Title
Training Produces Suboptimal Adaptation to Redundant Instrument Failure in a Simulated Human-Machine Interface

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Ons, Bart
Storms, Gert
Vanpaemel, Wolf

Publication Date
2005

Peer reviewed

Training Produces Suboptimal Adaptation to Redundant Instrument Failure in a Simulated Human-Machine Interface

Rob Youmans (ryouma1@uic.edu)
Stellan Ohlsson (stellan@uic.edu)

Department of Psychology, 1007 W. Harrison St.
University of Illinois, Chicago
Chicago, IL 60607 USA

Abstract

The operation of machines typically requires attention to instruments that signal the state of the machine. One safeguard against instrument failure is to provide backup instruments, but this works only if the operators react to failure by switching attention to the backups. Little is known about the effect of negative outcomes or feedback on attention allocation. In two experiments, we demonstrate that prior training causes operators of a simulated machine to adapt to instrument failure by changing to a suboptimal decision rule rather than by reallocating attention to a different information channel. The results raise theoretical questions and warn interface designers not to overrate backup instruments.

Introduction

The operation of complicated machines involves the entire cognitive architecture: perception, attention allocation, working memory, skill acquisition, decision making, problem solving and so on. We focus on how machine operators utilize available information about a machine in a situation in which the instruments they have been trained to use suddenly begin to provide inaccurate information that leads to poor task performance, when a second, valid, source of information is available. A fully rational operator would respond by switching to the valid information source.

Three cognitive observations suggest that the response of human operators is more complex. First, feedback and negative outcomes show that some task component has failed, but not necessarily which one. In prior work, we proposed a theory of how people learn from negative outcomes when the learner knows that the cause of the negative outcome is a fault in his or her task strategy (Ohlsson, 1996), but in human-machine interactions, the learner also has to consider the possibility that the machine is failing. How do individuals decide which component is responsible?

Second, training produces biases and automaticity that might interfere with rapid adaptation to changing task demands (Allport, A., Styles, E. A., & Hsieh, S., 1994). Although quickly switching from one task set to another might subjectively seem to progress smoothly and effortlessly, empirical evidence strongly suggests that switching between even simple task sets can be quite difficult (e.g., De Jong, 2000). To what extent are people limited by prior experience when faced with the need to adapt to changing task conditions?

Third, attention allocation is only partially under deliberate and intentional control. Little is known about the relation between feedback, negative outcomes, and attention allocation. Under which circumstances will a negative outcome alter attention allocation as opposed to other task components?

We investigate these questions with the help of a simulated human-machine interface in which the degrading of one set of instruments poses a need to re-allocate attention.

A Simulated Machine Interface

In our simulation, participants assume the role of the operator of a juice factory. Two *input containers*, Tank A and Tank B, were shown on the upper left side of a computer screen, connected with pipes to a *mixing tank* shown to the right; see Figure 1.

On the lower half of the screen was the gauge equivalent of the color information. Here, three realistic looking temperature gauges representing Tanks A, B, and the Mixing Tank were displayed; see Figure 1.

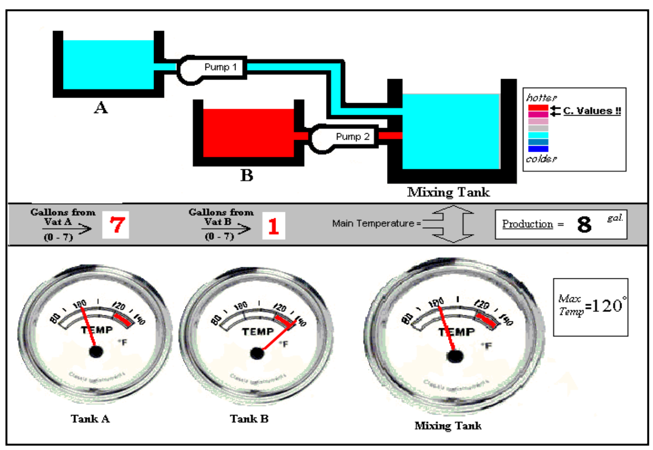

Figure 1: Example of factory interface. *Note.* Factory interface was in color.

