

This article was downloaded by: [Copenhagen University Library]  
On: 31 October 2012, At: 07:47  
Publisher: Taylor & Francis  
Informa Ltd Registered in England and Wales Registered Number:  
1072954 Registered office: Mortimer House, 37-41 Mortimer Street,  
London W1T 3JH, UK



## Ergonomics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/terg20>

### Use of an automatic tracker as a function of its reliability

SHARON M. MCFADDEN, BARRY L. GIESBRECHT & CHERYL A. GULA

Version of record first published: 10 Nov 2010.

To cite this article: SHARON M. MCFADDEN, BARRY L. GIESBRECHT & CHERYL A. GULA (1998): Use of an automatic tracker as a function of its reliability, *Ergonomics*, 41:4, 512-536

To link to this article: <http://dx.doi.org/10.1080/001401398186982>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Use of an automatic tracker as a function of its reliability†

SHARON M. MCFADDEN\*, BARRY L. GIESBRECHT‡ and CHERYL A. GULA

Defence and Civil Institute of Environmental Medicine, 1133 Sheppard Ave  
West, PO Box 2000, North York, Ontario M3M 3B9, Canada

*Keywords:* Automated controllers; Human-computer interaction; Trust; Reliability.

The present paper reports two studies investigating the use and usefulness of an automated tracker as a function of its reliability. The participants' task was to update the position of several targets when new information was received about the current position of existing targets, new targets, and noise signals. They could update the position of each target manually or they could assign one or more targets to an automatic tracker (AT), which used the same information available to the participant to update the position of the targets it was responsible for. In the first study, the reliability of the AT was varied from totally unreliable to very reliable. Participants' use of the AT and system performance increased as the reliability of the AT increased. However, actual use of the tracker was not a simple function of its reliability. Instead, use appeared to be a function of both AT reliability and the participants' ability to do the task manually. The second study examined system performance and use of an AT as a function of task difficulty (number of targets that had to be tracked) and initial reliability of the AT (high versus moderate reliability). When the task was more difficult, most participants continuously assigned targets to the AT independent of its reliability. There was also a significant correlation between AT use and the percentage of targets tracked. Some participants in the lower task difficulty condition did make less use of the AT, if they received the less reliable AT first. The results of these studies differ somewhat from previous research with automatic controllers, which have found that participants do not tend to use an automated system unless it is extremely reliable. Possible reasons for the difference were task difficulty and the fact that, with this system, participants retained ultimate control of the tracking task. If the AT did not track a target, the participant had the opportunity to handle that target manually. One finding that was consistent with the literature was that when the automated tracker was reliable, some participants failed to recognize errors made by it even with feedback and sufficient time to correct the errors.

### 1. Introduction

Advances in technology and the processing power of computers have made it possible for operators to detect and track more targets on sensor systems and tactical displays. As the number of potential targets increases, it becomes more difficult for an operator to monitor all of them. One possible method to reduce workload and to

---

†Some of the work reported in this paper was presented at the 12th Congress of the IEA, Toronto, Ontario, Canada, 14-18 August, 1994.

‡Currently a graduate student in the Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

\*Author for correspondence.

increase the number of targets that can be monitored simultaneously is to provide the operator with an automated tracking system. An automated tracker (AT) mimics the operator by scanning the incoming signals to determine which ones could represent the latest position of existing targets delegated to it.

It is inherently impossible to develop a perfect automated tracker where perfection is defined as 100% accuracy in tracking assigned targets. The automated tracker operates by examining all the signals in a predefined area around the last known position of each target that it has been assigned and choosing the signal with the highest probability of being the current position of that target. If this area is relatively small, fast moving or erratic targets will probably be lost. If the area is large, signals from other targets or non-targets could be associated with a given target.

Based on the above limitations, it is unlikely that a tracker will be designed that will replace an operator. The next best goal is to develop a useful tracker, one that will reduce operator workload and improve system performance over that achievable by the operator or the AT alone.

An examination of the literature suggests that it may not be a simple task to develop an AT that operators will use effectively and efficiently. Automated controllers, of which the AT is an example, are becoming an integral part of many complex computer-based systems. They are implemented with the aim of reducing human error and workload and improving system performance. However, recent research suggests that automated controllers may not meet these goals either because the human does not use them or because of failure to consider human capabilities and task requirements (Rouse and Morris 1986). For example, Muir and Moray (1989) and Lee and Moray (1992) have shown that the use of automated controllers varies as a function of the operator's trust in them. Unless a system is perceived to be reliable, an operator may not use it to its full potential.

Even when an automated controller is perceived as being reliable, individuals may still use it suboptimally in order to maintain some feeling of control over the system (Morris *et al.* 1988, Weisgerber and Savage 1990). For example, Morris *et al.* (1988) provided participants with an automatic aid for identifying targets on water. The aid was superior to the human in open water conditions and poorer than the human in narrow channels. In one condition, participants could allocate the identification task whenever they wished. In a second condition, the allocation of the task to the aid or the human was automated. Most participants preferred to control the allocation task even though performance was superior with automatic allocation.

In other cases, automated controllers may not be used because of the time and/or effort involved in engaging them. For example, participants in a study by Kirlik (1993) were required to fly a helicopter and monitor the activities of four other helicopters. The helicopter could be flown manually or under autopilot. The secondary task involved entering strings of commands from time to time. It was anticipated that the participants would use the autopilot (which controlled the helicopter less efficiently than was possible under manual operation) when they had to handle the secondary task. None of the participants followed this strategy. Kirlik attributed this result, in part, to the fact that it usually took longer to engage the autopilot than to complete a component of the secondary task.

If the automated system is used consistently, other problems can arise. Automated controllers frequently change the role of the operator from system controller to system monitor. The human is retained primarily to intervene if the

automated system encounters an unfamiliar situation or fails. However, humans are notoriously poor at long-term monitoring and are likely to miss errors made by the automated system. In addition, it is difficult for humans to retain skills that they do not use (Rouse and Morris 1986). Thus, the more they rely on the automated system, the less probable it is that they will be able to handle unfamiliar situations.

The results of these studies have strong implications for the potential usefulness of an automatic tracker. As indicated above, if the search area of the AT is too small, operators are likely to perceive them as being unreliable. On the other hand, a large search area can lead to false alarms that the operator may miss. If automatic trackers are to be useful, we need a better understanding of how people use them under differing levels of reliability and what the impact of use and reliability is on system performance. Thus, the current experiments investigated the use of an automatic tracking system as a function of its reliability.

Two experiments are reported here. In the first experiment, four different levels of reliability were investigated. The aim was to determine how use of an automated tracker would vary as a function of its reliability as well as the impact that level of reliability had on system performance. Based on the existing literature, it was predicted that participants would under-use the AT except at the highest level of reliability. Although participants varied in their use of the AT, they all made some use of it even under the lowest level of reliability (above baseline). This deviation from the expected results was attributed to the difficulty of the task and to participants' expectations about the reliability of the AT on any given run. To investigate these conclusions, a second experiment was carried out in which the authors attempted to manipulate both task difficulty and expected reliability.

## 2. Automatic Tracking System (ATS)

The ATS is a simulation of a target tracking task developed to study human use of automated controllers. The task presented by the ATS is to detect and then track the location of various targets (e.g. vessels) over time using information about the current location of a set of signals that is presented at regular intervals. Some of the signals are due to the targets being tracked, some to targets that have not yet been detected, and the remainder are spurious signals (noise signals) that appear from time to time. An automated tracker (AT) can be activated to assist the user in the task. The user activates the AT by assigning it one or more targets that he or she has already detected. The AT then attempts to update the position of those targets whenever new signal information is provided.

An example of the ATS screen is shown in figure 1. The functions that the user must carry out to complete the task are performed through a series of tracking display, signal table, and function button selections. All selections are carried out via a mouse-controlled cursor (position cursor over marker, signal, or function button and push mouse button). Invalid selections are indicated by a beep or no action following a selection.

