Association for Information Systems
AIS Electronic Library (AISeL)

SIGHCI 2009 Proceedings | Special Interest Group on Human-Computer Interaction

2009

Automated and Participative Decision Support in Computer-aided Credibility Assessment

Matthew L. Jensen
University of Oklahoma, mjensen@ou.edu

Paul Benjamin Lowry
Brigham Young University, paul.lowry.phd@gmail.com

Jeffery Jenkins
University of Arizona, jljenk@gmail.com

Automated and Participative Decision Support in Computer-aided Credibility Assessment

Matthew L. Jensen
Price College of Business
University of Oklahoma
mjensen@ou.edu

Paul Benjamin Lowry
Marriott School of Business
Brigham Young University
Paul.Lowry.PhD@gmail.com

Jeffery Jenkins
MIS Department
University of Arizona
jljenk@gmail.com

ABSTRACT

History has shown that inaccurate assessments of credibility can result in tremendous costs to businesses and society. This study uses Signal Detection Theory (SDT) to improve the accuracy of credibility assessments through combining automated and participatory decision support. Participatory decision support is also proposed to encourage acceptance of the decision aid's recommendation. A new hybrid decision aid is designed to perform automated linguistic analysis and elicit and analyze perceptual cues (i.e., indirect cues) from an observer. The results suggest that decision aids that collect both linguistic and indirect cues perform better than decision aids that collect only one type of cue. Users of systems that collect linguistic cues experience improved credibility assessment accuracy; yet, users of systems that collect both types of cues or only indirect cues do not experience higher accuracy. However, collecting indirect cues increases the user's acceptance of decision-aid recommendations.

Keywords

Credibility Assessment, Signal Detection Theory, Linguistic Analysis, Indirect Cues Elicitation, Decision Support Systems

INTRODUCTION

Credibility is very difficult for people to assess correctly in face-to-face (FtF) interactions (Bond and DePaulo, 2006). Yet, it is a critical capability that is foundational to effective communication and decision making. *Credibility* is the believability of a source due to message recipients' perceptions of the source's trustworthiness and expertise (Metzger et al., 2003). Credibility is influenced by receiver characteristics, source characteristics, message characteristics, and cognitive-processing routes (Chaiken and Maheswaran, 1994). Estimates of others' level of credibility are often misplaced. Therefore, the primary goal of assessing credibility is ensuring credibility is properly attributed—meaning credibility is given when a source's message is true and accurate.

Despite the importance of accurate credibility assessment, research has repeatedly shown that most people are overly trusting when evaluating incoming messages (Levine et al., 1999). A recent meta-analysis investigating human-assessment ability demonstrated that when people are faced with equal numbers of truthful or deceptive messages, they could distinguish truthful messages from deceptive ones at an accuracy rate of 54%, only marginally better than chance (Bond and DePaulo, 2006). To help address this issue, several researchers have tried to improve credibility assessments by using decision aids. Recently, research has investigated new unobtrusive methods of assessing credibility (Jensen et al., in press). These credibility assessments rely on observable behaviors to detect many cues that are normally difficult for humans to detect. Examples of such aids include automated language processing and analysis tools (Zhou et al., 2004a). Recommendations produced by these decision aids typically fall between 70% and 80% accuracy (Zhou et al., 2004b).

There are two limitations to unobtrusive decision aids: (1) users often do not accept a decision aid's recommendation, despite the aid's potential to improve the users' accuracy; and (2) diagnostic, perceptual measures of credibility have not been incorporated into the decision aid. Perceptual measures have been shown to improve credibility assessment accuracy (Vrij et al., 2001, Vrij et al., 2004). However, these measures are currently identifiable only by humans and thus have received little attention in designing decision aids. To address these issues, this study uses Signal Detection Theory (SDT) to design a hybrid expert system that both analyzes the structure and content of messages (i.e., direct, linguistic cues) and elicits perceptual information from an interaction observer (i.e., indirect cues).

BACKGROUND LITERATURE

Humans face a number of difficulties when attempting to assess the credibility of a source. First, in making assessments, people tend to rely on behaviors that are not diagnostic of deception (The Global Deception Research Team, 2006). Further, people typically adopt a heuristic-based approach for judging credibility. This phenomenon is termed *truth bias* (McCornack and Parks, 1986). While such heuristic labeling is done rapidly, it frequently undermines one's ability to detect possible deception.

Finally, people are limited in their information processing capabilities (Newell and Simon, 1972). Their typical focus on small subsets of non-diagnostic cues and the manifestation of biases are symptoms of these limitations.

