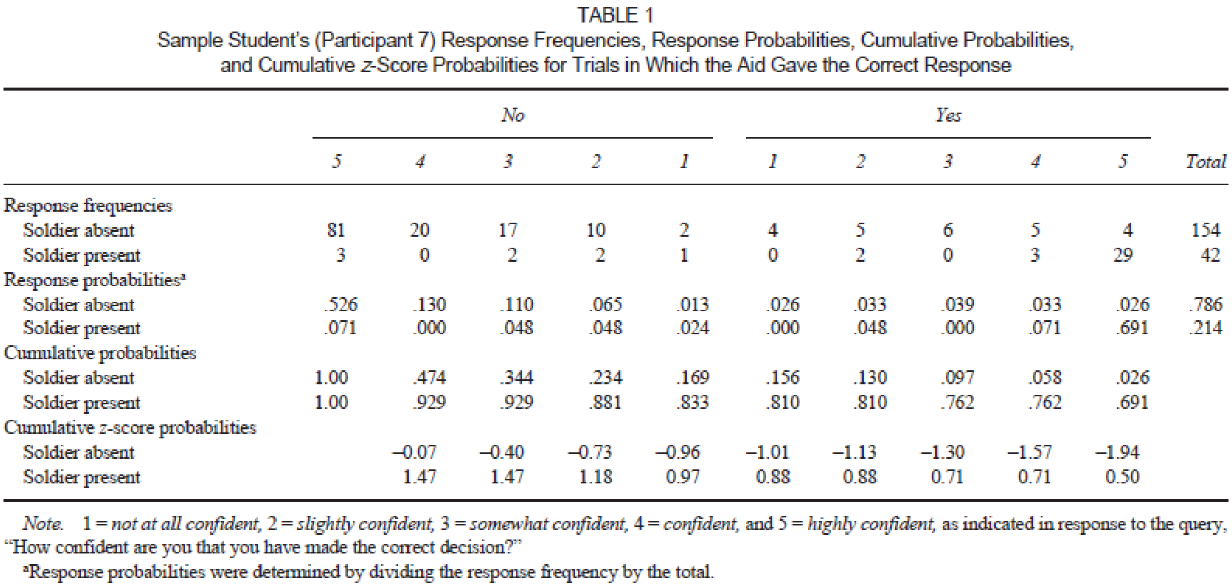

TABLE 1
Sample Student's (Participant 7) Response Frequencies, Response Probabilities, Cumulative Probabilities, and Cumulative *z*-Score Probabilities for Trials in Which the Aid Gave the Correct Response

	No					*Yes*					
	5	*4*	*3*	*2*	*1*	*1*	*2*	*3*	*4*	*5*	*Total*
Response frequencies											
Soldier absent	81	20	17	10	2	4	5	6	5	4	154
Soldier present	3	0	2	2	1	0	2	0	3	29	42
Response probabilities[a]											
Soldier absent	.526	.130	.110	.065	.013	.026	.033	.039	.033	.026	.786
Soldier present	.071	.000	.048	.048	.024	.000	.048	.000	.071	.691	.214
Cumulative probabilities											
Soldier absent	1.00	.474	.344	.234	.169	.156	.130	.097	.058	.026	
Soldier present	1.00	.929	.929	.881	.833	.810	.810	.762	.762	.691	
Cumulative *z*-score probabilities											
Soldier absent		–0.07	–0.40	–0.73	–0.96	–1.01	–1.13	–1.30	–1.57	–1.94	
Soldier present		1.47	1.47	1.18	0.97	0.88	0.88	0.71	0.71	0.50	

Note. 1 = *not at all confident*, 2 = *slightly confident*, 3 = *somewhat confident*, 4 = *confident*, and 5 = *highly confident*, as indicated in response to the query, "How confident are you that you have made the correct decision?"

[a] Response probabilities were determined by dividing the response frequency by the total.