### 2.1. Tracking display

On the left half of the screen is the tracking display. It defines the user's field of view. The purpose of the display is to enable the user to keep track of the position of the targets that he or she is responsible for. The user's own ship is in the centre of the two concentric circles. Four classes of target markers may appear in the display: manual-unassociated, manual-associated, AT-unassociated, or AT-associated. The markers

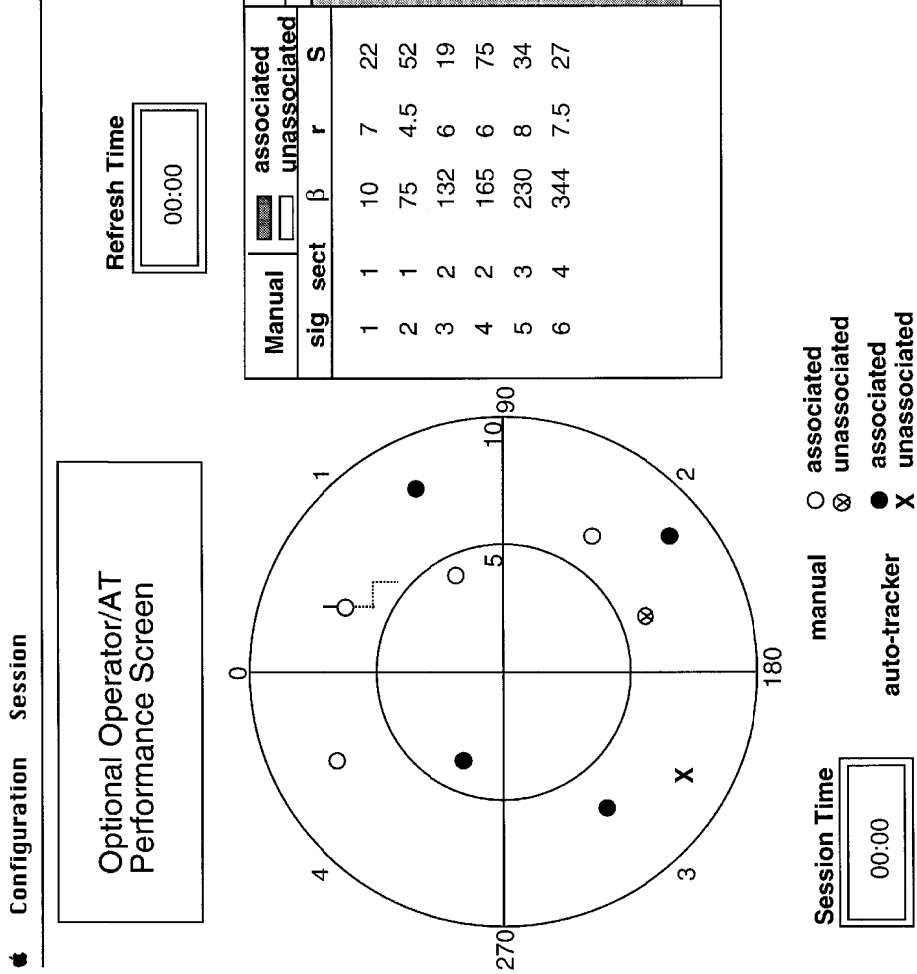


Figure 1. A schematic of the display for the Automatic Tracking System (ATS). The 'X' for the unassociated-AT marker appears as a white X in a black circle on the actual screen.

appear as small circles. The manual markers (white circles) indicate targets that are currently the responsibility of the user and the AT markers (black circles) indicate targets that are the responsibility of the AT. If a circle has an X in it, the marker does not have a signal currently associated with it. If specified by the experimenter, each marker will have a track associated with it that traces out the last 'n' positions occupied by that marker ( $n$  is defined by the experimenter) and a short line indicating the projected direction and speed of the target associated with that marker.

Actions that can be performed in the tracking display are marker selection and marker/signal de-selection (position cursor anywhere in the display where there is not a marker).

## 2.2. Signal tables

To the right of the tracking display is a signal table. It provides the user with signal information including: signal number ( $\text{sig} = 1 - n$ ), the sector the signal is coming from ( $\text{sect} = 1 - 4$ ), the signal's angle of deviation from true north ( $\beta = 0 - 360^\circ$ ), the radial distance of the signal from the user ( $r = 1 - 10$ ), and signal strength ( $S = 1 - 100$ ). There are two tables—manual and AT. The manual table shows the set of signals the user must handle and the AT table maintains current signal information for targets (or non-targets) assigned to the AT (black target markers without X only).

User-actions that can be performed in the signal table are: signal selection, scrolling of signal table, and switching between manual and AT signal table (position cursor in box labelled manual in figure 1).

Two types of signals appear in the signal table. Signals generated by targets (target signals) and signals generated by non-targets (noise signals). The strength, radius, sector, and angle of target signals are calculated using parameters for that target that were pre-set by the experimenter (see §2.6) and the length of time since that target was initiated. The position and direction of noise signals are randomly determined and they exist only for one to two updates. Also, their strength is randomly determined with the exception that it never exceeds 20.

The number of target and non-target signals that appear in the signal tables on each update is a function of a set of predefined parameters. These parameters are the maximum number of targets that must be tracked on each update, the rate at which new targets will be added, the maximum number of non-target signals that can appear during each update, and the probability that each non-target signal will appear.

## 2.3. Function buttons

To the right of the signal table are seven function buttons whose purpose is to provide the user with control over signal-target manipulation. The action that can be performed here is button selection. Each control button has inherent prerequisites that must be met before the button is selectable. Buttons that are not selectable are greyed. The prerequisites, actions, and events associated with each button are shown in table 1.

## 2.4. Timers

Two timers are present on the display. A session timer resides in the lower left corner of the display and an update timer resides immediately above the signal table. The session timer shows the total time remaining in a given scenario or run and the

Table 1. Prerequisites, actions, and events associated with function buttons in figure 1.

Function	Prerequisite	Action	Event
Add	Unassociated signal	Select a signal from the manual signal table; select Add button	Target marker appears on the tracking display and selected entry in signal table highlighted
Associate	Unassociated manual target marker and unassociated signal	Select an unassociated signal and unassociated target marker; select Associate button	X removed from target marker and signal table entry highlighted
Deassociate	Associated manual target marker	Select either a manually associated target marker or the associated signal entry; select Deassociate button	Entry in signal table no longer highlighted and target marker has X in it
Remove	Target marker	Select a manual or AT target marker; select Remove button	Target marker disappears from display. If it was associated with a signal, signal entry appears in manual signal table and is not highlighted
Assign	Manual target marker	Select a manual target marker from the display; select Assign button	Target marker changes from white to black; signal entry moves from manual to AT signal table
Deassign	Unassociated or associated AT target marker	Select an AT target marker; select Deassign button	Target marker changes from black to white; if marker was associated with a signal, signal entry moves from AT to manual signal table
Undo	Previous user action	Select Undo button	Results of previous action change back

update timer shows the time remaining before the information in the signal table will be updated. The maximum values for both timers are defined by the experimenter.

### 2.5. Feedback window

At the discretion of the experimenter, a feedback window appears above the tracking display. It provides the user with information on how well he or she is handling the target and noise signals in the manual signal table. Similarly, when the AT signal table is displayed, the same window shows the AT's score. Feedback is presented in graphical format and shows performance (0–100) over the 10 most recent updates. The scores are a function of the percentage of targets that the user or AT has successfully updated and the percentage of noise signals incorrectly added during each update period.

### 2.6. Targets

Target sets or scenarios are usually generated in advance and stored in files that are then loaded at the start of a run. Each target in the scenario is defined in terms of its

type, starting distance from the centre of the tracking display, angle of deviation from true north, initial direction, speed, and type of path traced out. The possible values for these parameters are shown in table 2. Additional parameters must be defined for the circle, zigzag, and experimenter-defined paths. For circles, a centre point and radius are required. For the zigzag, the extra parameters are the length of the zig and the zag and the angles between the zig and zag and the zag and zig portions of the path. The experimenter-defined path is an extension of the zigzag in that the experimenter defines a series of lengths and angles that the target will follow.

Each target in the set is added in turn, at a rate specified by the experimenter, to a list of active targets until the maximum number has been reached. The initial location of the target is defined by its starting distance and its angle of deviation from true north. Its initial strength is a function of type, distance, direction, and speed. Subsequent locations and strengths are determined on the basis of distance, speed, current direction, and type of path. If the target's path takes it outside the area of the tracking display, it is removed from the list of active targets and the next target in the set is added to that list.