Despite the difficulty with credibility assessment, there is reason to believe that humans can effectively contribute to a human-computer system of credibility assessment. Perceptual measures are among the strongest cues to deception (DePaulo et al., 2003) and humans are uniquely capable of evaluating them. Such perceptual cues are generated by manual behavioral coding where trained coders observe an interaction and record perceptions about what they observed. These perceptual cues are difficult to automatically approximate because they represent global assessments of a whole interaction, span multiple channels, and require semantic understanding of verbal messages. Building on the success of behavior coders' ability to identify cues highly correlated with truthful and deceptive messages, Vrij et al. developed and successfully tested methods for rapidly eliciting information from interaction observers (Vrij et al., 2001, Vrij et al., 2004). They term this elicitation the collection of "indirect cues" and hypothesize that assessing credibility via more indirect means would result in higher assessment accuracy.

Computer-Based Assessment Capabilities

A recent thrust in credibility assessment research has been the development of new, unobtrusive assessment methods based on observable behavior. These new methods are a significant departure from past attempts at machine-aided credibility assessment, which attempts have consistently targeted physiological indicators of stress and arousal. One area that has received attention during the development of unobtrusive credibility assessment decision aids is automated language processing and analysis (Zhou et al., 2004a). In most interactions, language is the mechanism through which deceptive messages are sent and received. Researchers have long sought to identify cues deceivers exhibit or strategies they use so that, when present, deception can be identified. A few manual credibility assessment methods have been developed as a result, but these methods all require trained reviewers to meticulously examine suspected statements for extended periods of time (Vrij, 2000).

There have been various attempts to construct computer-based decision aids to capture and analyze message characteristics and present recommendations concerning the credibility of the message (e.g., Zhou et al., 2004b). These aids have attempted to approximate manual credibility assessment methods in an automated setting and generally focus on categories of credibility cues such as passive voice, self-reference, negative statements, generalizations, uncertainties, temporal details, spatial details, and affective details (Zhou et al., 2004a). In contrast to more gestalt indirect cues, linguistic cues are very granular in nature (e.g., means and ratios of parts of speech) and require significant processing capability to monitor. Decision aids that utilize linguistic cues have consistently exceeded the assessment capabilities typically seen among unaided observers (e.g., Zhou et al., 2004b) and can significantly extend the capabilities of users.

THEORY AND HYPOTHESES

SDT is applied in scenarios where an individual (or group) is given a sensory stimulus and tries to discern signal from noise in the stimulus. SDT recognizes that individuals may have a difficult time discerning between a signal (e.g., deception) and noise (e.g., non-deception) that are present simultaneously in judgment tasks. SDT asserts that in every detection scenario two measurable and separate elements exist that allow individuals to discriminate between signals and noise: (1) the criterion used to make the decision as to whether a stimulus is signal or noise, and (2) sensitivity to the sensory stimulus (Green and Swets, 1966, Stanislaw and Todorov, 1999).

A key component of SDT is that individuals have a *decision variable* by which they determine whether a signal exists. Applied to our context, this decision variable would be something akin to suspicion. Each assessor would have a threshold for the decision variable, which, if exceeded, would indicate that deception is present. Clearly, the decision variable that is used, how it is measured, and the criterion that is used are perceptual and subjective.

SDT also proposes two mechanisms whereby decisions may be improved and it is by these mechanisms that credibility assessment may be enhanced by a decision aid. The first mechanism proposed by SDT to improve decisions is to create more separation between the signal distribution and the noise distribution. In our context, this is accomplished by basing the decision variable on more diagnostic cues or features. For this understanding, we must turn to research on deception and credibility. No characteristic or cue is completely diagnostic and reliable, but some are more diagnostic and reliable than others. Separation between the signal and noise distributions is accomplished by increasing the diagnostic ability of existing features or increasing the number of cues that provide unique diagnostic ability.

The second is proper placement of the criterion. It is through the placement of the criterion that biases become evident and this is especially pertinent in credibility assessment where people are generally disposed to characterize the messages that they receive as truth. A conservative criterion, prevents actual deceptive messages from being classified as deception even though the receiver has some level of suspicion. Proper placement of the criterion is accomplished by examining past values of the decision variable for occurrences of known deception and truth and then setting the criterion so that false negatives and false positives are minimized.

