



Accountability and automation bias[†]

LINDA J. SKITKA

*Department of Psychology (M/C 285), 1009 BSB, 1007 W. Harrison St.,
University of Illinois at Chicago, Chicago, IL 60607-7137, USA.
email: lskitka@uic.edu.*

KATHLEEN MOSIER

*San Francisco State University, Department of Psychology, San Francisco State University,
1600 Holloway Ave, San Francisco, CA 94132, USA*

MARK D. BURDICK

*San Jose State University Foundation/NASA Ames Research Center, AF: 262-4,
Morrett Field, USA*

(Received 9 July 1999 and accepted 25 July 1999)

Although generally introduced to guard against human error, automated devices can fundamentally change how people approach their work, which in turn can lead to new and different kinds of error. The present study explored the extent to which errors of omission (failures to respond to system irregularities or events because automated devices fail to detect or indicate them) and commission (when people follow an automated directive despite contradictory information from other more reliable sources of information because they either fail to check or discount that information) can be reduced under conditions of social accountability. Results indicated that making participants accountable for either their overall performance or their decision accuracy led to lower rates of “automation bias”. Errors of omission proved to be the result of cognitive vigilance decrements, whereas errors of commission proved to be the result of a combination of a failure to take into account information and a belief in the superior judgement of automated aids.

© 2000 Academic Press

KEYWORDS: automation; accountability; vigilance; decision making; bias.

1. Accountability and automation bias

Increasingly sophisticated technical advances have changed the decision-making environment in many workplaces. Although introduced with the explicit goal of reducing human error, automated systems sometimes have the paradoxical effect of simply

[†] This work was supported by a grant from NASA Ames Research Center #NCC 2-837. An earlier version of this manuscript was presented at the American Psychological Society meetings in San Francisco, July 1996.

changing the kinds of errors operators are likely to make. A series of recent studies have documented two classes of errors that can emerge in highly automated decision-making environments: (1) *omission errors*, defined as failures to respond to system irregularities or events because automated devices fail to detect or indicate them; and (2) *commission errors*, that occur when people incorrectly follow an automated directive or recommendation, because they do not verify it against other available information, or in spite of contra-indications from other sources of information of which they are aware (e.g. Mosier, Skitka & Korte, 1994; Mosier, Skitka, Heers & Burdick, 1998; Skitka, Mosier & Burdick, 1999; see also Layton, Smith & McCoy, 1994; Molloy & Parasuraman, 1996; Parasuraman & Riley, 1997 for related research).

Research investigating how people use automated decision aids has led to the conclusion that the presence of automated decision aids can at a minimum lead to misuse or abuse of these aids (e.g. Parasuraman & Riley, 1997), and at worst, can lead to biased decision-making processes (Mosier *et al.*, 1998; Skitka *et al.*, 1999). The goal of the present study was to explore the extent to which social accountability could ameliorate decision-making problems observed in highly automated work environments and to shed further light on the psychological underpinnings of automation bias.

2. Automation bias

Experimental evidence of automation bias leading to commission errors was provided by a full-mission simulation in the NASA Ames Advanced Concepts Flight Simulator (Mosier, Palmer & Degani, 1992). During one phase of flight, professional airline crews received contradictory fire indications. An auto-sensed checklist suggested that the crew shut down the #1 engine, but traditional engine parameters indicated that the #1 engine quickly recovered and that the #2 engine was actually more severely damaged and was beginning to fail. Turning off the #1 engine therefore would leave the crew without a fully functional engine. Regardless, 75% of the crews in the auto-sensing condition shut down the #1 engine. Only 25% with a traditional paper checklist and no automated aid did likewise. Analysis of the crews' audiotapes also indicated that crews in the automated condition tended to discuss much less information before coming to a decision to shut down the engine, suggesting that automated cues can short-circuit full information search.

Skitka *et al.* (1999) also found evidence of omission and commission errors in automated monitoring tasks with non-pilot samples. In their study of students' performance on a low-fidelity part task that required gauge and progress monitoring in addition to a tracking task, evidence of both commission and omission errors emerged. Participants in the automated decision aid condition learned that the automated aid was highly, but not perfectly, reliable but that other system indices (gauges, etc.) were always 100% valid and reliable. On 6 out of 100 events, the automated aid failed to announce that a system event occurred that required the participants to respond (an omission error opportunity). Comparison of performance on these same six events across automated and non-automated decision aid conditions revealed that participants in the automated condition missed on average 41% of these events, whereas those in the non-automated condition on average missed none of them ($M = 0.18$).

Similar evidence of bias emerged with respect to commission errors. Given that non-automated system indices were 100% valid and accurate, following an automated directive when it contradicted non-automated system indices represented evidence of a bias in decision making. Participants in the automated condition received 6 automated directives that were contra-indicated by other system indices. Analysis revealed that on average, people had a commission error rate of 65%. Over 20% of the sample made commission errors on all six events; less than 1% made no commission errors.

There are also a variety of anecdotal examples that suggest that people may use automated decision aids in biased ways, and that automation bias can have serious consequences. For example, the crash of an Eastern Airlines aircraft in 1972 may have been the result of an omission error. During the time immediately prior to crashing, the crew was preoccupied trying to determine why the landing gear indicator did not light when the gears were dropped. The autopilot was set to hold an altitude of 2000 ft, but was accidentally disengaged by a nudge to the control column. The crew did not know the autopilot had disengaged until prompted by Air Traffic Control to check their altitude: by that time they had descended to 30 ft above ground level and it was much too late to make a correction (Billings, 1996). The crew in this example had allocated responsibility for maintaining altitude to the automation, and subsequently neglected to check whether it was operating correctly until prompted by ATC.

