December 2007

Computer-aided Credibility Assessment by Novice Lie-Catchers

Matthew Jensen
University of Arizona

Judee Burgoon
University of Arizona

Jay Nunamaker, Jr.
University of Arizona

Follow this and additional works at: http://aisel.aisnet.org/amcis2007

Recommended Citation
http://aisel.aisnet.org/amcis2007/159

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Computer-aided Credibility Assessment by Novice Lie-Catchers

Matthew L. Jensen, Judee K. Burgoon, Jay F. Nunamaker, Jr.

Center for the Management of Information
University of Arizona
mjensen@cmi.arizona.edu

Abstract

History is replete with men and women who have caused a great deal of harm and damage through skillful deception. Even though effects of deceptive actions have garnered much recent attention, deception is not a new problem. Researchers have been fascinated with deception and with credibility assessment for centuries, yet humans perform poorly when assessing credibility (Bond et al. 2006). This work presents a prototype system that unobtrusively identifies kinesic and linguistic cues that may indicate deception. This research explores improving assessment accuracy through merging improved human capabilities with system use. System use was found to significantly improve assessment ability. Training in credibility assessment was found to weakly improve assessment ability.

Introduction

Deception is commonplace in human communication. People typically lie an average of one to two times a day (DePaulo et al. 1998). While most lies are innocent in nature, some lies have enormous consequences. History is replete with men and women who have caused a great deal of harm and damage through skillful deception. Additionally, recent terrorist actions against numerous countries and the resultant efforts to increase national and international security have brought deception and its effects to the forefront of public attention.

Even though effects of recent deceptive actions have garnered much attention, deception is not a new problem. Researchers have been fascinated with deception and with credibility assessment for centuries. Perhaps one reason that deception and credibility assessment have attracted so much attention through the years is that humans are inherently poor at accurately assessing credibility. Numerous studies have noted that people typically identify deception with accuracy only slightly better than chance (approximately 54%) (Bond et al. 2006; Kraut 1980; Vrij 2000). This poor performance is not limited to laypersons, but is also found in professional lie-catchers such as police officers and federal law enforcement officers (Vrij 2000).

Coupled with poor accuracy is poor calibration of confidence in credibility assessment judgments (DePaulo et al. 1997). Confidence in judgments is of considerable importance in deception detection as it may affect the attentiveness of the lie-catcher, the lie-catcher’s verification efforts, and misallocation of time and resources as erroneous judgments are made.

To combat the effects of low accuracy and poorly calibrated confidence in credibility assessment, researchers have developed numerous tools to assist in assessment. These methods take advantage of physiological or behavioral traits that appear (or disappear) in conjunction with deception. Perhaps the most familiar tool used in credibility assessment is the polygraph. Other methods of credibility assessment include brainwave analysis (Farwell et al. 1991), functional magnetic resonance scanning (fMRI) (Ganis et al. 2003), thermal imaging (Pavlidis et al. 2002), linguistic analysis (Zhou et al. 2004), statement validity assessment (Vrij 2000), and behavioral analysis (Meservy et al. 2005a; Meservy et al. 2005b).

While many methods exist to differentiate deception from truth, all methods are tied together by one trait: they rely on a human operator to make the final judgment. As Burgoon and colleagues (2005) stated, assessing credibility “defies a
fully automated solution. A more promising approach is to integrate improved human efforts with automated tools..., the end goal being a system that singles out individuals for further scrutiny in a manner that reduces false positives and false negatives.”

Computer-based tools have a number of characteristics that are important in credibility assessment. They can be ever-vigilant, handle repetitive and complex tasks, and are easily calibrated (Fitts 1951). However, tools require a human operator to control what is analyzed, determine what is relevant, and interpret the results. Humans have abilities that are important in credibility assessment that are extremely difficult for computer-based tools to reproduce. Humans excel at improvising and are flexible, they recall relevant facts at appropriate times, and are proficient in judgment (Fitts 1951). However, humans suffer from decreasing vigilance over time, cognitive laziness (Fiske et al. 1991), and a bias-based assessment approach (Levine et al. 1999). A marriage of human and computer-based tool abilities has been touted as one solution to the credibility assessment problem (Meservy et al. 2005b; Zhou et al. 2004). More deceptive cues can be monitored by a human using a computer-based tool, as shown in Figure 1. This coverage may provide greater ability to accurately assess credibility and properly calibrate confidence in that assessment.