In this study, we use a decision aid to perform a linguistic analysis to extract cues directly from messages. Linguistic analyses have been shown to provide diagnostic cues of deception. In addition, the decision aid used in this study solicits from users and processes indirect cues based on observed behaviors. Past research has found these perceptual measures to also be diagnostic indicators of credibility. Further, the decision aid is able to properly establish its decision criterion to maximize its overall assessment accuracy based on noise and signal distributions of the decision variable.

H1: A decision aid using both linguistic analysis and indirect cue elicitation will produce recommendations that will be more accurate than the judgments of an unaided observer.

While the direct and indirect cues the decision aid uses have been shown to be diagnostic, they are collected at different levels (granular vs. gestalt) and draw on differing characteristics. The direct cues are extracted solely from the message itself; however, indirect cues may consider not only the message but also the characteristics of the message source. Thus,

H2: A decision aid using both linguistic analysis and indirect cue elicitation will produce more accurate recommendations than a decision aid using only one of these components.

Decision aids that produce recommendations based on direct cues utilize theoretically sound, diagnostic cues of credibility to augment the cognitive capacity of users. With automated analysis of linguistic cues, the decision aid can automatically extract and analyze diagnostic cues in an unbiased fashion. Unaided credibility assessors would not have the cognitive capacity to track these diagnostic cues in real time, let alone analyze them in an unbiased fashion.

H3: Use of the decision aid implementing linguistic analysis will improve an observer's assessment accuracy.

Unique to this study is the implementation of indirect cue elicitation in a decision aid. Although the danger exists that observers will perpetuate their biases and suspicions through their indirect cues scoring, indirect cue elicitation appears to be a valid method to collect diagnostic, perceptual measures from interaction observers. This is in contrast with unaided assessment, where observers are left to determine for themselves the linkage between observed cues and the level of credibility. A decision aid can elicit diagnostic, indirect cues based on perceptions of source and message cues, reliably evaluate the cues, and present the user with an interpretable recommendation.

H4: Use of the decision aid implementing indirect cue elicitation will improve an observer's assessment accuracy.

Building on previous hypotheses, using both types of cues should provide the most accurate recommendations. The recommendations produced by this aid should positively influence the user's assessment accuracy the most.

H5: Use of the decision aid that implements linguistic analysis and indirect cue elicitation will improve an observer's assessment accuracy more than use of an aid implementing only one component.

Although the automated analysis of direct cues is anticipated to increase accuracy by augmenting the user, the increase in accuracy may be partially negated due to a reluctance to accept the decision aid's recommendation. This has been a significant area of concern noted in past research on aided credibility assessment (Jensen et al., 2009). In answer to this concern, we posit that an additional benefit provided by participatory computer-aided credibility assessment is an increased likelihood that the recommendation will be accepted by the user. The method of collecting direct, linguistic cues is fully automated and does not require user oversight or allow evaluation. In contrast, the users have a very active, participatory role in providing indirect cues. The users understand where the indirect cues came from and have a basis for evaluating the cues and, by extension, the recommendation based on the cues. They may also feel some ownership in the recommendations as they were source of the cues. Therefore,

H6: Users will accept the recommendation of the decision aid more frequently when the aid contains the indirect cue elicitation component.

METHOD

A controlled laboratory experiment was conducted to test the hypotheses. Upon arriving at the lab, participants were seated at a computer and randomly assigned to one of four conditions: Unaided, indirect cues only (IDC-only), linguistic analysis only (LA-only), and both indirect cues and linguistic analysis (IDC-LA). The IDC-LA condition tests the full functionality of the human-computer assessment system. Each participant viewed an orientation video that provided a brief description of the decision aid and reported accuracy rates of past validation efforts of linguistic analysis. Following the orientation, the participants viewed 10 randomly ordered interactions. After viewing an interviewee, the participant had access to the decision aid (if applicable) and then provided a credibility assessment consisting of a judgment (guilty or not guilty of cheating), level of deception, and level of confidence).

The experiment involved 167 participants recruited from an upper-division business course at a large southwestern university. The mean age of the participants included in this study was 21.4, mean years of secondary education were 3.3, and of all the participants, 45% were female and 55% were male. The stimulus materials for this study came from a previous experiment that collected high-stakes, unsanctioned deceptive and truthful interactions during an interview (Levine et al., 2006).

ANALYSIS

The mean raw accuracy rates of the decision aid, unaided users, and the number of participants who contributed indirect cues are shown in Table 1. Using message-feature mining, the LA-only aid correctly characterized six out of the ten interviews and all participants viewed the same ten interviews. Therefore, the raw accuracy of recommendations produced by the LA-only decision aid was 60% for all participants.

