Technology Dominance in Complex Decision Making: The Case of Aided Credibility Assessment

Article *in* Journal of Management Information Systems · July 2010

DOI: 10.2753/MIS0742-1222270108 · Source: DBLP

Paul Benjamin Lowry
Virginia Polytechnic Institute and State University

Judee K Burgoon
The University of Arizona

Jay F. Nunamaker Jr
The University of Arizona

Technology Dominance in Complex Decision Making: The Case of Aided Credibility Assessment

Matthew L. Jensen

Price College of Business

University of Oklahoma

mjensen@ou.edu

Paul Benjamin Lowry

Information Systems Department

Brigham Young University

Paul.Lowry.PhD@gmail.com

Judee K. Burgoon

Center for the Management of Information Systems

University of Arizona

jburgoon@arizona.edu

Jay F. Nunamaker, Jr.

Center for the Management of Information Systems

University of Arizona

nunamaker@arizona.edu

ABSTRACT: Decision aids have long been an important source of help in making structured decisions. However, decision support for more complex problems has been much more difficult to create. Decision aids are now being developed for very complex problems, and their effects among low and high task-knowledge individuals are still being explored. One such task is credibility assessment in which message recipients or observers must determine a message's veracity and trustworthiness. Credibility assessment is made difficult by lack of constraints, hidden or incomplete information, and mistaken beliefs of the assessor.

The Theory of Technology Dominance (TTD) proposes that technology is most effectively applied in intelligent decision aids when an experienced user is paired with a sophisticated decision aid. This work examines TTD in the complex task of credibility assessment. To assist in credibility assessment, we created a decision aid that augments the capabilities of the user—whether novice or professional. Using hypotheses based on TTD, the decision aid was tested using high-stakes deception in recorded interviews and involved both student (novice) and law enforcement (professional) users. Both professionals and novices improved their assessment accuracy by using the decision aid. Consistent with TTD, novices were more reliant on the decision aid than were professionals. However, contrary to TTD, there was no significant difference in the way novices and professionals interacted with the system, and the decision aid was not more beneficial to professionals. Novices and professionals frequently discounted the aid's recommendations, and in many cases professionals did not view explanations when the decision aid contradicted their assessments. Potential reasons for these findings, as well as limitations and future research opportunities, are discussed.

KEY WORDS: decision making, decision aids, credibility, credibility assessment, deception, deception detection, Theory of Technology Dominance (TTD)

INTRODUCTION

Supporting complex decision making has been one of the primary goals of decision aid designers for many years. It has long been acknowledged that humans have limited cognitive capabilities [51, 60]. Decision aid designers have attempted to extend and augment the capabilities of humans through assisted decision making and automated support to make solutions to complicated or tedious problems attainable [26]. Examples of such decision aids are common among past and present knowledge-based and decision-support systems research; examples include those of medical diagnosis [11], piloting an aircraft and supervising a nuclear power plant [50], auditing [9], and filing tax returns [52]. Such decision aids provide many benefits in addressing problems that are extremely difficult or tedious for people to solve on their own. The application of decision support systems is expanding further, as aids are now being developed for very complex problems.

Other recent examples include decisions to address problems that were previously out of reach. These include AI-based recommender systems [e.g., 28, 70, 73], which are being used for many innovative applications such as helping low-income families with school choice [70], increasing consumer online purchases [32], recommending movies [73], advanced end-user assistance in writing powerful database queries [49], and so on. Other approaches to help users with decisions have included making the presentation and summary of data more simple, such as the case with the use of “executive dashboards,” which can help executive and management users to more quickly identify trends and problematic enterprise exceptions and to better allocate resources [5, 57]. Despite the tremendous progress and wide application of intelligent decision aids, their effects among low and high task-knowledge users are yet to be fully explored.

One area of computer-aided decision making that is receiving a significant amount of attention is credibility assessment. Credibility assessment occurs when interaction participants or observers determine the veracity and trustworthiness of messages exchanged during an interaction. The ability to properly assess credibility is critical for a whole host of meaningful interactions. However, unaided credibility assessment is a very complex task that is particularly difficult for humans to perform because of the lack of constraints, hidden or incomplete information, and mistaken beliefs of the assessor. Empirical studies suggest that without the use of a decision aid, most people correctly identify deception about 54% of the time when deception and truth are equally likely [6, 37, 67]. This accuracy level is only slightly, although statistically better than flipping an unbiased coin to determine deception or truth [6]. When base rates are altered to reflect fewer deceivers and more truth tellers, accuracy in deception detection falls even lower [40]. Moreover, this poor accuracy rate is not limited to untrained, nonprofessional individuals. Groups such as state and federal law enforcement officers also exhibit poor accuracy during unaided credibility assessment [67].

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 [27]. In daily interactions, the majority of human communication is truthful, thus all incoming communication is initially labeled as true. Some researchers believe that estimates of credibility assessment ability are actually overstated because people identify truth correctly much more often than deception [41]. 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* and can be observed in law enforcement environments [24]. Under this bias, incoming messages are assigned the initial label of deceptive, which results in an overestimation of others' deceit.

Another explanation for poor judgment accuracy is mistaken reliance on nonverbal behaviors that do not distinguish between truth and deception [67]. Although most people believe there are telltale behaviors that differentiate truth tellers from deceivers, they commonly rely on stereotypical behaviors such as gaze aversion [64] that are unreliable when separating truth from deception [20, 67]. Training has been shown to produce a slight but significant improvement in detection accuracy [29]; however, accuracy rates still hover around chance and rarely surpass a raw accuracy rate of 65% [39, 67].

Compounding the problem of poor accuracy is the potential for people to feel overconfident in their ability to assess credibility. A high level of confidence would be desirable if coupled with high accuracy; however, most people are poorly calibrated in what they think they detect and what they actually detect [18, 42]. The level of confidence is of considerable importance in credibility assessment because it directly affects the attentiveness of the assessor, the assessor's verification efforts, and misallocation of time and resources as erroneous judgments are made [67].

To counter many of the difficulties associated with unaided credibility assessment, researchers are developing computer-based methods of assessment that unobtrusively analyze observable behaviors for signs of deceit [33, 48, 74]. Such a decision support system may provide a number of beneficial functions in credibility assessment that extend users' capabilities, yet the effects of a decision aid on individuals with high and low task knowledge are unknown. To address this knowledge gap, we built a prototype decision aid to assist with the complex task of credibility assessment, with the intent of improving assessments made by both novice and experienced lie-catchers. For theoretical grounding, we turned to the Theory of Technology Dominance (TTD) first proposed by Arnold and Sutton [4] as a basis for investigating the effects such a decision aid may have on novice and professional users.

After describing the issues and difficulties individuals face during credibility assessment, this paper describes TTD and proposes hypotheses about the differences and similarities in the judgment accuracy of professionals and novices who use a decision aid during aided credibility assessment. Next, we describe a study wherein professionals and novices assess the credibility of interviewees in a high-stakes, unsanctioned deception situation. Finally the results are discussed and potential limitations of the study are outlined.

BACKGROUND

This section summarizes past research pertaining to aided credibility assessment by professionals and novices. We first explain the complexities of credibility assessment and how expertise itself does not always make a significant difference in assessment accuracy. These issues then point to our explanation of how decision aids can help in this form of complex decision making.

Complexity of Credibility Assessment

Several attributes characterize complex decisions that take place in natural environments. Orasanu and Connolly [53] suggest several characteristics of problems that significantly impact the problems' complexity but are routinely ignored in decision-making research. Table 1 lists these characteristics along with explanations for each.

Table 1. Characteristics of Problems that Increase Complexity

Characteristic	Description
Ill-structured problems	No single accepted procedure to use in solving the problem Many open constraints
Uncertain dynamic environments	Incomplete and imperfect information Environmental changes during the decision window
Shifting, ill-defined, competing goals	Goals may change during the decision window Goals may be in conflict with each other Goals may increase or decrease in importance and need attention during the decision window
Action/feedback loops	Decisions concern a series of events Series of events covers information gathering and incremental decision making Difficulty attributing effect to the cause
Time stress	Frequent high time stress Loss of vigilance Use of heuristic decision making
High stakes	Possible substantial losses from poor performance
Multiple players	Multiple players each have their own goals and interests that must be considered Distributed decision making requires coordination and cooperation
Organizational goals and norms	Guidance and general goals applied by the organization Individual actions take place within overarching organizational contexts

Complex problems are frequently characterized by a lack of structure, or ill structure. Reitman [55] described ill-structured problems as problems that have many unspecified constraints. A problem's constraints assist the decision maker in understanding what an acceptable solution to the problem might be and how the decision maker might arrive at that solution. Thus, ill-structured decisions can be better solved as constraints are introduced. Simon [61] believed that the boundary between well-structured and ill-structured problems is fluid and not subject to formalization. He listed several characteristics that describe well-structured problems such as an ability to evaluate a proposed solution and the ability to represent an initial problem state, the goal state, and other states that may be reached or considered. But, he argued that many problems that might be considered well-structured have ill-structured components to them and vice versa.