Each input tank contains a certain amount of liquid at a certain temperature. The factory is operated by adding some amount of liquid from Tank A and some amount from Tank

B into the Mixing Tank. The mixture instantly becomes ready-to-sell juice. The amount and temperature of the juice is determined by the amount and temperature of the previous content of the Mixing Tank, the added input from Tank A and the added input from Tank B. Once a participant entered their inputs, the simulation was animated; the colored liquid was shown flowing through pipes into the Mixing Tank, and the Mixing Tank's color and gauge responded appropriately to new input. The new temperature in the Mixing Tank indicated the results of participants' inputs after each addition, and accurately represented how liquids mix. The task of the operator was to decide how much liquid was to be added to the Mixing Tank from Tanks A and B, with the goal of maximizing the production of juice without overheating the facility, a type of trade-off situation. Once the two inputs were added, the resulting state of the Mixing Tank was computed and displayed [1], and the operator could make the next decision about how much liquid to add from either input tank.

As shown in Figure 1, the display was divided into two sections by a thick gray line. On the top half of the screen is the section that we will refer to as the *color information* portion of the screen. On the bottom half of the screen is the section that we will refer to as the *gauge information* portion of the screen. In the beginning of each experiment, both the color and gauge information portions of the screen provide the necessary information to successfully perform the mixing task. For example, in the color portion, the temperatures in all tanks are represented by the liquids shades of blue or red. These colors were chosen as somewhat naturalistic indicators of heat (e.g., many bath and kitchen fixtures represent temperature, where blues are cool and reds are hot). To help ensure that participants understood these color values, a color-to-temperature guide was on screen at all times. The colors were divided into three shades of red, three shades of blue, with one shade of gray in between. Deep red was the hottest temperature, and deep blue was the coolest, with gray representing the average of the blue and red extremes. In the gauge information portion of the screen, a red 'needle' in each gauge indicated those same temperatures.

It was the temperature of the Mixing Tank that was most important in the simulation. Although the input liquids could be hot or cold, the Mixing Tank could not accommodate extreme heat. If temperatures in the Mixing Tank ever rose above a certain point, the juice was ruined. This point was represented by the two deepest shades of red (colors), and the two highest needle readings (gauges), and are referred to as the *critical temperature*. If the content of the Mixing Tank reached this temperature, then the pasteurization process was spoiled. Tanks A and B could safely hold liquids across the entire range of possible temperatures; only the Mixing Tank had this temperature restriction.

In both sections, the temperatures of the liquids in Tank A, Tank B, and the Mixing Tank are indicated. In normal operation mode, the factory may be operated on the basis of either the color or gauge information; these two information sources are redundant. Because the colors and gauges present identical information, the task can be solved equally well on the basis of either.

The simulated instruments were implemented so that they could be made to malfunction at a determined point in the simulation. When malfunctioning, both sets of instruments still displayed temperature values for the three tanks, but either the gauges or colors became inaccurate. Only one source of information became inaccurate, so the operator always had the option of discontinuing use of the failing source, and switching to the other.

Experiment 1

In Experiment 1, we examined participants' success rates for running the factory as a function of type of training and type of failure. Participants were either trained to use the colors or the gauges, and they experienced either no failure, failure in the type of instrument they had trained on, or failure in the instrument they had not been trained to use. If the participants responded to the failure of one source of information by switching attention to the other source, their performance might exhibit a brief decrement, a switch cost, and then quickly return to pre-failure performance levels. However, reluctance to switch from a failing source of information would, in this design, lead to rising Mixing Tank temperature, and eventually, an overheating of the system.

Method

Participants

Participants in this study were 181 undergraduate psychology students from the University of Illinois at Chicago. The participants were randomly assigned to each of six groups; five groups of 30, and one group of 31.

Design

The study was a fully crossed 2 (training: color, gauges) by 3 (failure type: none, colors, gauges) design.

Procedure

Informed consent and debriefing were done off-line, but in the experiment participants interacted with a computer. To clarify the procedure, we differentiate between the practice sessions and the three experimental rounds that followed.

Practice Sessions. The practice sessions were designed to teach participants how to operate the simulation successfully. Depending on the participants' experimental condition, the practice session began with instructions for using either the color or gauge information, but not both. Participants were told how Tanks A and B related to the Mixing Tank, and were told repeatedly that their goal was to use the factory to produce as much juice as possible, without overheating the Mixing Tank. Next, participants operated a partial version of the factory simulation that displayed only that portion of the screen that they had been trained on. For example, if a participant was in a color training condition,

[1] TEMPcurrent = [(14 * TEMPprior.) + (7 * TEMPa) + (7 * TEMPb)] / 3, rounded to the closest whole value.

then in the practice session they did not see the gauges, and those trained on gauges could not see any representation of the color portion of the screen during practice. The result was that participants practiced with only one of the sources of information, but never both.