### 2.7. Task

Initially, the tracking display is empty and the manual signal table contains information about a set of signals. The user must determine which of these signals are due to targets and 'add' these targets to the tracking display. It is then possible to 'assign' the target to the AT. Thereafter, at regular intervals (determined by the update rate), a new set of signals is presented to the user. When this occurs, the update timer is reset and an X is put in all of the target markers to indicate that their position no longer reflects the locations of the signals in the signal table. At the same time, the AT attempts to perform associations between the new signals and the target markers that it has been assigned. The signals it does not associate with a target appear in the manual signal table. The target markers it does not associate with a signal will appear as a white X in a black circle.

Every time the signal table is updated, the user's task is to determine which signals are due to targets, to 'associate' these signals with existing target markers (i.e. track the targets by updating their position on the display), or, if no suitable marker exists, to 'add' the signal to the tracking display. If the signal represents the current location of a marker that was assigned to the AT, the user must 'deassign' the marker, 'associate' the signal with the marker and then, if desired, 'assign' the marker to the AT again. The user can also remove any target markers that remain unassociated. This usually occurs when a target moves beyond the area of interest, or when the user thinks that a noise signal has been added incorrectly to the tracking

Table 2. Range of values of target parameters.

Parameter	Values
Target type	Surface non-military, surface military, subsurface nonmilitary, subsurface military
Initial radius	0 to 10
Initial angle	0 to 360°
Initial direction	North, NW, West, SW, South, SE, East, NE
Speed	> 10
Type of path	Point, line, circle, zigzag, experimenter-defined



display. However, users may also choose to remove markers for targets that they have not been able to update in the previous few periods.

The user should also check that the associations made by the AT are correct. If the user thinks the AT has incorrectly associated a signal with a target marker he or she can 'deassign' the marker, 'deassociate' the signal from its current marker, 'associate' the target marker with a different signal, and 'reassign' the target to the AT.

### 2.8. Automatic Tracker (AT)

The AT functions by comparing the position (radius and angle) of each target marker assigned to it with the position of each unassociated signal in the signal table. The AT uses an algorithm to compare the distance between the positions of a given marker and signal and assigns a probability that is a function of that distance. The signal with the highest probability is associated with the marker provided the probability exceeds some predefined minimum. This minimum, called the association threshold, is defined by the experimenter. Different association thresholds can be assigned for high- and low-strength signals so that the AT has a greater probability of losing low-strength signals. To reduce the likelihood that an incorrect signal will be associated with a target marker, only signals within a specified area around the target marker receive probabilities greater than zero. This area is defined by an ellipse centred on the marker. The parameters of the ellipse are defined by the experimenter.

The larger the ellipse, the greater the likelihood that the AT will associate a signal with a target and, at the same time, the greater the likelihood that an incorrect signal or noise signal will be associated with a marker. The higher the association threshold required, the smaller the likelihood that a signal will be associated with a given target and the greater the likelihood that the signal associated with that marker is the correct one.

For a given size of ellipse and threshold, the performance of the AT will be affected by the number of target and non-target signals in the signal table, the type of path traced out by the targets, and their speed. The more signals on the display, the more likely that the signal with the highest probability will not be the correct signal. The faster a target moves, the more likely that successive signals from that target will fall outside the area searched by the AT. Similarly, targets tracing out complex paths such as zigzags and circles are more likely to generate signals that do not meet the AT's criterion of acceptance.

## 3. Experiment 1

Of primary interest was to examine the effect that AT reliability would have on performance. To do so, it was necessary to systematically vary the number of targets that the AT would track and to make the task difficult to carry out manually. As discussed above, the reliability of the AT could be varied either by varying the size of the area searched or by varying the threshold for accepting a match. Based on the results of pilot studies, it appeared that performance of the AT could be varied most consistently and effectively by varying the size of the area searched. Three different areas were selected that, based on the results of the pilot studies, would produce a low, moderate, and high reliability AT, respectively.

Task difficulty could be varied by giving the participant more targets to track than he or she could update manually or making it difficult to discriminate target

signals from non-target signals. Since a target could only be assigned to the AT if the participant had initially detected it, the task was set up so that it was difficult for the participant to handle all of the targets manually before new information was presented. In addition, the scenarios did include targets whose strength was initially less than 20 — making it difficult to discriminate them from noise signals. The update rate was set at 40 s to allow a sufficient quantity of data to be collected within a reasonable time frame. This is more frequent than would be found with an operational tactical display where the update rate is probably of the order of minutes. However, it is not out of line with the update rate for other systems that might also employ an automated tracker such as a passive sonar system or a target tracking system on a tank. Based on the results of earlier pilot studies, the authors determined that, with training, most participants could update the position of about eight targets manually in 40 s, if no non-target signals were present in the signal table. With an average of six non-target signals, it was possible for participants to handle about six targets in 40 s.

Two control conditions were run. In both, the area the AT could search was set small enough that the AT never tracked a non-stationary target (nil reliability). The main control or baseline condition was identical to the test conditions in all other respects while the second nil reliability condition had a longer update period. Since task difficulty for a given set of scenarios was, in part, a function of update period, the authors wanted to assess the effect of providing a longer update period. However, it should be remembered that the distance a target travels is also a function of the update period. Thus, the change in the signal's position between successive update periods was greater.

### 3.1. Method

3.1.1. *Participants:* A total of eight participants, five males and three females, took part in the first experiment. Four of the males and one of the females were military personnel. The remainder were civilian in-house personnel or university students. The age range was 19–41 years and all participants had normal or corrected-to-normal vision. They were compensated for their participation.

3.1.2. *Apparatus:* The experiment was run on a Macintosh IIfx (Toronto). Stimuli were displayed on a 13-in. RGB monitor (Apple Canada, Inc., Toronto,) and interaction with the computer was by means of a cursor controlled mouse.

3.1.3. *Target sets:* Three different target sets or scenarios were created. Each contained a selection of straight line, zigzag, and circular target paths. Most of the targets had a strength greater than 20 for at least 75% of the time. The paths of the targets were selected to keep the number of collisions or near collisions small. The maximum number of non-target signals that could appear in a given update was 11 and the probability of each non-target signal being generated was 80%. In practice, the number of non-target signals on each update period averaged around six.

3.1.4. *Conditions:* The experiment was a completely within-participant design. All the participants completed three test runs, one with each scenario, under five different conditions. Four of the conditions varied in terms of the size of the area around the target marker that the AT would look for a signal. The four areas were

2, 20, 40, and 60 (nil, low, moderate, and high reliability). These numbers make sense only within the context of the experimental control system in that they are a function of the numerical values employed within the ATS for other parameters such as the radius of the tracking display. They do not bear any relation to real units of measurement. The fifth condition used an area of 2 and an update period of 50 s.

3.1.5. *Procedure*: Initially participants read a copy of the experimental protocol, which described the background to the research, the task, and the potential risks involved in their participation. They also had the opportunity to ask the experimenter any questions regarding the study before agreeing to participate and signing an informed consent form. They were then introduced to the task and had the opportunity to familiarize themselves with the stimulus configuration and the task and to practice all of the actions associated with manual tracking shown in table 1. A list of possible actions and their prerequisites was located on a copy holder beside the monitor at all times for easy reference. After the participants reported that they were comfortable with the interface, training was initiated.

Throughout the training and test sessions, participants were seated in an adjustable seat at the normal working distance from the computer screen (approximately 50 cm). The task was carried out in normal ambient illumination with the screen oriented to keep glare on the screen to a minimum. Each session usually lasted 1 to 1.5 h. Breaks were given between runs to reduce visual, mental, and postural fatigue.

The feedback window was always available in both the training and test sessions. A solid line traced out the last five positions (if available) of the target marker and a short line indicated the projected direction and speed of the target marker based on the last two updates. Except for the first two runs of the first training session, a random number of noise signals were added on each update.

The training sessions were designed to introduce the participants to all aspects of the task, to ensure that they were completely familiar with the interface, the task, and how the AT functioned, and then to give them practice with scenarios similar in difficulty to those in the test sessions.

