THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, *10*(1), 85–97

Automation Bias and Errors: Are Crews Better Than Individuals?

Linda J. Skitka
Department of Psychology
University of Illinois at Chicago

Kathleen L. Mosier
Department of Psychology
San Francisco State University

Mark Burdick
San Jose State University Foundation and NASA Ames Research Center
Moffett Field, CA

Bonnie Rosenblatt
Department of Psychology
University of Illinois at Chicago

The availability of automated decision aids can sometimes feed into the general human tendency to travel the road of least cognitive effort. Is this tendency toward "automation bias" (the use of automation as a heuristic replacement for vigilant information seeking and processing) ameliorated when more than one decision maker is monitoring system events? This study examined automation bias in two-person crews versus solo performers under varying instruction conditions. Training that focused on automation bias and associated errors successfully reduced commission, but not omission, errors. Teams and solo performers were equally likely to fail to respond to system irregularities or events when automated devices failed to indicate them, and to incorrectly follow automated directives when they contradicted other system information.

Requests for reprints should be sent to Linda J. Skitka, Department of Psychology (M/C 285), University of Illinois at Chicago, 1007 W. Harrison Street, Chicago, IL 60607–7137. E-mail: lskitka@uic.edu

A series of recent studies has identified two classes of errors that commonly emerge in highly automated decision environments: (a) omission errors, defined as failures to respond to system irregularities or events when automated devices fail to detect or indicate them, and (b) commission errors, which occur when people incorrectly follow an automated directive or recommendation, without verifying it against other available information or in spite of contradictions from other sources of information (e.g., Mosier, Skitka, Heers, & Burdick, 1997, 1998; Mosier, Skitka, & Korte, 1994; Skitka, Mosier, & Burdick, 1996). These errors are hypothesized to be a consequence of "automation bias," the use of automation as a heuristic replacement for vigilant information seeking and processing. Errors related to automation bias have been documented in Aviation Safety Reporting System reports (Mosier, Skitka, & Korte, 1994) and in laboratory studies using professional pilots as well as students (Mosier et al., 1997; Skitka et al., 1996). In all of the events documented, information necessary to avoid the errors was available to the decision maker but was either not identified or not checked.

Many characteristics of automation are supportive of, if not conducive to, its use as a heuristic. In "glass cockpit" aircraft, automated systems are highly reliable and accurate, qualities that encourage their use as a short cut in performing tasks and making decisions (see Mosier et al., 1998). Additionally, many features of highly automated aircraft discourage vigilant monitoring of system functioning and cross-checking of information. Sarter and Woods (1997), for example, recently documented problems in adequately monitoring automation, including a tendency of some pilots to focus solely on the primary flight display for information rather than scanning a wider array of instruments. Adequate monitoring is hindered by the "fragmentation of feedback," that is, the fact that relevant information is distributed more widely among cockpit indicators than in conventional aircraft, making the traditional scan pattern insufficient. In addition, monitoring is hindered by the confusion between displays of "commanded" versus "actual" flight configurations. Poor feedback also results in mode confusion or the belief that the aircraft is in one automated configuration (mode) rather than another, which could be another prelude to error (e.g., Woods & Sarter, 1998).

Virtually all of the research to date has examined the emergence of automation bias in the context of a single decision maker working in isolation. Across studies, this research reveals that participants working alone (and regardless of whether they are pilots or students) make errors of omission and commission on more than half the occasions in which they have an opportunity to do so. In most highly automated environments, however, more than one person is available to perform tasks and monitor systems. What remains an unanswered question is whether a second crewmember can act as a guard against automation bias.

At first glance, it would seem that omission and commission errors would be less likely in a two- than a one-person crew. When two people are monitoring system events, it would seem to double the chances that they would detect a system anomaly, even if it were not detected by an automated decision aid. Indeed, many cockpit procedures are created on this premise. Moreover, doubling the number of people would also seem to enhance the probability that at least one of them will notice if the automated decision aid gives a recommendation that is inconsistent with other system indexes.