The participants were asked to produce the maximum volume of juice per *trial*. Each trial consisted of two judgments about how much liquid to input, one amount from Tank A (judgment 1) and one from Tank B (judgment 2), into the Mixing Tank. Participants indicated how much juice, from zero to seven gallons, should be entered into the Mixing Tank by typing the appropriate digit on the keyboard. Juice production accumulated across trials. In this paper, a series of trials is referred to as a *round*.

If participants drove the Mixing Tank into the critical temperature range during practice, a warning appeared, and the system paused for 8 seconds. Otherwise, practice ended when participants had produced 300 gallons of juice. After this, the participants were instructed with respect to (but did not practice with) the second source of information that had been absent, and were told that this new portion of the screen conveyed the same information as that with which they had just practiced with, albeit in different form. P*articipants were told that if they were having trouble using one of the two sources of information in any part of the simulation, they should switch to the other source.* Following the practice session, participants had a chance to ask questions, and then began the experimental rounds.

No failure conditions. Following the practice session, participants in the no failure condition were asked to operate the factory for three rounds. In each round, both color and gauge portions of the screen were presented simultaneously and participants were asked to produce 150 gallons of juice as quickly as possible without overheating the Mixing Tank. If the Mixing Tank was heated past the critical temperature, the round ended and a failure display was presented indicating that the participant had overheated the system.

Color-failure conditions. Following the practice session, participants in the color-failure condition were asked to operate the factory for three rounds. The rounds operated just as they did in the no failure condition, except that in the color-failure condition the color information became highly misleading to the participants half way through each round. Specifically, after roughly half of each round was complete, the color information stopped matching those temperatures displayed by the gauges, and instead indicated that the liquid in Tanks A and B were shades of cool blue.

Gauge-failure conditions. Following the practice session, participants in the gauge-failure condition were asked to operate the factory for three rounds. The rounds operated just as they did in the no failure condition, except that in the gauge-failure condition the gauge information became highly misleading to the participants half way through each round. Specifically, after roughly half of each round was complete, the gauge information stopped matching those temperatures displayed by the gauges, and instead indicated that the liquids in Tanks A and B were cool temperatures.

In both failure conditions, the moment at which the mismatch between the sources of information occurred will be referred to as the *failure point*. The failure point was the moment during a round at which a successful participant that is basing his or her decisions on the failing source of information should recognize that the information has failed, and switch to the other source of information. In the failure conditions, the failure point always occurred when a participant had produced between 70 and 90 gallons of juice *and* the Mixing Tank was a gray color or less. If these conditions did not occur, then the failure point occurred at 90 gallons of production. These criteria helped to ensure that, after the failure point, there was still leeway for the Mixing Tank's temperature to increase.

Feedback. In all conditions, three forms of feedback were displayed for the participants operating the system; two while operating, and one after each round was over. The first form was the sudden mismatch between the temperature representations of Tanks A and B as represented by the colors and by the gauges. For example, in the color-failure condition, after the failure, the colors always incorrectly appeared as shades of blue, while the gauges continued to accurately reflect the temperatures of Tanks A and B. Thus, if participants examined the color and gauge representations of Tank A and B, their respective temperatures would no longer match. This type of feedback could be called *online feedback*, since its onset was during each experimental round, rather than beforehand or afterwards; see Figure 2. Once the failure occurred, online feedback persisted until the end of the round.

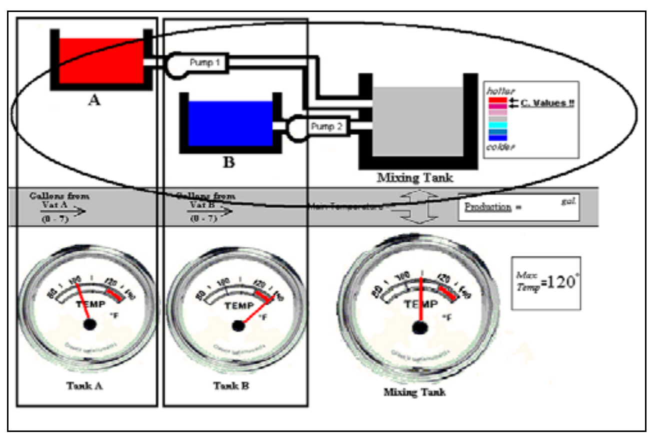

Figure 2: The two rectangles highlight the first form of online feedback, the mismatch between the color and gauge information. The oval highlights the second form of online feedback, the mismatch between what a participant would expect to happen to the Mixing Tank when the color information is used, and the actual outcome whereby the Mixing Tank's temperature increases rather than decreases. *Note.* Factory interface was in color.