Condition	*N*	Overall Accuracy of Aid Recommendation *(SD)* [%]	Hit Rate (SD) [%]	False Alarm Rate (SD) [%]
Unaided[a]	41	51.7 *(11.6)*	26.3 *(9.2)*	24.6 *(9.5)*
LA-only[b]		60.0	30.0	20.0
IDC-only	43	48.1 *(13.0)*	26.5 *(10.2)*	28.4 *(10.0)*
IDC-LA	42	62.4 *(4.8)*	32.1 *(4.7)*	19.8 *(4.7)*

Table 1. Accuracy rates of the decision aid

To compare the accuracy rates of the conditions, three *t*-tests were performed. To control for inflated type-I error, a Bonferroni correction for repeated tests was adopted. First, the accuracy rate of the human-computer system was compared to the accuracy rate of individuals in the unaided condition. In support of H1, the IDC-LA condition produced a recommendation accuracy rate that was significantly higher that the accuracy rate of unaided individuals ($t_{(81)} = 5.49$, $p < .001$). To test if the IDC-LA aid exceeded the performance of the LA-only aid, a one-sample *t*-test was performed with 60% as the value of comparison. The accuracy rate of recommendations in the IDC-LA condition exceeded the LA-only condition ($t_{(41)} = 3.186$, $p = .003$). The accuracy rate of the recommendations in the IDC-LA condition exceeded accuracy rate in the IDC-only condition ($t_{(83)} = 6.68$, $p < .001$), supporting H1.

Hypotheses 3–5 test users' assessment accuracy when using the different versions of the decision aid. To test H3, a two-way Analysis of Covariance (ANCOVA) was performed with IDC use and LA use as independent variables and accuracy as the dependent variable. No assumptions of parametric statistical tests were violated in this test. Covariates included years of secondary education, gender, and age. However, none of the covariates exerted a significant influence on assessment accuracy. Therefore, the model was reformulated to a standard two-way Analysis of Variance (ANOVA) and excluded the covariates. The raw accuracy rates are shown in Table 2. The users in the LA-only condition demonstrated improvement in assessment accuracy ($F_{(1,163)} = 7.112$, $p = .008$). This finding supports H3. In contrast, there was no significant effect on accuracy for users in the IDC-Only condition, and the interaction effect of IDCxLA was not significant. These findings fail to support H4 and H5.

Condition	*N*	Overall Accuracy of Aid Recommendation *(SD)* [%]	Hit Rate (SD) [%]	False Alarm Rate (SD) [%]
Unaided[a]	41	51.7 *(11.6)*	26.3 *(9.2)*	24.6 *(9.5)*
LA-only[b]	41	55.1 *(14.0)*	28.3 *(9.5)*	23.2 *(9.9)*
IDC-only	43	47.7 *(13.4)*	23.7 *(9.5)*	26.0 *(10.0)*
IDC-LA	42	55.2 *(13.5)*	28.8 *(9.9)*	23.6 *(9.6)*

Table 2. Accuracy rates of the decision aid users.

The acceptance of the decision aid's recommendation was tested via a one-way ANCOVA with IDC use as the independent variable and percentage of agreement as the dependent variable. Again, parametric testing assumptions were not violated and years of secondary education, gender, and age were included in the model as covariates. The versions of the decision aid that elicited indirect cues from the user had a greater number of recommendations accepted ($F_{(1,121)} = 13.49$, $p < .001$). Interestingly, younger participants seemed more likely to accept the recommendations of the decision aid ($F_{(1,121)} = 3.15$, $p = .078$).

DISCUSSION

The results suggest that a system becomes more diagnostic when both indirect cue elicitation and linguistic analysis are instantiated in the decision aid. The decision aid is more diagnostic with both components than when it has only one component (H2) and the performance of the decision aid exceeds that of the unaided observer (H1). However, only the users of the decision aid employing direct, linguistic cues showed a significant improvement over unaided users (H3). The elicitation of indirect cues alone did not improve accuracy (H4). Further, the users who were using the decision aid instantiating both linguistic analysis and indirect cue elicitation did not demonstrate a corresponding improvement in accuracy (H5). However, elicitation of indirect cues did encourage more acceptance of the decision aid's recommendations (H6). Thus, the accuracy of the decision aid improved through the consideration of direct and indirect cues, but that accuracy improvement did not transfer sufficiently to the users of the aid—the ones who are ultimately responsible for assessment.