Similarly, the 1983 Korean Airlines plane that was shot down by Soviet fighters' problems can also be traced in part to a lack of crew vigilance in monitoring of automated systems. In this case, the crew selected a magnetic heading and followed it throughout the flight rather than coupling the navigational system's inertial reference system to the autopilot. The flight did not follow its originally planned flight path, but rather maintained the selected magnetic heading until it was shot down. The crew was relying entirely on automation that had been inappropriately set up and never checked their progress manually, allowing the flight to stray well into Soviet airspace ("Analysis of Flight Data", 1993). A similar failure to monitor a satellite-based navigational system caused the Panamanian cruise ship *Royal Majesty* to run aground in 1995, causing \$7 million in damages and lost revenue (Azar, 1998). In sum, automation bias appears to be both a real and potentially serious problem in highly automated decision-making contexts.

3. Accountability

Accountability demands sensitize decision-makers to the need to construct compelling justifications for their choices. Considerable research has found that increasing social accountability can successfully ameliorate a broad array of cognitive biases, including primacy effects (Tetlock, 1983), the fundamental attribution error (Tetlock, 1985a), overconfidence effects (Tetlock & Kim, 1987), as well as the "sunk cost" effect (Simonson & Nye, 1992). Increasing accountability for decisions before the decision process begins creates pressure for decision-makers to become more cognitively complex—i.e. to employ more multi-dimensional, self-critical and complex information processing strategies and to put more effort into identifying appropriate responses than when accountability demands are not present (Tetlock, 1985b).

Accountability creates pressure to attend to more information and to also integrate this information in more sophisticated ways. Specifically, accountable as compared to non-accountable participants use more complex rules in choosing among response options (McAllister, Mitchell & Beach, 1979; Tetlock, 1983); show increased self-awareness of the determinants of their judgments (Cvetkovich, 1978; Hagafors & Brehmer, 1983); process persuasive messages in detail rather than rely on surface appraisals (Chaiken, 1980) and are more discriminating and responsive to evidence when evaluating others (Rozelle & Baxter, 1981; Tetlock, 1985b).

Accountability, however, does not always lead to more functional decision-making outcomes. For example, Tetlock and Boettger (1994) found that accountability, although related to higher levels of cognitive complexity, was also related to postponing and buck-passing on decisions that could have saved lives. Other research has found that making people accountable for decision-making processes improved judgment, but making them accountable for decision-making outcomes did not (Doney & Armstrong, 1996; Siegel-Jacobs & Yates, 1996).

Depending on the circumstances, then, accountability can promote thoughts devoted to self-*justification*, thereby cutting off situation awareness or a broader consideration of decision alternatives. Under other circumstances, accountability pressures lead people to be more self-*critical* and to take into account a broader range of alternatives. One kind of thinking is likely to lead to rehearsal of the same decision outcome and why it is right; the other kind of thinking will lead to more integratively complex consideration of a wider range of variables and options. To the extent that automation bias is the result of cognitive appraisal and processing problems, rather than a motivation to impression manage or a belief in the relative authority of automated devices, it is likely to be a member of the class of biases amenable to amelioration by social accountability.

To date, accountability manipulations have been used exclusively in contexts where some kind of social judgment was the criterion of interest. The present study extends what we know about accountability effects by investigating whether accountability has an impact on human *performance*, not just human *judgment*. Human performance is defined differently depending upon the goals of a given performance setting. To what extent does accountability affect all parts of task performance (e.g. speed, accuracy), vs. only specific types of task performance? Moreover, what are the trade-offs involved in making people accountable for various aspects of their performance? For example, does making people accountable for greater integration of information lead to decrements in terms of response time to important events or off-loading of other tasks? Can making people generally accountable for their performance ameliorate the tendency toward automation bias, or is it necessary to make people accountable on a specific task-by-task basis?

The present study was designed to compare different kinds of accountability instructions (general vs. specific) on people's performance in a multi-task work environment. Of specific interest were the following questions: (a) can omission and commission errors be reduced by social accountability manipulations, (b) how effective are general vs. specific accountability instructions in reducing automation bias effects; (c) what are the trade-offs in making people accountable for different aspects of their performance and finally (d) what forms the psychological foundation of omission and commission errors—a

