EFFECTS OF ATTRIBUTE AND GOAL FRAMING ON AUTOMATION RELIANCE AND COMPLIANCE

Frank C. Lacson, Douglas A. Wiegmann, and Poornima Madhavan
Aviation Human Factors Division -- University of Illinois, Urbana-Champaign
Champaign, IL

Presentation of an aid's reliability may have mixed results on utilization due to differences in the way the reliability is framed, as well as its effects on reliance versus compliance. Using an aided signal detection task, the current study compared effects of attribute framing (80% correct versus 20% incorrect) and goal framing (focusing on hits and correct rejections versus misses and false alarms) on sensitivity, response bias, compliance, and reliance. Sensitivity and response bias results indicated few differences between aided groups. Information framing did lead to differences in compliance, but not reliance compared to an unframed group. However, reliance differences occurred when comparing positive and negative framing group. These results indicate that the manner in which information about the reliability of a diagnostic aid is presented can have significant, albeit subtle, effects on automation utilization.

INTRODUCTION

Signal detection tasks are as common as looking for fruit defects at the grocery store or finding an open teammate on the basketball court. Also, signal detection tasks are present in domains containing uncertainty and high costs for consequences. For example, the airport screener looking for inappropriate items in luggage, the doctor searching for a tumor on an X-ray image, or the soldier determining the presence of a combatant in unfamiliar terrain must perform their tasks with a high degree of reliability. Technological solutions for improving detection accuracy, such as automation, have been proposed for these difficult tasks. However, users of automated aids tend to under-trust and under-utilize aids that are imperfectly reliable (i.e., less than 100% accurate).

Informing users of the actual reliability of the aid may lead to a more appropriate calibration of trust and dependence. Providing operators with information on automation reliability or on the types of errors made by automation have been shown mixed results in automation utilization (Dzindolet, Pierce, Beck, & Dawe, 2002; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Lacson, Wiegmann, & Madhavan, 2003; Parasuraman & Riley, 1997). These mixed results may occur due to the presentation of aid reliability information in terms of accuracy or errors (Dzindolet et. al., 2002; Lacson et. al., 2003) or by different presentations of a task goal (Nygren, 1997).

A second possible reason for these mixed results may involve the manner in which the information concerning automation reliability is framed or presented to operators. Research on choice framing shows that if risky alternatives are framed in terms of possible gains, people tend to be risk averse, whereas if the same information is framed in terms of possible losses, people tend to be risk seeking (Tversky & Kahneman, 1981). Specifically, *attribute framing* effects occur when an object's attribute is manipulated – examples involve the framing of probabilistic information, such as the likelihood that an aid will be correct or make an error (Levin, Schnittjer, & Thee, 1988; Levin, Schneider, & Gaeth, 1998). *Goal framing* effects can also occur when mentioning the positive or negative consequences of carrying out a behavior or strategy, such as the cost of false alarms vs. the benefits of a hit (Levin et. al., 1988).

Studies on the applications of framing have been conducted to explore effects of time pressure (Svenson & Benson, 1993), dynamic decision-making (Nygren, 1997), and naturalistic decision-making contexts (Perrin, Barnett, Walrath, and Grossman, 2001). However, only a few studies (Dzindolet et. al., 2002; Lacson et. al., 2003) have explored the effects of attribute framing in the context of repeated automation use. Additionally, much research has not yet shown the distinction between two types of automation utilization: automation *compliance* (an operator decides to act when an alert sounds) versus automation *reliance* (an operator decides not to act when an alert is silent) (Meyer, 2001). Perhaps information about an aid's reliability can have subtle effects on the type of utilization strategy adopted by users, thereby diluting any overall positive impact that information about the reliability of may have in a given context.

The current study expands on previous research on automation by examining the effects of both attribute framing and goal framing on operators' compliance and reliance on a signal detection aid. Specifically, a positive frame describes the aid as "80% accurate" and includes a task goal of "making as many correct decisions as possible." A negative frame describes the aid as "20% inaccurate" and includes a task goal of "avoiding as many incorrect decisions and possible." A neutral frame combines the information present in the positive and negative framed groups while an uninformed frame does not contain any aid reliability descriptions or task goals. All frames are compared to participants that manually perform the signal detection task. Decision feedback is given and a payoff structure differently weighs the values and costs of hits, misses, correct rejections, and false alarms.