However, research suggests that even if a second person does not increase the odds of detecting more events, his or her mere presence may have an impact on a given operator's behavior. *Social facilitation* is defined as improvements in performance produced by the mere presence of others, whether these others are an audience or co-actors (Allport, 1920; Triplett, 1898). For example, Allport (1920) asked participants to write down as many word associations as they could think of for different words. Using the same participants in an alone versus with-others condition, he found that 93% of the participants could generate more alternative meanings in the presence of others. Similar effects emerged with animal studies involving feeding behavior and mazes; animals performed these tasks faster in the presence of other animals than when alone (Chen, 1937; Gates & Allee, 1933; Ross & Ross, 1949). Although most research reveals facilitating effects for the presence of others on measures of performance (e.g., Allport, 1920), some research has revealed that the presence of others can also lead to decreased performance, especially on unfamiliar or complex tasks (e.g., Pessin, 1933).

Zajonc offered a solution to this seeming inconsistency with a drive theory of social facilitation (Zajonc, 1965). According to drive theory, the presence of others has a nondirectional effect on people's behavior. The nondirectional component implies that the presence of others does not influence *what* type of behavior people engage in (e.g., performance enhancing or debilitating behavior), only that people's motivation, or drive, to behave in a particular way will be enhanced in the presence of others. Situational cues direct what people do, and social facilitation intensifies this response. Social facilitation may therefore lead to an increase or decrease in performance, depending on what the dominant response is in the social context. Individuals, then, are more likely to emit dominant responses in the presence of others than when alone, and performance is either enhanced or impaired depending on the match of the dominant response to the performance being measured (for reviews, see Geen, 1989; Geen & Gange, 1977). Assuming that vigilant performance is a dominant response for professional pilots, the presence of a second crewmember should heighten vigilance and improve performance.

However, there is also some evidence that indicates that having another decision maker or system monitor will lead to increased performance. The presence of other people is often found to be distracting to performance (Sanders & Baron, 1975). In addition, there is considerable evidence that increasing the number of people responsible for a task leads to social loafing, or the tendency of individuals to exert less effort when participating as a member of a group than when alone (Ingham, Levinger, Graves, & Peckham, 1974; Williams, Harkins, & Latané, 1981; for reviews, see Karau & Williams, 1993; Karau, Williams, & Kipling, 1995). Most directly relevant to the goals of this study was research that indicated that social loafing is not restricted to simple motor tasks but also applies to cognitive tasks. Harkins and Szymanski (1989) found that people working in three-person groups generated only 75% as many uses for a common object as they did when working alone. These same groups also made more than twice as many errors on a vigilance task of detecting brief flashes on a computer screen than they did when working alone.

Sharing responsibility for system monitoring tasks and decision making with a computer may have psychological parallels to sharing tasks with humans. Given that people treat computers that share task responsibilities as a "team member" and show many of the same in-group favoritism effects for computers that they show with people (Nass, Fogg, & Moon, 1996), it may not be surprising to find that diffusion of responsibility and social loafing effects (or social facilitation) might also emerge in human–computer or human–automation interaction. To the extent that task responsibilities are shared with computerized or automated decision aids, people may well diffuse responsibility for those tasks to those aids and feel less compelled to put forth a strong individual effort. Adding another crewmember to the mix may therefore act to continue to dilute both crewmembers' sense of primary responsibility for system monitoring, in addition to the dilution already created by sharing these responsibilities with automated monitoring aids.

The goal of this article was not to test these social psychological theories, per se, but was instead to use social psychological theorizing and research as a source of ideas for why or why not the presence of a second crewmember might help guard against automation bias. Just as theories from cognitive psychology and ergonomics can inform display design, social psychological theories can help inform how we should design the human side of work environments, training, and procedures.

This study was therefore designed to explicitly compare performance across one- and two-person crews to determine whether the presence of another person either reduced, increased, or had no impact on the tendency to make omission and commission errors in automated contexts. If attending to system states, utilizing all information, and responding correctly to system events are dominant responses in the kinds of domains we have been studying, we should expect that the presence of a second crewmember will decrease errors by increasing each one's motivation to maintain vigilance with respect to automated systems. If, on the other hand, the use of automation as a heuristic is the dominant response, we might expect automation bias and associated errors to be facilitated in a crew setting and exacerbated by the tendency toward social loafing.