The first training session was composed of five runs or scenarios of 10 min each. Participants were required to update a maximum of four targets every 30 s. Noise signals were introduced on the third scenario and targets whose signal strength fell below 20 on some updates appeared in scenario 5.

In the second session, the use of the AT was demonstrated and participants were given an opportunity to practice the operations associated with using it. They then repeated the last three scenarios from session 1 but with the option of assigning the targets to the AT. The size of the AT search area was 60. On the fourth run, the number of targets to be tracked increased to six, the update interval to 40 s, and the session time to 15 min.

On day 3, the participants repeated the last two scenarios of day 2. They were then given a scenario with eight targets, no non-targets and no AT in order to test their ability to use the interface. They then completed one more scenario with the AT in which they had to track the position of eight targets. The other parameters were the same as the six-target run. On the last day of training, participants completed four scenarios similar in form to the last scenario on day 3.

All the participants completed at least four training sessions. However, the training process was flexible in that if participants appeared to be having difficulty at any point during the training period they would be given extra practice at that level before moving on to the next level. When this happened the training period was extended to five sessions.

The training period was followed by five test sessions of three scenarios each. The order of the runs during the test sessions was randomized for each participant with the constraint that the participant never carried out the same scenario twice in a session and never repeated exactly the same condition in any one session. Thus, participants were exposed to each scenario and completed runs on three of the five conditions during each session. During each test run, participants were required to track a maximum of eight targets over a 20-min period. The signal table was updated every 40 s except for the second control condition which had an update rate of 50 s. Target signals were added two at a time until eight target signals were appearing in each update period. If a target signal exceeded the range of the tracking display, it was replaced by the next target in the scenario.

At the beginning of each session, participants were reminded that the performance window would be present at all times, that they were to detect and/or track the position of eight targets on each update to the best of their ability, and that they could use the AT as much or as little as they wished. They were not informed about how reliable or unreliable the AT would be on each run.

*3.1.6. Performance measures:* A large number of performance measures were collected. These included direct measures such as the actions (add, remove, associate, assign, etc.) during each update period and the response times, and indirect measures such as participant, AT, and system (participant plus AT) hit rates, false alarm rates, and error rates for the different types of errors (misses, lost targets, and misassociations). These measures were analysed overall and over time (across update periods).

### 3.2. Results

Most of the results deal with system performance, primarily because it is overall system performance rather than the performance of the human or the AT that interests us. Moreover, with some measures, it was difficult to determine when to attribute actions to the human and to the AT. For instance, if the human misassociated a signal to a target and then passed the target over to the AT which continued to misassociate signals from target A to target B, the error on subsequent updates was attributed to the AT. However, possibly it should have been attributed to the human.

*3.2.1. AT reliability:* By definition, AT reliability, or the percentage of targets assigned to the AT that it was able to track, increased as the size of the area that the AT searched increased. However, unless participants assigned every target to the AT on every update, these percentages were not available directly from the data. This never happened because of task difficulty and participant strategy. Thus to calculate AT reliability, the percentage of targets assigned to the AT that were correctly updated was plotted against targets assigned to the AT for each participant, scenario, and reliability level. Straight lines were fitted to these data points for each scenario and reliability level. These fits were used to predict the number of targets

that would have been tracked if eight targets had been assigned to the AT on each update period. The values calculated by this method were 53, 70 and 78% under the low, moderate and high reliability conditions, respectively.

In certain cases, using targets tracked correctly by the AT could lead to an underestimation of perceived if not actual reliability. If the signal from one target is associated with the marker from another and vice versa on a given update and the AT continues to track the two targets successfully, all subsequent updates will be labelled as errors. However, the participant may perceive the AT as operating reliably. Thus for the purpose of determining the perceived reliability of the AT, it may be preferable to use targets tracked rather than targets tracked correctly. Using that criterion, the percentages of targets the AT was capable of tracking were 55, 76, and 85%.

3.2.2. *AT use*: AT use was assessed by looking at the percentage of targets assigned to and tracked by the AT. The results averaged across all participants and scenarios are shown in figure 2. A repeated measures ANOVA indicated that the differences in AT use were significant across conditions ( $F(3,21) = 99.62, p < 0.01$ ). There was also a significant difference across scenario ( $F(2,14) = 5.01, p < 0.05$ ), and a significant interaction between condition and scenario ( $F(6,42) = 4.61, p < 0.05$ ).

As can be seen, the mean data are very close to the percentage of targets that the AT could have been expected to track. However, the picture changes when one looks at AT use relative to expected use for individual participants (figure 3). The first four participants (lower use group) assigned fewer targets to the AT than it could be expected to track. The remaining participants repeatedly assigned targets that the AT could not track consistently (higher use group). An analysis of variance indicated

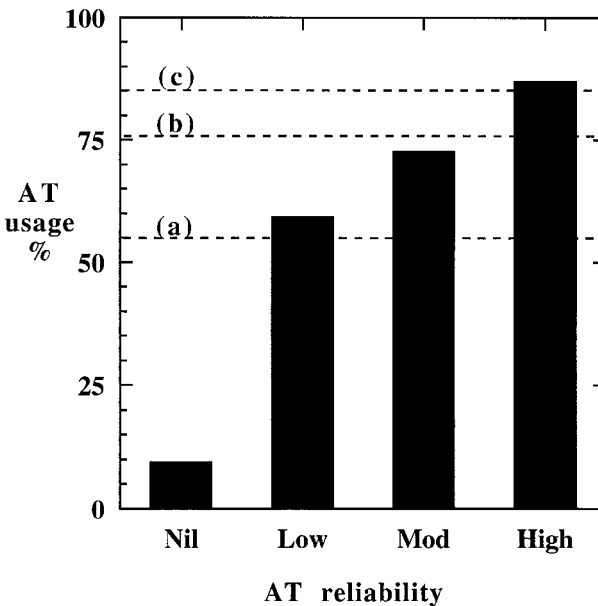


Figure 2. Percentage of targets assigned to and tracked by the AT under each level of AT reliability. The dashed lines show the percentage of targets that the AT would be expected to track in the (a) low, (b) moderate, and (c) high reliability conditions.

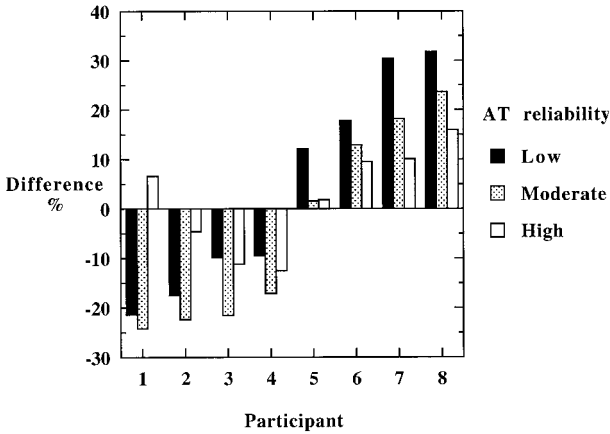


Figure 3. Difference between the percentage of targets assigned to and tracked by the AT and the percentage of targets the AT would be expected to track as a function of AT reliability and participant. The results for each condition have been subtracted from the percentage of targets the AT would have been expected to track in that condition (the dashed lines in figure 2).

that the use of the AT by these two groups was significantly different ( $F(1,6) = 65.73$ ,  $p < 0.01$ ). Given this significant difference in AT use, the remaining performance measures were analysed as a function of usage as well as condition.

The form of the difference in AT usage is seen when one looks at AT use for each group across trials (figure 4). The lower use group would start off by assigning most of the targets to the AT, but as the run progressed their use of the AT decreased. The slope of usage for the higher use group, on the other hand, was essentially flat varying only across reliability level.