METHOD

Participants

Participants consisted of 100 undergraduate and graduate students at the University of Illinois, Urbana-Champaign. Eight dollars was given upon completion of the one-hour experiment.

Procedure

Participants performed 100 signal detection trials consisting of 80 noise trials and 20 signal plus noise trials. The order of signal and noise trials was identical across participants. Additionally, participants were not informed about the signal base-rate of .20. The noise letters, A, R, Q, and V, were each displayed 250 times. For the 20 signal trials, 5 signal letters (X) were also placed on the screen. After a 1.25 second delay, the screen was cleared and a diagnostic aid provided a recommendation on whether a signal letter X was present on the screen. For each trial, the probability of the aid making a correct recommendation (hit or correct rejection) was .80 and likelihood of the aid making an incorrect recommendation (false alarm or miss) was .20. This recommendation was given as: 'the aid indicates that a TARGET was PRESENT' or 'the aid indicates that a target was NOT PRESENT.'

Most groups received a recommendation from a diagnostic aid (reliability = .80) on whether or not a signal letter was present on the screen. A confidence rating was collected before feedback from that particular trial was presented. Afterwards, the computer program indicated whether a hit, miss, false alarm, or correct rejection had occurred. Participants were given 20 points for each "hit" and 5 points for each "correct rejection." A deduction of 40 points occurred for each "miss" and 10 points for each "false alarm." A running point total was presented at the start of the next trial.

Before the experiment, participants were assigned to one of five groups (see Table 1), each having a unique attribute and goal frame of the decision aid description (n=20 for each group). Participants in all groups were told to stress accuracy over speed.

Group	Attribute frame	Goal frame
Positive	Aid is "80% accurate and makes ~80 correct decisions"	Maximize hits and correct rejections
Negative	Aid is "20% inaccurate makes ~20 incorrect decisions"	Minimize misses and false alarms
Neutral	Positive + Negative group	Positive + Negative group
Uninformed	None	None
Unaided	N/A	N/A

Table 1. Attribute and goal frame descriptions.

Participants were notified that the final responsibility rested in their own decisions, regardless of the aid's recommendation. A single-factor, between-participants design was used for this experiment. The factor consisted of the attribute and goal framing type present in the instruction set and interface – five levels of this factor were present: positive frame, negative frame, neutral frame, no frame ('uninformed'), and unaided.

RESULTS

Results on sensitivity (d') show that only the neutral group had a higher sensitivity score than the aid ($t(19) = 2.87$, $p < .02$). However, no significant differences occurred between groups ($F = 0.44$, $p = .78$).

Beta (response bias) for the unaided group was higher (more conservative) than the aided groups: positive ($t(38) = 2.66$, $p < .02$), negative ($t(38) = 2.71$, $p < .02$), neutral ($t(38) = 1.97$, $p = .06$), and uninformed ($t(38) = 2.15$, $p < .05$). The conservative response bias in the unaided group led to a greater number of misses and a lower number of false alarms compared to the aided groups. However, no significant differences existed between the aided groups, though beta for the aided groups tended to be slightly higher for than the aid. Compliance rates were measured by the probability of the operator indicating 'yes, a target was present' given that the aid also indicated 'yes, a target is present,' or $p(Oy|Ay)$. Results indicated that the positive ($t(38) = 2.36$, $p < .05$), negative ($t(38) = 3.22$, $p < .01$), and neutral ($t(38) = 1.67$, $p = .10$) groups had a higher p(Oy|Ay) value than the uninformed group.

Hence, the positive ($t(38) = 2.12$, $p < .05$) and negative ($t(38) = 3.33$, $p < .01$), and neutral ($t(38) = 1.78$, $p = .08$) groups also exhibited a higher likelihood of committing a false alarm given the aid false alarmed [p(Ofa|Afa)] than the uninformed group. Nonetheless, the opposite was also true. The positive ($t(38) = 1.93$, $p = .06$) and the negative ($t(38) = 2.54$, $p < .02$) groups were less likely to miss a target given the aid "hit" the target [p(Om|Ah)] than the uninformed group.

Reliance rates were measured by the probability of the operator indicating: 'no, target absent' given that the aid also indicated 'no, target absent,' or p(On|An). Results indicated that the negative group had a greater p(On|An) than the positive group ($t(38) = 2.51$, $p < .02$). Also, neutral group had a significantly higher p(On|An) than the positive group ($t(38) = 1.87$, $p = .07$).