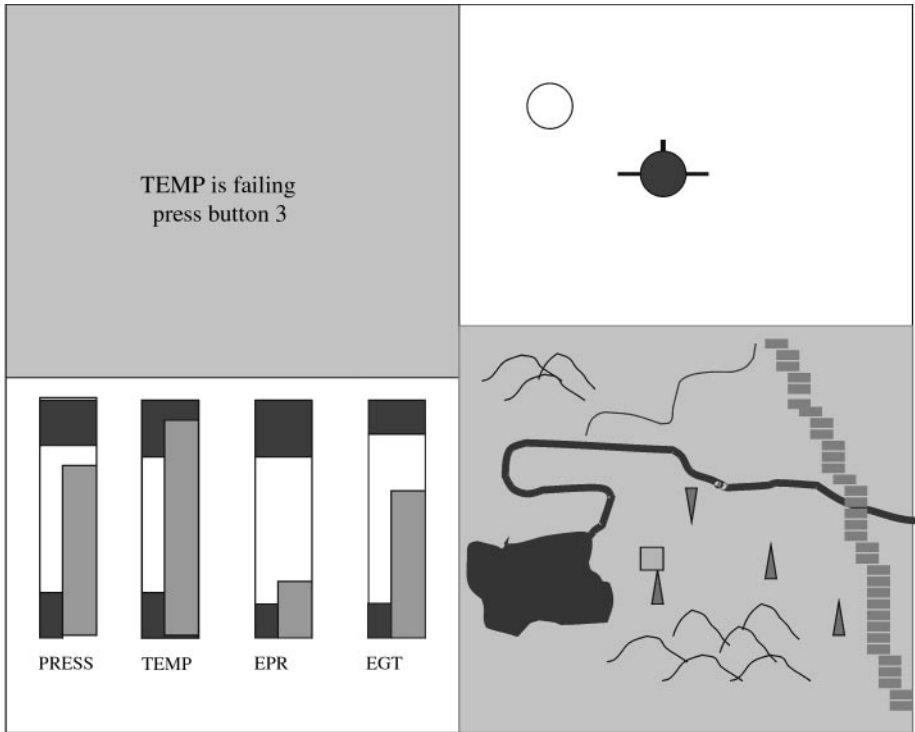


FIGURE 1. W/PANES task display.

belief in the relative authority/superiority of computerized decision aids, cognitive laziness or some combination of both?

4. Method

4.1. PARTICIPANTS

181 undergraduates participated in partial fulfillment of course requirements.

4.2. THE EXPERIMENTAL TASK

Participants' primary task was to complete five computerized "flight" trials using Workload/PerformANce Simulation software (W/PANES, NASA Ames Research Center, 1989). This program presents participants with a set of tasks designed to simulate the types of monitoring and tracking responsibilities involved in flying a commercial aircraft. It has been used successfully to create an environment in which errors of omission and commission can and do emerge (e.g. Skitka *et al.*, 1998) and at rates that parallel those of professional pilots in NASA Ames mini-advanced concept flight simulator (Mosier *et al.*, 1998).

W/PANES presented participants with a task display with four quadrants on a 14" color monitor. One quadrant was devoted to a tracking task, a second to a gauge

monitoring task and a third to monitoring progress across a top-down view of a map (see Figure 1; note that although the figure is black and white, participants' screen was in color). The fourth quadrant was reserved for messages and directives from an automated monitoring aid (AMA).

1. *The tracking task.* Participants used a two-axis joystick to keep an ownship symbol (the circle with a line through it, represented in the top-right quadrant of Figure 1) aligned with a moving circular target. The target circle moved as a function of the disturbance imposed by a sum of sine's algorithm. Therefore, participants' goal was to keep the target circle centered around the ownship symbol by following the motion of the target circle with the joystick, compensating for movements away from the center in heading (horizontal) and altitude (vertical).
2. *Waypoints.* In addition to having to maintain tracking performance, participants were also required to monitor their location on a map (presented in the lower-right quadrant of Figure 1). A pink square representing the ownship traveled from left to right across the map. Red and blue triangles on the map represented "waypoints". Each time their ownship symbol passed one of these waypoints, participants were trained to push a button on a response box to "simulate radioing their location to a communications tower". More specifically, when the bottom-right corner of the pink square touched the top of the triangle, the participant either pushed the Beta frequency button (if the triangle was blue) or the Alpha frequency button (if the triangle was red). Response buttons were labeled with red or blue dots to facilitate correct responses.

There were four waypoint events on each trial. The layout and location of the waypoints varied across trials.

3. *Gauge events.* The lower-left quadrant displayed four analog slide-rule gauges that were used for the gauge-monitoring task. Each gauge had upper and lower red zones. When a gauge indicator went into a red zone (indicators moved continuously, but stayed within normal ranges the vast proportion of the time), participants were trained to press the gauge's corresponding button on their response box. If three gauges went into a red zone at the same time, participants were trained to push a button to reset the gauges. Gauges were programmed to return to a "safe" value after 10 s in the red zone if no button was pressed (i.e. if the participant missed the event). The order of gauge events and when they occurred were randomized across the five trials.
4. *Automated monitoring aid (AMA).* The top-left quadrant was reserved for AMA announcements. The AMA notified participants of events (e.g. a gauge moving into the red zone), and recommended courses of action (e.g. push button 3). The AMA was described as highly but not perfectly reliable, whereas all other system indices were described as 100% reliable and accurate.
5. *Automation verification.* Participants were provided with a secondary display, presented on a separate 14" monitor that they could use to verify AMA assessments and directives. The second screen presented participants with a matrix of cells that opened to reveal information with a mouse-click (the Mouselab Decision Laboratory software, Johnson, Payne, Schkade & Bettman, 1991). To verify a directive, participants mouse-clicked open one to three cells to find the verification