**3.2.3. Percentage of targets tracked:** Averaging across all participants, the number of targets tracked successfully for the four different levels of reliability was 66% (for the base-line condition when the AT would not track any targets), 80%, 89%, and 92% (for low, moderate and high reliability, respectively). A repeated measures ANOVA supported the observed change in percentage of targets tracked across conditions ( $F(3,21) = 29.5$ ,  $p < 0.01$ ). There was also a difference in percentage of targets tracked across the scenarios ( $F(2,14) = 29.4$ ,  $p < 0.01$ ), but the interaction was not significant ( $F(6,42) = 1.76$ ,  $p < 0.2$ ). A pair-wise comparison (Tukey Studentized Range Test) of the different levels of AT reliability indicated a significant difference between the base-line condition and the remaining levels and between the low reliability condition and the moderate and high reliability levels ( $p < 0.05$ ).

The impact of AT reliability on percentage of targets tracked as a function of AT usage is shown in figure 5. As can be seen, performance was similar for the lower and higher use groups except for the baseline condition where the lower use group tracked 20% more targets on average. The pattern shown in figure 5 is supported by an analysis of variance, which showed no overall effect of usage on targets tracked but a significant interaction between usage and condition ( $F(3,18) = 9.58$ ,  $p < 0.01$ ).

**3.2.4. Response time:** The response time measure was the mean time in the update period at which the last action was taken. As shown in figure 6, this time decreased

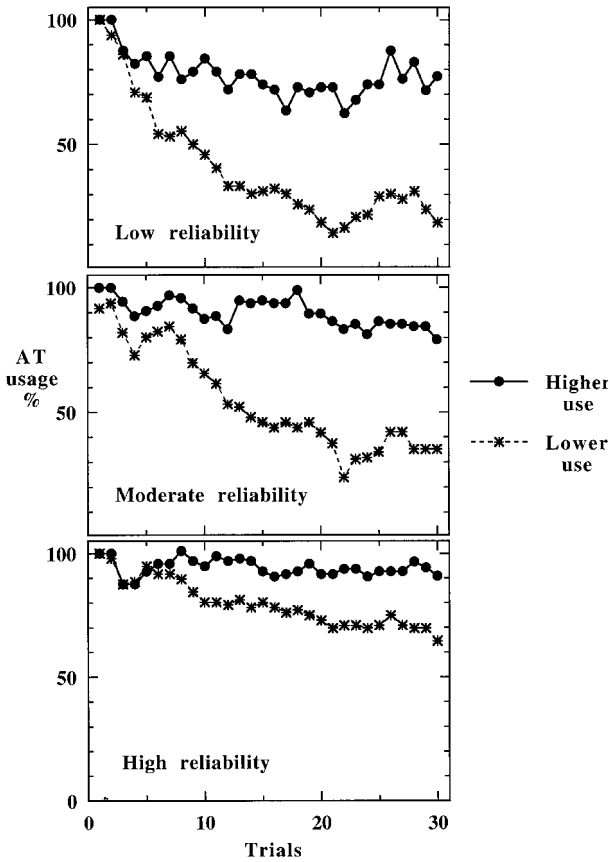


Figure 4. AT usage across trials averaged over participants using AT more than expected and participants using AT less than expected for three levels of AT reliability.

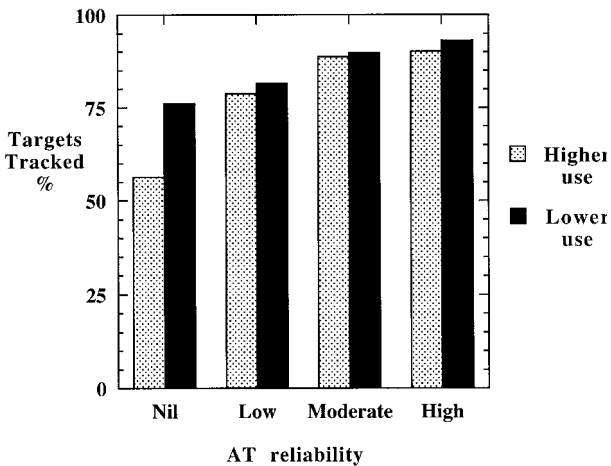


Figure 5. Percentage of targets tracked successfully as a function of AT reliability and use of AT. Includes targets tracked manually and by AT.

from a mean of 35 s in the base-line condition to less than 20 s in the high reliability condition. Response times, especially in the base-line condition, are constrained by the fact that participants had a maximum of 40 s to update all targets and might spend several seconds comparing the location of a marker with the location of the signals. Despite this constraint, the data were not highly skewed. A 2-way repeated measures ANOVA showed a significant decrease in response time across the four levels of reliability ( $F(3,21) = 121.79, p < 0.01$ ). There was also a difference in response time across scenarios ( $F(2,14) = 73.71, p < 0.01$ ) and a significant interaction ( $F(6,42) = 10.28, p < 0.01$ ). The interaction could be attributed to the fact that response times were similar across all scenarios under the base-line and low reliability condition. Under the remaining conditions, participants took less time in scenario 3 than in the other two scenarios. The higher use group took somewhat less time than the lower use group at the medium and high reliability conditions, but overall there was not a significant difference between the two groups  $F(1,6) = 1.31, p < 0.3$ ).

**3.2.5. Errors:** Errors can be classified into missed, lost, and misassociated targets and false alarms. As shown in figure 7, changes in error rates across the different levels of reliability were not similar for the four types of errors. Lost and missed targets decreased significantly as AT reliability increased ( $F(3,21) = 12.6, p < 0.01$  for lost targets and  $F(3,21) = 10.28, p < 0.01$  for missed targets). Number of missed targets also varied as a function of scenario ( $F(2,14) = 67.0, p < 0.01$ ) and there was a significant interaction between scenario and condition ( $F(6,42) = 3.44, p < 0.05$ ).

Overall, misassociations were not significantly different as a function of reliability ( $F < 0.3$ ). As shown in table 3, manual misassociations tended to decline with increasing reliability while AT misassociations increased especially in the higher use group. Scenario had a significant effect on the number of misassociations ( $F(2,14) = 9.81, p < 0.01$ ), with the number being considerably higher in scenario 2 as shown in the second part of table 3.

The last type of error shown in figure 7 is false alarms. Unlike the other types of errors, false alarms are a function of the number of noise signals on the display

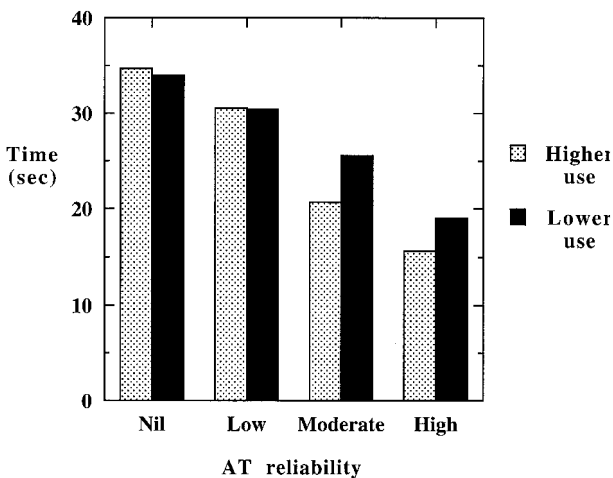


Figure 6. Time taken per trial as a function of AT reliability and use of AT.



rather than the number of targets. False alarms were significantly different both as a function of condition ( $F(3,21) = 3.85$ ,  $p < 0.05$ ) and scenario ( $F(2,14) = 43.29$ ,  $p < 0.01$ ). A pairwise comparison indicated that the difference in false alarms as a function of condition was due primarily to the small number in the moderate reliability condition relative to the base-line condition and the difference as a function of scenario was due to the relatively small number in scenario 3 ( $p < 0.05$ ).

Analyses of the effect of usage on the error measures indicated a significant difference for misses ( $F(1,6) = 8.58$ ,  $p < 0.05$ ) only. In addition, there was a significant interaction between usage and condition in the number of misses ( $F(3,18) = 9.43$ ,  $p < 0.01$ ) and in the number of lost targets ( $F(3,18) = 6.45$ ,  $p < 0.05$ ).

