Cognitive Anchoring on Self-Generated Decisions Reduces Operator Reliance on Automated Diagnostic Aids

Poornima Madhavan and **Douglas A. Wiegmann,** University of Illinois at Urbana-Champaign

Automation users often disagree with diagnostic aids that are imperfectly reliable. The extent to which users' agreements with an aid are anchored to their personal, self-generated diagnoses was explored. Participants (N = 75) performed 200 trials in which they diagnosed pump failures using an imperfectly reliable automated aid. One group (nonforced anchor, n = 50) provided diagnoses only after consulting the aid. Another group (forced anchor, n = 25) provided diagnoses both before and after receiving feedback from the aid. Within the nonforced anchor group, participants' self-reported tendency to prediagnose system failures significantly predicted their tendency to disagree with the aid, revealing a cognitive anchoring effect. Agreement rates of participants in the forced anchor group indicated that public commitment to a diagnosis did not strengthen this effect. Potential applications include the development of methods for reducing cognitive anchoring effects and improving automation utilization in high-risk domains.

INTRODUCTION

The use of automated diagnostic aids is becoming increasingly common within a variety of complex systems, such as aviation, nuclear power, and health care. The ultimate effectiveness of these aids, however, is seldom realized because operators tend to undertrust imperfectly reliable automation, resulting in its disuse (Parasuraman & Riley, 1997; Wickens & Hollands, 2000). In the case of automated diagnostic aids, such disuse is often manifested in the form of operator disagreements with the aid, even when aided performance is statistically more accurate than unaided performance (Wiegmann, Rich, & Zhang, 2001).

Disuse of automated aids can occur even when systems are opaque (i.e., provide little or no information to the operator upon which to base a diagnosis; Wiegmann, 2002; Wiegmann et al., 2001). For example, handheld analyzers are now commonly used to help airline luggage inspectors detect the presence of hidden contraband such as narcotics and explosives in passenger luggage. Given that luggage is generally opaque, operators must rely almost entirely on the automation to perform this analysis. Although such analyzers are fairly accurate, they do have a tendency to produce false alarms. Consequently, operators generally undertrust these devices and have reportedly ignored analyzer alarms (Parasuraman, Hancock, & Olofinboba, 1997). Anecdotally, operators formulate an opinion concerning the likelihood that a particular passenger might be dangerous and then base their interpretation of the automated alarm based on this hypothesis – a phenomenon known as *cognitive anchoring.*

Cognitive Anchoring

A great deal of empirical evidence in cognitive and social psychological research suggests that when people are faced with discrepant information after choosing a certain hypothesis, they have a strong tendency to bias their belief revisions in favor of the initially chosen hypothesis (Einhorn & Hogarth, 1982; Hogarth & Einhorn, 1992; Tversky & Kahneman, 1974; Wickens & Hollands, 2000). They often behave as if they have attached a "mental anchor" to the

Address correspondence to Douglas A. Wiegmann, University of Illinois, Aviation Human Factors Division, 1 Airport Rd., Savoy, IL 61874; dwiegman@uiuc.edu. **HUMAN FACTORS,** Vol. 47, No. 2, Summer 2005, pp. 332–341.

initial hypothesis and do not easily shift it to the alternative, giving the impression that "first impressions are lasting." Likewise, operators of imperfectly reliable automation might attempt to circumvent automation errors by generating their own independent decisions about system states prior to receiving information from the diagnostic aid. Consequently, there may be a strong a tendency for operators to "cognitively anchor" their final decisions onto their initial hypotheses, thus increasing the probability that operators will disagree with an aid that provides alternative information, regardless of the objective accuracy of the aid's diagnoses.

Self-anchoring. Decision makers have an implicit tendency to spontaneously generate their own independent hypotheses about states of the world based on available raw data and to anchor their final decisions onto these initial hypotheses (see Tversky & Kahneman, 1974). However, opaque systems afford automation users little or no access to raw data on which to base diagnostic decision making. Operators' hypotheses about system states are at best intuitive guesses or "hunches." The extent to which users of these opaque systems will independently generate (or guess) their own hypotheses about system states and cognitively anchor their agreements with an automated diagnostic aid onto these initial "hunches" is still unknown.