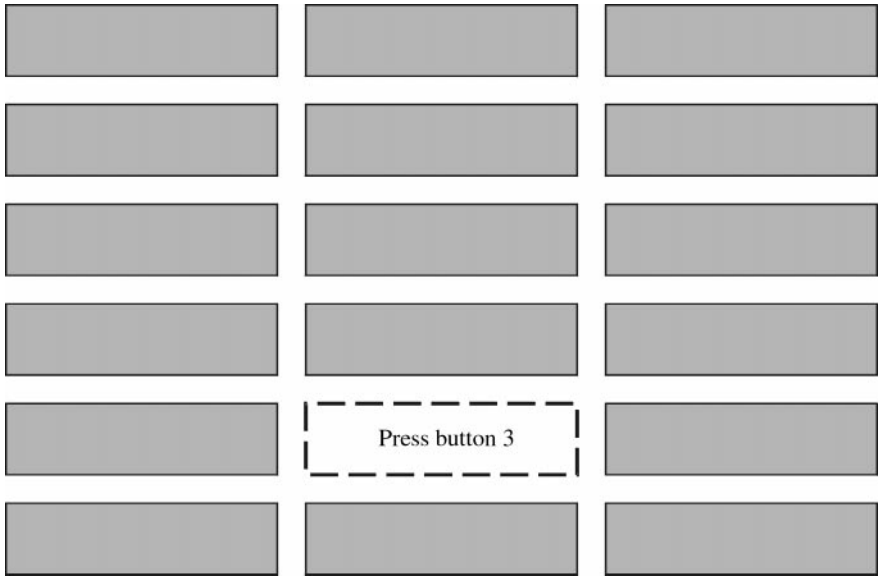


FIGURE 2. Auxiliary verification display.

information (see Figure 2). Verification level was therefore a within-subjects manipulation of how difficult and time-consuming it was to verify any given directive.

6. *Accountability.* The critical manipulation of the present study was whether participants were made accountable for their performance, and the specific nature of what they were made accountable for. The control condition was no accountability, and there were four accountability conditions. Non-accountable participants were led to believe that their performance was not being recorded, and were told nothing of a post-experimental interview.† All accountable participants, in contrast, believed that performance data was being collected, and that their task was to maximize either their overall performance (tracking, accuracy in response to events and response time to events), or their performance on a specific task—accuracy in response to gauge and waypoint events, their response time to events or their tracking. All accountability conditions included the instruction that participants would also have to explain and justify their strategies in meeting these performance goals following the session.

† Pilot testing indicated that accountability can be a challenging variable to manipulate. Making people feel accountable in experimental settings turned out to be very easy; participants naturally think their behavior is being evaluated and potentially judged and they reported high levels of accountability. These combination of steps were taken after pilot testing indicated that they were persuasive in eliminating evaluation concerns without inviting people to “goof off”. As seen in Table 1, the non-accountability condition did not lead people to make more errors than some of the accountable conditions, supporting the contention that our non-accountable participants were still engaged in the task.

4.3. DEPENDENT MEASURES

1. *Omission errors.* Participants received prompts from the AMA about system events on all but 4 out of the 50 events. The AMA failed to warn participants of a gauge reset situation twice and similarly failed to notify participants of two waypoint events. A failure to respond to these events was coded as an omission error.
2. *Commission errors.* The AMA gave an erroneous warning about an “impending” three-gauge event on six occasions, and instructed participants to push the reset button. On these events, two gauges were in the red zone, with a third threatening to move in that direction but not there. Ten seconds after the AMA directive, all three gauges moved completely into the safe zone.

If participants followed the automated directive they would be making an error because the three gauges were not in fact in the red zone. However, if participants verified the AMA directive on the Mouselab screen, they would be told the proper button press for a two-gauge situation (a situation for which they had not been given explicit training about a response). Recall that participants were trained that the gauge and waypoint indicators and verification information would always be 100% reliable, but that the AMA was not 100% reliable or valid. Following AMA information that contradicted what the gauges represented, or that had not been confirmed by the verification procedure, would therefore represent a biased response. A response was scored as a commission error only if participants followed the AMA directive; doing nothing or giving the response recommended by the verification procedure were not coded as commission errors.

3. *Verification behavior.* Whether participants verified AMA directives (every AMA directive could be verified with the Mouselab grid) was used as a measure of complete information assessment or search. The difficulty and time it took to obtain verification information varied from event to event in the form of how many cells had to be opened to retrieve it, from one to three cells. Directives were given for each event except the four omission error opportunities, but even omission error events could be verified. Verification could therefore range from 0 to 50 across all trials.
4. *Response time to gauge and waypoint events.* Response times to gauge and waypoint events were measured to millisecond accuracy.
5. *Tracking performance.* Tracking performance was measured as the weighted average of the linear deviations between the tracking circle and ownship symbol (for altitude error) and degrees (for heading error). Low scores on this measure therefore represented better performance, and remaining closer to the target, than high scores.
6. *Perception of the experiment.* Post-experimental questionnaires included checks for whether manipulations created the expected social accountability effects, in addition to items that tapped perceived task difficulty, confidence, the extent to which computers make better judgments than people do and belief in the reliability of the automated decision aid. Questions were phrased as statements to which participants indicated their relative degree of agreement on 1 (strongly disagree) to 7 (strongly agree) scales (see Table 4 for items besides manipulation checks).

4.4. PROCEDURE

Participants were run individually in a cubicle with two computer monitors, a computer mouse and a response box with buttons and a joystick. Each participant was trained on the display and W/PANES tasks. After being given complete verbal instructions and addressing any questions, participants were given three practice trials of 5 min each. Participants were allowed to repeat training trials until they achieved comfortable mastery.