The second form of feedback was also online, and consisted of the mismatch between the color representations of the temperature in Tanks A and B, and the effects of adding this juice to the Mixing Tank. In short, the second form of feedback was the illogical way in which the supposedly cool contents of Tanks A and B *increased* the temperature of the Mixing Tank. Like the first form of feedback, this inconsistency persisted until the end of the round, but unlike the first form, this feedback actually made it possible to readily deduce that the color information had become unreliable, while the gauge information maintained its integrity.

The third form of feedback was the outcome feedback that followed each round. In short, participants were told when they failed or succeeded in producing 150 gallons of juice after each round. Thus, in addition to responding to both forms of online feedback *within* rounds, participants also had the opportunity to respond to feedback *between* rounds. For example, a participant in the color-failure condition could adapt to the failure of color information by reflecting on that failure after the first round, and deciding to go with the gauge information in the following rounds. If this decision were made, then in subsequent rounds the participant would succeed, and he or she could deduce that the error was located in the color information, even without reacting to the online feedback.

Results

Our analysis of Experiment 1 focused on the success or failure of participants in the simulation. Specifically, we examined what conditions led to successful operation of the factory simulation, and what conditions led to failure.

No Failing Information Source

In conditions in which no failures of any information source occurred, participants were generally successful at the task. As shown in Table 1, participants that trained on color information succeeded 66% of the time in Round 1, 75% in Round 2, and 82% of the time in Round 3. Similarly, participants trained on gauges succeeded at the task 66% of the time in Round 1, 62% in Round 2, and 69% in Round 3. None of the differences between the color and gauge conditions were significant. In short, regardless of which information source (colors or gauges) participants practiced on, they were able to produce 150 gallons of juice without overheating the system around 66% of the time on their first try, and improved thereafter.

Practice with Stable Information Source

In conditions in which a failure occurred in the information source that was not practiced, participants were generally successful at the task. As shown in Table 1, participants that trained on color information and experienced failing gauges succeeded 60% of the time in Round 1, 73% in Round 2, and 70% of the time in Round 3. Similarly, participants trained on gauges that experienced color failures succeeded at the task 70% of the time in Round 1, 67% in Round 2, and 70% in Round 3. None of the differences between the color and gauge conditions were significant, nor were any of the differences between those participants who practiced with a stable information source and those in the control conditions.

In short, in both the color and gauge failure conditions, there were no significant differences between participants who encountered no failure and those who practiced on the stable information source. For example, if a participant practiced on color information, and gauges subsequently failed, their performance was not affected. Likewise, the performances of participants who practiced with the gauges and experienced color failure were not affected. Because participants in the no failure condition performed identically to participants who practiced on a stable information source, all further comparisons were made between participants who practiced with a stable information source and those who practiced with a failing information source.

Practice With Failing Information Source

As shown in Table 1, this experiment detected large significant differences in the performance of participants who had practiced with an information source that was stable versus those participants who practiced with a failing source of information.

Table 1: Number of participants from Experiment 1 who successfully completed the simulation by condition.

Round	Condition	Color Practice	Gauge Practice
1	Colors Fail	2 (6%)***	21 (70%)
	Gauges Fail	18 (60%)	7 (23%)***
	No Failure	19 (66%)	19 (66%)
2	Colors Fail	5 (16%)***	20 (67%)
	Gauges Fail	22 (73%)	12 (40%)*
	No Failure	22 (76%)	18 (62%)
3	Colors Fail	11 (36%)**	21 (70%)
	Gauges Fail	21 (70%)	15 (50%) ^{ns}
	No Failure	24 (83%)	20 (69%)

Note. Where indicated, significant differences reflect comparisons between groups who practiced with a failing information and groups who practiced on the non-failing information. * $p < .05$; ** $p < .01$; *** $p < .001$

Round 1. Participants trained on either color or gauge information succeeded in Round 1 of our simulation differently depending on which source of information failed. When color information failed, only 6% were able to successfully complete the round, which was significantly less than the corresponding condition's 60% success rate when training occurred on the stable information source, χ^2 (1, N = 61) = 19.84, $p < .001$. Likewise, only 23% of participants who were trained on gauge information succeeded when gauge information failed, compared with 70% when training occurred on the stable information source, χ^2 (1, N = 60) = 11.28, $p < .001$.