Separate analyses of variance (ANOVAs) were completed on the error (detection sensitivity), slope, and bias data. No differences between conditions were found (a) for errors--without aid, M = .16 (n = 21); 60% aid, M = .16 (n = 22); 75% aid, M = .13 (n = 22); 90% aid, M = .12 (n = 24); and overall, M = .13; or (b) for slope--without aid, M = .45; 60% aid, M = .47; 75% aid, M = .53; and 90% aid, M = .49. The bias data indicated that, in all conditions, there was a bias to respond positively, which differed across conditions, F(3,85) = 5.48, p < .01. Results from the Tukey honestly significant difference (HSD) test for unequal sample sizes revealed that the mean bias for those working without an automated aid was greater than for those working with an aid that was correct 60% or 75% of the time: without aid, M = .64; 60% aid, M = .23; and 75% aid, M = .11. Given the task, the pattern of response bias is to be expected because there were far fewer slides in which the target was present than absent. Thus, to make an error, the automated aid was more likely to create a false alarm than a miss. The more errors (e.g., the 60% aid vs. 90% aid), the more likely the aid would make a false alarm decision. Relying on the automated aid, participants would also be more likely to indicate that the target was present. Thus, bias would be greater for those relying on less accurate aids than those working without an aid or with more accurate aids.

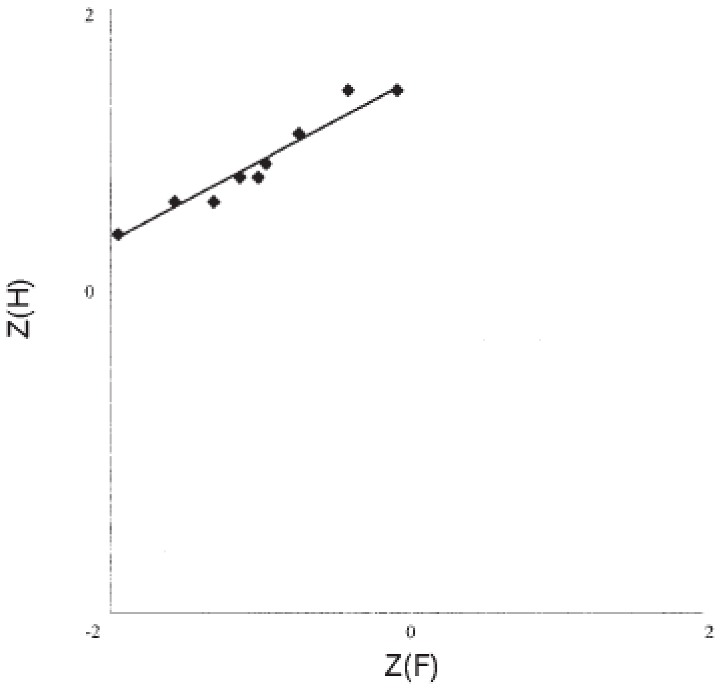

Figure 2 Sample receiver operator characteristic for Participant 7

Misuse and Disuse

To specifically examine misuse and disuse, the same process as previously described was conducted on two subsets of the data for the three conditions under which the automated aid provided information to the participant. One subset of the data consisted of all trials in which the aid provided correct information; the second subset consisted of all trials in which the aid provided incorrect information. As described previously, three dependent measures of error, slope, and bias were determined for each participant for each of the two subsets of data. (See Table 2 for means and standard deviations of the dependent variables for each condition.)

Misuse, or overreliance on automation, was defined operationally as the p(error|aid error); disuse, or underutilization of automation, was operationally defined as the p(error|aid correct). A 3 (automation reliability: 90% vs. 75% vs. 60% aid) x 2 (automation recommendation: correct or incorrect) ANOVA performed with the error data indicated a main effect of automation recommendation, F(1,65) = 31.43, p < .01. The mean proportion of errors was greater when the automated aid provided an incorrect decision than when the automated aid provided a correct decision, p(error|aid error) = .27, p(error|aid correct) = .13. Therefore, participants relied on the aid's decisions more than they ignored them. The main effect for automation reliability and the interaction were not significant. Thus, regardless of the reliability of the automated aid, misuse was more likely than disuse.