Participants in the accountability conditions were informed of their major performance objective at the end of their training by an experimenter dressed professionally and in a white lab coat; were told that they would be interviewed at the end of the experiment by the experimenter and another expert; that the purpose of the interview would be for them to explain and justify their strategies for maximizing their performance during the experiment and that their interview would be taped for later in depth evaluation (none of the participants were in fact interviewed).

Participants in the non-accountable condition had a “student” experimenter (one of the same two experimenters who ran the accountability conditions, but dressed in jeans and T-shirt), who told participants that the data recorder on the computer was not working (a disk drive was left hanging by wires out of the computer to facilitate this story, that the experimenter was always working on when participants showed up at the lab). Participants were informed that they would nonetheless be run through the session so that they could get the experimental credit, and because their questionnaire data tapping how they felt about the task would be useful even without their performance data. Accountable participants’ instructions also emphasized anonymity, and no mention was made of a post-experimental interview.

After completing their training and receiving the accountability manipulation, participants completed five experimental trials, with a 5 min break after the third trial. There were 50 events across the five trials, i.e. gauge events and waypoint checks that required responses from the participant. At the end of the experimental trials, participants completed a short questionnaire, were completely debriefed and allowed to leave.

5. Results

5.1. MANIPULATION CHECKS

A variety of measures assessed whether participants perceived the experiment consistently with the experimental manipulations. Two questions tapped the extent to which participants believed that they were accountable for their performance—whether participants felt their individual performance was being observed, and separately, whether their individual performance was completely anonymous. Both measures varied as a function of accountability condition, $F(4, 176) = 14.38, p < 0.001$, and $F(4, 176) = 13.24, p < 0.001$. Further analysis indicated that participants who were accountable for their overall performance, accuracy, tracking or quick response did not differ in their perception that their performance was being observed, $F(3, 176) = 0.14, n.s.$, or was anonymous $F(3, 176) = 0.49, n.s.$ However, all accountable participants agreed more that their performance was being observed ($M = 5.93$) than participants in the non-accountable condition ($M = 3.59$), $F(1, 179) = 58.03, p < 0.001$. Similarly, participants in

TABLE 1
Average number of commission (range: 0–6) and omission errors (range: 0–4) as a function of accountability condition

	Commission errors	Omission errors
Accountable for:		
Overall performance	1.24a (1.23)	1.15a (1.21)
Quick response	2.03b (2.03)	1.44a (1.46)
Accuracy	1.32a (1.32)	1.08a (1.28)
Tracking	3.14c (2.51)	2.08b (1.38)
Non-accountable	1.92b (2.06)	1.95b (1.30)

Note: Means with non-common subscripts within columns were significantly different at $p < 0.05$, numbers in parentheses are standard deviations.

the non-accountable condition indicated much higher agreement that their performance was anonymous ($M = 5.05$), than participants in any of the accountability conditions ($M = 2.45$), $F(1, 179) = 52.09$, $p < 0.001$. In sum, manipulation checks supported that the accountability manipulations created the intended perceptual set on the part of participants.

In addition to checking the viability of the accountability manipulation, recall data indicated that 98% of the participants correctly recalled that the AMA was not always 100% reliable, but that the gauges, waypoints and verification indices were. Analysis revealed that the results were consistent regardless of whether participants correctly recalled these details or not, and so all participants were retained for analysis.

5.2. DOES ACCOUNTABILITY SUCCESSFULLY AMELIORATE AUTOMATION BIAS?

Of particular interest was whether social accountability led to a reduction in commission and omission errors in highly automated decision-making contexts. Results indicated that when appropriately channeled, accountability did indeed lead to lower rates of both commission and omission errors. Participants in the accountable for overall performance and response accuracy conditions showed lower rates of both commission and omission errors than participants in the other conditions (accountable for quick response, tracking or non-accountable). These results are described in greater detail below.

1. *Commission errors.* Analysis revealed a main effect for accountability on the tendency to make commission errors, $F(4, 176) = 4.45$, $p < 0.01$ (see Table 1 for details). Tukey HSD tests indicated that participants in the accountable for tracking condition made the most commission errors, followed by participants in the

TABLE 2
Average number of times participants verified automated directives (range: 0–50)

	Verifications
Accountable for:	
Overall performance	40.15a (14.38)
Quick response	31.61 (18.72)
Accuracy	39.14a (16.05)
Tracking	26.17b (19.83)
Non-accountable	34.41 (17.53)

Note: Means with non-common subscripts were significantly different at $p < 0.05$, numbers in parentheses are standard deviations.

non-accountable and accountable for quick response conditions. Participants in the accountable for overall performance and accuracy conditions did not differ in number of commission errors, and made significantly fewer commission errors than participants in the other conditions.

In sum, these results indicated that accountability decreased the tendency to make commission errors under specific conditions, i.e. when participants were accountable for their overall performance or accuracy. Participants accountable for their tracking performance (a non-monitoring task) were especially likely to make commission errors.

