
AFRL-RI-RS-TR-2013-148

# ACTIVE AUTHENTICATION USING COVERT COGNITIVE INTERROGATION GAMES

SOUTHWEST RESEARCH INSTITUTE

*JUNE 2013*

FINAL TECHNICAL REPORT

***APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED***

STINFO COPY

**AIR FORCE RESEARCH LABORATORY**
**INFORMATION DIRECTORATE**

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

# NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2013-148 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

**/ S /**

TODD CUSHMAN
Work Unit Manager

**/ S /**

WARREN H. DEBANY, JR.
Technical Advisor, Information
Exploitation & Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

# REPORT DOCUMENTATION PAGE

*Form Approved*
**OMB No. 0704-0188**

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| JUNE 2013 | FINAL TECHNICAL REPORT | JUN 2012 – MAR 2013 |

**4. TITLE AND SUBTITLE**

ACTIVE AUTHENTICATION USING COVERT COGNITIVE INTERROGATION GAMES

**5a. CONTRACT NUMBER**
FA8750-12-C-0177

**5b. GRANT NUMBER**
N/A

**5c. PROGRAM ELEMENT NUMBER**
61101E

**6. AUTHOR(S)**

Jenifer Wheeler, Denise Varner, John Carrola

**5d. PROJECT NUMBER**
ATAU

**5e. TASK NUMBER**
SW

**5f. WORK UNIT NUMBER**
RI

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

**PRIME**
Southwest Research Institute
6220 Culebra Road
San Antonio, TX 78238

**SUB**
Sentier Strategic Resources,LLC
401 Congress Avenue, Suite 1540
Austin, TX 78701

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Research Laboratory/RIGA
525 Brooks Road
Rome NY 13441-4505

Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, TX 78701

**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFRL/RI

**11. SPONSOR/MONITOR'S REPORT NUMBER**
AFRL-RI-RS-TR-2013-148

**12. DISTRIBUTION AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Southwest Research Institute® (SwRI®) developed a novel method for authenticating a computer user's identity by deploying covert games disguised as anomalous computer functionality. The method applies game theory principles to allow users to develop unique strategy paths for playing the imperceptible games. By examining users' subconscious game playing strategies, the team captured discriminatory information without sophisticated contextual analysis or explicit communicative behavior. This approach complements many other cognitive fingerprint detection methods and could be used to help increase the overall accuracy of a system of combined biometric modalities. The technical approach for this research project included three major phases: Analysis, Design and Development, and Final Evaluation.

**15. SUBJECT TERMS**
Active Authentication, Game Theory, Cognitive fingerprint detection

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | **TODD CUSHMAN** |
| U | U | U | UU | 31 | 19b. TELEPONE NUMBER (*Include area code)* **N/A** |

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# ABSTRACT

Southwest Research Institute® (SwRI®) developed a novel method for authenticating a computer user's identity by deploying covert games disguised as anomalous computer functionality. The method applies game theory principles to allow users to develop unique strategy paths for playing the imperceptible games. By examining users' subconscious game playing strategies, the team captured discriminatory information without sophisticated contextual analysis or explicit communicative behavior. This approach complements many other cognitive fingerprint detection methods and could be used to help increase the overall accuracy of a system of combined biometric modalities. The technical approach for this research project included three major phases: Analysis, Design and Development, and Final Evaluation.

The SwRI team conducted an analysis by interviewing and surveying 10 representatives of the target audience about their computer habits and preferences. Results indicated that typical audience members generally use Microsoft® Office products, are comfortable troubleshooting minor problems and installing their own software, and are not likely to call for help when facing minor annoyances on their computers. The team applied game theory principles to design an inverse Prisoner's Dilemma game in the form of a modal window resembling Windows alerts. Three pilot tests with 33 participants indicated the modal window game remained covert to all participants; they viewed the interruptions as issues with computer functionality. Participants also tended to perseverate toward one type of strategy, which prompted the team to develop two game theoretical simulation tools to further distinguish the user types and allow more patterns to emerge within the range of most variation. The final evaluation yielded a Receiving Operating Characteristic (ROC) curve with an average d' value of 0.909 with a true positive rate of 0.80. None of the participants detected the game, although many reported strategizing when they chose how they would respond to the game interruptions. After analyzing the data, we identified 10 attributes that characterized all participants. These attributes were compared in a similarity matrix and then reduced through Multi-Dimensional Scaling (MDS) techniques. Finally, a four-dimensional subspace of the variables was identified and used to construct a ROC curve. The required 80% hit rate was exceeded. The false alarm rate was higher than desired; however, evidence indicates this can possibly be reduced further research.