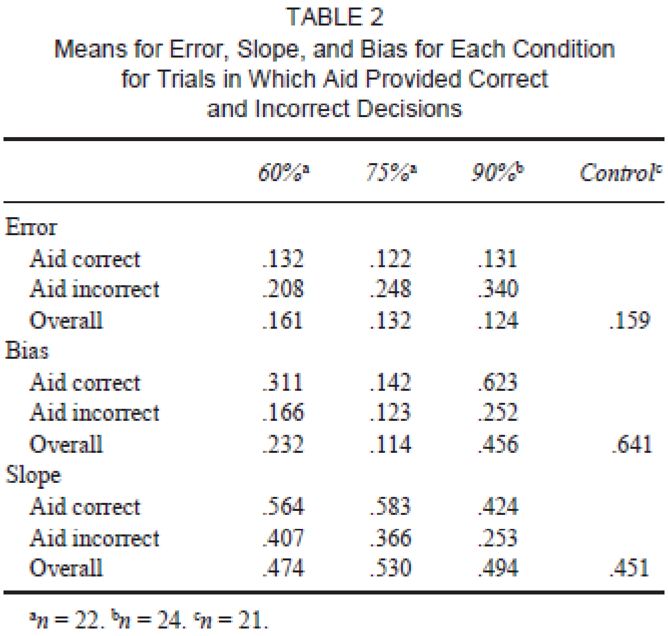

TABLE 2
Means for Error, Slope, and Bias for Each Condition for Trials in Which Aid Provided Correct and Incorrect Decisions

	60%[a]	*75%*[a]	*90%*[b]	*Control*[c]
Error				
Aid correct	.132	.122	.131	
Aid incorrect	.208	.248	.340	
Overall	.161	.132	.124	.159
Bias				
Aid correct	.311	.142	.623	
Aid incorrect	.166	.123	.252	
Overall	.232	.114	.456	.641
Slope				
Aid correct	.564	.583	.424	
Aid incorrect	.407	.366	.253	
Overall	.474	.530	.494	.451

[a]*n* = 22. [b]*n* = 24. [c]*n* = 21.

A 3 (automation reliability: 90% vs. 75% vs. 60% aid) x 2 (automation recommendation: correct or incorrect) ANOVA conducted on the slope data indicated an effect for automation recommendation, F(1,65) = 10.86, p < .02. The mean slope of the ROC plotted in standardized coordinates when the automated aid was incorrect was .26, compared with .52 when the aid was correct. Thus, the ratio between the variance associated with the target-present (target plus noise) distribution and the variance associated with the target-absent distribution (noise) was greater when the automated aid was incorrect than when the aid was correct. The main effect for automation reliability and the interaction were not significant.

A 3 (automation reliability: 90% vs. 75% vs. 60% aid) x 2 (automation recommendation: correct or incorrect) ANOVA conducted on the bias data revealed a significant effect for automation reliability, F(2,65) = 4.11, p < .03. A Tukey HSD test for unequal sample sizes indicated that those working with an aid that was correct 90% of the time were significantly more positively biased than those working with an aid that was correct 75% of the time: 75% aid, M = .02, and 90% aid, M = .48.

Slide Difficulty

To examine the hypothesis that participants relied on the decision-making aid only for difficult slides, each slide in which the target was present was assigned a difficulty score based on the number of people working without a decision-making aid who incorrectly stated that the target was absent. A three-way split was performed on the difficulty scores. Based on this split, each slide was placed into one of three categories of difficulty: high, medium, and low. Next, two probabilities were calculated for those working with aids that were correct 60%, 75%, and 90% of the time: (a) the probability of correctly stating that the target was present when the aid correctly informed the participant the target was present, and (b) the probability of correctly stating the target was present when the aid incorrectly informed the participant the target was absent. If participants relied on the aid only on difficult slides, then p(error|aid error) = p(error|aid correct) for the easy slides, but p(error|aid error) > p(error|aid correct) for the difficult slides. Table 3 presents these two probabilities for each condition under each level of slide difficulty. Regardless of slide difficulty, participants were more likely to misuse than disuse the automated aid. Given that the probabilities are in the same direction as in the overall analysis and, according to the Simpson paradox, that the presence of a third variable will lead to changes of direction of probabilities (Hintzman, 1980), slide difficulty did not affect participants' response strategies.