Round 2. Participants trained on either color or gauge information succeeded in Round 2 of our simulation differently depending on which source of information failed. When color information failed, 16% were able to successfully complete the round, compared with 73% when training occurred on the stable information source, χ^2 (1, N = 61) = 20.22, $p < .001$. Likewise, only 40% of those trained on gauge information succeeded when gauge information failed, compared with 67% when training occurred on the stable information source, χ^2 (1, N = 60) = 4.92, $p < .05$.

Round 3. Participants trained on color information succeeded in Round 3 of our simulation differently depending on which source of information failed. When color information failed, only 36% were able to successfully complete the round, compared with 70% when training occurred on the stable information source, χ^2 (1, N = 61) = 7.04, $p < .01$. However, though the same trend was detected for participants who were trained on gauge information, the difference between the 50% of those who were trained on gauge information was not significantly different from the 70% when training occurred on the stable information source, χ^2 (1, N = 60) = 2.50, *ns*.

Discussion of Experiment 1

The results showed a decrease in performance when an instrument failed that the participants had been trained to use. Contrary to expectation, the participants did not recover after such failures in the first round, in spite of the two types of feedback the simulation provided. Perhaps more surprisingly, participants did not recover even after two additional rounds. Put differently, after completing an average of 31 trials per round, for an average of 93 trails, only 36% of participants trained to use the color indicators succeeded, as compared to 70% of those with the same training with the failing gauges and 83% of those with the same training and no equipment failure. The recovery rate for the participants trained on the gauges was better, 50% as compared to 70% and 69%, but the main finding of Experiment 1 is clear: The participants experienced difficulties in switching to the alternative information channel.

Experiment 2

What was the nature of the difficulty in returning to no failure performance levels? After all, the participants only had to look at a different part of the screen to gain the same information as they had gained previously. The answer turns on the nature of the participants' adjustment to the instrument failure. The obvious hypothesis is that they (a) noticed the negative outcome, (b) interpreted it as some failure of their customary information source, and (c) accessed the other, and less familiar, source. This hypothesis makes two predictions: First, that reaction times, the time from the end of one trial to the first input in the next, should increase sharply when the failure is introduced. Second, that this increase should gradually diminish over trials.

Method

Participants

Participants were 36 undergraduate psychology students from the University of Illinois at Chicago. They were randomly assigned to each of four groups, with 9 per group.

Materials/tasks

The same simulation as in Experiment 1 was used in Experiment 2, except that the program was modified to record participant reaction times on each trial. In Experiment 2, reaction times were measured from the presentation of Tank A and B's new values at the beginning of each trial, to the participant's first response.

Procedure and Design

The same procedure as in Experiment 1 was used in Experiment 2. Because no differences were found between participants who practiced on a stable information source and those in conditions where no information source failed, the later condition was not run in Experiment 2. The design was therefore a fully crossed 2 (training: color, gauges) by 2 (failure: color, gauges) design.

Results

Performance levels

The participants' success rates in Experiment 2 were not significantly different from participants' performance in Experiment 1. Participants rarely succeeded when the information source they had practiced on failed, usually succeeded when the information source that they had not practiced with failed, and all conditions improved across rounds as they did in Experiment 1.

Table 1: Number of participants from Experiment 2 who successfully completed the simulation by condition.

Round	Condition	Color Practice	Gauge Practice
1	Colors Fail	1 (11%)	5 (56%)
	Gauges Fail	5 (56%)	2 (23%)
2	Colors Fail	2 (22%)	6 (67%)
	Gauges Fail	5 (56%)	4 (44%)
3	Colors Fail	4 (44%)	8 (89%)
	Gauges Fail	8 (89%)	6 (67%)

Reaction Times

Reaction times in the three trials before the failure point and the three trials after were averaged within each condition, for each of the three rounds in Experiment 2. No significant differences emerged in any of the four conditions across the trials; see Figure 3. The time it took participants to make a decision about how much liquid to send into the Mixing Tank remained consistent contrary to expectation.