2. *Omission errors.* Omission errors also varied significantly as a function of accountability condition, $F(4, 176) = 4.36$, $p < 0.01$ (see Table 1 for details). Tukey HSD comparisons indicated participants in the non-accountable and tracking conditions made more omission errors than participants in the accountable for overall performance, accuracy or quick response conditions. No other differences were significant. Therefore, these results indicated that accountability for overall performance and accuracy led to a lower omission error rate than did accountability for responding quickly, tracking or non-accountable instructions, a result that paralleled the effects of accountability on commission errors.
3. *Verification behavior.* If accountability affects people's performance at least in the overall performance and accuracy conditions by encouraging people to be more cognitively vigilant, we should see an increase in verification behavior in these conditions as well. To investigate this hypothesis, we examined whether participants were more likely to verify automated cues in the overall performance and accuracy conditions, relative to the other accountability and non-accountability conditions.

Analysis revealed a significant effect for accountability condition on the extent to which participants verified AMA directives, $F(4, 176) = 4.12$, $p < 0.01$. As seen in

TABLE 3
Weighted average of linear distance off target in the tracking task as a function of accountability condition

	Average time off target
Accountable for:	
Overall performance	76.24a
Quick response	84.95a
Accuracy	81.68a
Tracking	56.31b
Non-accountable	90.88a

Note: Means with non-common subscripts were significantly different at $p < 0.05$.

Table 2, participants in the accountable for overall performance and accuracy conditions were indeed more likely to verify directives than participants in the other conditions, $F(1, 176) = 4.03$, $p < 0.05$. Tukey comparisons also indicated that verification behavior was the lowest in the tracking condition. Taken together, these results indicated that being accountable for accuracy and overall performance did lead to higher levels of vigilance, and that being accountable for tracking led participants to shed verification behavior. No significant effects for level of verification (how many cells had to be opened) emerged.

5.3. TRACKING PERFORMANCE AND RESPONSE TIME

Although appropriately channeled accountability can have the effect of reducing commission and omission errors, is this at the expense of performance in other areas? Specifically, does being more vigilant and verifying automated cues to reduce omission and commission errors on monitoring tasks necessitate decreases in tracking performance and increases in response time? The answer proved to be no; accountability did not lead to performance trade-offs on non-monitoring tasks or lead to slower response times to events. Being made accountable for tracking improved tracking performance, but being non-accountable or accountable for other tasks did not lead to a decrement in tracking performance. There were no significant differences in response times to events across conditions. These analyses are described in more detail below.

Analysis of tracking performance as a function of accountability condition revealed a significant effect, $F(4, 172) = 2.67$, $p < 0.05$ (see Table 3 for details). Tukey HSD tests indicated that participants in the accountable for tracking performance condition performed the tracking task better (were closer over time to the target) than participants in other conditions. No other differences were significant, although the means suggested that rather than leading to poorer tracking performance, participants in the non-tracking accountability conditions if anything showed better tracking performance than their non-accountable peers (see Table 3).

TABLE 4
Correlations with omission and commission error rates and how much participants completely verified cues

	Commission errors	Omission errors	Complete verifications
I was comfortable with the experimental task	- 0.16*	- 0.21*	0.15*
I found the tasks involved in the study to be difficult	0.12	0.07	- 0.11
I felt nervous about my performance	0.07	0.12	- 0.01
Tracking was more important than the other tasks	0.21*	0.36**	- 0.09
Responding quickly was more important than the other tasks	0.26**	0.12	- 0.26**
I am confident about the quality of my performance	- 0.23**	- 0.14	0.28**
I am comfortable justifying my strategies for maximizing my performance in this study	- 0.19**	- 0.20**	0.20**
Computers make better judgments than people on many tasks	0.02	0.10	0.04
Computers take into account more information than people	0.15*	- 0.02	- 0.03
The automated directive always notified you on what you were supposed to do	- 0.06	0.03	0.02

Note: * $p < 0.05$; ** $p < 0.01$.

In addition to measuring tracking performance, response time to events was measured to millisecond accuracy. No significant differences emerged across conditions in response time to gauge events, $F(4, 172) = 1.08$, n.s., or waypoint events, $F(4, 172) = 1.56$, n.s.

In combination, then, being accountable for overall performance and accuracy did not result in performance decrements in either tracking or response time to events. Being made accountable for tracking improved tracking performance. Being accountable for responding quickly to events did not lead to better response times than those not accountable for responding quickly, but neither did being accountable for other tasks yield performance decrements in response times to events.

5.4. POSSIBLE PSYCHOLOGICAL MECHANISMS UNDERLYING AUTOMATION BIAS

The fact that social accountability led to decreases in omission and commission errors, and increased verification behavior, strongly supports the idea that automation bias is the result of using automation as a heuristic replacement for more vigilant and complete information search and processing. To further understand automation bias, we analysed

participants' subjective assessments of the experiment as they related to the tendency to make omission or commission errors.

As seen in Table 4, higher omission and commission error rates were associated with less comfort with the experimental task; participants' belief that tracking was more important than other tasks; and less comfort with justifying strategies for performance. In addition to these shared patterns of findings, increased rates of commission errors (but not omission errors) were also associated with the belief that it was more important to respond quickly, less confidence about the quality of one's performance and of particular interest, the belief that computers take into account more information than people.

Correlations with verification behavior paralleled these results. Participants were more likely to completely verify if they were comfortable with the experimental task, confident about the quality of their performance and felt comfortable justifying their strategies for maximizing their performance. They were less likely to completely verify if they felt it was more important to respond quickly than accurately. In all cases, error and verification rates were uncorrelated with perceived reliability of the automated system, task difficulty, nervousness about performance or a belief that computers make better judgments than people do.