In the case of credibility assessment, there are many elements of assessment that lack structure and, thus, make credibility assessment a complex task. First and foremost, there are a large number of behaviors people can potentially focus on during credibility assessment. Yet, despite popular opinion to the contrary, there are no “silver bullets” in detecting deception. There are no completely diagnostic behaviors that one can observe to form consistently accurate credibility assessments [67]. Many behaviors have been correlated to deception, but few have been shown to be diagnostic across studies [20, 63]. As DePaulo et al. [20] conclude, “behavioral cues that are discernible by human perceivers are associated with deceit only probabilistically” (p. 106). Further, research is still probing many important moderators that play a critical role in how deceptive behavioral displays manifest; such moderators include identity relevance, motivation of deceivers, ability to plan deception, interactivity, and duration of deception [e.g., 20, 63, 67]. Additionally, cultural identity [7], motives and attitudes about deception [58], and interaction medium [31] play critical roles in how credibility may be assessed by a receiver or observer. The effect of all of these factors, taken together in a natural setting, renders the problem of credibility assessment particularly complex. Many of the variables relevant in credibility assessment have been determined, but linkages between them are still vague or undetermined. Moreover, the fleeting nature of behaviors diagnostic of deception contributes to a problem with a large number of unspecified constraints.

Credibility assessment tasks are also replete with uncertain and dynamic environments. The aforementioned link between deception and outward behavioral displays is a prime example. Behavioral features that are diagnostic for one individual may not be for another individual. Moreover, depending on the context of the interaction being assessed, the person doing the assessment may know very little about the potential deceiver (such is the case in our study). In cases where uncertainty is great, assessors may turn to heuristic-based methods of evaluating credibility. Heuristics that are prevalent in credibility assessment include basing assessments on assumed prevalence of truth, visual cues, demeanor, and expectancy violations [13].

Shifting and conflicting goals and action/feedback loops also manifest in the way deception frequently is perpetrated. Buller and Burgoon [12] offer a dynamic view of deception in their interpersonal deception theory where deceivers are constantly updating their tactics and estimations of their deceptions' effectiveness. Receivers maintain their own levels of suspicion and may opt to press or probe a deceiver if they are suspicious. This results in a sort of cat-and-mouse game between deceivers and receivers where strategies, goals, suspicion, and perceptions of suspicion all interplay during the interaction. A receiver or observer who intends to assess the credibility of a potential deceiver bears the responsibility to track all of these factors throughout the interaction.

Depending on the context, those assessing credibility may face considerable time pressure and high-stakes outcomes, which are at times in conflict. For example, those who operate border crossings, check points, or screening stations are faced with a constant flow of large numbers of people. Processing these people quickly and efficiently is highly desirable. Yet, the desire for expediency may be at odds with the thoroughness necessary for effective screening. In addition, high-stakes outcomes are frequently encountered in credibility assessments, and especially those that occur in a law enforcement or security context. Obviously, there can be significant ramifications for a deceiver if he or she is deceiving in an attempt to avoid punishment for illegal actions. Likewise, the person performing the assessment also faces high stakes because the results of detaining an innocent person or freeing one who is guilty of a crime can be very costly.

Credibility assessment always involves multiple individuals, with at least one deceiver and at least one receiver. Each individual’s goals, interests, and intentions may vary considerably. In a law enforcement or security scenario, these goals, interests, and intentions may be heavily influenced by organizational forces (e.g., rules, procedures, laws). All of these characteristics combine to paint credibility assessment as a difficult problem, particularly due to its uncertainty, dynamism, and lack of structure.

Effects of Expertise

The effects of expertise in complex task performance have shown mixed results in previous research. One stream of research suggests that expertise allows individuals to form more complex models and identify solutions to problems more rapidly [38]. Further, such abilities positively influence complex task performance [16]. Another stream of research suggests that for complex, ill-structured problems, expertise makes little difference [36, 45]. Johnson [35] attempts to reconcile these seemingly conflicting findings by pointing out that in cases where the focus of the research is the process by which experts come to a conclusion, expert performance tends to exceed that of novices. In cases where decision outcomes are the focus, expert performance roughly approximates that of novices. Shanteau [59] offered another explanation and suggested that task characteristics may be the reason for these inconclusive findings. Devine and Kozlowski [22] continued this work and showed that expertise increased accuracy in a computer-based task only when the task was well structured. When the task was ill structured, expertise made little difference.

In the task of credibility assessment, with a few exceptions, expertise has been shown to make little difference. Traditionally, expertise has been determined by status as a professional lie-catcher, such as a police officer or federal law enforcement agent [67]. In the majority of studies that compare the performance of professional lie-catchers to novice lie-catchers (traditionally undergraduate students), the results have been consistent: no difference exists in assessment accuracy rates [14, 21, 46, 68, 69].[1] In one noted exception, experience as a professional law enforcement officer did contribute to greater assessment accuracy when judging actual criminal interviews [42]. This finding may indicate that assessment performance may be better in a realistic setting; however, the mean professional assessment accuracy in this case was 65%. Ekman and O'Sullivan [25] presented evidence suggesting that agents from the United States Secret Service and Central Intelligence Agency could assess credibility more accurately than could others. However, others have publically questioned the validity of this finding [8].

Decision Aid in Credibility Assessment

To counter many of the difficulties associated with unaided credibility assessment, researchers are developing computer-based methods of assessment that unobtrusively analyze observable behaviors [33, 48, 74]. A decision aid may serve a number of beneficial functions in

---

[1] In response to these findings, the term *professionals* is used to describe those with professional experience in assessing credibility throughout the rest of this paper. Research and work outside the realm of credibility assessment may term these individuals as "experts." However, we feel the term *professional* is more precise.

credibility assessment. First, it can extend the sensory and processing capabilities of both novices and professionals to examine behaviors that have been shown to be diagnostic of deception in previous research [e.g., 33, 34]. These capabilities are particularly useful because combinations of observable behaviors have been shown to differentiate deceivers from truth tellers [15]; yet, observers may miss these behaviors, not know to attend to these behaviors, or not have the cognitive ability to observe, track, and monitor these behaviors as they occur. By capturing and analyzing behaviors, a decision aid enforces a constraint on the problem of credibility assessment and, thus, adds structure to the task. A decision aid also helps to mitigate much of the working memory failures that have plagued high task-knowledge individuals in the past [e.g., 1].

A decision aid may also effectively manage action/feedback loops that are common in credibility assessment and note any changes in behavior that appear as a result of a change in strategy or revealed suspicion [47]. Time stress may also be alleviated because of the increased processing capabilities of the decision aid. However, because of the difficulty of assessing credibility, there are some distinct limitations on the operation of a decision aid that also merit discussion. First, because of the probabilistic link between behaviors and deception, the decision aid may provide faulty recommendations. Faulty recommendations are a major risk to any potential user, both in terms of detriment to accuracy [62] and in future usage behavior [65]. Second, there are characteristics of credibility assessment that a decision aid is ill-suited to address. For example, the ramifications of high-stakes credibility assessment may be more effectively considered by the user and not by the decision aid.

THEORY AND HYPOTHESES

Several theories provide explanatory power in the use and adoption of technology in decision-making tasks(e.g., Technology Acceptance Model [17]; Technology Transition Model [10]; and various integrating views [66, 71]). However, we sought a theoretical model that describes relationships between constructs that are critical in credibility assessment—namely, task complexity and task expertise. The TTD describes the key constructs and relationships between these constructs and decision aid reliance. Furthermore, the TTD was developed specifically to examine the effects of intelligent decision aids, which makes it applicable in exploring aided credibility assessment. Therefore the TTD was adopted to better understand the effects of introducing a decision aid during credibility assessment.