Automation utilization strategies are often influenced by the interaction of the user's perceived reliability of the automated aid and self-biases (Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001; Lee & Moray, 1992; Riley, 1989). In general, humans tend to underestimate the reliability of imperfect automated aids (see Dzindolet, Pierce, Beck, & Dawe, 2002; Wiegmann et al., 2001). This lowered perceived reliability interacts with the operator's self-biases to lead to the choice of a particular automation utilization strategy. Operators' self-biases may interact differently with perceived reliability depending on whether these biases pertain to specific tasks or are more generic in nature. For example, in the context of opaque systems, the choice of a utilization strategy might be influenced by general biases about one's abilities to generate intuitive guesses or take risks, rather than by one's specific ability to perform the task in the presence of sufficient raw data. Consequently, the tendency to self-generate anchors is likely to be a function of both the perceived reliability of the aid and self-confidence in one's own abilities in a generic sense. A lower perceived reliability of the aid is likely to produce a greater tendency to prediagnose system states prior to receiving information from an aid and, hence, is more likely to produce cognitive anchoring. In contrast, higher estimates of self-confidence and risk-taking tendencies may also increase hypothesis generation and result in a cognitive anchoring effect.

Forced anchoring. Under some conditions, individuals might be required to publicly commit to their diagnosis prior to receiving additional information from other sources. Such might be the case with a physician speaking with a patient during an initial consultation or a luggage inspector declaring to a passenger the detection of contraband in order to justify opening the piece of baggage. Public expressions of a diagnosis can produce cognitive dissonance, or feelings of discomfort, when information that contradicts the original hypothesis is provided (Festinger, 1957). In order to reduce dissonance, an individual often becomes more convinced that his or her original choice or belief was correct and expresses a greater preference for the original belief (Oshikawa, 1970). According to Brehm and Cohen (1962), public commitment to a belief makes it extremely difficult to nullify this dissonance-arousing commitment. Consequently, procedures that require automation users to document their diagnosis prior to receiving information from the aid might produce a strong anchoring effect and subsequent disagreements with an aid when the aid provides an alternative hypothesis. This forced anchor effect may be much stronger than is the hypothesized self-anchoring effect described earlier.

Purpose of the Present Study

The purpose of the present study was manifold. The first three objectives were to examine (a) automation users' tendencies to generate their own independent hypotheses about system states given a completely opaque system, (b) their propensity to anchor their agreements with a diagnostic aid onto these self-generated hypotheses, and (c) the extent to which this self-anchoring is related to perceived aid reliability and self-confidence.

To address these issues, we gathered participants' (n = 50) self-ratings of abilities and risk-taking tendencies. The participants then performed a series of trials in which they repeatedly diagnosed the validity of pump failures using only information provided to them by an automated diagnostic aid. They provided their diagnosis of the failure only after receiving information from the aid and then rated their confidence in their diagnosis. Participants then completed a questionnaire that asked them to indicate their perceived reliability of the aid and to estimate the number of trials, if any, during which they generated a diagnosis of the pump failure prior to receiving the feedback from the aid. Based on past and recent models of automation trust (Dzindolet et al., 2001; Lee & Moray, 1992; Riley, 1989) and cognitive anchoring (Einhorn & Hogarth, 1982), we hypothesized that (a) participants would have an implicit tendency to guess or "prediagnose" system failures prior to receiving information from the diagnostic aid; (b) post hoc estimates of prediagnosis would be negatively correlated with participants' agreements with the aid, indicating a self-anchoring effect; and (c) prediagnosis estimates would be negatively correlated with participants' perceived reliability of the aid and positively correlated with their self-ratings of ability and risk-taking tendencies.

Another objective of the present study was to explore whether requiring automation users to publicly document their diagnosis prior to receiving information from the aid would produce a stronger anchoring effect than would self-generated anchors. To accomplish this objective, we had an additional group of participants (n = 25) perform the same series of trials. However, on each trial, participants in this group were required to indicate their own diagnosis before receiving the aid's diagnosis. They were then given a second opportunity to revise their diagnosis after receiving information from the aid.