Our contrary finding of H4 merits additional discussion. Our work exposes a potentially dangerous scenario where users are accepting the recommendations of a decision aid where the decision aid's recommendations are not improving their assessment accuracy. The reasons for the users' poor performance in utilizing indirect cues may stem from the following: deficiencies in the user and deficiencies in the system. Both potential deficiencies are discussed below.

The first possible explanation behind indirect cue failure is that the questions eliciting the indirect cues were somehow faulty or not diagnostic. This conclusion contradicts what has been shown in past research: complexity, engagement, plausibility, uncertainty, cooperativeness, anxiety, and affect have all been shown to be highly diagnostic across varying conditions.

The users may not have found the recommendation very helpful, indicating weakness in the interface design or of the content layout. However, this explanation for the poor performance is difficult to support because the recommendations were accepted by the users in the large majority of judgments.

There are a number of explanations for the poor performance of the IDC-only condition that stem from the user. First, the users may not have properly understood the questions eliciting the indirect cues. This is likely, given that there is wide variation in indirect cue scores for the same interviewee. This variation indicates potential reliability problems with participants' understanding of the questions and it is problematic because the variation in question responses results in variation in the decision aid's recommendation. The provision of explanations was an attempt at attenuating this effect by encouraging common definitions of key terms during elicitation. However, those in the IDC conditions did not view more explanations than the users in the LA-only condition.

An additional difficulty that the users faced was separating their judgments from the indirect cues that they observed. This problem was observed during the selection and pilot testing of the indirect cue items where participants would make an assessment and then ensure that all of their responses to the indirect cue questions matched their assessment. It may be unreasonable to expect that observers are capable of scoring indirect cues in an unbiased balanced fashion, when they must also provide a veracity judgment of what they observe.

Finally, the observers may not have appreciated the difficulty and level of effort required to properly assess credibility. As mentioned previously, observers easily fall into the trap of heuristic-based assessment techniques (e.g., decision rules such as "believe everyone"). However, such decision rules may be more complex and involve the cues than were elicited by the decision aid. Thus, the simple decision rules the users had may have been supplanted by other simple decision rules suggested by the decision aid.

REFERENCES

1. Bond, C. F. and DePaulo, B. M. (2006) *Personality and Social Psychology Review,* **10,** 214-234.
2. Chaiken, S. and Maheswaran, D. (1994) *Journal of Personality and Social Psychology,* **66,** 460-473.
3. DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K. and Cooper, H. (2003) *Psychological Bulletin,* **129,** 74-118.
4. Green, D. and Swets, J. (1966) *Signal Detection Theory and Psychophysics,* Wiley, New York, New York, USA.
5. Jensen, M. L., Lowry, P. B., Burgoon, J. K. and Nunamaker, J. F. (2009) In *Hawaii International Conference on System Sciences*Kona, HI.
6. Jensen, M. L., Meservy, T. O., Burgoon, J. K. and Nunamaker, J. F. (in press) In *Group Decision and Negotiation*.
7. Levine, T. R., Kim, R. K., Park, H. S. and Hughes, M. (2006) *Communication Monographs,* **73,** 243 - 260.
8. Levine, T. R., Park, H. S. and McCornack, S. A. (1999) *Communication Monographs,* **66,** 125-144.
9. McCornack, S. A. and Parks, M. R. (1986) In *Communication Yearbook*(Ed, McLaughlin, M. L.) Sage, Beverly Hills, CA.
10. Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R. and McCann, R. M. (2003) *Communication Yearbook,* **27,** 293-336.
11. Newell, A. and Simon, H. A. (1972) *Human problem solving,* Prentice-Hall, Inc., Englewood Cliffs, NJ.
12. Stanislaw, H. and Todorov, N. (1999) *Behavior Research Methods, Instruments & Computers,* **31,** 137-149.
13. The Global Deception Research Team (2006) *Journal of Cross-Cultural Psychology,* **37,** 60-74.
14. Vrij, A. (2000) *Detecting lies and deceit: The psychology of lying and implications for professional practice,* John Wiley & Sons, Chichester.
15. Vrij, A., Edward, K. and Bull, R. (2001) *Legal and Criminological Psychology,* **6,** 185-196.
16. Vrij, A., Evans, H., Akehurst, L. and Mann, S. (2004) *Applied Cognitive Psychology,* **18,** 283-296.
17. Zhou, L., Burgoon, J. K., Nunamaker, J. F. and Twitchell, D. P. (2004a) *Group Decision and Negotiation,* **13,** 81-106.
18. Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K. and Nunamaker, J. F. (2004b) *Journal of Management Information Systems,* **20,** 139-166.