The purpose of this study is to explore the accuracy and confidence levels of a human-computer system of credibility assessment. The effects of the credibility assessment tool will be explored by recording the human-computer relationship through repeated assessment tasks and by manipulating access to an experimental prototype. Training the users receive will also be manipulated.

**Literature review**

Human ability in credibility assessment is poor; empirical studies suggest that most humans identify deception about 54% of the time (Bond et al. 2006; Kraut 1980; Vrij 2000). This accuracy level is only slightly better than flipping an unbiased coin to determine deception or truth. Even more troubling is that this poor accuracy rate is not limited to untrained, non-professional lie-catchers. Groups such as state and federal law enforcement officers also succumb to poor accuracy in deception detection (Vrij 2000).

Several possible explanations exist for this consistent inaccuracy in judging deception. One is the well-documented phenomenon called the truth bias wherein people routinely assume that the messages they are receiving from others are truthful (McCornack et al. 1986). Some researchers believe that the truth bias is a manifestation of various decision heuristics (Levine et al. 1999). In daily interactions, the vast majority of human communication is truthful. Thus, to save cognitive
effort, incoming communication is regarded as truthful. However, in situations where humans are exposed to large numbers of lies, a contrasting bias has been observed. This bias has been termed the lie-bias or Othello bias (Ekman 1985). Under this bias, truthful communication is judged to be deceptive; that is, there is an overestimation of others’ deceit.

Another explanation for poor judgment accuracy is mistaken reliance on behaviors or cues that do not distinguish between truth and deception (Vrij 2000). Most people believe that behaviors can differentiate truth-tellers from deceivers; however, they rely on behaviors such as gaze aversion which have been shown to be unreliable when separating truth from deception (Vrij 2000). Training has been shown to have a slight effect on detection accuracy; however, total accuracy rates still hover around chance and rarely surpass 65% (Levine et al. 2005; Vrij 2000).

In addition to poor accuracy, humans seem to have a high confidence level concerning their ability to detect deception (DePaulo et al. 1997). Overconfidence is frequently encountered when studying human decision making under uncertainty (Russo et al. 1992; Russo et al. 2002). However, the level of confidence is of considerable importance in deception detection as it directly affects the attentiveness of the lie-catcher, the lie-catcher’s verification efforts, and misallocation of time and resources as erroneous judgments are made. Confidence levels vary a great deal among untrained, non-professional lie-catchers (e.g. student experimental subjects) (DePaulo et al. 1997). The confidence level increases dramatically for those with some training in deception detection or who are employed to identify deceit (Kassin et al. 1999; Vrij 2000). In complex decision making aside from deception detection, extensive experience and training calibrated confidence in performance (Kruger et al. 1999); however, the exact relationship between deception detection accuracy and confidence is still under examination. Research has shown that there is very little correlation between deception detection ability and confidence in that ability (DePaulo et al. 1997).

**Computer-based methods for deception detection**

Researchers at the University of Arizona and Rutgers University are developing computer-based methods of deception detection which unobtrusively analyze kinesic and linguistic behavior in search of deceptive cues. The methods analyze the movements and linguistic properties of communication from one person engaged in a recorded face-to-face interaction. The methods utilize supervised learning techniques from manually prepared training sets to detect patterns in the kinesic and linguistic channels. Abbreviated descriptions of the kinesic and linguistic analysis methods are shared below.

**Kinesics**

Empirical evidence suggests that deceivers’ head and hands move differently than truth-tellers’. Two recent meta-analyses conclude that there is a significant decrease in the amount of illustrating deceivers do in comparison to truth-tellers (DePaulo et al. 2003; Vrij 2000). Illustrating gestures are those gestures which normally accompany speech. They can include Iconics, Metaphorics, Beats, and Cohesives in the McNeill (1992) classification. Illustrating gestures can represent semantic content in speech, can emphasize certain points, or can designate a relationship between ideas in speech. For example, an illustrator may be seen when one individual is giving directions to another and demonstrates a left turn with her hand. This gesture would represent the semantic content of the phrase “go left.”

DePaulo et al. (2003) also point out that deceivers display significantly more chin raises than truth-tellers. They also observed less but non-significant undifferentiated head movement in deceivers; however, Buller et al. (1994) found that deceivers show significantly less total head movement than truth-tellers.