Performance of participants in this forced anchor group was compared with the performance of participants in the nonforced anchor group by dividing participants in the nonforced anchor group into high (n = 25) and low (n = 25) self-anchor conditions based on their self-reported tendency to prediagnose pump failures. Based on the theories of public commitment (Brehm & Cohen, 1962) and cognitive dissonance (Festinger, 1957), we hypothesized that (a) participants in the forced anchor group would disagree with the diagnostic aid more often than participants in both the high- and low self-anchor conditions; and (b) participants in the forced anchor group would be more confident in their diagnoses and would provide lower estimates of the aid's reliability in order to rationalize their higher disagreement rates, thereby reducing their feelings of cognitive dissonance.

METHOD

Participants

A total of 75 undergraduate and graduate students from the University of Illinois at Urbana-Champaign completed all phases of the experiment. Participation was voluntary, and all participants were paid $6.00/hr for their participation. The total participation time did not exceed 1.5 hr.

Tasks and Procedures

Participants began the experiment by completing a preexperimental questionnaire that asked them to rate their problem-solving and decision-making abilities, as compared with an average college student, using a Likert scale that ranged from –4 (*worse than most*) to 0 (*about the same*) to 4 (*better than most*). Participants also rated their risk-taking tendencies as compared with other college students using a scale that ranged from –4 (*less risky than most*) to 0 (*about the same*) to 4 (*riskier than most*).

Participants then performed 200 trials of a computer simulation task that required them to repeatedly diagnose the validity of pump failures within a waste-processing facility. The simulation was developed using Visual Basic for MS-DOS and presented using a desktop computer equipped with a 22-inch (55.9-cm) color monitor and standard keyboard. The control panel in the computer simulation contained two radial-dial pressure gauges and an alarm indicator. At the onset of every trial the gauge bars appeared to move over a range of readings and the alarm light was shown as green. After a delay of 5 s, the system appeared to fail as the alarm light changed from a steady green to a flashing red. The gauges stopped moving, and the text "SYSTEM FAILURE" was displayed above the alarm light. A true system failure occurred in half of the trials (i.e., 50% true failure rate). Following the system failure, the program presented an automated diagnostic aid to help the user determine if the failure was true or false. The nonforced anchor group ($n = 50$) was instructed to consult the aid immediately upon the onset of the pump failure alarm and to provide their diagnosis of the failure only after receiving information from the aid. Activating the diagnostic aid was followed by the appearance of an interim screen with the message "Please wait," which remained on the screen for approximately 5 s, followed by the diagnostic aid's conclusion as to whether or not the pump had really failed. After receiving the aid's diagnosis, these participants input their answer using the Y key to indicate that "yes," the system had really failed, or the N key to indicate that "no," the system had not failed.

For the forced anchor group ($n = 25$), the activation of the aid was followed by the appearance of the question "While waiting for the aid's diagnosis, IN YOUR OPINION has the system really failed (y/n)?" After the participant had input a response, the screen with the message "Please wait" appeared for 5 s, followed by the diagnostic aid's conclusion. After receiving the aid's diagnosis, these participants were again prompted to input their diagnosis using the Y key to indicate that "yes," the system had really failed, or the N key to indicate that "no," the system had not really failed.

Participants in both groups indicated their confidence in their final diagnosis on a Likert scale ranging from 1 (*no confidence*) to 5 (*very confident*). Feedback was given as to whether the diagnosis was correct, and the score was updated by awarding 10 points for a correct diagnosis or subtracting 10 points for an incorrect answer. For all participants, the likelihood of the aid presenting either a correct (hit or correct rejection) or an incorrect diagnosis (miss or false alarm) was .80 and .20, respectively. Prior to testing, participants were assigned to either the nonforced or forced anchor groups using quasi-random assignment procedures (Cook & Campbell, 1979). No participants were given any information about the true (80%) reliability of the aid.

Postexperimental Questionnaire

After the completion of 200 trials, participants were administered a brief postexperimental questionnaire that required them to estimate the diagnostic aid's actual reliability using a percentage scale that ranged from 0 to 100. In addition, participants in the nonforced anchor group were asked to estimate the number of trials (0–200) during which they had mentally made their own diagnosis before receiving the aid's diagnosis.