The positive group had a lower likelihood of missing a target when the aid missed [p(Om|Am)] than did the negative ($t(37) = 2.62$, $p < .01$) or neutral group ($t(37) = 2.85$, $p < .01$). In contrast, the positive group had a significantly greater likelihood of committing a false alarm given the aid correct rejected [p(Ofa|Acr)] than did both the negative ($t(38) = 1.96$, $p = .06$) and neutral ($t(38) = 1.67$, $p = .10$) groups.

DISCUSSION

Manual performance

As expected, manual target detection in this task was difficult due to time restrictions and a low signal letter to distracter letter ratio on signal trials. On trials where a signal letter was present, there were only 10 signal letters versus 1,000 noise letters. The mean hit rate of the unaided group was close to chance (.55) and the mean false alarm rate (.07) was near zero, resulting in a d' value of 1.80. In contrast, the low signal to noise ratio resulted in participants adopting a relatively conservative response bias (mean beta = 7.29), which is consistent with other studies in this area. Nonetheless, the values and costs present in the payoff structure should have led participants to adopt a response bias that resembled the optimal beta of 1.0. It appears that the signal rate rather than the payoff structure had larger effect in setting the unaided participants' beta perhaps because the completion of the task was not contingent on actual performance accuracy.

Overall automation effects

Some participants were given an aid that had a hit rate of .80, a false alarm rate of .20. Hence, the sensitivity of the aid (d' = 1.68) was relatively equivalent to manual performance. However, given the payoff structure of the task, the hit and false alarm rate of the aid did equate to a less conservative and optimal response bias (beta = 1.0). Furthermore, given the type of aid was a form of Stage 2 automation that did not provide any specific attention-grabbing cues, we hypothesized that the aid would have a bigger impact on response bias than on sensitivity.

As expected, aided participants did not have a higher sensitivity compared to the unaided group. Although these participants had a higher hit rate, they also had a higher false alarm rate. This indicates that aided participants generally followed the aid's recommendations, resulting in a shift in response bias towards the aid. However, participants did not agree with the aid on every trial since their beta value was greater than the aid's beta. The beta value for the aided groups fell in-between the aid and the unaided group. Apparently, the aid's imperfect reliability reduced participants' trust in the aid to the point where they did not completely depend on the aid.

Participants tended to comply with the aid less than they relied on the aid. This was expected since the aid was generally more likely to declare: "yes, signal present" compared to the participants. The aid's relatively high false alarm rate compared to manual performance may have contributed to this difference.

Reliability information and framing effects

Participants assigned to the neutral group were given reliability information in terms of accuracy and error rate. Likewise, these participants were given a goal to maximize correct responses and to minimize incorrect responses. Providing this reliability information in a neutral format led to a slight improvement in sensitivity compared to the performance of the aid alone. Having a similar hit rate, but a lower false alarm rate than the aid led to this sensitivity difference. Giving participants reliability and goal information led to similar response bias shifts as the uninformed group – beta was lower than unaided performance, but not optimal.

However, the inclusion of aid reliability and goal information led to subtle differences in response patterns. Specifically, compliance rates for the neutral group were greater than the uninformed group. Increased compliance rates also led to more occasions of errors of omission – participants in the neutral group were more like to false alarm with the aid compared to the uninformed group. Since the overall false alarm rate between the neutral and uninformed groups were not significantly different, one can conclude that providing neutral-framed information led to a greater percentage of aid-induced false alarms. Reliance rates between these two groups did not differ.

For participants in the positive frame group, the reliability and goal information was split into accuracy rates and to maximize hits and correct rejections. Participants assigned to the negative frame group were given the information in terms of error rates and to minimize misses and false alarms. Specific frames led to similar sensitivity and response bias effects as the neutral frame.

The isolation of reliability and goal information into specific frames led to even greater increase in compliance rates compared to the uninformed group. Additionally, these participants were less likely to miss the target when the aid had detected the target. In contrast, reliance tended to be differently affected by the manner in which reliability information was presented. Negatively framed information tended to increase reliance, whereas positively frame information tended to decrease reliance. Hence, the negative framed group was more likely to miss a target given the aid missed, but they were less likely to false alarm given the aid correct rejected.