TTD was originally developed to understand accounting professionals' actions as they used intelligent decision aids in common accounting functions (e.g., assisting in ongoing concern decisions [2] and calculating tax liability [52]). The theory provides conditions under which individuals will become more or less reliant on decision aids. Key influences on reliance are task experience, task complexity, familiarity with the decision aid, and cognitive fit between user and decision aid [4]. Figure 1 summarizes the key relationships specified by TTD.

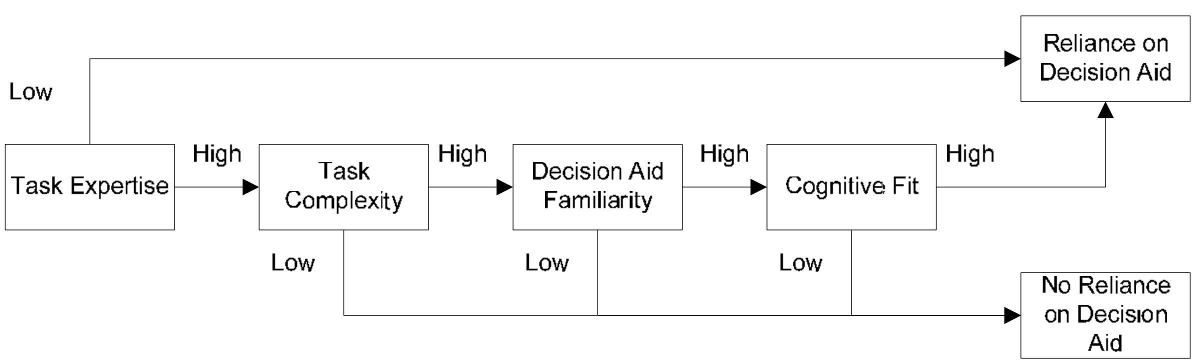

Figure 1. The Theory of Technology Dominance [adapted from 4]

Task expertise is defined by Arnold and Sutton [4] as "the level of experience a decision maker has with respect to the completion of a given decision task and the degree to which the decision maker has formed strategies for completing or solving the task" (p. 180). As shown in Figure 1, when users of a decision aid have low task knowledge, they are predicted to rely heavily on the decision aid for guidance, regardless of the task. This link is intuitive because individuals who lack the skills necessary to perform a particular task are more likely to seek assistance. However, for those with high task expertise, task complexity and decision aid characteristics of familiarity and cognitive fit must first be considered before predicting the level of reliance. If the complexity is low, there is no need for those with high task expertise to seek help from a decision aid. TTD treats task complexity separately from the decision aid characteristics and posits that there is a positive relationship between task complexity and reliance [4]. High task complexity encourages reliance on the decision aid, but only if potential users are familiar with the aid and have high cognitive fit with the decision aid. If familiarity and cognitive fit are high, experienced users will be required to expend less cognitive effort in reviewing and accepting the recommendation of an aid [4].

One of the most important propositions of TTD is that when novices are paired with an intelligent decision aid, there exists the potential for poor decisions because novices over-rely on the aid when forming their decisions. Arnold and Sutton state:

> *The mismatch between the user and the intelligent decision aid (in terms of expertise) makes the decision maker susceptible to dominance by the technology. While this dominance is related to overreliance on the intelligent decision aid by the decision maker, the concern here is that the degree of overreliance will be strong enough to have a significant deleterious affect [sic] on the user's decision. [4] (p. 184)*

TTD posits that the most gains from an intelligent decision aid are to be had by pairing the aid with an individual with high task experience. Such a pairing minimizes the risk of technology dominance by the decision aid—because experts are less apt to over-rely on an aid than will novices—and avoids the long-term problem of de-skilling of the aided users [4]. The decision aid should help the user with high task knowledge to extend information processing capabilities [3]. High task experience should allow a number of benefits including proper application of the decision aid, understanding of decision aid limitations, and increased robustness to decision aid failure [4].

Hypotheses

Although both novices and professionals may believe they can accurately assess the credibility of others, novices are more likely to recognize they have little formal training or professional experience in performing such assessments. TTD suggests that novices, because of their lack of task-related knowledge, will tend to defer their own judgments in favor of the decision aid's recommendation [4]. This phenomenon has been demonstrated in past empirical work, but this research has taken place in the comparatively well-structured contexts of ongoing concern decisions [2, 3] and tax compliance decisions [44, 52]. The novices may also treat the decision aid as a source of authority and, thus, be more likely to adopt its recommendation [54]. We hypothesize an extension of this finding in the complex task of aided credibility assessment where novices will defer their own judgments and rely on the recommendation of the decision aid more often than will professionals.

H1: Novices' assessments will be less different from the aid's recommendations than will professionals' assessments.

In complex problems, a decision aid is not likely to be infallible. Although information processing limits on decision aids are much greater than limits on humans, some of what may be relevant in solving a complex problem may be ignored by the decision aid (e.g., a "broken-leg" feature [45]). Because novices are more likely to rely heavily on decision aids when addressing complex problems, the potential exists for novices to be unprepared for a situation that was unexpected during the design of the decision aid. Thus, TTD suggests that when novices use a decision aid, there is increased potential for poor or faulty decisions because of their greater reliance on a decision aid, as opposed to the decisions of experts [3, 4].

Although risk may exist in novices' making poor aided decisions in other complex and ill-structured contexts, these findings may not extend to the context of credibility assessment. Some have argued that any shortcomings in the operation, design, or function of the system should be considered and countered by a skilled user [4]. For example, if a decision aid should fail, the user should be able to step in and adequately take over [56]. Yet, as noted, research has repeatedly shown that unaided credibility assessment performance by novices and professionals is generally poor [6]. If true, and if a decision aid were to fail, novice users would be no worse off than unaided professionals in assessing credibility. For a decision aid to be detrimental to a novice's credibility assessment, its shortcomings would need to be so severe as to undermine the already poor performance of unaided assessment.[2]

We posit that in the context of credibility assessment, a decision aid should provide many advantages to a novice. First, a decision aid brings a systematic approach to credibility assessment that may counter many of the biased-based approaches novices use to assess

[2] It is important to note that only overall assessment accuracy is being discussed here (as it has been in past deception and credibility research). There may be differences in the types of errors professionals and novices make, and these errors may have different associated costs. Thus, on the basis of error cost, there may be added danger from aided credibility assessment by novices.

credibility. An aid would also extend the information processing capabilities of a novice because more behavior is systematically analyzed for signs of deceit. For example, in past research investigating computer-aided credibility assessment by novices, use of the decision aid was linked to large improvements in assessment accuracy [33]. However, the novice is left to ensure the proper operation and application of the decision aid, a task the novice may not be capable of doing. Although inappropriate reliance by novices on the decision aid may occur with adverse consequences, we predict that in the context of credibility assessment, this reliance will improve assessment accuracy by novices overall.

H2: Aided credibility assessment by novices will be more accurate than unaided assessment by novices.

Consistent with TTD, we predict that professionals will be initially reluctant to rely on the decision aid. As hypothesized by TTD, when confronted by a decision aid, professionals with high task knowledge will attempt to understand how the decision aid arrives at its recommendation in order to determine if the recommendation is suitable for acceptance [4]. Gregor and Benbasat [30] suggested that professionals will use explanations to resolve anomalies and to verify the performance of the decision aid. This phenomenon has garnered empirical support in other research [2, 43]. Consistent with Arnold et al. [2], we expect that as perceived anomalies are explained and resolved and the professional user has a chance to evaluate the performance of the decision aid, the professional user will accept and rely upon more recommendations from the decision aid.

H3a: Professionals are more likely than novices to access explanations generated by the decision aid.

H3b: Professionals who use explanations generated by the decision aid are more likely to adhere to the decision aid's recommendation.

As professionals examine the operation of the decision aid through usage of the explanations and adoption of the decision aid's recommendations, they should experience many of the same benefits in aided credibility assessment that novices experience. The decision aid should help in countering existing biases and should extend the information gathering and processing capabilities of the professionals. This effect has been seen with other aided forms of credibility assessment, such as the polygraph [67].

H4: Aided credibility assessment by professionals will be more accurate than unaided assessment by professionals.