To make use of previously noted behavioral cues to deception, kinesic analysis seeks to automatically detect these cues. Kinesic analysis utilizes a tracking method developed by Computational Biomedicine Imaging and Modeling Center (CBIM) at Rutgers University (Lu et al. 2005). The method extracts hand and face regions using the color distribution from a digital image sequence. A three-dimensional look-up-table (3-D LUT) is prepared to set the color distribution of the face and hands. This 3-D LUT is created in advance of any tracking using skin color samples. After extracting the hand and face regions from an image sequence, the system computes elliptical “blobs” identifying candidates for the face and hands. The 3-D LUT may incorrectly identify candidate regions which are similar to skin color; however, these candidates are disregarded through fine segmentation and comparing the subspaces of the face and hand candidates. Thus, the most face-like and hand-like regions in a video sequence are identified. From the blobs, the left hand, right hand and face can be tracked continuously. A complete technical description of the kinesic analysis is beyond the scope of this paper; however, the interested reader is directed to (Lu et al. 2005). A single frame of a video segment subjected to kinesic analysis is shown in Figure 2. Sample features that are tracked through time are shown in Figure 3. Kinesic analysis calculates over 100 features (in addition to the features shown in Figure 3). Additional examples of features include distances between the head and hands, amount of time the hands are in an approximation of the torso region, and angle of declination of the head. A full listing of the features used in kinesic analysis is available (Meservy et al. 2005a). Kinesic analysis summarizes the features available through blob tracking and then uses the summaries for classification of the interaction.
Linguistic analysis is capable of analyzing linguistic features of interactions in search of cues that may indicate deception. These features are derived from transcripts of an interaction and are created using a method called message feature mining. Message feature mining is a method for classifying messages as deceptive or truthful based on content-independent message features (Zhou et al. 2003).

Empirical evidence suggests that deceivers may use language differently than truth-tellers. Deceivers have been shown to have shorter talk time, share fewer details, and display elevated uncertainty (DePaulo et al. 2003). Deceivers may also demonstrate less expressivity, less language diversity, and less complexity (Zhou et al. 2004; Zhou et al. 2003). Numerous linguistic features may be used to gauge the characteristics of messages. For example, both word counts and verb counts may be used to judge message lengths and average sentence lengths and average word lengths may be used to judge language complexity. A complete listing of relevant linguistic features useful in deception detection can be found in (Zhou et al. 2003).

Feature extraction is accomplished by first transcribing an interaction of interest. Automatic transcription is not current part of linguistic analysis; however such functionality is technologically feasible. Features are extracted from the transcription via a part of speech tagger (POS). Currently the method utilizes the General Architecture for Text Engineering (Cunningham 2002) for text parsing and the Whissell dictionary (Whissell et al. 1986) for categorizing affective words. After extraction, the features are used to classify the message as truthful or deceptive.
Reliability of kinesic and linguistic methods

The reliability of the kinesic and the linguistic methods of credibility assessment is currently being explored. Various experiments have shown that reliability rates vary between 60-90% (Burgoon et al. 2007; Meservy et al. 2005b; Qin et al. 2004) and field tests are currently ongoing. The variation in reliability may be the result of a host of influential factors including: environmental constraints, the ability to track human movement, variation of human behavior during various interactions, motivation of liars to succeed, and cultural and social factors that may influence behavior. Therefore, researchers expect the reliability of the methods to change as they are used in different environments (Swets 1986).

Monitoring behavior and linguistic properties has many practical and operational advantages. First, the monitoring can take place unobtrusively and without the knowledge of the person being examined via a video camera and a microphone (Meservy et al. 2005b). This is in contrast to many other forms of deception detection that require the use of sensors attached to the body (e.g., polygraph). Behavior monitoring has been shown to retain its accuracy even when the video frame rate falls (Meservy et al. 2006). Further, behavior monitoring and linguistic analysis can be easily merged with other methods of deception detection for increased accuracy. Specifically, voice analysis and thermal imaging methods might be used in conjunction with kinesic and linguistic methods and the combined system would still retain its unobtrusive qualities.

Training in credibility assessment

There is currently debate concerning the effectiveness of training in credibility assessment. Some researchers have found that training has had a mixed effect on credibility assessment capability. However, much of the uncertainty about the effects of training can be attributed to inconsistent and poorly constructed studies (Frank et al. 2003). More recent research indicates that training has been shown to have a positive, significant effect on credibility assessment (Levine et al. 2005; Vrij 2000). Additionally, computer-based training in deception detection has also been shown to be effective in improving credibility assessment (George et al. 2004). A recent meta-analysis of numerous studies involving deception training found that training had a positive significant effect when groups with training were compared to groups without training (Frank et al. 2003). Some researchers also believe that current research in lie-detection actually underestimates the ability of human lie-detectors (Frank et al. 2003).