3.2.6. *Effect of length of update period:* The results for the 40 s update period indicated that participants handled approximately 5.3 targets during each update

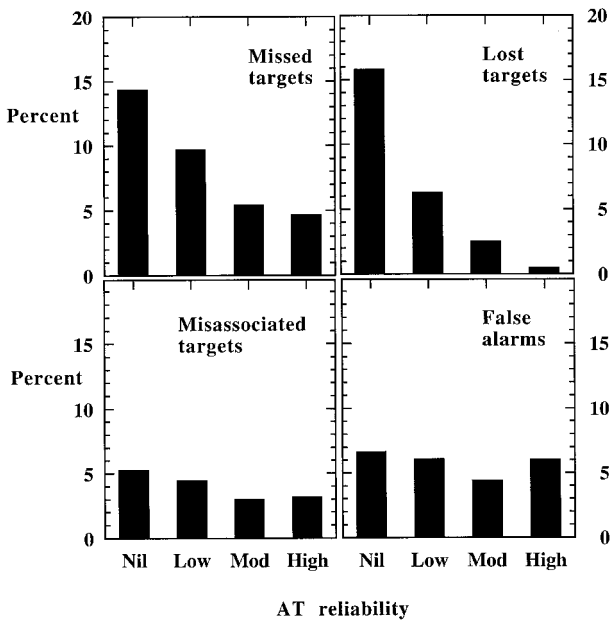


Figure 7. Percentage of missed, lost, and misassociated targets, and false alarms averaged over all participants and scenarios.

Table 3. Percentage of manual and AT misassociations in the high use and low use group as a function of AT reliability averaged across all scenarios and for scenario 2 only.

Scenario	Usage	Source	Nil	Low	Moderate	High
All	High	Manual	4.8	3.0	0.7	0.2
		AT	0.0	1.8	3.4	4.3
	Low	Manual	5.7	3.8	1.5	0.6
		AT	0.0	0.4	0.3	1.2
2	High	Manual	2.9	5.1	1.6	0.6
		AT	0.0	1.3	7.6	10.9
	Low	Manual	6.0	5.1	1.7	0.8
		AT	0.0	0.2	1.1	2.3

period. On that basis, it would be expected that they could handle at least one additional target on average in the extra 10 s. In fact, they only handled 0.3 extra targets in the extra time. The only significant effect of the additional time was an increase in the false alarm rate ( $F(1,7) = 13.47, p < 0.01$ ).

### 3.3. Discussion

As expected, use of the AT increased as AT reliability increased. However, only one participant (1 in figure 3) showed the expected pattern of AT use; namely, under-use of the lowest reliability AT and over-use of the highest reliability AT. The remaining participants either consistently under-used or consistently over-used the AT. However, even those that under-used the AT never abandoned it completely except in the control conditions. One possible explanation for this pattern of use of the AT can be found in some recent work on trust and reliability by Lee and Moray (1992). They found that people's use of an automated controller in a process control task was not only a function of their trust in the automation but also a function of their self-confidence in their ability to carry out the task manually. In the present task, it was almost impossible to consistently track all eight targets manually. Even if the AT tracked two or three targets for a few update periods at a time, it reduced the effort required of the participants and could improve their performance. Additional evidence for this hypothesis is that the participants who tended to under-use the AT did better in the control condition than the participants who over-used it. They could already handle the task reasonably well and presumably did not find as much of a need for the AT.

Another possible reason for the extensive use of the AT was that participants never lost overall control when they assigned targets to the AT. Rouse and Morris (1986) concluded that people were more likely to use an automated controller if they had some discretion in using it and if they were able to intervene when necessary. In the present experiment, the participants could monitor the AT's decisions and change them if they wished. If the AT lost a target, the participant had an opportunity to update that target without being penalized for the AT's failure. Participants also had complete control of which and how many targets they assigned to the AT.

Finally, AT usage, by the lower use group in particular, was possibly influenced by the methodology. At the beginning of a run, participants were not told how reliable the AT would be. They only knew that on some runs it could be very reliable and on others it was useless. Thus as shown in figure 4, participants tended to start out by assigning all of the targets that they found to the AT. The high use group continued to reassign targets while the low use group tended to use the AT less and less. This initial use of the AT under all levels of reliability tended to inflate average use, especially in the low and moderate reliability conditions.

The primary benefit of the AT was to reduce the number of lost and missed targets. With the AT handling some of the targets, the participant had time to update the remainder and to search for new targets. On the other hand, missed targets did not disappear completely and misassociations and false alarms remained relatively stable. The continued occurrence of misses and misassociations even with a highly reliable AT explains the insignificant improvement in targets tracked in going from the moderate to high reliability condition. The residual missed targets might be expected. If a new target had a relatively low strength, participants would be unable to clearly discriminate it from a non-target signal. However, when the AT was

working, they could have overcome this problem by adding all or most of the signals, assigning them to the AT, and then removing the ones the AT failed to track on the next update. The relatively low number of false alarms suggests that this strategy was not widely employed by any of the participants.

The stability in the percentage of misassociations was due primarily to a failure to monitor the AT adequately. As shown in table 3, AT misassociations tended to increase with reliability especially in the higher use group as well as in scenario 2. Misassociations usually occurred in clearly defined situations where two targets passed close to one another. This occurred relatively rarely and the best examples were in the second scenario. In such situations, either the human or the AT could become confused as to which signal went with which target. However, if the targets were misassociated, the effect could be detected by checking the performance monitoring system and often could be seen in a strange deviation in the history of the target's movements on the target display. One would expect misassociations to decline especially given the fact that the participants were receiving feedback. Moreover, in the moderate and high reliability conditions, they usually had time to check for errors. This finding supports previous observations that when an automated controller is perceived as being reliable, the human operator often fails to monitor it adequately and misses errors made by the controller (Rouse and Morris 1986).

3.3.1. *AT use*: Purely by chance, one-half of the participants tended to under-use the AT and one-half to over-use it. This division provided an opportunity to compare the effects of strategy on the various performance measures. In most cases, the differences between the two groups on most performance measures were not significant. This could have been due to the small number of participants in each group. However, an examination of the data suggests that for the most part the mean scores of the two groups were similar. The primary difference between the two groups was that the lower use group tracked more targets in the control condition and had fewer AT-initiated errors in the high reliability condition. The difference in targets tracked for the two groups is supported by the interaction between usage and condition for targets tracked, misses, and lost targets. This pattern of results would suggest that the AT can compensate for inability to carry out the task manually. However, the compensation had its price in this study; namely, these same participants tended to monitor AT performance less closely. This failure to monitor could be due to a poorer understanding of the dynamics of the situation. However, more systematic study of participants with different levels of expertise and with a wider range of scenarios would be required before a definitive statement could be made.

The similarity of the two groups on other performance measures may also have been a function of the task. While the task was difficult to carry out manually, it was relatively easy when the AT was reliable. Thus with the AT handling some of the targets, participants had sufficient time to either handle some targets manually or reassign targets to the AT that it could not track. To evaluate the cost associated with these different strategies, it would be necessary to look at performance with higher target loads.

3.3.2. *Performance across the different scenarios*: With some of the performance measures, there was a significant difference in performance as a function of scenario and in some cases a significant interaction between scenario and condition. The

impact of scenario on number of misassociations has already been discussed. With the other measures, the difference was usually between the third scenario and the other two, probably due to the fact that the signal strengths of the targets were more consistently above 20 in that scenario. The result was that, in scenario 3, the percentage of targets tracked was significantly higher while misses and false alarms were significantly lower. In addition, use of the AT was higher in that scenario. The higher signal strength was of greatest benefit in the baseline and low reliability conditions when participants were pressed for time. This would explain the interaction between scenario and condition that occurred with misses and hits. This pattern of results underlines the impact that the characteristics of the scenario can have on performance in complex tasks and the importance of using a range of scenarios and carefully testing them prior to the actual experiment.

3.3.3. *Increase in update period:* It might seem surprising that increasing the duration of the trial was not beneficial. However, participants presumably handled the easiest targets first — those with high signal strength. They then went on to the more difficult targets in the time that remained. The more difficult targets were the ones that could be confused easily with non-targets. Thus, signals added or associated in the extra time would have a greater probability of being noise signals. Support for this behaviour is indicated in the significant increase in false alarm rate with the longer update period. There is another reason why false alarm rate may have increased: the distance the targets moved was a function of their speed. Thus, targets moved further in the 50-s interval making it potentially more difficult to accurately track a target, especially if there were a lot of non-targets with similar signal strength.