RESULTS

Overview of Analyses

The analysis data consisted of two phases. The first phase examined the effects of self-anchoring on participants' agreements with the diagnostic aid and included data only for participants in the nonforced anchor group. The second phase examined the effects of public commitment on cognitive anchoring and included data for both the nonforced anchor and forced anchor groups. Analyses were performed using SPSS 10.0.1 for Windows, and a p value of .05 or less (two tailed) is implied when effects are reported as significant. Effect sizes for between-group comparisons are reported using either Cohen's d, with values of .20, .50, and .80 reflecting small, medium, and large effects, or Cohen's f, with values of .10, .25, and .40 reflecting small, medium, and large effects, respectively (Cohen, 1988).

Data screening. The data were examined to check for outliers and fit between data distributions and assumptions of multivariate analysis (Tabachnik & Fidell, 2001). The variables were examined separately for the nonforced anchor and forced anchor groups. There were no significant outliers in the data. "Prediagnoses estimates" were originally square root transformed to reduce the slight positive skew in the distribution of scores. However, the transformation did not significantly alter the results. Therefore, for clarity purposes, the untransformed raw data were retained for further analysis.

Phase 1: Nonforced Anchor Group

Agreement probabilities. Participants in the nonforced anchor group disagreed with the automated aid on an average of 14% of the trials, resulting in an average agreement probability of .86. Agreement probabilities varied considerably across participants, ranging from a low of .45 to a high of 1.00. Three participants never disagreed with the aid throughout the experiment. Table 1 presents the correlations between agreement rates and other key variables collected in this study for participants in the nonforced anchor group.

Post hoc estimates of prediagnosis. Participants estimated that they tended to prediagnose system failures on an average of 53, or 26.5%, of the trials. These estimates varied considerably, ranging from a low of 0 to a high of 200 (100%) trials. Eight participants reported that they did not generate any prediagnoses at all, whereas 1 participant reported generating prediagnoses on all 200 trials. Prediagnosis estimates were significantly and negatively correlated with agreement scores ($r = –.594$, $p < .001$). Thus an increase in the tendency to prediagnose system failures was associated with a significant decrease in agreements with the aid.

Perceived reliability. Participants, on average, estimated the diagnostic aid to be about 76.3% reliable. Perceived reliability was significantly positively correlated with agreement rates ($r = .295$, $p < .05$). However, reliability estimates were not significantly correlated with estimates of prediagnosis.

Individual difference measures. Self-stated ratings of risk-taking tendencies correlated significantly with both prediagnosis estimates ($r = .324$, $p < .05$) and agreement rates ($r = –.353$, $p < .05$). However, participants' ratings of decision-making and problem-solving skills did not correlate significantly with either prediagnosis estimates or agreements.

Decision confidence. Decision confidence was calculated for all 50 participants on agreement trials and for 47 participants on disagreement trials (3 out of 50 participants never disagreed with the aid). Confidence ratings on trials in which participants agreed with the aid ($M = 3.23$, $SD = 0.84$) were generally higher than on trials in which they disagreed with the aid ($M = 2.94$, $SD = .84$), $t(46) = 2.78$, $p < .01$, $d = .35$. Confidence ratings on agreement and disagreement trials, however, were not significantly correlated with agreement rates, prediagnosis estimates, or perceived reliability of the aid. Confidence ratings were positively correlated with ratings of both risk-taking tendencies and decision-making ability (see Table 1).

Predicting agreement scores. We performed multiple regression analyses to explore the contribution of participants' prediagnosis estimates in predicting their agreements with the aid. The first analysis utilized a stepwise, forward-entry regression procedure to determine which variables accounted for the greatest amount of variance in agreement scores. The predictor variables included prediagnosis estimates, perceived reliability, confidence ratings on agreement and disagreement trials, and estimates of

TABLE 1: Means, Standard Deviations, and Correlations Between Objective and Subjective Measures in the Nonforced Anchor Group ($n = 50$)

Variable	Pre-diag	Agree	Conf-agree	Conf-disag	Reliability	Risk	Prob	Dec
Mean	0 .53	0.86	3.23	2.94	76.3	4.52	1.64	1.52
(*SD*)	(0.53)	(0.15)	(0.84)	(0.84)	(16.07)	(2.2)	(1.5)	(1.7)
Pre-diag		–.594**	.101	.172	–.084	.324*	.149	.008
Agree			.012	–.123	.295*	–.353**	.067	–.083
Conf-agree				.633	.056	.253*	.071	.282*
Conf-disag					.018	.261*	.068	.237*
Reliability						–.143	–.013	–.308**
Risk							.304**	.320**
Prob								.566**
Dec.								