In addition to exploring the impact of the number of crewmembers, this study also investigated the efficacy of several possible interventions. It was hypothesized that training participants to explicitly verify automated functions and directives against other available sources of system information could reduce errors. Therefore, in addition to a control condition that involved system training and instructions, another condition instructed participants that they must verify automated directives as part of task procedures. Finally, to address the question of whether training participants specifically about the tendency of people toward automation bias and the need to guard against errors of commission and omission would ameliorate these errors, a third training condition was included that explicitly informed participants about these errors and how they could be prevented (e.g., through monitoring system states even when not prompted to do so by the automation and by verifying automated functions and directives). Finally, we explored whether a system display enhancement might help guard against automation-related errors and particularly commission errors by including a display prompt, "VERIFY," with each automated action.

PARTICIPANTS

One-hundred forty-four students from a large Midwestern university received partial course credit for their participation in the study, yielding 48 two-person crews and 48 one-person crews.

OVERVIEW

Participants performed monitoring and tracking tasks in a low-fidelity, computerized flight simulation program. Participants did these tasks under conditions that varied as a function of (a) "crew," that is, whether they worked alone or with another person (two levels); (b) one of three levels of training (training that instructed participants that they *could* verify automated directives, training that emphasized that they *must* verify automated directives, or training that included instruction about errors people tend to make in automated contexts and how they can be avoided, as well as instructions that they could verify automated directives); and (c) whether participants received a prompt to verify automated directives each time they received a directive or were not prompted to verify. In sum, the study represented a 2 × 3 × 2 (Crew × Training × Prompt to Verify) three-way between-subjects experimental design. The dependent variables of interest were the number of omission and commission errors participants made across these conditions. An Automated Monitoring Aid (AMA) detected and announced all but 6 of 100 events that required responses, creating six opportunities for participants to make omission errors (i.e., failing to detect an event if not explicitly prompted about it by the

AMA). Similarly, the AMA gave an inappropriate directive six times (e.g., indicating that a gauge was in a red zone when in fact it was not), providing six opportunities for commission errors (i.e., following an AMA directive even when other indexes indicated that the directive was not appropriate). Therefore, the actual reliability of the AMA was 88%.

TASKS

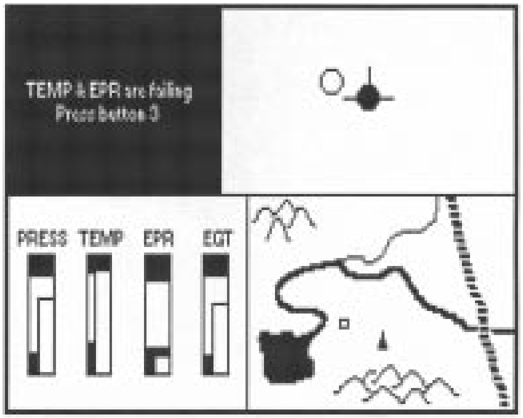
Participants' primary task was to complete four flights or trials using the Workload/PerformANcE Simulation software (W/PANES) developed at NASA Ames Research Center (NASA Ames Research Center, 1989). This program presents participants with a set of tasks designed to simulate the types of monitoring and tracking tasks involved in flying commercial aircraft. Participants were exposed to four quadrants of information using a 486/33 Personal Computer, and 14-in. color monitor (see Figure 1).

The Tracking Task

Participants used a two-axis joystick to keep their own-ship symbol (the circle with lines through it in the top right quadrant of Figure 1) aligned with a moving circular target by following the motion of the target circle with the joystick, compensating for movements away from the center in heading (horizontal) and altitude (vertical). This task ran continuously throughout each of the four trials and required the greatest consistent attention from the participant.

Waypoints

In addition to maintaining tracking performance, participants were also required to monitor their position on a map (the lower right quadrant of Figure 1). A pink square representing own-ship traveled from left to right across the map. Red and blue triangles on the map represented waypoints. Participants were trained to push a button each time the own-ship symbol passed one of these waypoints (presumably to simulate radioing their location to a communications tower). 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 waypoints on each trial. The layout of the map and the location of the waypoints varied across the four trials.



TEMP & EPR are failing
Press button 3

Figure 1 Primary task display.

Gauge Events

The lower-left quadrant displayed four analog slide-rule gauges that were used for a gauge monitoring task. Each gauge had an upper and lower red zone. When any gauge went into a red zone (Gauge indicators moved continuously, but stayed within the normal range the vast proportion of the time.), participants were instructed to press the gauge's corresponding button on their response box; if all three gauges went into a red zone at the same time, participants were trained to push a "gauge reset" button. Gauges were programmed to return to a "safe" value after 10 sec in the red zone if no button was pressed (i.e., if the participant missed an event). There were 17 single gauge events and four reset events for each trial. The timing and order of gauge events were randomized across the four trials.