In sum, there was some support for the notion that commission, but not omission errors are associated with a belief in the superiority of computerized judgment over human judgment. However, there was no reliable association between a belief in the relative authority of computers and shedding of verification behavior. Taken together with the other results of the study, the vast proportion of evidence seems to be pointing to the fact that automation is used as a decision-making short cut that prematurely shuts down situation assessment more than a fundamental belief in the relative authority of automated aids.

6. Discussion

Taken together, the results of the present study indicate that people do not always use cues in rational ways. Although participants are made explicitly aware that their gauges always provided 100% reliable and accurate information, and their automated decision aids were not perfectly reliable, participants nonetheless were likely to miss events if not specifically prompted about them by the AMA, and followed AMA directives even when they contradicted gauge recommendations. Although research comparing performance under automated versus non-automated decision aid contexts shows some increase in overall performance for those with an automated decision aid (see Skitka *et al.*, 1999), there remain some reasons to be concerned about omission and commission errors in automated contexts. Automated decision aids are often used in contexts where errors are especially costly in the form of human life (e.g. aircraft, hospital intensive care units, nuclear power plants). Although automated aids may on average reduce overall error rates even when they are highly reliable, it would be desirable in these contexts to also guard against the potential problems introduced by a tendency of the AMAs to short-circuit situation awareness. Increased social accountability may be an answer in some settings.

Adding to a considerable literature that indicates that accountability can often ameliorate different cognitive biases, our results indicated that appropriately channeled accountability can reduce the tendency to make errors of omission and commission, and lead to improved task performance. Previous research has found that accountability can lead to greater cognitive complexity and improved human judgment; this study indicates that it can also positively affect performance on monitoring and tracking tasks.

Accountability for overall performance or decision accuracy had the expected effects: Participants in these conditions made fewer omission and commission errors than participants who were non-accountable, or who were accountable for responding quickly to events or for their tracking performance. Participants who were accountable for their overall performance or accuracy were both more attentive, as evidenced by catching events even when they were not prompted about them by the automated monitoring aid, and more thorough information seekers, as evidenced by the increased use of verification procedures to prevent making commission errors. The increase in verification behaviors in particular is important. Although previous research has indicated that accountability pressures lead to increased cognitive complexity as determined by thought protocol analysis (e.g. Tetlock, Skitka & Boettger, 1989), the present study indicated that accountability pressures can also lead decision-makers to seek out additional information before making a decision. In contexts where high situation awareness is associated with better decision-making, increased social accountability for performance may be especially useful.

It is interesting to note that although accountability for overall performance and accuracy reduced the frequency of omission and commission errors, this did not happen at the expense of response time or performance on other tasks. Participants in these conditions responded as quickly and performed as well or better than non-accountable participants on the tracking task. Only those made explicitly accountable for their tracking performance significantly outperformed participants in the other conditions. There was no particular gain in response time when participants were made accountable for responding quickly, nor was there any cost in response time to events when participants were accountable for something else, such as accuracy.

Taken together, these results point to the applied utility of social accountability in some decision-making contexts. Adding various forms of accountability for decision-making may have the benefit of increasing performance in various settings, especially those for which there is a high need for vigilance and/or situation awareness.

Other results shed some light on the psychological underpinnings of how automation leads to errors in decision-making. There were some correlations between a belief in the superiority of computers as decision-makers and commission errors, but the magnitude of the correlation was relatively small. Although a belief in the superiority of computerized judgment was a strong predictor of who did and did not follow a computerized recommendation that contradicted all other decision-relevant information on the scenario study of Skitka and Mosier (1994), that study explicitly provided all task-relevant information and there was no need for participants to be particularly vigilant to get it. When task relevant information is more difficult to obtain, issues of vigilance and cognitive laziness emerge as much stronger predictors of the extent to which people rely on automated aids than does belief in the relative authority of automated aids. Consistent with the cognitive miser explanation of automation bias, social accountability only

led to increased vigilance and verification behavior when people were made to feel accountable for monitoring behavior and accuracy, and not when they were made to feel accountable for task not requiring increased vigilance or information gathering (e.g. tracking).

Further research needs to be done to explore how to mitigate the tendency to shed responsibility for task monitoring when people have an automated decision aid, and to guard against possible declines in skill levels as people increasingly rely on automation to do manual tasks (e.g. an autopilot). Parasuraman and his colleagues have found that adaptive automation can at least prevent the problem of overreliance on automated aids leading to a decline in manual skills (Parasuraman, Mouloua & Molloy, 1997), that could also help keep operators alert and engaged. Adaptive automation takes over tasks only when operators are overloaded, but gives them back to the human operator during periods of lower workload. This approach represents a creative intervention that maximizes the strengths of automated aids—they assist people during periods of high workload and free cognitive resources to be focused on procedures that may be less appropriate to assign to automation—by that takes into account the potential costs of highly automated systems in the form of potential skill decay or complacency. Other research has begun to explore the efficacy of explicit training about the tendency to use automation in biased ways, with some encouraging effects with respect to ameliorating at least the tendency to make commission errors (Skitka, Mosier, Burdick & Rosenblatt, 1998).