Note. Pre-diag = prejudgments made by the self-anchor group; Agree = agreement rates; Conf-agree = confidence on agreement trials; Conf-disag = confidence on disagreement trials; Risk = self-ratings of risk-taking tendencies; Prob = self-ratings of problem-solving abilities; Dec = self-ratings of decision-making abilities.

**$p < .01$, *$p < .05$ based on a two-tailed test.

ability and risk-taking tendencies (Table 2). Results revealed that prediagnosis estimates were the strongest predictor of agreement scores, followed by perceived aid reliability. No other variables significantly contributed to the model, which accounted for 39.8% of the variance in agreement scores.

The second regression analysis utilized a hierarchical regression approach (Cohen & Cohen, 1983) in which perceived reliability, confidence ratings, and self-ratings of ability and risk-taking tendency were entered into the equation first, prior to prediagnosis estimates. This procedure allowed for the unique contribution of prediagnosis estimates in predicting agreement scores to be assessed after accounting for variance explained by these other variables (see Table 3). Before prediagnosis estimates were entered into the equation, confidence on agreement and disagreement trials, perceived reliability, and risk-taking scores were all significant predictors of agreement, resulting in a regression model that accounted for 27.2% of the variance in agreement scores. When prediagnosis scores were entered into the equation, risk-taking scores remained a significant predictor, and problem-solving scores also emerged as a significant contributor to the prediction of agreement scores. However, prediagnosis estimates still explained the largest amount of variance in participants' agreement scores. The resultant regression model accounted for 52.1% of the variance in agreement scores.

Predicting estimates of prediagnosis. In order to determine which variables were the strongest predictors of prediagnosis estimates, we conducted sequential regression analyses by the method of backward elimination, in which perceived reliability, self-ratings of ability, and confidence ratings were entered into the equation. Consistent with the correlational analysis presented previously, these regression analyses indicated that participants' rating of their risk-taking tendencies was the only significant predictor of prediagnosis scores. The best resultant regression model accounted for only 9% of the variance in prediagnostic estimates.

Phase 2: Forced Anchor Versus Self-Anchor

In order to compare the performance of participants who reported self-generating prediagnoses with that of participants who were forced to prediagnose (publicly commit) pump failures, we performed a median split on prediagnosis estimates, dividing participants in the nonforced anchor group into low self-anchor ($n = 25$) and high self-anchor groups ($n = 25$). The mean prediagnosis score for the low and high self-anchoring groups was 15 ($SD = 15.26$) and 91 ($SD = 51.48$), respectively. Anchoring condition (forced anchor, low self-anchor, and high self-anchor) was treated as the between-groups factor in subsequent analyses.

For the purpose of analysis, disagreement rates and corresponding agreement scores for the forced anchor group were computed using only trials during which there was an opportunity for disagreement (i.e., the aid generated a diagnosis that differed from participants' preliminary diagnosis). This method differed slightly from that used to calculate agreement scores for the self-anchor groups, which involved calculating agreement across all 200 trials. Ideally, agreement scores for the self-anchor group

TABLE 2: Summary of Stepwise Forward-Entry Regression Analysis for Variables Predicting Agreement Probabilities in the Nonforced Anchor Group

	r^2	sr^2	pr^2	Significance
Step 1				
Prediagnosis	.3147	.3147	.3147	.000
$R^2 = .353$; model significance: $p < .005$				
Step 2				
Prediagnosis	.3147	.3170	.2798	.000
Reliability	.1176	.1218	.0835	.017
$R^2 = .413$; $\Delta R^2 = .060$; model significance: $p < .005$				

Note: r^2 = variance in agreement accounted for with other predictors; sr^2 = variance in agreement *uniquely* accounted for; pr^2 = percentage of variance *not* accounted for by other predictors.