Self-Report Data

ANOVAs were performed to compare the responses to items on the questionnaire among participants in the four conditions. Those working with an aid that was correct 75% of the time did not enjoy the task as much as those working without an aid or with the aid at other levels of reliability (i.e., 90% or 60%), F(3,86) = 2.72, p < .05: without aid, M = 7.13; 90% aid, M = 6.86; 60% aid, M = 6.83; and 75% aid, M = 5.77, on a scale ranging from 1 (very little) and 9 (a great amount). No other differences among participants in the four conditions emerged. Table 4 presents the means and standard deviations of each item for each condition. Of interest was the response to the questionnaire item given only to participants provided with an aid, "To what extent did you use the information from the contrast detector to make your decision? I used the contrast detector ..." Participants were given a 9-point scale ranging from 1 (very little) to 5 (in between) to 9 (a great amount). Responses did not differ based on the reliability of the automated aid; the average response was 4.09. Thus, participants indicated they used it some but not a great amount.

TABLE 3
Probabilities of Incorrectly Identifying the Soldier Given the Aid's Decision Accuracy by Condition and Slide Difficulty

	Aid Misleads Participant p(Incorrect/Aid Incorrect)			*Aid Correctly Guides Participant p(Incorrect/Aid Correct)*		
Difficulty Level	60%	75%	90%	60%	75%	90%
High (of the 21 people working without an aid, less than 8 correctly identify the soldier)	.722	.736	.700	.648	.481	.614
Medium (of the 21 people working without an aid, 8–14 correctly identify the soldier)	.320	.403	.460	.306	.210	.280
Easy (of the 21 people working without an aid, more than 14 correctly identify the soldier)	.118	.107	.109	.076	.053	.051

TABLE 4
Means and Standard Deviations for Responses to Survey Items for Each Condition

	Without Aid[a]		*Aid Correct 60%*[b]		*Aid Correct 75%*[c]		*Aid Correct 90%*[d]		*Total*[e]	
Survey Item	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
How much effort did you put into finding the target? (from *very little* to *a great amount*)	7.69	1.00	8.17	0.94	8.05	0.95	7.95	1.28	8.03	1.03
How would you rate your performance? I was correct: (from *more often than most other students* to *less often than most other students*)	4.13	1.54	5.04	1.61	4.18	1.79	4.91	1.55	4.65	1.65
How quickly did you perform your task? I was: (from *quicker than most other students* to *less quick than most other students*)	3.96	1.57	4.57	1.50	4.09	1.63	4.33	1.32	4.23	1.51
How much did you enjoy performing your task? (from *very little* to *a great amount*)	7.13	1.54	6.83	1.92	5.77	1.85	6.86	1.49	6.66	1.76
With more time, do you think you would have made more correct decisions? (from *very few more* to *very many more*)	7.38	2.10	8.57	0.95	7.50	2.04	7.52	1.63	7.74	1.78
To what extent did you use the information from the contrast detector to make your decision? I used the contrast detector: (from *very little* to *a great amount*)			4.17	2.42	4.32	2.48	3.76	2.76	4.09	2.52

[a]*n* = 24. [b]*n* = 23. [c]*n* = 22. [d]*n* = 21. [e]*n* = 90.

DISCUSSION

Although overall performance was not affected by the presence or reliability of the automated aid, this is not the first study to find such results. Parasuraman, Molloy, and Singh (1993) and Singh, Molloy, and Parasuraman (1997) found that participants were not more likely to use more reliable automated aids than less reliable aids.

Participants were more likely to make an error by overly relying on the automated aid than by ignoring it. This pattern existed regardless of the reliability of the automated system and regardless of the difficulty of the slide. Because misuse was found to lead to inappropriate use of automated systems more than disuse, researchers should examine procedures to reduce misuse. What caused participants to misuse their automated system? A general framework of automation use, which posits that social, cognitive, and motivational processes combine to influence automation use, may be useful in answering this question (Dzindolet, Beck, Pierce, & Dawe, 2001; see also Figure 3).