#### 4. Experiment 2

The second experiment was designed to explore some of the issues raised by the first experiment; namely, the impact of task difficulty and perceived reliability on use of the AT. In the first experiment, it was found that participants tended to make some use of the AT even when it was relatively unreliable. This result was at odds with other findings which show that participants tend not to use an automated system if it is perceived as being unreliable (Lee and Moray 1992, Muir and Moray 1989). One possible explanation was task difficulty. It was not possible to track eight targets manually under the conditions imposed in experiment 1. Thus, any assistance the AT could provide was beneficial. In the second experiment task difficulty was varied by having one-half of the participants monitor six targets and the remainder eight targets.

A second possible reason that participants used a relatively unreliable AT was that they did not know how reliable it was on any given run. In other studies on reliability and trust, participants are usually given considerable practice with a reliable controller and then the automated system is degraded unexpectedly (Lee and Moray 1992, Muir and Moray 1989). Thus in the second study, reliability was varied in a more systematic and less obvious way. Participants obtained 2 days of experience with the high reliability AT used in the first experiment, followed by 2 days with a less reliable AT or vice versa. The size of the area that the AT searched was consistent across all 4 days and identical to the area used in the high reliability condition in experiment 1. Instead, the threshold for accepting a signal as representing the next position of a target was varied from low to moderate. If the size of the area the AT searches is decreased, the number of targets that the AT

tracks consistently tends to decrease. Participants can usually determine which targets are tracked consistently and which are not and respond accordingly. When threshold is varied, this pattern tends to be less consistent. It was thought that such inconsistency might lead to rejection of the AT. Thus, it was hypothesized that a decrease in AT reliability would lead to a reduction in use of the AT and that this reduction would be seen primarily amongst those participants that had to track six as opposed to eight targets.

#### 4.1. Method

The method was identical to that in experiment 1 with the exceptions noted below.

4.1.1. *Participants*: Twenty participants, 12 females and 8 males, between the ages of 19 and 38 years participated in the experiment. All reported normal or corrected-to-normal vision. Two were military participants and the remainder were recruited from either in-house personnel or the nearby university community.

4.1.2. *Conditions*: The experiment was a  $2 \times 2$  factorial design. The two factors were number of targets that had to be tracked and the order in which the two levels of AT reliability were presented. This resulted in four different conditions—reliable first with six targets, reliable first with eight targets, reliable second with six targets and reliable second with eight targets. The reliability of the AT was varied by changing the association thresholds for high- and low-strength signals. For the higher reliability signals the association threshold was set to 0.1 for signals with a strength greater than 10 and at 0.6 for lower strength signals. For the lower reliability AT, the association threshold was set at 0.6 and 0.9, respectively. In both cases the area was set at 60, which was the area used in the high reliability condition in the previous experiment.

Participants were assigned randomly to one of the four conditions (five participants per condition) and they completed a total of 12 test runs—six with one level of AT reliability followed by six with the other level of AT reliability.

4.1.3. *Procedure*: As in the first experiment, there were four training sessions. These differed from the previous experiment somewhat in order to give participants more experience with carrying out the task manually. On the first day, participants completed five runs of 10 min duration. On each run, they had to update a maximum of four targets during each 30-s period. Noise signals were introduced on the third run and low strength target signals on the fifth run. On the second day, participants completed four, 15-min runs. In the first three runs, they had to update a maximum of six targets and on the fourth, eight targets every 40 s. The third day was identical to the second except that they completed two runs with six targets followed by two runs with eight targets. Participants were introduced to the AT on the fourth and last day of training. They were briefed on how to use the AT, how the AT determined which signals to associate with which targets and they were informed that they could use it as much or as little as they wished. They then carried out the same four runs as on day 3, but they had the option of assigning targets to the AT. The AT was set up with the same parameters as in the high reliability condition.

Following training, each participant participated in four test sessions. In the first two, AT reliability remained constant and was a function of the condition the

participant was in. Sessions 3 and 4 were identical to the first two except that AT reliability was changed. Thus, participants in the reliable-first, six-target condition had to update six targets every 40 s and had the option of using the AT which was set up with an area of 60 and association probabilities of 0.1 and 0.6 in the first two sessions and 0.6 and 0.9 in the last two. Participants completed the same set of six scenarios under each level of AT reliability, but in a different random order. No feedback was provided during the test sessions.

#### 4.2. Results

The same performance measures were analysed as in the first experiment. Since somewhat different scenarios were used, the actual reliability of the AT was somewhat different. The values calculated were 89% when the association threshold was 0.01/0.06 and 60% when the association threshold was 0.06/0.09.

Table 4 shows how performance varied as a function of the number of targets the participants had to track, the association probability, and the order in which participants received the two levels of reliability. The results have been averaged across participants and scenarios. As can be seen, performance varied as a function of both AT reliability and number of targets. The order in which the participants carried out the different AT conditions had no effect on any of the measures of performance ( $F < 1$ ). Significantly more targets were tracked successfully when the participants had fewer targets to track ( $F(1,32) = 7.41, p < 0.05$ ), and when the AT was more reliable ( $F(1,32) = 14.57, p < 0.01$ ). Participants also took significantly less time to handle the targets when the AT was more reliable ( $F(1,32) = 101.38, p < 0.01$ ). The most frequent type of error was misses followed by lost targets. Misassociations were minimal; however, the scenarios had been designed to minimize the possibility of targets passing close to one another.

Although mean AT usage varied as much as 20% across the various conditions, there was not a significant effect of target number ( $F(1,32) = 1.73, p < 0.2$ ) or reliability ( $F(1,32) = 3.87, p < 0.06$ ) on AT usage. The reason for this can be seen in figure 8. AT use varied more across participants than across conditions. Two participants in the reliable first conditions decreased their use of the AT by 50% or greater when they were faced with the lower reliability AT and one participant's AT usage increased by 30% when it became more reliable. The remaining 17 participants, however, made relatively consistent use of the AT across both levels of reliability. Other than the exceptions noted above, the only difference across the

Table 4. AT use, targets tracked, misses, lost targets, and time taken as a function of number of targets tracked, order in which AT reliability varied, and reliability of AT.

Targets tracked	Reliable AT	Association threshold	AT use (%)	Targets tracked (%)	Misses (%)	Lost targets (%)	Time (s)
6	First	0.1	94.0	91.9	4.5	0.3	14.9
		0.6	79.7	87.5	6.8	3.1	24.7
	Second	0.1	80.0	96.2	2.0	0.5	16.0
		0.6	77.6	91.9	3.7	3.0	25.1
8	First	0.1	97.5	92.4	3.2	0.3	13.0
		0.6	83.2	82.1	7.9	6.5	28.4
	Second	0.1	95.2	91.2	3.6	0.3	11.3
		0.6	83.8	80.3	11.6	3.8	29.0

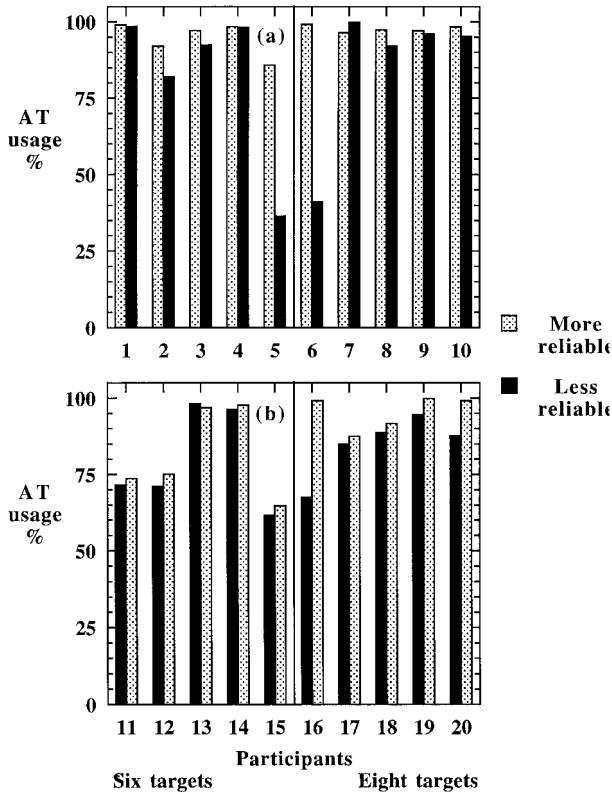


Figure 8. AT usage as a function of whether participants received more reliable AT first, reliability of AT, number of targets, and participant. Participants in (a) received the reliable AT first, and participants in (b) received the unreliable AT first.

four groups was that three of the five participants in the six-target, unreliable-first condition used the AT less than 75% of the time.