TABLE 3: Summary of Hierarchical Regression Analysis for Variables Predicting Agreement Probabilities in Nonforced Anchor Group

	r^2	sr^2	pr^2	Significance
Step 1				
Conf-agree	.0042	.0806	.0640	.020
Conf-disag	.0128	.0045	.0346	.083
Reliability	.1108	.0835	.0666	.018
Problem solving	.0023	.0272	.0204	.181
Decision making	.0121	.0004	.0003	.863
Risk taking	.1030	.1218	.1011	.004
R^2 = .202; model significance: $p < .005$				
Step 2				
Conf-agree	.0007	.0376	.0188	.224
Conf-disag	.0151	.0034	.0016	.720
Reliability	.1176	.0605	.0309	.121
Problem solving	.0024	.1176	.0635	.028
Decision making	.0190	.0306	.0151	.274
Risk-taking	.1475	.1246	.0068	.024
Prediagnosis	.3147	.3226	.2275	.000
R^2 = .483; ΔR^2 = .281; model significance: $p < .001$				

Note. Conf-agree = confidence on agreement trials; Conf-disag = confidence on disagreement trials; r^2 = variance in agreement accounted for with other predictors; sr^2 = variance in agreement *uniquely* accounted for; pr^2 = percentage of variance *not* accounted for by other predictors.

should have been calculated in the same way as for the forced anchor group. However, there was no way to identify the trials in which participants in the self-anchor groups privately generated prediagnoses, so different methods were used to compute agreement scores for these groups.

Agreement probabilities. Participants in the forced anchor group disagreed with the automated aid on an average of 19% of the trials, resulting in an average agreement probability of .81 (SD = .14). Agreement scores in the forced anchor group were compared with agreement scores of participants in the low and high self-anchor groups using a one-way between-groups analysis of variance (ANOVA). Results revealed a significant effect for treatment group, $F(2, 72) = 7.92$, $p < .005$, $f = .47$. Post hoc t tests revealed that agreement scores were significantly lower in the high self-anchor (M = .78, SD = .18) than in the low self-anchor group (M = .93, SD = .08), $t(48) = 3.81$, $p < .001$, $d = 1.1$. Agreement scores in the forced anchor group (M = .81, SD = .15) were also significantly lower than agreement scores in the low self-anchor group, $t(48) = 3.76$, $p < .001$, $d = 1.09$, indicating a cognitive anchoring effect. Overall agreement probabilities did not differ significantly between the forced anchor and high self-anchor groups, $t(48) = .51$, $p = .62$, $d = .15$, indicating that public commitment did not strengthen the anchoring effect.

Perceived reliability. Results of a one-way between-groups ANOVA on reliability estimates revealed a significant difference among the reliability estimates made by the three groups, $F(2, 72) = 3.82$, $p < .05$, $f = .33$. Post hoc tests revealed that participants in the forced anchor group had significantly lower and consequently less accurate reliability estimates (M = 66.93, SD = 13.62) than did the low self-anchor (M = 76, SD = 16.55), $t(48) = 2.12$, $p < .05$, $d = .61$, and high self-anchor groups (M = 77.4, SD = 13.24), $t(48) = 2.76$, $p < .01$, $d = .8$. The low self-anchor and high self-anchor groups did not differ significantly in their reliability estimates, $t(48) = 0.33$, $p = .74$, $d = .1$.