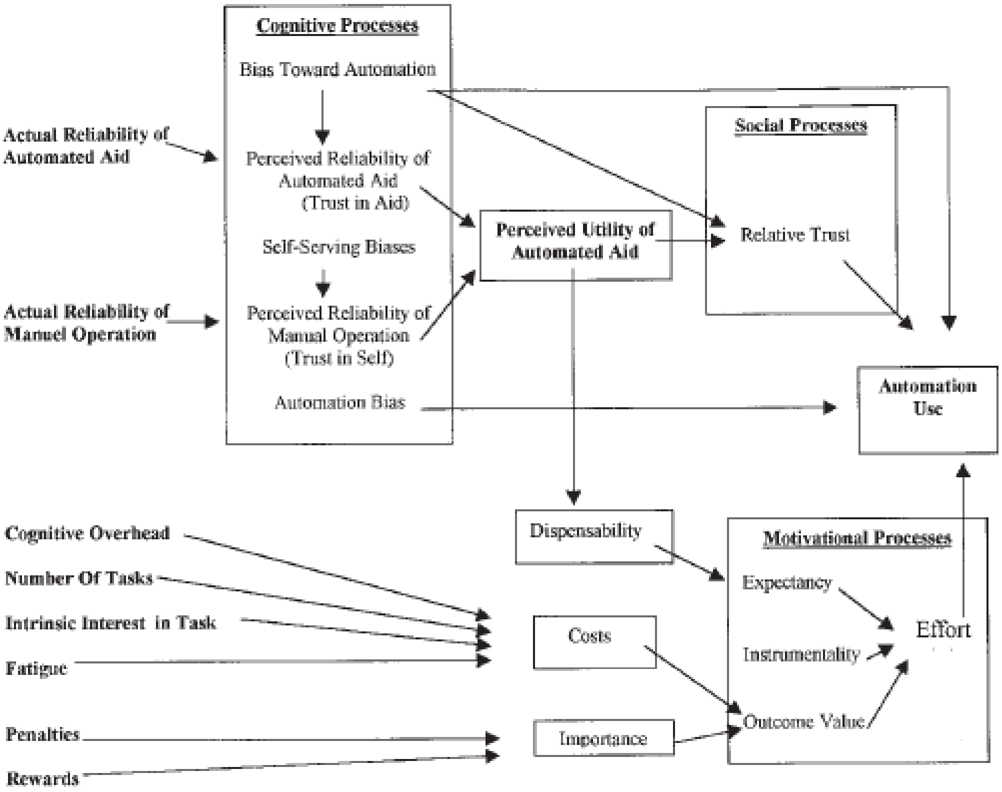

Figure 3 Framework to predict automation use (Dzindolet, Pierce, Beck, & Dawe, 1999).

Social Processes

Misuse is likely to occur when operators view the automated aid as an expert and place great trust in it (Mosier & Skitka, 1996). According to Dzindolet et al.'s (2001) framework, relative trust and automation use are determined from the outcome of a comparison process between the perceived reliability of the automated aid (trust in aid) and the perceived reliability of manual control (trust in self). The outcome of the decision process, termed the perceived utility of the automated aid, will be most accurate when the actual ability of the aid and the actual ability of the manual operator are compared. Unfortunately, the actual reliability of the aid and of the manual operator are unlikely to be accurately perceived by the operator. In reality, errors and biases are likely to occur, and the larger the errors and biases, the more likely misuse (and disuse) will occur.

Errors occur due to self-serving biases of the human operator because, without feedback, human operators are likely to overestimate their manual ability. For example, Dzindolet et al. (1999) found that participants erroneously assessed their performance to be superior to that of their automated aids, which led to disuse of the automated systems. Of course, this would not account for the misuse found in this study.

Another type of error occurs when human operators estimate the performance of their automated aid because a bias toward automation leads many people to predict near-perfect performances from automated aids (Dzindolet et al., 1999). This may be due, in part, to a demand characteristic created when the human operator and automated aid are paired, and human operators (soldiers and experimental participants) may feel obligated to use an automated aid issued to them. However, informing participants that their aids were incorrect 40%, 25%, and 10% of the time should have decreased the effect of this bias.

In addition, the reliability of the automated aid may be inaccurately assessed when participants do not clearly understand how the automated system works and, thus, have no basis for understanding why the aid made a mistake. Cohen, Parasuraman, and Freeman (1998) hypothesized that, to encourage appropriate use of automated aids, users must be trained to recognize situations in which the aid is (and is not) likely to supply them with accurate decisions. Perhaps even participants working with an aid that is correct 90% of the time distrusted the aid because they did not understand why it made errors. This disuse, however, was likely to be counteracted by cognitive or motivational processes, which led to misuse.