# 1 SUMMARY

The purpose of this project was to design and develop a novel method for establishing the cognitive fingerprints of computer users by deploying games disguised as anomalous computer functionality. The Southwest Research Institute® (SwRI®) team conducted an analysis by interviewing and surveying 10 representatives of the target audience about their computer habits and preferences. Results indicated that typical audience members are intermediate computer users who work with Microsoft Office products, are comfortable troubleshooting minor problems and installing their own software, and are not likely to call for help when facing minor annoyances on their computers. The team applied game theory principles to design an inverse Prisoner's Dilemma game in the form of a modal window resembling Windows® alerts. Three pilot tests with 33 participants and a larger final evaluation with 62 participants indicated that the modal window game remained covert to all participants, who viewed the interruptions as issues with computer functionality. Participants also tended to perseverate toward one type of strategy, which prompted the team to develop two game theoretical simulation tools to further distinguish the user types and allow more patterns to emerge within the range of most variation. The final evaluation yielded a Receiving Operating Characteristic (ROC) curve with an average d' value of 0.909.

# 2 INTRODUCTION

The SwRI team developed a novel method for capturing and discriminating aspects of the cognitive fingerprint, contributing to the authentication of a user's identity. The approach validates user identity by deploying covert games disguised as anomalous computer functionality. The method, which applies game theory principles, allows authenticated users to develop unique strategy paths for playing the games, even if they are imperceptible. While there are numerous potential analytical techniques for collecting human-computer interaction data designed to distinguish one user from another, our approach found that additional discrimination is possible by encoding behavioral responses to covert, game-like tasks. By examining users' subconscious game playing strategies, our approach captures discriminatory user information without sophisticated contextual analysis or explicit communicative behavior. This approach complements many other cognitive fingerprint detection and classification methods and could be used to increase the overall accuracy of a combined biometric modality system.

To develop and evaluate this unique approach, SwRI teamed with Sentier Strategic Resources, LLC to combine SwRI's experience in behavioral modeling, educational software development, and learning science, with Sentier's extensive experience in cognitive psychology and human subjects testing. After conducting an analysis, the team applied game theory principles to design a modal window game and conduct three pilot tests and a final evaluation.

The technical approach for this research project included three major phases:

- **Analysis**: Collecting behavioral information and developing a persona of a typical user.
- **Design and Development**: Determining which types of covert game-like interactions will most accurately discriminate users with the least amount of disruption.
- **Final Evaluation**: Testing the efficacy of the system with a large group of participants to calculate the system's rate of hits, misses, correct rejections, and false alarms.

The results of these phases are described in this report.

# 3 METHODS, ASSUMPTIONS, AND PROCEDURES

## 3.1 Analysis

The SwRI team conducted user surveys focused on computer use habits, computer experience and skill, interaction styles, and work environment characteristics. The survey participants were SwRI employees (not associated with this project) in San Antonio, TX. Based on conversations with Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) representatives, the team determined that SwRI's work environment is very similar to a government or Department of Defense (DoD) office environment, and that SwRI employees would adequately represent DoD analysts for the purpose of the analysis. Data were collected in two phases. First, 10 participants completed a comprehensive survey (via face-to-face interviews) about their computer habits and skills. Participants then completed brief, twice-daily online surveys related to their most recent computer interactions (over the past three to four hours) for one week. They reported which applications they had opened and for how long, the types of problems they had encountered, messages they received from the computer, etc.

The results of the analysis indicated that the survey participants were generally intermediate-level computer users. They efficiently use their computers and applications to complete regular tasks, are willing and able to respond to basic error messages and other computer prompts, and are willing to do basic troubleshooting and installation tasks. The following sub-sections provide a summary of the survey results.

### 3.1.1 Work Environment

The survey participants generally spend extended periods of time engaged with their computers. They tend to use a small number of applications intensively (e.g., Microsoft Office applications), suggesting they have some mastery over these applications and are likely to notice changes to their normal working conditions in these applications. They are likely familiar enough with these applications to notice unobtrusive information displays at the periphery of each window. They are typically comfortable responding to computer prompts within these applications, troubleshooting basic problems, and searching for new ways to complete tasks. Most participants reported having two monitors and enough screen space. They reported commonly feeling rushed to complete their work, and may notice changes in computer speed, clutter (they may ignore it), and other issues.