Decision confidence. Confidence ratings were examined using a 3 × 2 mixed ANOVA in which group (low self-anchor, high self-anchor, and forced anchor) was the between-groups factor and response type (agree vs. disagree) was the within-subjects factor. The results of this analysis revealed a significant main effect of response type, $F(1, 69) = 18.32$, $p < .001$, $f = .6$, as well as for group, $F(2, 69) = 3.32$, $p < .05$,

$f = .33$. On average, confidence ratings were higher on agreement trials ($M = 3.15$, $SD = .87$) than on disagreement trials ($M = 2.81$, $SD = .88$), $t(71) = 4.24$, $p < .001$, $d = .39$. However, contrary to cognitive dissonance theory, confidence ratings in the high self-anchor group ($M = 3.31$, $SD = .73$) were significantly higher than in the forced anchor group ($M = 2.78$, $SD = .86$), $t(48) = 2.32$, $p < .05$, $d = .66$, and in the low self-anchor group ($M = 2.86$, $SD = .74$), $t(48) = 2.13$, $p < .05$, $d = .61$. Low self-anchor and forced anchor groups did not differ significantly in their confidence estimates, $t(48) = 0.35$, $p = .725$, $d = .1$.

Individual difference measures. Self-ratings of problem-solving skills, decision-making abilities, and risk-taking tendencies were analyzed using a series of one-way between-groups ANOVAs. Analysis of risk-taking ratings revealed a significant effect of experimental group, $F(2, 72) = 3.22$, $p < .05$, $f = .3$. Consistent with the correlational analyses presented previously, participants in the high self-anchor group had higher self-ratings of risk-taking tendencies ($M = 5.2$, $SD = 2.02$) than did participants in the low self-anchor group ($M = 3.84$, $SD = 2.17$), $t(48) = 2.29$, $p < .05$, $d = .65$. However, given that participants were randomly assigned to either the nonforced or forced anchor groups, neither the high nor the low self-anchor group differed significantly from the forced anchor group ($M = 4.64$, $SD = 1.44$) on self-ratings of risk-taking tendencies. Groups also did not differ significantly on ratings of problem-solving or decision-making skills.

DISCUSSION

The results of the present experiment suggest that even under conditions of extreme system opacity, automation users may tend to guess at the cause of a system failure and then anchor their agreements with a diagnostic aid onto these initial "hunches." Most participants periodically prediagnosed system failures prior to receiving a diagnosis from the aid, and a greater propensity to self-generate diagnoses was associated with significantly lower levels of agreement with the aid. Furthermore, prediagnosis estimates were found to be a better predictor of agreement rates than were all other personal and trust variables collected in this study. Together, these findings support the cognitive anchoring hypothesis that automation users who make their own independent diagnoses before receiving information from an aid are often inclined to prefer their initial diagnoses to those provided by the aid.

Based on frameworks of automation trust (see Dzindolet et al., 2001), we expected a lower level of perceived aid reliability to produce a greater tendency to prediagnose system failures, whereas higher estimates of self-abilities were predicted to increase the likelihood of hypothesis generation. However, neither estimates of the aid's reliability nor self-estimates of problem-solving ability reliably predicted participants' prediagnosis estimates. Nonetheless, participants' self-rated risk-taking tendencies were related to an increase in participants' proclivity to prediagnose system failures and subsequently disagree with the aid. Given that there was no additional information upon which to base a diagnosis, participants disagreeing with aid might be better conceptualized as "gambling" rather than problem solving (Wiegmann, 2002). Hence risk-taking tendencies were the best predictor of the tendency to prediagnose system failures. However, additional research is needed to examine participants' actual conceptualization of the task.

Public Commitment and Cognitive Anchoring

According to cognitive dissonance theory, public expressions of a belief can produce cognitive dissonance when information that contradicts the belief is provided subsequently (Festinger, 1957). In order to reduce dissonance, an individual expresses a greater preference for the original belief (Oshikawa, 1970). Therefore, requiring automation users to publicly express their diagnosis prior to receiving information from the aid should encourage cognitive anchoring. Indeed, participants who were required to indicate their diagnosis prior to consulting the aid, on average, disagreed with the aid more often than did participants who reported a below-median level tendency to self-generate their own diagnoses. However, contrary to cognitive dissonance theory (Festinger, 1957), participants who reported a propensity to self-generate prediagnoses disagreed with the aid equally as often as those who were required to publicly document their diagnoses. On one hand, perhaps the forced anchor manipulation utilized in this study was not "public" enough to create a strong dissonance effect. On the other hand, the self-anchoring effect might be much stronger than was originally expected.