Use of the AT did tend to be correlated with targets tracked. There was a significant correlation between AT usage and AT targets tracked at both levels of reliability in the eight target conditions ( $R = 0.8$ ,  $p < 0.01$  with the higher reliability AT and  $R = 0.87$ ,  $p < 0.01$  with the lower reliability AT) and at the lower level of reliability with six targets ( $R = 0.68$ ,  $p < 0.05$ ). This result differs from experiment 1 where there was no significant difference in percentage of targets tracked between the lower and higher use groups. In fact, the lower use group tracked slightly more targets on average.

#### 4.3. Discussion

It had been predicted that use of the AT would drop if participants were exposed to a less reliable AT than they were accustomed to and that this drop would be a function of task difficulty. The data provided little support for this hypothesis. In most cases, a reduction in AT reliability did not lead to a reduction in frequency of use of the AT even when the participants had fewer targets to monitor. The mean differences shown in table 4 were primarily due to one participant in each of the reliable-first

conditions. Number of targets did tend to affect the use of the AT in the unreliable-first conditions; however, the effect was not consistent. Five of the ten participants used the AT as much as possible in the six-target conditions, independent of its reliability.

The only consistent pattern between those that relied on the AT and those that were more conservative was that the more conservative participants tracked fewer targets, especially in the eight-target conditions. However, it is difficult to tell whether they tracked fewer targets because they used the AT less, or because they were less proficient at the task and therefore did not have time to reassign all the targets.

If anything, participants tended to make more use of the AT in this experiment under the lower reliability conditions than they did in the previous experiment, with 83% of the targets being assigned to or tracked by the AT on average. This compares with 55% and 73% in the low and moderate reliability conditions, respectively, in the first experiment.

One possible reason for this difference is the method used to vary AT reliability. In the first experiment reliability was varied by changing the size of the area the AT would search for suitable signals. This method resulted in some targets being tracked consistently and others not at all. In the current experiment, the association threshold was changed. Using this method, the probability that a target would be tracked or not tracked on a given update was more idiosyncratic. The impact of the two methods is illustrated in figure 9. It shows the mean number of targets tracked per update as a function of the mean number of targets assigned to the AT per update. The closer a data point falls to the diagonal line the more effective the participant has been in assigning targets that the AT will track. As can be seen, this occurs more frequently in the low and moderate reliability conditions in the first experiment. It could be that participants in the first experiment were better at assigning targets that the AT would track consistently. However, it is more probable that the AT tended to track targets less consistently in the lower reliability conditions in the second experiment.

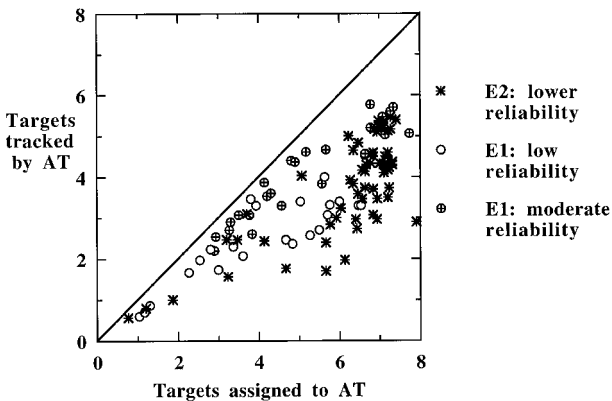


Figure 9. Number of targets tracked by the AT as a function of number of targets assigned to the AT per update for each scenario and participant. The data are from the low and moderate reliable AT in experiment 1 (E1) and the eight target condition with the less reliable AT in experiment 2 (E2).



## 5. General discussion

These studies were an initial attempt to understand how people will use an automated tracker as a function of its reliability. As stated in the introduction, studies on the use of automated controllers and decision aids frequently find that people do not use them unless they perceive them to be reliable. Even then, they often continue to rely at least partially on their own ability. On the other hand, if automated controllers are perceived as being reliable, operators often fail to monitor them adequately. In this study, all of the participants made at least some use of the automatic tracker under all levels of reliability. Even when faced with an unexpected change in AT reliability, as in experiment 2, most participants continued to use the AT to the same extent that they had originally.

The results of these experiments are consistent with the hypothesis that participants' use of an automated controller is a function of their perceived ability to do the task manually (Lee and Moray 1992). In experiment 1, participants who tracked the most targets when the AT was not available made the least use of the AT when it was available. In experiment 2, conservative use of the AT was found primarily in the group of participants who had only six targets to monitor and who had the less reliable AT initially.

The relatively high use of the AT by all participants in both experiments could also be partly attributed to the fact that the participants always retained ultimate control. If the AT failed to track a target or made an error, the participant could correct it. However, there is little direct evidence for this explanation beyond the fact that most participants did not tend to track some of the targets manually, even when they had the time. In the experiment by Morris *et al.* (1988), participants did tend to track some of the targets manually even when it was to their advantage to let the automated system handle them.

Evidence that participants often fail to adequately monitor a reliable automated controller was found primarily in experiment 1. Some participants failed to notice AT misassociations despite the availability of feedback. However, these errors were only missed by participants who made extensive use of the AT. This finding would suggest that participants should receive substantial training and continuing experience with handling the task manually. Alternatively, one might consider giving additional exposure to scenarios where these types of AT errors are likely to occur. Further research is required to systematically investigate these different approaches.

## 6. Conclusions

Two experiments were run to assess the use and usefulness of an automated tracker as a function of its reliability. Use of the AT increased as a function of its reliability; however, many participants in both experiments made extensive use of a relatively unreliable AT. Moreover, participant's use of the AT did not decrease when its reliability decreased unexpectedly. This finding differed from the results of previous experiments which have found that participants tend to under-use an automated controller unless it is extremely reliable and that they decrease their reliance on an automatic controller if they perceive that its reliability has dropped. The discrepant pattern of results in the present experiments were attributed to task difficulty and to the fact that participants could handle targets that the AT failed to track. Participants who made extensive use of the AT were less likely to correct AT-initiated errors, other than lost targets, even when feedback was provided. This

behaviour was consistent with previous observations that humans are not good at monitoring automated systems that they perceive to be reliable.

The results of the current experiments suggest that operators are likely to use an automated tracker even if it is only moderately reliable. However, these experiments employed a relatively limited range of conditions. Further research is required to substantiate these findings with larger numbers of targets and a wider range of scenarios. The results also suggest that some people are better at monitoring an automated controller than others. It would be useful to investigate whether training or experience could have an impact on an individual's ability to detect system errors such as false alarms and misassociations.

### References

- KIRLIK, A. 1993, Modeling strategic behaviour in human-automation interaction: why an 'aid' can (and should) go unused, *Human Factors*, **35**, 221–242.
- LEE, J. and MORAY, N. 1992, Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, **35**, 1243–1270.
- MORRIS, N. M., ROUSE, W. B. and WARD, S. L. 1988, Studies of dynamic task allocation in an aerial search environment, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-18**, 376–388.
- MUIR, B. M. and MORAY, N. 1989, Operators' trust in and use of automatic controllers, *Proceedings of the Annual Conference of the Human Factors Association of Canada*, Mississauga, Ontario: (Human Factors Association of Canada), 163–166.
- ROUSE, W. B. and MORRIS, N. M. 1986, Understanding and enhancing user acceptance of computer technology, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-16**, 965–973.
- WEISGERBER, S. A. and SAVAGE, S. F. 1990, Operator ship classification using an Automatic Target Recognition (ATR) system in conjunction with Forward-Looking Infrared (FLIR) imagery, NWC TP 7101, Naval Weapons Center, China Lake, CA.