Automated Monitoring Aid

The top-left quadrant contained AMA announcements. The AMA notified participants of events, and recommended specific courses of action. Participants were told that the AMA was highly reliable but was not 100% accurate in detecting when they needed to respond to waypoints or to gauge events (i.e., it might fail to notify them of critical events or could give inappropriate action directives). It was emphasized that the participant's job was to perform correctly and accurately and that the information presented to them by their gauges and maps would always be 100% accurate.

Across the four trials, there were six omission and six commission error opportunities. Specifically, the AMA failed to prompt the participant about an event on six occasions (creating opportunities for omission errors) and gave inappropriate directives on six occasions (creating opportunities for commission errors). The AMA therefore had an actual reliability of 88% across all trials.

One- Versus Two-Person Crews

In the one-person crew, the pilot alone was responsible for all the W/PANES tasks. In two-person crews, 1 participant was randomly assigned to the pilot role and the other to the role of copilot. Pilots and copilots were instructed on all tasks and were told that they were both responsible for monitoring W/PANES system events. The pilot was responsible for actually doing the tracking task and for making all button pushes in response to events. Copilots were given a set of 75 two-digit multiplication problems that they were told were being used to simulate the nonflight responsibilities of copilots (e.g., navigation and communication) and were trained that their responsibilities were to assist the pilot with monitoring W/PANES for system events and to also accurately complete as many math problems as possible.

Procedure

Participants were recruited to participate either singly or in pairs and were instructed on the W/PANES tasks. All participants were told that the AMA was not 100% accurate, whereas the gauges and waypoint events depicted on the screen were always reliable and accurate. Participants were told that they were responsible for responding correctly, regardless of whether they were correctly prompted by the AMA to respond or if the AMA gave an inappropriate recommendation.

In the normal training condition, participants were told that they *could* check the appropriate indicators to verify the directives provided by the AMA. In the verification training condition, participants were told they *must* verify all directives issued by the AMA by checking the various primary indicators (e.g., the map or the gauges). In the automation bias training condition, participants were told that there was a tendency in decision makers to use automated aids when they have them, as a substitute for vigilant monitoring and cross-checking of information, and that these kinds of decision aids can lead to two specific kinds of errors. Omission errors were explained as failures to respond to events when not explicitly prompted about the event by the AMA. Commission errors were explained as resulting from failing to verify AMA directives and the tendency to follow AMA recommendations even when the gauges or waypoints indicated that the AMA is not correct. As in the verification training condition, participants in the automation bias condition were instructed to verify all directives.

After verbal instructions and a demonstration of W/PANES, participants were given a 5-min practice session. For the experimental sessions, half of the participants received AMA directives that included a verification prompt. Specifically, the word “Verify” was added to any directives they received from the AMA. The other half of the participants received no prompt to verify the AMA directives. Order of trial blocks was balanced to control for possible order effects.

RESULTS

Omission Errors

Descriptively, 51% of the participants made one or more omission errors, and almost 30% made three or more. On average, participants made 1.85 omission errors out of a total of six possible errors (*SD* = 1.81), regardless of experimental condition.

To what extent did having a second crewmember, training, or prompts to verify influence the number of omission errors participants made on the W/PANES task? A 2 × 3 × 2 (Crew × Training × Verify Prompt) between-subjects analysis of variance revealed no significant effects of these variables on the number of omission errors. Because the lack of statistically reliable effects could be due to the fact that sample sizes were relatively low per cell (minimum $n = 8$ per cell), effect sizes were also calculated. The effect size for the crew and prompt main effects and the crew by training interaction were all <.001, suggesting that even if sample size was increased, no effects for these variables would be detected. The training main effect, and Crew × Prompt, Training × Prompt, and Crew × Training × Prompt interactions all had effect sizes <.03. Other analysis calculated the necessary sample size per cell that would be needed to have an 80% chance of detecting an effect of this size (see Keppel, 1991, p. 77). The results of this analysis indicated that it would take another 104 participants per cell to have the statistical power to detect effects of this size. In short, to the extent that crew (one- or two-person), training, or prompts affected the tendency to make omission errors, these effects are extremely small and probably of no practical use as interventions.