A key purpose of this study is to compare the computer-aided performance of novices and professionals in the context of credibility assessment. As noted previously, there is conflicting evidence comparing the performance of high task-knowledge individuals to low task-knowledge individuals. Thus, it is unclear whether the experience of professional lie-catchers should be a benefit or a hindrance in aided credibility assessment. On one hand, professional lie-catchers have shown no significantly greater ability to distinguish deception from truth than do novices. This is consistent with other research evaluating the performance of experts in complex decision making in natural environments [e.g., 36, 45]. From this perspective, one might expect professionals and novices to perform equally well in computer-aided credibility assessment. Because both novices and professionals are on equal footing during unaided assessment, the introduction of a decision aid should have a similar effect on them both.

Conversely, TTD posits that professionals will produce better decisions because the capabilities of the decision aid and experience of the professional user are effectively paired [4].

Thus, the professional may treat the decision aid as an "electronic colleague" [4]. A professional in this position would be able to understand the limitations of the aid, consider relevant information outside the aid's scope of consideration, and judge the operation of the decision aid and, therefore, apply it correctly.

Thus, in the face of conflicting theoretical and empirical evidence, our final area of inquiry is framed as a nondirectional, exploratory hypothesis. This hypothesis examines whether there will be a difference in assessment accuracy between professionals and novices.

H5: There is a difference in aided credibility assessment accuracy between novices and professionals.

METHOD

To test these hypotheses, an experiment was conducted in which novices and professionals used a prototype decision aid to assess credibility. This experiment is an extension of the work reported in [33] and followed the same general experimental procedure.

Experimental Task

After arriving at the laboratory, professional and novice experiment participants were given a brief survey and were instructed about the operation and expected performance of the decision aid they would be using. Then they viewed a series of 10 video segments depicting a dyadic interview and had to identify which segments contained deception. Each participant had access to the decision aid during the task. During the orientation, sample deceptive videos were described and shown to the participants so they would be familiar with the context of the videos they would be judging. The persons interviewed in the videos were unknown to the participants.

The interviews for this study were provided by a previous experiment that was conducted at a large Midwestern university whose original purpose was to examine the behavior of deceivers under high stakes and to identify behaviors that might differentiate between true deceivers and false confessors (Levine et al. 2006). The interviewees were confronted about actual and unsanctioned cheating on a task that gave cash rewards for correct responses. To understand the deceivers' motivation and deception context, a short description of the interviews is provided below.

Following the Exline procedure [27], undergraduate students from an introductory communication course were invited to participate in a study. The students 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 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 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 viewing the answers or by refusing. After a few minutes, the experimenter returned and finished asking the trivia questions. After all the trivia questions were complete, the participant was brought to an interview room where he or she was interviewed. There, the interviewer confronted the participant with a structured interview to find out if he or she had cheated. All participants were interviewed by a single interviewer, and the interviewer posed the same questions to all participants.

The professionals and novices viewed 10 interviews in which each interviewee denied allegations of cheating. Five interviews were deceptive and five interviews were truthful. After viewing each interview, the professional and novice participants submitted an initial judgment without the assistance of the decision aid. After submission of the initial judgment, the participants then had access to the recommendation and explanations of the decision aid. At this point, the participants submitted a final judgment. Each initial and final judgment consisted of whether the participant believed the interviewee was guilty of cheating, the level of deception (0-10), and level of confidence (0-10).

Familiarity with Decision Aid and Cognitive Fit

As familiarity and cognitive fit are expressly included in TTD, they were given careful consideration during the study. Participants were introduced to the decision aid and stimulus videos through an in-depth orientation. The development, function, and operation of the decision aid were described, and participants viewed a sample interview. They also were instructed on how to access explanations. The participants were informed that the decision aid would correctly assess credibility 60-80% of the time, giving them an idea about the performance of the decision aid. This accuracy rate is consistent with other research investigating credibility assessment from linguistic and kinesic behavior [15, 48, 75]. Finally, the participants performed the same task with the decision aid 10 times, and none of the participants experienced any noticeable problems using the decision aid.

Cognitive fit was considered during the design and development of the decision aid. Individuals frequently form, albeit incorrectly, credibility assessments based on others' behavior, so capturing and systematically examining behavior for signs of deceit should not have been unfamiliar to the participants. The layout of the aid was clear and concise, with results and a recommendation provided on the left side of the screen. The explanations were also carefully crafted to be accessible to users. Following research practice in expert- and knowledge-based systems [2], researchers in credibility and deception reviewed the explanations provided by the aid to ensure clarity and correctness.

Participants

A total of 31 novice participants took part in the experiment and were motivated to participate by course credit. In addition, novices who demonstrated assessment accuracy rates in the top 10% were awarded a $10 cash prize. The mean age of the novice participants was 21.5, mean years of secondary education was 3.5, and of all the participants, 45% was female.

Additionally, 20 professional participants were recruited from the United States Secret Service; city police departments; the Internal Revenue Service, Criminal Investigation Division; and the Bureau of Alcohol, Tobacco, Firearms, and Explosives. All participants had conducted interviews and interrogations during criminal investigations. The mean age of the professional participants was 39.0, mean years of secondary education was 5.0, and of all the participants, 10% was female. The professionals' average number of years of law enforcement experience was 12.0. As an incentive for participation, professional participants were offered free training in deception detection after the conclusion of the experiment.

Decision Aid

The decision aid used in this experiment, called the Behavioral Analysis Prototype (BAP) [33], is a prototype intelligent decision aid that presents the results of automated linguistic and kinesic analyses. Kinesic analysis extracts and examines features diagnostic of deception from the way a person moves [48]. Linguistic analysis extracts and examines features diagnostic of deception from the structure and pattern of a person's language [74, 75]. Linguistic and kinesic features have been shown to be diagnostic of deception in previous research [20] and have been used in law enforcement environments [67]. The BAP presents the results of both kinesic and linguistic analysis and produces a combined score by equally weighting the linguistic and kinesic scores. The BAP also produces a confidence measure presenting the strength of classification. The function of the BAP is an improvement on the intelligent decision aids described previously in literature: namely, there is no interactive function for gathering information from the user. To our knowledge, the BAP is the first attempt at an intelligent decision aid designed to support the decision processes of credibility assessment based on kinesic and linguistic cues.

The prototype contains natural language explanations of the cues that are included in the kinesic and linguistic analyses. These explanations were manually generated for this study, but followed a consistent format and could feasibly be automatically generated in future prototypes. The explanations were designed to fill similar roles of explanations from other expert systems, namely: definition, rule-trace, and justification explanations [e.g., 23, 72]. All explanations are provided after the analysis of an interview is complete (no feed-forward explanations are available). A complete description of the prototype is beyond the scope of this paper; however, the interested reader is directed to other work [15, 33]. The interface as well as sample explanations from the BAS are shown in Figures 2 and 3.