Perhaps positive frame participants focused more on obtaining "hits," as they were instructed. A strategy to obtain more hits may have caused a behavior to indicate: "yes, signal present" even when the aid has indicated: "no, signal absent." Since the false alarm rates for all other groups besides the positive group were lower than the aid, the previously mentioned strategy may also have caused participants to purposely indicate "yes, signal present" even when both aid and the participant's viewing did not detect the target letter.

Framing and the perfect automation schema

Providing information about the aid's imperfect reliability, regardless of frame type, increased compliance rates compared to a non-framed, or uninformed, group. However, an improvement of reliance rates for the framed groups was not as evident. One reason for these effects involves the violation of the "perfect automation schema" (Dzindolet et. al., 2002). Participants generally have a bias towards automation and expect near perfect performance from aids. As a result, dependence drops drastically and recovers slowly when the aid makes an error. This change is dependence by providing aid reliability information occurs because participants primarily use the aid to help detect targets (obtaining hits) instead of helping to not detect a target (obtaining correct rejections). These findings are consistent with the findings in Lacson et. al. (2003) – providing framed information led to increased aid dependence.

Dependence benefits of the attribute framing of aid reliability are most evident in the negative frame group. Participants in the negative frame group have the most reduced expectation of "perfect" automation due to the emphasis on of the aid's error rate. Although the negative group tended to have a higher compliance and reliance than the other framing groups, a trend also existed for the negative group to have a higher degree of inappropriate reliance and inappropriate compliance. These instances of inappropriate reliance and compliance are most dangerous when the aid causes the participant to make an error without the participant's knowledge. Since sensitivity and response bias for the negative group were not different from the other framed groups, these slightly different response patterns may have a subtle effect on performance.

Future areas of study

Attribute framing effects may become smaller or larger depending on how far the aid reliability level is from 50% or 100%. For example, framing effects may be more pronounced at high reliabilities (99% correct / 1% incorrect) versus low reliabilities (55% correct / 45% incorrect). These results would have implications for systems that have different reliabilities for signal trials and noise trials (the aid's hit rate and the correct rejection rate were identical in this study). For a given attribute frame (80% correct / 20 % incorrect), framing effects may differ in size between hit rate and correct rejection rate. As a result, it may be appropriate to include reliability information for hits/misses, but not for correct rejections/false alarms.

Conclusion

The manner in which information about the reliability of an aid is presented can subtly influence operators' utilization strategies. Such effects may have a positive impact on utilization if implemented properly. However, there may also be a negative impact on system performance if information about aid reliability is employed in a haphazard manner. Still, the slightly superior performance of the neutral group in this study (operators given both positive and negative information) suggests that providing more information about the nature of an aid's errors and reliability may be the best solution to improving automation utilization. For tasks that resemble luggage screening, the opportunity for an aid to shift the operator's response bias towards optimality may prove to be its greatest asset.

REFERENCES

Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., & Beck, H.P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58(6),* 697-718.

Dzindolet, M.T., Pierce, L.G., Beck, H.P., & Dawe, L.A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors, 44(1),* 79-94.

Lacson, F. C., Wiegmann, D.W., & Madhavan, P. (2003, August). Effects of instructional framing on trust and reliance. American Psychological Association Conference. Toronto, Canada.

Levin, I.P., Schneider, S.L., & Gaeth, G.J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behvior and Human Decision Processes, 76(2),* 149-188.

Levin, I. P., Schnittjer, S.K., & Thee, S.L. (1988). Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology, 24(6),* 520-529.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43(4),* 563-572.

Nygren, T.E. (1997). Framing of task performance strategies: Effects of performance in a multi-attribute dynamic decision making environment. *Human Factors, 39(3),* 425-437.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors, 29(6),* 695-706.

Perrin, B.M., Barnett, B.J., Walrath, L., & Grossman, J.D. (2001). Information order and outcome framing: An assessment of judgment bias in a naturalistic decision-making context. *Human Factors, 43(2),* 227-238.

Svenson, O., & Benson, L. (1993). Framing and time pressure in decision making. In O. Svenson & J. Maule (Eds.) *Time pressure and stress in human judgment and decision making,* (pp. 27-40). New York: Plenum.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and psychology of choice. *Science, 211(4481),* 453-458.