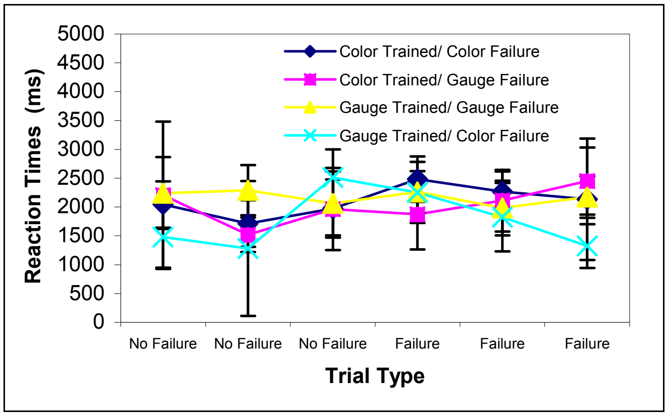

Figure 3: Participant's average reaction times for the three trials preceding and the three trials following information failures by condition in Round 1. Due to space constraints, highly similar null results, found in Rounds 2 and 3, are not here shown.

Discussion of Experiment 2

The reaction time data were not consistent with the hypothesis that the failure to recover, observed in both experiments, was due to the need to think through the implications of the negative outcomes. Casual observation of participants' performances suggests a different interpretation: The participants reacted to each round's failure by entering smaller amounts of input liquid; see Figure 4. This had the effect of prolonging the number of trials in a round before overheating the Mixing Tank, but at the cost of requiring more trials to reach the target of 150 gallons. In short, participants in the experimental conditions significantly reduced their average input into the simulation across rounds by roughly one unit between Rounds 1 and 2, $t(8) = 2.40$, $p < .05$, and again between Rounds 2 and 3, $t(8) = 3.02$, $p < .05$.

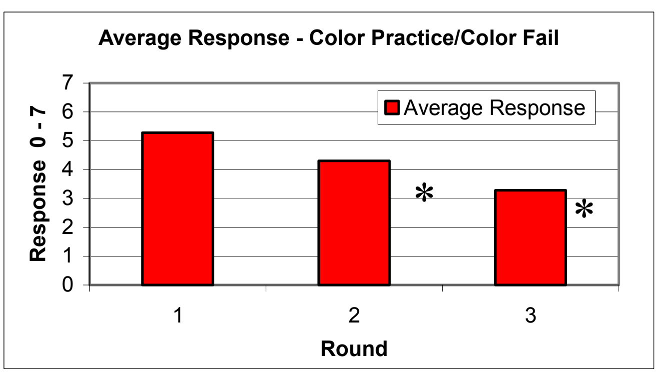

Figure 4. Participants' decisions on how much liquid to send to the Mixing Tank across rounds. *$p < .05$

General Discussion

When two sources of information are available and one of them fails it seems as if a fully rational operator would simply switch to the other one. Because our operators were explicitly told to switch to the alternate source of information if having trouble, doing so seems an obvious strategy to adopt. However, negative outcomes are not easy to interpret, training creates biases that might interfere with noticing behavior, and attention allocation may not be fully under voluntary control. In Experiments 1 and 2, we observed a drastic performance decrement in terms of success rates when the information source a participant had been trained on failed, and we also observed an unexpectedly slow recovery of these success rates across the three experimental rounds. However, the hypothesis that these issues were due to a struggle to think through the online feedback and how to adapt to it within each round was not supported by the reaction time data from Experiment 2. Even in the face of repeatedly failed rounds, participants' *within* round reaction time data did not support the notion that they became more sensitive to *noticing* failure points, let alone reacting to them. Instead, it appears that the participants adapted to their failures *between* rounds, by switching to an unexpected and suboptimal decision strategy: By lowering input into the Mixing Tank, users prolonged the number of trials until failure, while not addressing a key variable, namely maximum production in each trial. Participants did so regardless of the instructed strategy to switch to the other source of information when in trouble. We know of no cognitive theory that would predict that re-allocation of attention would be such an unattractive option in this situation. Could it be that the automaticity of practice on one source of information not only creates a reluctance to switch to others, but also a failure to *react* to the environments outside of these automatic procedures? If this is true, then our results strongly caution against assuming that providing backup instruments is an effective safeguard against instrument failure.

Acknowledgements

The authors thank Justin Oesterreich for his many valuable contributions to this project. We also acknowledge Tim Miura, Ben Jee, James Pellegrino, and Jenny Wiley for their helpful suggestions, and Dustin Rhodes for his help in setting up our lab and deep appreciation of Wonder Woman. Finally, we would like to thank Richard Mueller for his insightful comments on amateur pilot training.

References

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press.

De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes* (pp. 357-376). Cambridge, MA: MIT Press.

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 103, pp. 241-262.