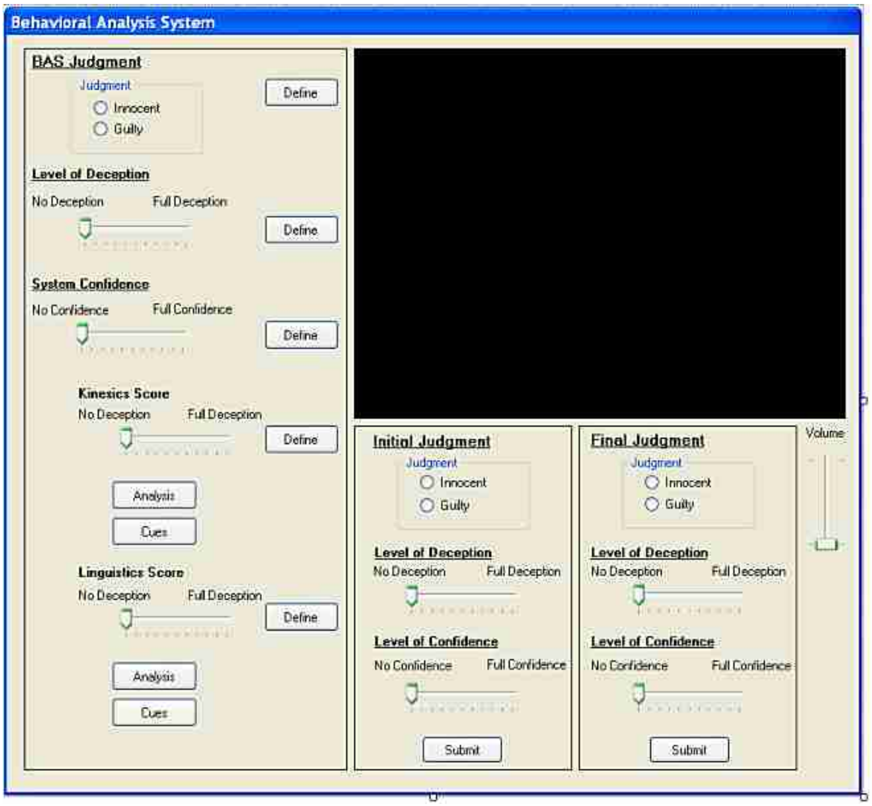

Figure 2. Screen Shot of the Behavioral Analysis System

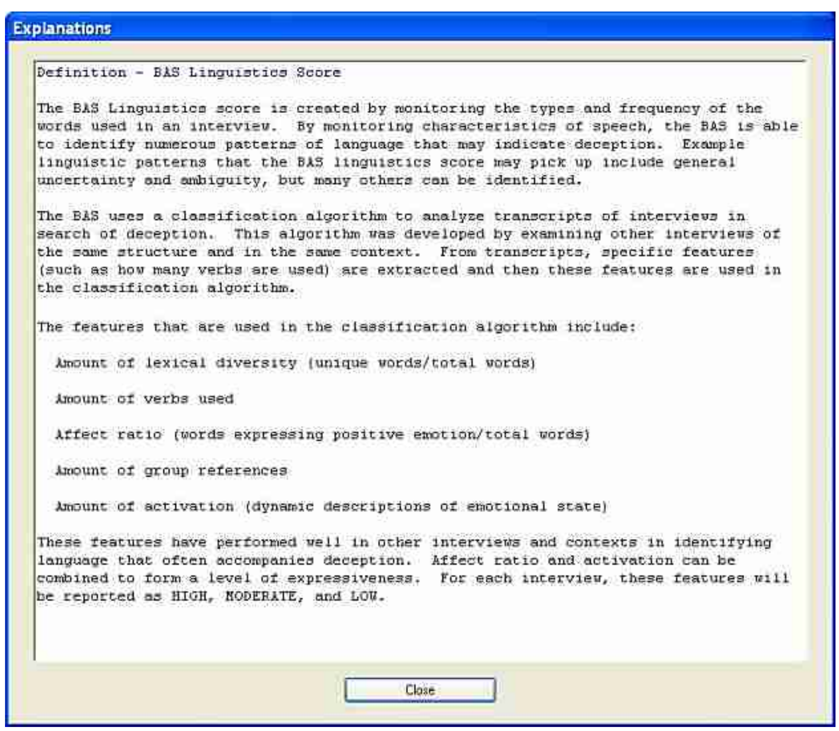

Figure 3. Definition Explanation for the Linguistic Deception Score

ANALYSIS

To examine the differences in use, reliance, and accuracy among the novice and professional participants, a series of statistical analyses was employed. In all analyses investigating accuracy improvements and reliance, comparisons of the level of deception (0-10) were performed to be consistent with the finding that credibility assessment is more accurately and realistically captured by using continuous measures and by avoiding dichotomous, nominal variables (e.g., deception/truth) [14, 19]. Table 2 presents the mean number of explanations used by professionals and novices and the number of times the recommendation of guilty or innocent was accepted by the participant. Table 3 shows how the level of deception recorded by participants changed between the initial and final assessments and illustrates the reliance of the participants on the decision aid. Table 4 shows accuracy rates as a function of the difference between actual deception levels and estimated deception levels.

Table 2. Explanation Usage and Recommendation Acceptance

Participants	N	Definitions *(SD)*	Rule Trace *(SD)*	Justification *(SD)*	Accepted Recommendations [%] *(SD)*
Novices	31	1.5 *(1.4)*	1.6 *(1.7)*	1.4 *(1.7)*	68.1 *(9.8)*
Professionals	20	2.0 *(1.6)*	2.8 *(2.8)*	2.0 *(2.2)*	61.0 *(16.8)*

Table 3. Level of Reliance on the Decision Aid

Participants	N	$\|BAP_D – Initial_D\|$ *(SD)**	$\|BAP_D - Final_D\|$ *(SD)***
Novices	31	3.6 *(.6)*	2.5 *(.7)*
Professionals	20	3.6 *(.7)*	3.2 *(.9)*

* Absolute difference between the BAP's deception score and deception level indicated in the initial judgment.

**Absolute difference between the BAP's deception score and deception level indicated in the final judgment.

Table 4. Accuracy Rates of Novices and Professionals

Participants	N	$\lvert$Actual – Initial$_D\rvert$ *(SD)* †	$\lvert$Actual – Final$_D\rvert$ *(SD)*‡
Novices	31	5.2 *(.8)*	4.5 *(.7)*
Professionals	20	5.0 *(1.0)*	4.6 *(1.0)*

† Absolute difference between the actual case and the deception level indicated in the initial judgment (Truthful interviewee, actual = 0; Deceptive interviewee, actual = 10).

‡ Absolute difference between the actual case and the deception level indicated in the final judgment (Truthful interviewee, actual = 0; Deceptive interviewee, actual = 10).

To perform robust analyses, a potential covariate was examined to avoid possible confounding results. Before the experiment, participants were asked in a survey about their comfort level with a computer. There were reported differences between professional and novice participants in comfort level with a computer ($t_{(49)} = 2.32$, $p = .025$); therefore, comfort level was included as a covariate in all tests comparing novices to professionals. Also, as a manipulation check, participants were asked about the difficulty of the task. It was anticipated that novices would find it more difficult than professionals, and this was the case ($t_{(49)} = 2.86$, $p = .006$).

To test the first hypothesis, a series of analysis of covariates was performed comparing professional and novice reliance on the decision aid. First, to ensure there was no systematic similarity between initial judgments of professionals and novices and the recommendation of the decision aid, the difference between the BAP's recommendation and the level of deception indicated in the initial judgment ($\lvert BAP_D$ – Initial$_D\rvert$) was examined. Assumptions of parametric statistical tests were met, yet there were no systematic differences ($F_{(1, 48)} = .23$, $p = .633$). Then, to determine the extent to which the decision aid influenced final judgments, the difference between the decision aid's recommendation and the level of deception supplied by novices and professionals during their final judgment ($\lvert BAP_D$ – Final$_D\rvert$) was examined. In this examination, greater reliance was indicated by a smaller difference between the recommendation and the level of deception supplied by the participants (i.e., the final judgment would more closely resemble the recommendation of the decision aid). Assumptions of parametric statistical testing were met, and the covariate was not significant. The difference was smaller for novices than for professionals ($F_{(1, 48)} = 10.33$, $p = .002$), indicating that novices were more reliant on the decision aid than were professionals. This finding supports H1.

To test the second hypothesis, a paired t-test was performed on data only from novices. This t-test compared the difference between the level of deception in the initial judgment and the actual condition of the interviewee ($|\text{Actual} – \text{Initial}_D|$) to the difference between the level of deception in the final judgment and the actual condition of the interviewee ($|\text{Actual} – \text{Final}_D|$). Improvement in assessment accuracy was indicated by a decrease in the difference between the judged and actual levels of deception (i.e., the final judgments more closely resembled the actual condition of the interviewee). Novice participants did demonstrate significant improvement in their assessment accuracy through use of the decision aid ($t_{(30)} = 7.89$, $p < .001$), supporting H2.

Hypotheses 3a and 3b concern the role of explanation usage among professionals and novices. As shown in Table 2, the number of explanations accessed by professionals and novices in each category was relatively small. Reasons for this finding are addressed in the discussion section. In addition, the distribution of the number of accessed explanations was skewed. A multivariate analysis of covariance with the numbers of accessed explanations in each category as the dependent variables was used. To address the skewness, a natural log transformation was performed on the numbers of accessed explanations. Thus, the analysis met assumptions of parametric statistical testing. However, professional status had no significant influence on the number of definition ($F_{(1, 48)} = 1.16$, $p = .287$), rule-trace ($F_{(1, 48)} = 1.55$, $p = .219$), or justification

($F_{(1, 48)} = .89$, $p = .349$) explanations. The covariate was also not significant. This finding fails to support H3a.

To determine whether the number of explanations accessed by professionals increased the acceptance of the decision aid's recommendation, a simple multiple linear regression was performed, with the difference between the decision aid's recommendation and the professionals' level of deception during their final judgment ($|BAP_D - Final_D|$) as the dependent variable. The independent variables were the transformed totals of explanations accessed in each category. The overall model was significant ($R^2 = .55$, $F_{(3, 16)} = 6.65$, $p = .004$); however, only the number of definition explanations was individually significant in the regression model ($\beta = -1.34$, $t_{(16)} = -4.37$, $p < .001$). This suggests that as professionals used definition explanations, their final judgments more closely approximated the decision aid's recommendation (i.e., the difference between them was smaller). This finding partially supports H3b.