Hypotheses

To examine the effects of joining improved human detection ability with a computer-aid, a set of hypotheses was developed to be tested experimentally. An experimental prototype was developed using kinesic and linguistic methods of deception detection that were previously described. Although validation studies are currently ongoing, the kinesic and linguistic methods have demonstrated accuracy rates from 60%-90%. This range is higher than the accuracy typically observed in unaided humans (54%). Therefore, those who use the experimental prototype are predicted to demonstrate higher accuracy in credibility assessment than unaided humans.

H1: Novice lie-catchers using the system demonstrate higher accuracy than unaided novices.

An effective system of credibility assessment also requires proper aligning of assessment confidence with assessment accuracy. A computer-based tool is not subject to many of the biases that hamper the ability of humans in detecting deception. The kinesic and linguistic methods present empirically-based credibility assessments and the rationale behind those assessments. Any human user is able to review the reasons behind the assessments and then reconcile the reasoning with his or her observations. The reconciliation should induce a more critical consideration of the basis for each assessment and is predicted to align the level of confidence with the accuracy level.

H2: Novice lie-catchers using the system show more alignment between assessment accuracy and assessment confidence than unaided novices.

To further improve accuracy in credibility assessment, training of the human user will be introduced. Within the context of computer-aided credibility assessment, training has two possible benefits. First, training in credibility assessment should assist the human user in correctly identifying suspicious behaviors. Second, the training should improve the ability of the human user to correctly interpret and weigh the results of the experimental prototype.

H3: Trained novice lie-catchers demonstrate higher accuracy in credibility assessment than untrained novices.
H4: Trained novice lie-catchers demonstrate more alignment between assessment accuracy and assessment confidence than untrained novices.

Description of methodology

To examine the effects of an experimental prototype and training in credibility assessment, an experimental study was performed. This study compared the performance of novice lie-catchers in a 2x2 design where training and system use were manipulated. The participants judged 10 interviews to determine if the interviewee was being truthful during the interview. Table I shows the treatment matrix and the number of experiment participants in each treatment.

<table>
<thead>
<tr>
<th>Novice Lie-catchers</th>
<th>No Training</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Use</td>
<td>N = 31</td>
<td>N = 31</td>
</tr>
<tr>
<td>Use</td>
<td>N = 30</td>
<td>N = 31</td>
</tr>
</tbody>
</table>

Table I Experiment design and sample sizes

Participants

Novice lie-catchers (N = 123) were selected from an upper-division Management Information Systems (MIS) courses at a large Southwestern university. The MIS course was required for all business majors (accounting, business economics, business management, entrepreneurship, finance, management information systems, marketing, and operations management) and public administration majors (criminal justice, human and health services, public management and policy). The participants were recruited to participate in the experiment by public announcements during class. The participants were awarded course credit for participation. In addition, the top 10% of in each experimental treatment was awarded a cash reward in the amount of $10 (USD). The mean age of the participants was 22, mean years of secondary education was 3.8, and of all the participants, 50.4% were female.

Experimental Prototype

The prototype in this experiment was an interface that presented the results from linguistic and kineic analysis. The experimental prototype was trained on representative interviews and then tested on the 10 interviews the participants saw. The system correctly classified eight out of the 10 interviews by averaging the scores from linguistic and kineic analysis. The kineic and linguistic analyses were conducted using feature extraction methods described previously and synchronous-enter logistic regression for classification. A single deceptive interview and a single truthful interview were misclassified by the experimental system.

The interface displayed a video-taped dyadic interview where only the interviewee was fully visible. For each interview, the interface required two sets of three measures: judgment of guilt or innocence, level of deception, and level of confidence. The interface appeared as shown in Figure 4.
Figure 4 Experimental interface

The order of interaction between the interface and user proceeded in the following order:
1. The user viewed an interview.
2. The user submitted his or her initial judgment.
3. The prototype revealed its results which included a level of deception, judgment, and confidence level along with explanations behind each of those conclusions.
4. The user submitted a final judgment.