In conclusion, the human operator is without a doubt an extremely critical agent in the human-automation system. Although human factors engineers pay considerable attention to human-centered display design and ergonomics, there needs to be greater consideration for how the introduction of automated systems might create changes in the operators' perception of the decision-making environment. This research points out the need for a clearer understanding of the psychological impact of automation on human operators, and for the development of decision-making contexts that take into account the strengths and weaknesses of both the human and the automation to better optimize performance.

References

- ANALYSIS OF FLIGHT DATA AND CABIN VOICE TAPES FOR KOREAN AIRLINES FLIGHT KE-007 (1993). *Aviation and Space Technology*, **138**, 17.
- AZAR, B. (1998). Danger of automation: it makes us complacent. *Psychological Monitor*, **29**, 18.
- BILLINGS, C. E. (1996). *Human-centered aircraft automation: Principles and guidelines*. Technical mem. no. 110381. NASA Ames Research Center, Moffett Field, CA.
- BREWER, M. B. & CRANO, W. D. (1994). *Social Psychology*. St. Paul, MN: West Publishing.
- CHAIKEN, S. (1980). Heuristic versus systematic information processing and the use of source versus message and cues in persuasion. *Journal of Personality and Social Psychology*, **39**, 752–766.
- CVETKOVICH, G. (1978). Cognitive accommodation, language, and social responsibility. *Social Psychology Quarterly*, **41**, 149–155.
- DONEY, P. M. & ARMSTRONG, G. M. (1996). Effects of accountability on symbolic information search and information analysis by organizational buyers. *Journal of the Academy of Marketing Science*, **24**, 57–65.
- FISKE, S. T. & TAYLOR, S. E. (1994). *Social Cognition*, 2nd edn. New York: McGraw-Hill.
- HAGAFORS, R. & BREHMER, B. (1983). Does having to justify one's decisions change the nature of the decision process? *Organizational Behavior and Human Performance*, **31**, 223–232.

- JOHNSON, E. J., PAYNE, J. W., SCHKADE, D. A. & BETTMAN, J. R. (1991). *Monitoring information processing and decisions: The MouseLab system*. Philadelphia: University of Pennsylvania, The Wharton School.
- LAYTON, C., SMITH, P. J. & MCCOY, C. E. (1994). Design of a cooperative problem-solving system for en-route flight planning: an empirical evaluation. *Human Factors*, **36**, 94–119.
- MCALLISTER, P. W., MITCHELL, T. R. & BEACH, L. R. (1979). The contingency model for the selection of decision strategies: an empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance*, **24**, 228–244.
- MOLLOY, R. & PARASURAMAN, R. (1996). Monitoring an automated system for a single failure: vigilance and task complexity effects. *Human Factors*, **38**, 311–322.
- MOSIER, K. L., PALMER, E. A. & DEGANI, A. (1992). Electronic checklists: implications for decision making. In *Proceedings of the Human Factors Society 36th Annual Meeting*, pp. 7–11. Santa Monica, CA: Human Factors Society.
- MOSIER, K. L., SKITKA, L. J., BURDICK, M. & HEERS, S. (1998). Automation bias: decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology*, **8**, 47–63.
- MOSIER, K., SKITKA, L. J. & KORTE, K. (1994). Cognitive and social psychological issues in flight crew/automation interaction. *Proceedings of the Automation Technology and Human Performance Conference*, Sage.
- NASA AMES RESEARCH CENTER (1989). *Window/panes: workload PerformANce simulation*. Rotorcraft Human Factors Research Branch, NASA Ames Research Center, Moffett Field, CA.
- PARASURAMAN, R., MOULOVA, M. & MOLLOY, R. (1997). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, **38**, 665–679.
- PARASURAMAN, R. & RILEY, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, **39**, 230–253.
- ROZELLE, R. M. & BAXTER, J. C. (1981). The influence of role pressures on the perceiver: judgments of videotaped interviews varying judge accountability and responsibility. *Journal of Applied Psychology*, **66**, 437–441.
- SIEGEL-JACOBS, K. & YATES, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, **65**, 1–17.
- SIMONSON, I. & NYE, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, **51**, 416–446.
- SKITKA, L. J. & MOSIER, K. L. (1994). Automation bias: *when, where and why?* Paper presented at the *Judgment and Decision-Making Society Annual Convention*, St. Louis, MO.
- SKITKA, L. J., MOSIER, K. L. & BURDICK, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies* (in press)
- TETLOCK, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, **46**, 74–83.
- TETLOCK, P. E. (1985a). Accountability: a social check on the fundamental attribution error. *Social Psychology Quarterly*, **48**, 227–236.
- TETLOCK, P. E. (1985b). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, **45**, 74–83.
- TETLOCK, P. E. (1991). An alternative metaphor in the study of judgment and choice: people as politicians. *Theory and Psychology*, **1**, 451–475.
- TETLOCK, P. E. & BOETTGER, R. (1989). Accountability? A social magnifier of the dilution effect. *Journal of Personality and Social Psychology*, **57**, 388–398.
- TETLOCK, P. E. & BOETTGER, R. (1994). Accountability amplifies the status quo effect when change creates victims. *Journal of Behavioral Decision Making*, **7**, 1–23.
- TETLOCK, P. E. & KIM, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, **52**.
- TETLOCK, P. E., SKITKA, L. J. & BOETTGER, R. (1989). Social and cognitive strategies for coping with accountability: conformity, complexity and bolstering. *Journal of Personality and Social Psychology*, **57**, 632–640.