To examine the fourth hypothesis, another paired t-test was performed on data only from professionals. Again, this t-test compared the difference between the level of deception in the initial judgment ($|Actual - Initial_D|$) and the actual condition of the interviewee to the difference between the level of deception in the final judgment and the actual condition of the interviewee ($|Actual - Final_D|$). Professional participants did demonstrate significant improvement in their assessment accuracy through use of the decision aid ($t_{(19)} = 2.83$, $p = .011$), providing support for H4. Thus, professionals' assessment accuracy improved through use of the decision aid.

The final hypothesis explores the difference in aided assessment accuracy among novices and professionals. This hypothesis was tested by a final analysis of covariates where the difference between the level of deception in the final judgment and the actual condition of the interviewee ($|Actual - Final_D|$) was compared between novices and professionals. This comparison yielded no significant difference ($F_{(1, 48)} = .07$, $p = .789$). The covariate was also not significant. As might be expected from inspecting Table 4, professionals did not derive any greater benefit to accuracy from using the decision aid than did novices.

DISCUSSION AND IMPLICATIONS

The findings reported in this paper examine the effect of an intelligent decision aid during the complex decision task of credibility assessment. To our knowledge, this is the first attempt at utilizing TTD in such a complex task as credibility assessment. The findings found in this paper are summarized in Table 5.

Table 5. Summary of Research Results

Hypotheses	Findings
H1: Novices' assessments will be less different from the aid's recommendations than will professionals' assessments.	Supported
H2: Aided credibility assessment by novices will be more accurate than unaided assessment by novices.	Supported
H3a: Professionals are more likely than novices to access explanations when using the decision aid.	Not supported
H3b: Professionals who use explanations are more likely to adhere to the decision aid's recommendation.	Partially supported (for definition explanations only)
H4: Aided credibility assessment by professionals will be more accurate than unaided assessment by professionals.	Supported
H5: There will be a difference in aided credibility assessment accuracy between novices and professionals.	Not supported

The main findings of this research have important implications for decision-aid users performing highly complex tasks such as credibility assessment. First, as is predicted by TTD, novices rely more heavily on decision aids than do professionals. Recommendations from decision aids represent attractive options for novices who have little formal task knowledge (e.g., training) on which to base their own assessments. In addition, novices are more likely to treat the recommendation of the decision aid as a source of authority and, thus, are more likely to be influenced by the aid's recommendation [54].

Although, in our study, novices were more likely to rely on the decision aid, it is interesting to note that novices did not conform their decisions completely to the decision aid's recommendations. When the decision aid provided a deception score, the novices felt comfortable enough to deviate from the recommendation by an average of ±2.5 points (out of 10). This suggests that the novices felt comfortable dismissing the aid's recommendation and weakens the possibility that novices would be dominated by the decision aid.

Incongruent with TTD, our findings suggest that novices benefited from using the decision aid to assess credibility. However, there are two characteristics of credibility assessment that may limit the generalizability of this finding. First, because the task of credibility assessment is so difficult, there is a low threshold to exceed for aided novices to demonstrate accuracy improvements. Next, the environment of this experiment was highly constrained. This may have favored the performance of novices because the decision aid operated correctly and was properly applied. During actual use in a natural setting, oversight and application responsibilities would fall to the novices and, in this setting, there may be the risk of a decrease in accuracy. Nevertheless, when the decision aid is operating correctly and is properly applied, aided novices showed an improvement in assessment accuracy. Thus, training novices in the operation and proper application of such a decision aid should make the accuracy improvements demonstrated in this research more likely in practice.[3]

The next pair of hypotheses tested the use of explanations by professionals and the effects use had on recommendation acceptance. There was no support for professionals using more explanations than novices did. This finding contrasts with other published work [e.g., 2] and suggests that professionals and novices interacted with the decision aid in a similar fashion. This contrasting finding may have occurred for a number of reasons. First, both novices and professionals accessed a relatively small number of explanations when using the decision aid, which may have introduced range restriction problems when the analyses were completed. Second, the relatively small number of professional participants may have underpowered the analysis, causing actual differences to be missed (see limitations section). However, this unsupported finding may also be a result of the complex nature of the task the novices and professionals were requested to perform and is consistent with other findings of this research.

The limited number of explanations accessed by the participants in this study merits additional attention as it highlights characteristics of credibility assessment that make this task particularly difficult. In general, professionals use explanations when a decision aid does something that is unexpected and/or is perceived as anomalous [30]. A recommendation that is contradictory to a professional's initial judgment should be sufficient to warrant additional investigation through accessing explanations to discover how the recommendation was formed. Yet, in our study, when the decision aid provided a deception score that was five or more points different than the initial judgment of the professionals, the professionals reviewed explanations only 61% of the time; the remaining 29% of the time, the professionals did not access a single

[3] This is currently done with other credibility assessment technologies such as the polygraph. Although the polygraph device is not an intelligent decision aid, it does extend the abilities of the user. Therefore, in preparing to use the polygraph, operators are instructed on the proper operation and application of the device.

explanation to understand the contradictory recommendation. These results first suggest that the professionals believed they were familiar enough with the decision aid to be able to evaluate the recommendation without any explanation. One might infer that in almost one third of the cases where the decision aid offered a recommendation inconsistent with the professionals' initial assessment, the professionals did not perceive an anomaly. They simply disregarded the recommendation and did not allow themselves to be influenced by the aid's explanations. This may also be indicative of the overconfidence individuals frequently have in credibility assessment.

There are several important implications of this finding. If the purpose of an intelligent decision aid and its explanations is to persuade a user to adopt the recommended course of action (perhaps using Toulmin's model for explanation development [e.g., 72]), the user must feel an information gap sufficient to warrant the time and effort needed to consult explanations. If individuals believe they can perform a complex task such as credibility assessment without the help of a decision aid, the explanatory capabilities and utility of the recommendation are nullified. Compounding this problem is that in complex tasks such as credibility assessment, the decision aid is not infallible. The user may suspect this and expect to disagree with the aid's recommendation periodically. Thus, the user is faced with the responsibility of discerning when the decision aid is correct and when it is not. This would make the task even more difficult, because the user would then need to assess credibility while evaluating the performance of the decision aid.

Although professionals used fewer explanations in this study, the explanations that they did use played an important role in how the recommendations were treated. Professionals who used definitional explanations tended to alter their final assessments and make them reflect more closely the recommendations from the decision aid. This finding is interesting because accessing definitional explanations has been associated with novice usage patterns in previous research [2]. As professionals adhered to the aid's recommendations, their assessment accuracy improved.

Both novices and professionals benefited from using the decision aid during credibility assessment, yet professionals did not receive more benefit from aid usage. In fact, with the exception of novices' increased reliance on the decision aid, the usage behavior of professionals and novices differed little. There was no difference in explanation usage, explanations the professionals did use have been associated with novice usage in the previous studies, and accuracy rates were not different. This finding addresses our final hypothesis and leads us to believe that in the complex task of credibility assessment, similar accuracy rates may be expected from novices and professionals. This conclusion may be unsettling; however, this is consistent with research in unaided credibility assessment and other complex tasks in which professionals (those labeled as experts) perform no better than novices. This finding fails to support TTD in considering how an intelligent decision aid will improve decisions of those with high and low task knowledge. In the task of credibility assessment, improvements were demonstrated, but no differential benefit was derived by those with high task knowledge.

The implications of this finding can be a slippery slope—one we attempt to traverse with care. If the same benefit is acquired by both novices and professionals, one might be tempted to ask, "Why require a high task-knowledge individual or any individual at all to operate this decision aid?" The second part of this question is the easiest to answer: to avoid obvious misuse and abuse of the decision aid, some human user must ensure the proper application and operation of the aid. The first part of the question is more nuanced. The task in this experiment placed professionals and novices on somewhat of an even footing and made many of the skills the professionals possess (e.g., interviewing and investigatory techniques) irrelevant. Although professionals may struggle just as much as novices to assess credibility from behavioral cues, professionals have a significant leg up on novices in a variety of peripheral, yet critical, areas surrounding credibility assessment. Some of these skills include the aforementioned interviewing and investigation techniques, the knowledge of and familiarity with the law, and the ability to reconcile known facts with statements given during an interview. All of these skills may give the professional user the ability to add structure and constraints to the task of assessing credibility. Thus, potential performance increases favoring professionals may come through these peripheral areas.