Commission Errors

On average, participants made 3.25 commission errors out of a possible 6 (*SD* = 1.88), and almost 80% made 2 or more commission errors. An examination of the number of commission errors as a function of crew size, prompts to verify, training, and trial order indicated that only training affected the number of commission errors participants made, $F(2, 84) = 3.64$, $p < .05$, $\omega^2 = 08$. In other words, 8% of the variance in commission errors could be accounted for by the training manipulation (an effect size that Cohen, 1977, would categorize as above a medium effect size). Tukey tests indicated that the group that was explicitly trained about automation bias and resultant omission and commission errors made fewer commission errors ($M = 2.59$, $SD = 1.72$) than either the could-verify training group ($M = 3.84$, $SD = 1.61$) or the must-verify group ($M = 3.31$, $SD = 2.12$).

No other effects reached traditional significance levels. The crew and prompt main effects, and Crew × Prompt, and Crew × Training × Prompt interactions each had effect sizes <.01. The Crew × Training interaction had an effect size of $\omega^2 = .03$, and the Training × Prompt interaction had an effect size of $\omega^2 = 04$. However, similar to the analysis conducted with respect to omission errors, analysis indicated that it would take increasing the sample size per cell by at least 100 for these effects to have an 80% chance of being detected. In short, even if the additional data were collected and these effects were found to be statistically significant, they would not be very practically meaningful as interventions because their effects are so very modest in size.

In sum, training about the problem of automation-related omission and commission errors helped reduce the number of commission errors made. However, explicit training to verify automated directives, having a second crewmember to help monitor system events, or being prompted to verify automated directives had no impact on the number of errors people made.

DISCUSSION

This study explored the extent to which automation bias remains a pervasive problem even (a) in the context of a two-person crew, (b) when participants were given explicit training about automation bias as a potential problem, (c) when participants were explicitly trained to verify automated directives, and (d) when participants were prompted to verify automated directives. Results revealed that none of these variables had an impact on the number of omission errors made, and only training affected participants' tendencies to make commission errors. Participants who were explicitly trained about the phenomena of automation bias were less likely to make commission errors than those who did not receive training about automation bias.

Perhaps the most important finding of this study is that the presence of a second crewmember did not guard against the phenomena of automation bias. Having a second person to monitor system events did not decrease the tendency of people to miss events if not specifically prompted by an automated decision aid that they were happening, and it did not increase the likelihood that they verified automated directives against other available information before responding to system events. It is important to note that having a second crewmember also did not increase people's tendencies to make omission or commission errors. Therefore, the results did not support the notion that automation bias represents a “dominant” response in highly automated decision-making contexts. According to Zajonc’s (1965) drive theory, if automation bias was in fact a dominant response, then people should show higher rates of omission and commission errors in two- than one-person crews. In contrast, if responding in a highly vigilant and situationally aware manner was a dominant response in these settings, then we should have observed lower rates of omission and commission errors in two- than in one-person crews. Results indicated that error rates remained the same, regardless of the number of people monitoring system states.

These results suggest at least two possible interpretations that require attention in future research: (a) Social facilitation does not influence behavior in highly automated settings, or (b) there may be important individual differences in “dominant responses” or decision-making styles in highly automated settings. The first explanation seems unlikely given that Harkins and Szymanski (1989) found that people made twice as many errors in a computerized vigilance task when working with others than when working alone. The individual difference explanation suggests instead that some people may be more prone to be naturally vigilant and others more prone to be “cognitive misers” (Fiske & Taylor, 1994). When in the presence of others, these dominant response styles or patterns might be enhanced but could cancel each other out when averaged across individual differences in dominant orientation. Future research could explore a broader consideration of individual differences as they relate to both monitoring behavior and the presence of a second crewmember.

The most encouraging result of this study was the fact that participants who were made explicitly aware of automation bias were less likely to make one class of errors (i.e., commission errors). One possible implication of this result is that early instruction on automation bias (i.e., before individuals have much experience with highly reliable automated systems) may be effective in instilling the practice of seeking verification information that confirms automated recommendations before taking action.