### 3.1.2 Microsoft Office Experience Level

Most survey participants use Microsoft Word daily, and all consider themselves experienced users. About half were comfortable modifying Word's interface (e.g., buttons displayed), and most use advanced features like Track Changes. Most were willing to use at least a few of Word's advanced features, suggesting that these types of interaction may be useful in game designs.

Participants reported using Microsoft Excel frequently and generally considered themselves to be intermediate users. Most use keyboard shortcuts for at least some common tasks. However, many of the advanced Excel features used by power users (e.g., macros, filtering, and add-in

functionality) are not used by the survey participants. Therefore, the typical users studied in this project may be less likely to notice unusual features or functionality in Excel.

### 3.1.3 Interaction Preferences

Survey participants generally use keyboard shortcuts for common tasks such as copying, pasting, and saving files, but less frequently for other tasks. They tend to prefer making choices with point and click options rather than text menus. They also tend to have multiple windows open at one time, but with many minimized, which may create game opportunities related to window visibility and clutter.

### 3.1.4 Customization

Participants generally use common customization options (e.g., desktop and taskbar shortcuts, wallpaper, and screen saver). They do not seem to use other less obvious customization options such as mouse pointer speed, double click speed, mouse button configuration, etc. They commonly use customization options, which may provide several options for game design, as users are likely to notice idiosyncrasies related to their preferred settings.

### 3.1.5 Technical & Troubleshooting

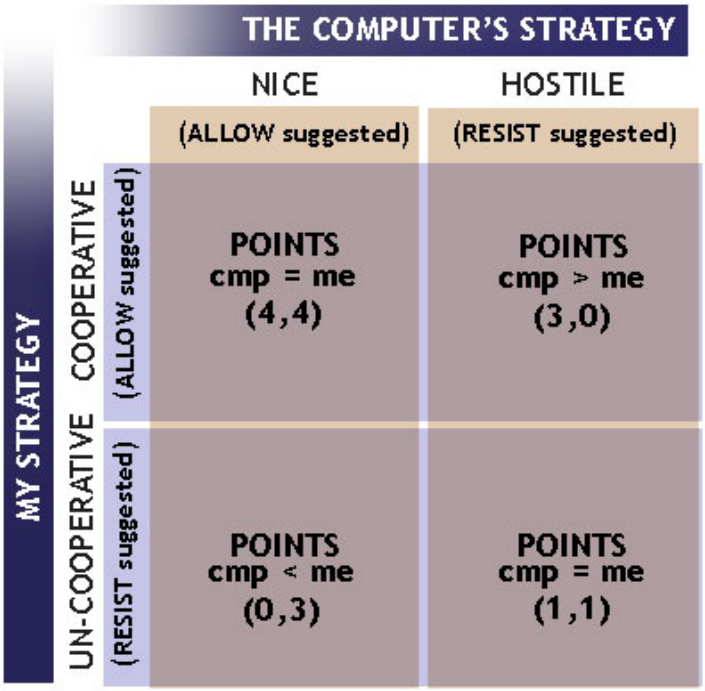
Although most survey participants reported some confidence and ability in performing technical and troubleshooting tasks, their experience is constrained to basic software installation, and some exposure to standard Windows interfaces such as the control panel. Few are comfortable with more difficult troubleshooting tasks such as evaluating or changing network settings, installing drivers, etc. Many reported managing add-ons and security settings in Internet Explorer. They also tend to be security conscious, often not allowing applications access to the internet when prompted, even for well-known applications. As a result, any troubleshooting tasks should be limited to very common tasks, and users are likely to ignore prompts to update software. It does appear, however, that advanced user settings within Internet Explorer might be useful for game design.

Warnings appeared in about 10% of application launches, common enough for everyone to be accustomed to them. Software update prompts occurred even more frequently. Survey participants tended to comply with these requests about half the time. Otherwise, they chose the option to be reminded later or closed the prompt without complying.

During the analysis period, survey participants noticed computer slowdowns, long load times, and confusing or annoying events. They reported very few issues with confusing messages. They also were not typically bothered by excessive scrolling or administrative prompts.