Limitations and Future Research

One significant limitation of this work is that the tasks were completed in an experimental setting. This was necessary to provide the desired amount of control. However, this setting does not allow for the additional complexity that is present in credibility assessment in an actual setting (e.g., including time pressure, organizational issues). The tight control of the setting may have hampered the performance of the professionals because many relevant skills the professionals possess were rendered useless (e.g., interrogation skills). Also, the motivation of those assessing credibility may have been lower than in a natural setting. Although the deceptive and truthful interviewees demonstrated a high degree of motivation, there were few inducements to motivate the professional and novice participants. The professional participants treated the task seriously (as it was an implied test of their ability as law enforcement officers), and the novice participants were incentivized with a cash reward. However, there were no significant consequences from improperly assessing credibility. Thus, an interesting future extension of this work might explore the usage patterns of credibility assessment aids by novices and professionals when outcome stakes are high. Related work might also consider the how the cost of misallocation of attention and overconfidence affect aid usage.

Another limitation of this work is that the participants did not receive immediate feedback on their performance during the task. Although immediate feedback rarely occurs in actual credibility assessment, immediate feedback on the accuracy of participants' assessments may have altered novice and professional usage behavior. Feedback on performance may play a role in the familiarity component of TTD as users learn how well they perform while using the decision aid. Future research along these lines could include testing the impact of feedback on usage behavior immediately following the feedback and after some time lapse.

Additionally, as previously mentioned, the number of professional participants was smaller than the number of novices, thus, our analyses involving professionals may be underpowered. This limitation may particularly affect the hypothesis investigating the differential use of explanations by professionals, because a larger sample size may have supported a significant difference. However, the significant differences demonstrated in the rest of the paper are indicative of large to very large effect sizes because only a small sample was used.

Finally, a key component of TTD is the de-skilling of individuals who use a decision aid. This research adds little insight into this critical issue. Although unaided credibility assessment accuracy has been shown to be very poor [6], blind overreliance on a decision aid is also undesirable. What is needed is an effective way to merge the capabilities that come through task experience with the functions of new decision aids in a manner that improves the assessment capabilities of both.

CONCLUSION

In this study, we created an advanced decision aid to help with the complex task of credibility assessment. We used TTD to propose that technology is most effectively applied in intelligent decision aids when the experience of the user is paired with the sophistication and capability of the decision aid. Hypotheses were proposed, based on TTD, and the decision aid was tested in a study that involved both novices and professionals in credibility assessment. Both professionals and novices improved their assessment accuracy through use of the decision aid. Consistent with TTD, novices were more reliant on the decision aid than were professionals. However, contrary to TTD, there was no significant difference in the way that novices and professionals interacted with the system, and the decision aid was not more beneficial for professionals. Novices and professionals frequently discounted the aid's recommendations, and in many cases professionals did not view explanations when the decision aid contradicted their assessments. Future research that can build on our findings includes examining variations in incentives for credibility assessment, providing feedback on assessment performance, and exploring additional ways task experience may benefit computer-aided credibility assessment.

ACKNOWLEDGMENTS

Portions of this research were supported by the Defense Academy for Credibility Assessment (Contract #W911NF-07-1-0378). The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official U.S. Government position, policy, or decision. The authors acknowledge Dr. Timothy Levine, who graciously provided the stimulus materials for this experiment, and Dr. Roberto Mejias, who facilitated access to experiment participants. We also appreciate edits and reviews performed by Jeffrey L. Jenkins and Laura Rawlins.

REFERENCES

1. Anderson, J. R. Skill acquisition: Compilation of weak-method problem solution. *Psychological Review*, 94, 2 (1987), 192-210.
2. Arnold, V.; Clark, N.; Collier, P. A.; Leech, S. A.; and Sutton, S. G. The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *MIS Quarterly*, 30, 1 (2006), 79-97.
3. Arnold, V.; Collier, P. A.; Leech, S. A.; and Sutton, S. G. Impact of intelligent decision aids on expert and novice decision-makers' judgments. *Accounting & Finance*, 44, 1 (2004), 1-26.
4. Arnold, V. and Sutton, S. G. The theory of technology dominance: Understanding the impact of intelligent decision aids on decision makers' judgments. *Advances in Accounting Behavioral Research*, 1, 3 (1998), 175-194.
5. Baye, M. R.; Gatti, J.; J., R.; Kattuman, P.; and Morgan, J. A dashboard for online pricing. *California Management Review*, 50, 1 (2007), 202-216.
6. Bond, C. F. and DePaulo, B. M. Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 3 (2006), 214-234.
7. Bond, C. F.; Omar, A.; Mahmoud, A.; and Bonser, R. N. Lie detection across cultures. *Journal of Nonverbal Behavior*, 14, 3 (1990), 189-204.
8. Bond, C. F. and Uysal, A. On lie detection "wizards". *Law and Human Behavior*, 31, 1 (2007), 109-115.
9. Bonner, S. E.; Libby, R.; and Nelson, M. W. Using decision aids to improve auditors conditional probability judgments. *Accounting Review*, 71, 2 (1996), 221-240.
10. Briggs, R. O.; Adkins, M.; Mittleman, D.; Kruse, J.; Miller, S.; and Nunamaker Jr., J. F. A technology transition model derived from field investigation of GSS use aboard the U.S.S. Coronado. *Journal of Management Information Systems*, 15, 3 (1998), 151 - 195.
11. Buchanan, B. G. and Shortliffe, E. H., *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*. Boston: Addison-Wesley Longman Publishing Co., 1984.
12. Buller, D. B. and Burgoon, J. K. Interpersonal deception theory. *Communication Theory*, 6, 3 (1996), 203–242.
13. Burgoon, J. K.; Blair, J. P.; and Strom, R. E. Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, 34, 4 (2008), 572-599.
14. Burgoon, J. K.; Buller, D. B.; Ebesu, A. S.; and Rockwell, P. Interpersonal deception: V. Accuracy in deception detection. *Communication Monographs*, 61, 4 (1994), 303-325.
15. Burgoon, J. K.; Twitchell, D. P.; Jensen, M. L.; Adkins, M.; Kruse, J.; Deokar, A.; Lu, S.; Metaxas, D. N.; Nunamaker Jr., J. F.; and Younger, R. E. Detecting concealment in transportation screening. *IEEE Transactions on Intelligent Transportation Systems*, 10, 1 (2009), 103-112.

16. Chi, M. T. H.; Glaser, R.; and Farr, M. J., The nature of expertise. Hillsdale, NJ: L. Erlbaum Associates, 1988.
17. Davis, F.; Bagozzi, R. P.; and Warshaw, P. R. User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35, 8 (1989), 982-1002.
18. DePaulo, B. M.; Charlton, K.; Cooper, H.; Lindsay, J. J.; and Muhlenbruck, L. The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, 1, 4 (1997), 346-357.
19. DePaulo, B. M.; LeMay, C. S.; and Epstein, J. A. Effects of importance of success and expectations for success on effectiveness at deceiving. *Personality and Social Psychology Bulletin*, 17, 1 (1991), 14-24.
20. DePaulo, B. M.; Lindsay, J.; Malone, B.; Muhlenbruck, L.; Charlton, K.; and Cooper, H. Cues to deception. *Psychological Bulletin*, 129, 1 (2003), 74-118.
21. DePaulo, B. M. and Pfeifer, R. L. On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16, 3 (1986), 249-267.
22. Devine, D. J. and Kozlowski, S. W. J. Domain-specific knowledge and task characteristics in decision making. *Organizational Behavior and Human Decision Processes*, 64, 3 (1995), 294-306.
23. Dhaliwal, J. S. and Benbasat, I. The use and effects of knowledge-based system explanations: Theoretical foundations and a framework for empirical evaluation. *Information Systems Research*, 7, 3 (1996), 342-362.
24. Ekman, P., *Telling Lies*. New York: W. W. Norton & Company, 1985.
25. Ekman, P. and O'Sullivan, M. Who can catch a liar? *American Psychologist*, 46, 9 (1991), 913-920.
26. Engelbart, D. C., Augmenting human intellect: A conceptual framework. Air Force Office of Scientific Research, Washington D. C. 1962.
27. Exline, R. V.; Thibaut, J.; Hickey, C. B.; and Gumpert, P., Visual interaction in relation to Machiavellianism and an unethical act. In. Christie; and Geis, Eds *Studies in Machiavellianism*. New York: Academic Press, 1970.
28. Fleder, D. and Hosanagar, K. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55, 5 (2009), 697-712.
29. Frank, M. G. and Feeley, T. H. To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31, 1 (2003), 58-75.
30. Gregor, S. and Benbasat, I. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 23, 4 (1999), 497-530.
31. Hancock, J., Digital deception: When, where and how people lie online. In. Joinson; McKenna; Postmes; and Reips, Eds *Oxford Handbook of Internet Psychology*. Oxford: Oxford University Press, 2007, pp. 287-301.
32. Huang, C.-L. and Huang, W.-L. Handling sequential pattern decay: Developing a two-stage collaborative recommender system. *Electronic Commerce Research & Applications*, 8, 3 (2009), 117-129.
33. Jensen, M. L.; Burgoon, J. K.; and Nunamaker Jr., J. F. Judging the credibility of information gathered from face-to-face interactions. *working paper*, (2009),
34. Jensen, M. L.; Lowry, P. B.; Burgoon, J. K.; and Nunamaker Jr., J. F., Unobtrusive computer-aided credibility assessment by novice and professional lie-catchers. in *Hawaii International Conference on System Sciences*. Kona, HI, 2009.