Cognitive Processes

Misuse may be the result of the automation bias (Mosier & Skitka, 1996, 1999), which is the result of the manner in which human operators process information provided by the automated aid. Rather than exerting cognitive effort to gather and process information, operators use the decision supplied by the automated system in a heuristic manner. Regardless of the reliability of the automated aid, the decision reached by the aid would influence the human operator's final decision. One way to reduce the automation bias, however, is to set the decision parameter so that it optimizes the detection and false-alarm rates of the soldier--system team. Sorkin and Woods (1985) reported that optimizing the decision parameter of the automated system (alone) can reduce the performance of the human--system team.

Motivational Processes

Diffusion of responsibility is a motivational process that may lead to misuse (Mosier & Skitka, 1996). When members work in a group, the responsibility for the group's product is diffused among all the group members (cf. Kerr & Bruun, 1983). Several researchers have thought of the human--computer system as a dyad or team in which one member is not human (e.g., Bowers, Oser, Salas, & Cannon-Bowers, 1996), leading to the conclusion that the human may feel less responsible for the outcome when working with an automated system than when working alone and may extend less effort.

One theory that has been successful in accounting for much of the findings regarding the diffusion of responsibility is Shepperd's (1993) expectancy-value theory. According to this theory, motivation is predicted from a function of three factors: expectancy, instrumentality, and outcome value. The first, expectancy, is the extent to which members feel that their efforts are necessary for the group to succeed. When members feel that their contributions are dispensable, or when one's individual contribution is unidentifiable or not evaluated, one is likely to take a free ride, or work less hard (Kerr & Bruun, 1983), leading to misuse of the automated system. Although dispensability should have varied with the reliability of the automated aid (e.g., participants paired with aids that were correct 60% of the time should have felt less dispensable than those paired with aids that were correct 90% of the time), identifiability and the likelihood of evaluation did not vary among conditions.

Instrumentality, the extent to which members feel that the group's successful performance will lead to a positive overall outcome, also is predicted to affect effort. Members who believe that the outcome is not contingent on the group's performance are less likely to work hard. Thus, misuse would be high among members who feel their group's performance is irrelevant.

Finally, the value of the outcome may affect misuse and disuse. Outcome value is the difference between the importance of the outcome and the costs associated with working hard to reach that outcome. Increasing the costs or minimizing the importance of the reward will lead members to put forth less effort, increasing the likelihood that they will rely on their automated aids. Costs vary with the intrinsic interest of the task, the number of other tasks one must perform, fatigue, and cognitive overhead. Importance of the outcome varies with personal importance of successfully completing the task and with the rewards and penalties of successful task completion. Because participants in this study earned extra credit in a class regardless of their task performance, the outcome value was relatively low. On the battlefield, however, the penalty of firing on a friendly soldier and not firing on an enemy target is so great that the outcome value is extremely high. As a result, misuse by the soldier in combat may not be as great as was found in the experimental setting. In fact, among such highly motivated people, disuse may become more of a problem than misuse, which is consistent with some interviews with Gulf War soldiers, who turned off their automated systems.

CONCLUSIONS

This was the first study to simultaneously measure disuse and misuse with an automated decision-making aid. Misuse was found to exceed disuse, even when the reliability of the automated system was very low. We hypothesize that the misuse may be due to three reasons: participants' overly trusting the automated system, the automation bias, and the modest outcome value.

Because of the widespread misuse of even unreliable automated systems, combat identification systems should be made as reliable as possible. It must be assumed that soldiers will rely on the combat identification decisions--even when the decisions are not highly reliable. In conjunction with this effort, perhaps training could reduce the likelihood of misuse. One training option may be to highlight when and why the system is likely to be unreliable so that soldiers can adjust their reliance on the system (e.g., Kirlik, Walker, Fisk, & Nagel, 1996) and then to give soldiers extensive experience using the system, as is now possible by using simulations such as the close combat tactical trainer or synthetic battlefields.

However, several limitations of this study beg for future research to be conducted in this area. First, research must be conducted in a more realistic, combatlike environment. Participants in this study performed only one task in a low-stress setting with few distractions, unlike the multitasked, chaotic, stressful environment of the battlefield. As a result, consequences for failures in this study were limited, whereas on the battlefield, penalties for incorrect decisions are often lethal. In addition, students teamed with the detector had the advantage of working with an aid of known reliability that remained constant throughout the experimental session. Although the meaning of a "friendly" signal is clear, an "unknown" response is ambiguous. To make rational use of the ICIDS or BCIS, soldiers must know the probability that an "unknown" signal does and does not indicate an enemy. Unfortunately, these probabilities change during battle in ways that are difficult to predict. For instance, the likelihood that an "unknown" response identifies an enemy may be very high until a large number of U.S. or allied forces without transponders enter the area. Battle damage to the transponders or antenna of friendly vehicles also will augment the frequency of "unknown" signals.