The experimental prototype contained natural language explanations of the cues that were included in the kinesic and linguistic analyses. These explanations were manually generated but followed a consistent format and could feasibly be automatically generated in future prototypes. Following research practice in expert and knowledge based systems (Arnold et al. 2006), an expert in deception detection reviewed the explanations provided by the system for clarity and correctness.

Stimulus materials

The stimulus materials for this study were provided by a previous experiment that was conducted at a large Midwestern university. The original purpose of the experiment was to examine the behavior of deceivers under high stakes (Levine et al. 2006). The stimulus materials consisted of 10 interviews. Five interviews were deceptive and five interviews were truthful.

Method of collection of the stimulus materials

Undergraduate students from an introductory communication course were invited to participate in a study for course credit. The participants were informed that the study concerned effective teamwork and deception was not mentioned in any experimental instructions. The participants were informed that they would be working in pairs to answer difficult trivia
questions and they were promised a large cash reward if they performed well on the trivia questions. Each participant was paired with a confederate and an experimenter entered the room and asked a number of obscure trivia questions. After a few of questions, the experimenter was called out of the room and left the set of trivia questions and answers in the room with the participant and confederate. The confederate then encouraged the participant to cheat and look at the answers. The participants self-selected their treatment group by either observing the answers or refusing to observe the answers. After a few minutes, the experimenter returned and finished asking the trivia questions. After all of the trivia questions were complete, each participant was brought to an interview area where he or she was interviewed. The participants were told that the interview would concern the role of teamwork in responding to the trivia questions. Instead, the interviewer confronted the participants with a structured interview to find out if they had cheated. All participants were interviewed by a single interviewer and the interviewer posed the same questions to all participants.

**Manipulations**

*Use* was operationalized by availability of the prototype. The *use* group had access to the system (See Figure 4) and the *nonuse* group did not. For the nonuse group, the interface excluded the results of the prototype and did not require an initial judgment. The participants in this condition only submitted one set of judgments including an indication guilt or innocence, level of deception, and confidence in their assessment.

The training manipulation occurred in two conditions: *training* and *no training*. Participants in the training condition participated in computer-based training that was designed for use within government agencies. The training focused on deceptive cues that are amenable to analysis by humans. The training covered behavioral cues in five broad categories: arousal, emotion, cognitive effort, memory, and communicator tactics. Those in the training condition watched a lecture-style video with embedded examples. Following the video lecture, the participants took a knowledge quiz that emphasized key concepts in the training. After completion of the knowledge quiz, the correct answers were revealed so the participants could identify any mistakes. Participants also took a judgment quiz where they viewed five video clips and had to determine if the person they were viewing was being deceptive. After answering, the participants were shown the correct answer along with explanations about deceptive or truthful cues that were exhibited.

**Analysis**

Multiple judgments from each participant necessitated the use of repeated measures MANOVA in a mixed two factor design. Repeated measures MANOVA has the benefit of greater power than normal MANOVA as the error is reduced by the inclusion of multiple measurements (Keppel et al. 2004).

To address H1 and H3, repeated measures MANOVA was performed on the number of correct final judgments for each condition. The mean accuracy rates by condition are shown in Table II.

<table>
<thead>
<tr>
<th></th>
<th>Novice Lie-catchers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Training</td>
</tr>
<tr>
<td>No Use</td>
<td>45%</td>
</tr>
<tr>
<td>Use</td>
<td>59%</td>
</tr>
</tbody>
</table>

**Table II Mean accuracy rates**

Tests of between-subject effects shows that novice participants who had access to the experimental system performed significantly better than those who do not have access to the system, $F(1, 119) = 23.451, p < .001$. This finding supports H1.

In contrast, there is no significant difference between subjects effect for training, $F(1, 119) = 1.313, p = .254$. This finding supposes that novice participants who had access to training did not perform better than those participants who did not have training. This finding does not support H3. The interaction effect of system x training on accuracy was not significant.

To measure the level of alignment between judgment accuracy and judgment confidence, a measure called the mean probability score was used. The mean probability score (MPS) has been used in past judgment research as measure of alignment between people’s probability estimates and actual occurrences of events (Yates 1990). The MPS is calculated by the squared difference between an individual’s probability estimate of an event occurring and its actual occurrence. The MPS produces a score between 0 and 1, with 0 being perfectly aligned. This measure was adapted from Yates (1990) and used with the measures of confidence the participants provided with every final judgment. The mean MPSs are shown in Table III.
Novice Lie-catchers

<table>
<thead>
<tr>
<th></th>
<th>No Training</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Use</td>
<td>.408</td>
<td>.368</td>
</tr>
<tr>
<td>Use</td>
<td>.288</td>
<td>.269</td>
</tr>
</tbody>
</table>

Table III Mean MPSs

Novice participants who have access to the experimental system demonstrated significantly better alignment between judgment accuracy and judgment confidence than those who do not have access to the system, $F(1, 119) = 37.976, p < .001$. This finding supports H2.