Although this study is limited by the fact that a student sample was used, previous research has found that omission and commission error rates are quite constant across student and professional pilot samples (Mosier & Skitka, 1996; Mosier et al., 1997, 1998). Other studies have found some main effect differences in automation use between student samples and professional pilots (i.e., pilots tend to rely on automation more than students in part task simulations) but no differences in overall usage patterns across automation failures (Riley, 1996). Riley, Lyall, and Wiener (1993) also found that pilots were more likely than students to inappropriately rely on an automated aid for a computer-game task after it had just failed, suggesting that habitual use may “bias their use of automation in a way that may be counterproductive” (p. 24). Given that we know that lay people and professional pilots have the same baseline tendencies toward making at least omission and commission errors (even if there are differences in overall preference for using automation as Riley, 1996, reports), using student samples to test different strategies that could be used to ameliorate these tendencies allowed us to build a sufficient sample size to be confident in the validity of our results. Current research is replicating this study using a pilot sample to more conclusively establish whether the presence of another person might ameliorate the tendency to make errors of omission and commission in actual aviation settings.

Taken together, these results suggest that early training about automation bias may be a valuable avenue to pursue to prevent problems of commission errors in glass-cockpit decision-making settings. In addition, the pattern of findings observed here hint that further exploration of individual differences in baseline propensities toward cognitive vigilance versus miserliness might be warranted.

ACKNOWLEDGMENT

This research was supported by Grant NCC 2–297 from NASA Ames Research Center.

REFERENCES

Allport, F. H. (1920). The influence of the group upon association and thought. *Experimental Psychology, 3,* 159–182.

Chen, S. C. (1937). Social modification of the activity of ants in nest-building. *Physiological Zoology, 10,* 420–436.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fiske, S. T., & Taylor, S. E. (1994). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Gates, M. F., & Allee, W. C. (1933). Conditioned behavior of isolated and groups of cockroaches on a simple maze. *Journal of Comparative Psychology, 15,* 331–358.

Geen, R. G. (1989). Alternative conceptions of social facilitation. In P. B. Paulus (Ed.), *The psychology of group influence* (2nd ed., pp. 16–31). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Geen, R. G., & Gange, J. J. (1977). Drive theory of social facilitation: Twelve years of theory and research. *Psychological Bulletin, 84,* 1267–1288.

Harkins, S. G., & Szymanski, K. (1989). Social loafing and group evaluation. *Journal of Personality and Social Psychology, 56,* 934–941.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology, 10,* 371–384.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65,* 681–706.

Karau, S. J., Williams, K. D., & Kipling, D. (1995). Social loafing: Research findings, implications, and future directions. *Current Directions in Psychological Science, 4,* 134–140.

Keppel, G. (1991). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice Hall.

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

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. D. (1997). Patterns in the use of cockpit automation. In M. Mouloua & J. Koonce (Eds.), *Human–automation interaction: Research and practice* (pp. 167–173). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. D. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 8,* 47–63.

Mosier, K. L., Skitka, L. J., & Korte, K. J. (1994). Cognitive and social psychological issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 191–197). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

NASA Ames Research Center. (1989). *Window/PANES: Workload/PerformANcE Simulation.* Moffett Field, CA: NASA Ames Research Center.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human–Computer Studies, 45,* 669–678.

Pessin, J. (1933). The comparative effects of social and mechanical stimulation on memorizing. *American Journal of Psychology, 45,* 263–270.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 19–36). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Riley, V., Lyall, E., & Wiener, E. (1993). *Analytic methods for flight-deck automation design and evaluation, phase two report: Pilot use of automation* (FAA Contractor Report No. DTFA01–91–C–00039). Minneapolis, MN: Honeywell Technology Center.

Ross, S., & Ross, J. G. (1949). Social facilitation of feeding behavior in dogs: I. Group and solitary feeding. *Journal of Genetic Psychology, 74,* 293–304.

Sanders, G. S., & Baron, R. S. (1975). Is social comparison irrelevant for producing choice shifts? *Journal of Experimental Social Psychology, 13,* 303–314.

Sarter, N. B., & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human Factors, 39*, 553–569.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1996, July). Accountability and automation bias. In E. Tetlock & J. Lerner (Chairs), *Accountability: The social context of cognition*. Symposium conducted at the Eighth Annual American Psychological Society Conference, San Francisco.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology, 9,* 507–533.

Williams, K., Harkins, S., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology, 40,* 303–311.

Woods, D. D., & Sarter, N. B. (1998). *Learning from automation surprises and "going sour" accidents: Progress on human-centered automation* (NASA Contract No. NCC 2–592). Moffett Field, CA: NASA Ames Research Center.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269–274.

Manuscript first received June 1998