Based on cognitive dissonance theory, we expected participants in the forced anchor group to be more confident in their diagnoses and to express lower estimates of the aid's reliability in order to reduce their feelings of cognitive dissonance. This hypothesis was partially supported. Participants in the forced anchor group had significantly lower estimates of the aid's reliability than did participants in the high self-anchor group. However, participants in the high self-anchor group had higher ratings of decision confidence than did participants in the forced anchor group. One possible explanation for these disparate findings is the fundamental attribution error (Harvey, Town, & Yarkin, 1981; Krull, 2001). Specifically, automation users may attribute the cause of their cognitive anchoring and agreements with an aid differently, depending on whether the reason for generating preliminary diagnoses is mandated or freely chosen. Users who freely choose to self-generate preliminary diagnoses may reduce dissonance by believing that they are more likely to be correct. The lack of a public commitment may not give them an opportunity to reassess the accuracy of their decisions, thereby inflating their confidence (Block & Harper, 1991). Conversely, users who are required via procedures to generate a preliminary diagnosis may reduce their cognitive dissonance by believing that the aid is unreliable.

Practical Implications

The results of the present study suggest that operators in opaque complex systems may sometimes guess at the cause of system states and anchor their agreements with an aid onto this initial "hunch," even if aided diagnosis is statistically more accurate than unaided diagnosis. Such a utilization strategy could have grave consequences in high-risk domains, such as aviation, defense, and health care, in which users may actively ignore information provided by automated aids in preference for their own incorrect decisions. However, the extent to which real-world users of automated aids will engage in cognitive anchoring is likely to be a function of a variety of factors, including their need to be in control of the situation, the availability of additional sources of information, and the consequences associated with correct and incorrect decisions (Dzindolet et al., 2002). The present study used a relatively contrived situation in which the consequences of wrong guesses (i.e., losing points) were not comparable to those faced by a physician, who cannot afford to misdiagnose an illness by incorrectly utilizing information from an automated aid, or by a naval commander who must accurately assess the source of a radar track before firing a missile. Therefore, additional research is needed to explore the generalizability of the present findings to more real-world scenarios before methods and procedures for ensuring appropriate automation utilization can be developed.

REFERENCES

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes, 49,* 188–207.

Brehm, J., & Cohen, A. R. (1962). *Explorations in cognitive dissonance.* New York: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). New York: Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation design and analysis issues for field settings.* Boston: Houghton Mifflin.

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,* 79–94.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology, 13,* 147–164.

Einhorn, H. J., & Hogarth, R. M. (1982). *Theory of diagnostic interference: 1. Imagination and the psychophysics of evidence* (Tech. Report No. 2). Chicago: University of Chicago, School of Business.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Harvey, J. H., Town, J. P., & Yarkin, L. K. (1981). How fundamental is the "fundamental attribution error"? *Journal of Personality and Social Psychology, 40,* 346–349.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24,* 1–55.

Krull, D. S. (2001). On partitioning the fundamental attribution error: Dispositionalism and the correspondence bias. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 211–227). Mahwah, NJ: Erlbaum.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35,* 1243–1270.

Oshikawa, S. (1970). Consumer pre-decision conflict and post-decision dissonance. *Behavioral Science, 15,* 132–140.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics, 40,* 390–399.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230–253.

Riley, V. (1989). A general model of mixed-initiative human-machine systems. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 124–128). Santa Monica, CA: Human Factors and Ergonomics Society.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wiegmann, D. A. (2002). Agreeing with automated diagnostic aids: A study of users' concurrence strategies. *Human Factors, 44,* 44–50.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2,* 352–367.

Poornima Madhavan is a postdoctoral research fellow in the Department of Social and Decision Sciences at Carnegie Mellon University. She received her Ph.D. in engineering psychology in 2005 from the University of Illinois at Urbana-Champaign.

Douglas A. Wiegmann is an associate professor of aviation human factors and associate head at the Aviation Human Factors Division within the Institute of Aviation at the University of Illinois at Urbana-Champaign, where he also holds appointments in the Department of Psychology and the Beckman Institute for Advanced Science and Technology. He received his Ph.D. in experimental psychology from Texas Christian University in 1992.

Date received: September 18, 2002
Date accepted: August 31, 2004