35. Johnson, E. J., Expertise and decision under uncertainty: Performance and process. In. Chi; Glaser; and Farr, Eds *The Nature of Expertise*. Hillsdale, NJ: Erlbaum, 1988, pp. 209-259.
36. Johnson, P. E.; Jamal, K.; and Berryman, R. G. Effects of framing on auditor decisions. *Organizational Behavior and Human Decision Processes*, 50, 1 (1991), 75-105.
37. Kraut, R. Humans as lie detectors. *Journal of Communication*, 30, 4 (1980), 209–216.
38. Larkin, J.; McDermott, J.; Simon, D. P.; and Simon, H. A. Expert and novice performance in solving physics problems. *Science*, 208, 4450 (1980), 1335-1342.
39. Levine, T. R.; Feeley, T. H.; McCornack, S. A.; Hughes, M.; and Harms, C. M. Testing the effects of nonverbal behavior training on accuracy in deception detection with the inclusion of a bogus training control group. *Western Journal of Communication*, 69, 3 (2005), 203–217.
40. Levine, T. R.; Kim, R. K.; Park, H. S.; and Hughes, M. Deception detection accuracy is a predictable linear function of message veracity base-rate: A formal test of Park and Levine's probability model. *Communication Monographs*, 73, 3 (2006), 243-260.
41. Levine, T. R.; Park, H. S.; and McCornack, S. A. Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs*, 66, 2 (1999), 125-144.
42. Mann, S.; Vrij, A.; and Bull, R. Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89, 1 (2004), 137-149.
43. Mao, J. and Benbasat, I. The use of explanations in knowledge-based systems: Cognitive perspectives and process-tracing analysis. *Journal of Management Information Systems*, 17, 2 (2000), 153-179.
44. Masselli, J.; Ricketts, R.; Arnold, V.; and Sutton, S. G. The impact of embedded intelligent agents on tax compliance decisions. *Journal of the American Tax Association*, (2002), 60-78.
45. Meehl, P. E., *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press 1954.
46. Meissner, C. A. and Kassin, S. M. "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior*, 26, 5 (2002), 469-480.
47. Meservy, T. O.; Jensen, M. L.; Kruse, J.; Burgoon, J. K.; and Nunamaker Jr., J. F., Automatic extraction of deceptive behavioral cues from video. In. Kantor; Muresan; Roberts; Zeng; Wang; Chen; and Merkle, Eds *Intelligence and Security Informatics: Proceedings of the IEEE International Conference on Intelligence and Security Informatics ISI 2005*, 3495. Berlin: Springer-Verlag, 2005, pp. 198-208.
48. Meservy, T. O.; Jensen, M. L.; Kruse, J.; Twitchell, D. P.; Tsechpenakis, G.; Burgoon, J. K.; Metaxas, D. N.; and Nunamaker Jr., J. F. Deception detection through automatic, unobtrusive analysis of nonverbal behavior. *IEEE Intelligent Systems*, 20, 5 (2005), 36-43.
49. Mirzadeh, N. and Ricci, F. Cooperative query rewriting for decision making support and recommender systems. *Applied Artificial Intelligence*, 21, 10 (2007), 895-932.
50. Mosier, K. L. and Skitka, L. J., Human decision makers and automated decision aids: Made for each other? In. Parasuraman; and Mouloua, Eds *Automation and Human Performance*, 1, *Human Factors in Transportation*. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, pp. 201-220.
51. Newell, A. and Simon, H. A., *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1972.

52. Noga, T. and Arnold, V. Do tax decision support systems affect the accuracy of tax compliance decisions? . *International Journal of Accounting Information Systems*, 3, 3 (2002), 125-144
53. Orasanu, J. and Connolly, T., The reinvention of decision making. In. Klein; Orasanu; Calderwood; and Zsambok, Eds *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing, Corp., 1993.
54. Parasuraman, R. and Riley, V. Humans and automation: use, misuse, disuse, abuse. *Human Factors*, 39, 2 (1997), 230-254.
55. Reitman, J. S., *Cognition and Thought*. New York: Wiley, 1965.
56. Rochlin, G. I., *Trapped in the Net: the Unanticipated Consequences of Computerisation*. Princeton, NJ: Princeton University Press, 1997.
57. Ruby, C. L.; Green, M. L.; and Miller, R. The operations dashboard: A collaborative environment for monitoring virtual organization-specific compute element operational status. *Parallel Processing Letters*, 16, 4 (2006), 485-500.
58. Seiter, J. S.; Bruschke, J.; and Bai, C. The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver-deceived relationship. *Western Journal of Communication*, 66, 2 (2002), 158-181.
59. Shanteau, J. Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, (1992), 252-252.
60. Simon, H. A. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, (1955), 99-118.
61. Simon, H. A. The structure of ill structured problems. *Artificial Intelligence*, 4, 3-4 (1973), 181-201.
62. Skitka, L. J.; Mosier, K. L.; and Burdick, M. Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 5 (1999), 991-1006.
63. Sporer, S. L. and Schwandt, B. Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, 13, 1 (2007), 1-34.
64. The Global Deception Research Team. A world of lies. *Journal of Cross-Cultural Psychology*, 37, 1 (2006), 60-74.
65. Tseng, S. and Fogg, B. J. Credibility and computing technology. *Communications of the ACM*, 42, 5 (1999), 39-44.
66. Venkatesh, V.; Morris, M. G.; Davis, G. B.; and Davis, F. D. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 3 (2003), 425-478.
67. Vrij, A., *Detecting lies and deceit: The psychology of lying and implications for professional practice*. Chichester: John Wiley & Sons, 2000.
68. Vrij, A. and Graham, S. Individual differences between liars and the ability to detect lies *Expert Evidence*, 5, 4 (1997), 144-148.
69. Vrij, A. and Mann, S. Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime & Law*, 7, 1 (2001), 119-132.
70. Wilson, D. C.; Leland, S.; Godwin, K.; Baxter, A.; Levy, A.; Smart, J.; Najjar, N.; and Andaparambil, J. SmartChoice: An online recommender system to support low-income families in public school choice. *AI Magazine*, 30, 2 (2009), 46-58.
71. Wixom, B. H. and Todd, P. A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, 16, 1 (2005), 85-102.
72. Ye, L. R. and Johnson, P. E. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, 19, 2 (1995), 157-172.

73. Zhang, Y.; Zhuang, Y.; Wu, J.; and Zhang, L. Applying probabilistic latent semantic analysis to multi-criteria recommender system. *AI Communications*, 22, 2 (2009), 97-107.
74. Zhou, L.; Burgoon, J. K.; Nunamaker Jr., J. F.; and Twitchell, D. P. Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: An empirical investigation. *Group Decision and Negotiation*, 13, 1 (2004), 81-106.
75. Zhou, L.; Twitchell, D. P.; Qin, T.; Burgoon, J. K.; and Nunamaker Jr., J. F. Toward the automatic prediction of deception: An empirical comparison of classification methods. *Journal of Management Information Systems*, 20, 4 (2004), 139-166.