It is extremely important that researchers discover how soldiers interpret the "unknown" decision reached by the system, especially if the automation bias plays a large role in causing misuse. If soldiers are likely to interpret the "unknown" signal to be synonymous with the enemy, then fratricide may not be substantially reduced, especially in an engagement with mixed units (i.e., combat identification and non--combat identification units). Future research should be conducted to explore options, such as system design or training, to reduce misidentification of unknowns as enemies and to lead users to appropriately rely on this type of automated information. This study demonstrates that simply notifying participants of the automated system's error rate before interacting with the system does not encourage appropriate automation use.

In conclusion, the results from this study suggest that human operators are not especially sensitive to the reliability of an automated decision-making aid. Regardless of the aid' s reliability, human operators are likely to rely on the decisions reached by the aid. Whether this is due to the automation bias, overestimation of the trustworthiness of the automated aid, motivational factors of the human operator, or some combination of these, the situation should be explored in future research.

ACKNOWLEDGMENT

Portions of these data were presented at the International Conference on Human Interaction With Complex Systems, Urbana, IL, May 2000.

REFERENCES

Bliss, J., Dunn, M., & Fuller, B. S. (1995). Reversal of the cry-wolf effect: An investigation of two methods to increase alarm response rates. Perceptual & Motor Skills, 80, 1231-1242.

Bowers, C. A., Oser, R. L., Salas, E., & Cannon-Bowers, J. A. (1996). Team performance in automated systems. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 243-263). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998,July). Trust in decision aids: A model and its training implications. Paper presented at the 1998 Command and Control Research and Technology Symposium, Monterey, CA.

Dorfman, D. D., & Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: Rating method data. Journal of Mathematical Psychology, 6, 487-496.

Dorfman, D. D., Beavers, L. L., & Saslow, C. (1973). Estimation of signal detection theory parameters from rating-method data: A comparison of the method of scoring and direct search. Bulletin of the Psychonomic Society, 1, 207-208.

Doton, L. (1996,Winter). Integrating technology to reduce fratricide. Acquisition Review Quarterly, pp. 1-18.

Dzindolet, M. T., Beck, H. P., Pierce, L. G., & Dawe, L. A. (2001). A framework of automation use (Rep. No. ARL-TR-2412). Aberdeen Proving Ground, MD: Army Research Laboratory.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (1999). Misuse and disuse of human and automated aids with a decision-making task. Manuscript submitted for publication.

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. Psychological Review, 87, 398-410.

Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. Journal of Personality and Social Psychology, 44, 78-94.

Kirlik, A., Walker, N., Fisk, A. D., & Nagel, K. (1996). Supporting perception in the service of dynamic decision making. Human Factors, 38, 288-299.

Knight, W. R., & Spencer, W. J. (1996). Assessment of the Battlefield Combat Identification System (BCIS) Limited User Test (LUT). Alexandria, VA: U.S. Army Operational Test and Evaluation Command.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35, 1243-1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human--Computer Studies, 40, 153-184.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. Psychological Bulletin, 107, 401-413.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, 11, 431-441.

Moes, M., Knox, K., Pierce, L. G., & Beck, H. P. (1999,March). Should I decide or let the machine decide for me? Poster session presented at the annual meeting of the Southeastern Psychological Association, Savannah, GA.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. Journal of Experimental Psychology: Applied, 6, 44-58.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mosier, K. L., & Skitka, L. J. (1999). Automation use and automation bias. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 43, 344-348.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency." The International Journal of Aviation Psychology, 3, 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39, 230-253.

Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. Psychological Bulletin, 113, 67-81.

Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? Psychological Bulletin, 80, 481-488.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. The International Journal of Aviation Psychology, 3, 111-122.

Singh, I. L., Molloy, R., & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: Role of display location. International Journal of Human--Computer Studies, 46, 17-30.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. Human--Computer Interaction, 1, 49-75.