Further, there is some evidence in support of training affecting alignment between judgment accuracy and judgment confidence, $F(1, 119) = 2.786, p = .098$. This finding weakly supports H4. The interaction effect of system x training on alignment was not significant. A summary of the hypotheses and findings is shown in Table IV.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: BAS improves judgment accuracy</td>
<td>Supported</td>
</tr>
<tr>
<td>H2: BAS improves alignment between confidence and accuracy</td>
<td>Supported</td>
</tr>
<tr>
<td>H3: Training improves judgment accuracy</td>
<td>Not Supported</td>
</tr>
<tr>
<td>H4: Training improves alignment between confidence and accuracy</td>
<td>Weakly Supported</td>
</tr>
</tbody>
</table>

Table IV Summary of hypotheses and findings

Discussion

Participants without access to the experimental prototype and without training in deception detection clearly demonstrated the problems that many who assess credibility face: their accuracy in detecting deception is low and their confidence in their performance is unjustifiably high. These experimental participants would have been better off flipping an unbiased coin to determine guilt or innocence, and their mean MPS would be classified as poor at best (Yates 1990).

Training does not improve judgment accuracy and may slightly improve alignment between confidence and judgment accuracy. The effects of training were more pronounced when used without the experimental system. At this juncture, it is important to note a limitation of this study: the subjects had limited training in deception detection. The training session lasted 22 minutes with approximately 10 minutes of practice. Past research has called into question instruction on credibility assessments in brief sessions before detection tasks (Frank et al. 2003).

Clearly, training used in this experiment was not meant to be an exhaustive course on deception; however the training contained valuable information that was useful in the deception task. The deceptive interviewees demonstrated many of the cues that were described in the training, but the novice lie-detectors did not catch these cues. Therefore, we do not discount the importance of training in deception detection, but rather acknowledge potential weakness in training methods. This conclusion is in line with conclusions reached by other researchers who investigate training in deception detection. Frank and Feeley (2003) have specified six criteria (Table V) that need to be fulfilled if training for a valid investigation of training in a credibility assessment task.

| Relevance – Training must be relevant to what subjects usually encounter |
| High Stakes – Lies must be consequential |
| Proper Training – Material must be scientifically based and effectively taught |
| Proper Testing – Proper assessment of training effects |
| Generalizable across situations – Training must be applicable to other situations |
| Generalizable across time – Training must be applicable to other times |

Table V Effective training criteria

Every effort has been made to conform to these standards of effective training; however, one may argue that the stakes of the lies observed were low and that the training was not properly conducted. Possible improvements in the training may include live lecture (with question and answer opportunities), more examples of deceptive and truthful interactions under high stakes, and greater repetition of key concepts (perhaps with multiple lectures).

System use produced the most robust findings with significant improvements in both accuracy and alignment between judgment accuracy and confidence. Interestingly, the system alone correctly classified 8 out of the 10 interviewees correctly, yet a novice using the system correctly classified only approximately 6 out of the 10 interviews correctly. This
finding underscores the difficulty that humans face when detecting deception and may invite thoughts of a completely automated detection system.

We do not condone the total exclusion of the human from the detection process, as any assessment system will have clear limitations. However, the benefits of coupling skilled human detectors with computer aids may only come as human responsibilities are clearly delineated and the functionality of the system is reliably understood.

Conclusion

This paper presents findings from the first attempt to study computer-aided deception detection that joined improved human capability and with an automated, unobtrusive detection system based on linguistic and kinesic analysis. It demonstrates the improvement in deception detection accuracy and proper calibration between judgment accuracy and judgment confidence that is attainable through the introduction of a system, yet many questions remain. Interesting future steps include analyzing the effects such a system may have on professional lie-catchers such as law enforcement officers and the effects improved training may have on detection accuracy.

References


Fitts, P.M. "Human engineering for an effective air navigation and traffic control system," Ohio State University Research Foundation Report, Columbus, OH.