## 3.2 Design and Development

Game theory research indicates the existence of optimal strategies for many simple games; given enough trials, people tend to adopt strategies depending on the game type, its pay-offs, and individual preferences (Binmore, 2007). For these reasons and because of the potential to experiment with different pay-off matrices, the team designed an inverse Prisoner's Dilemma (Micko, 2007) game for this study. As shown in Figure 1, this type of game is characterized by the inverse rank order of the pay-offs; active exploitation yields the highest pay-off, followed by mutual cooperation, then by mutual obstruction, and finally by passive exploitation. The points and pay-off structure are also shown in Figure 1.



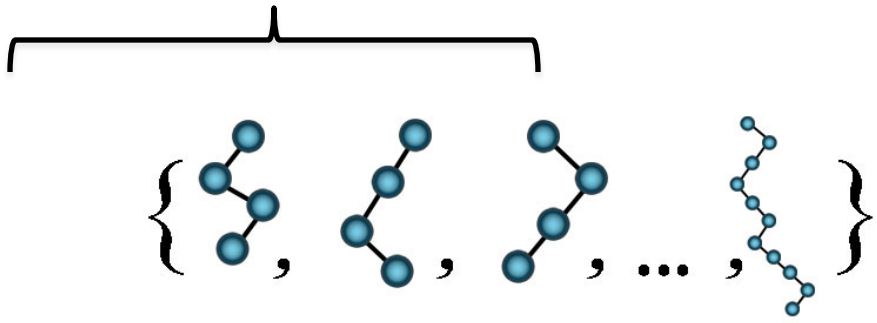



**Figure 1. Game Pay-off Matrix**

As indicated in Figure 1, the computer has the ability to present nice moves (offering fewer and less frustrating interruptions) or hostile moves (offering more frequent and disruptive interruptions) during the course of a game. User responses to each computer move can be either uncooperative or cooperative. A strategy is composed of the initial move and a rule for

subsequent moves (e.g., Tit for Tat). A user strategy is the sequences of moves he or she made in response to the computer's moves.

The points earned by the computer and the user depend on the responses employed. In the classic version of Prisoner's Dilemma, both participants are aware of the game, the other player, and the points earned. In this study's version, users are unaware the computer is playing a game and that they are scoring points through their responses. Therefore, the game is covert, although users unconsciously develop strategies for playing the game. Game pay-offs are designed with a clear optimal path (ending the game in four moves), several intermediate paths, and a slow path (ending in 14 moves).

Using input from the analysis, the team designed the Prisoner's Dilemma game interface as a modal window (See Figure 2). This interface offers two major advantages:

- The interface mimics many types of Windows alert messages, and typical users are accustomed to seeing these types of messages.
- The window is intrusive enough to cause users to respond, but is similar enough to other types of alerts that users are not likely to call for technical support.

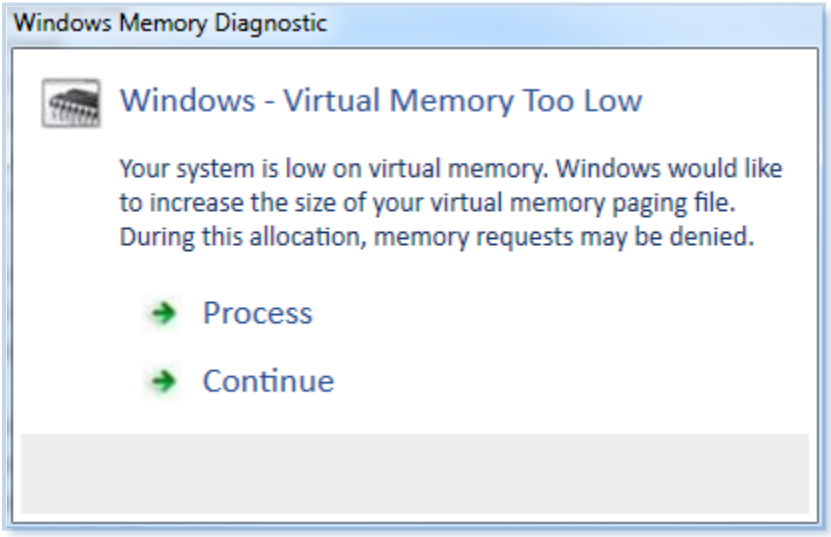



**Figure 2. Modal Window Game**

In this version of the game, user responses are considered cooperative if they select *Process* and uncooperative if they select *Continue*. Points are scored according to the matrix in Figure 1. A game is considered a series of moves that ends after 14 points are scored (either by the computer or by the user).

In addition, administrators have the ability to manipulate the game's strategies and parameters inside a control panel window (see Figure 3). This window was not visible to study participants.

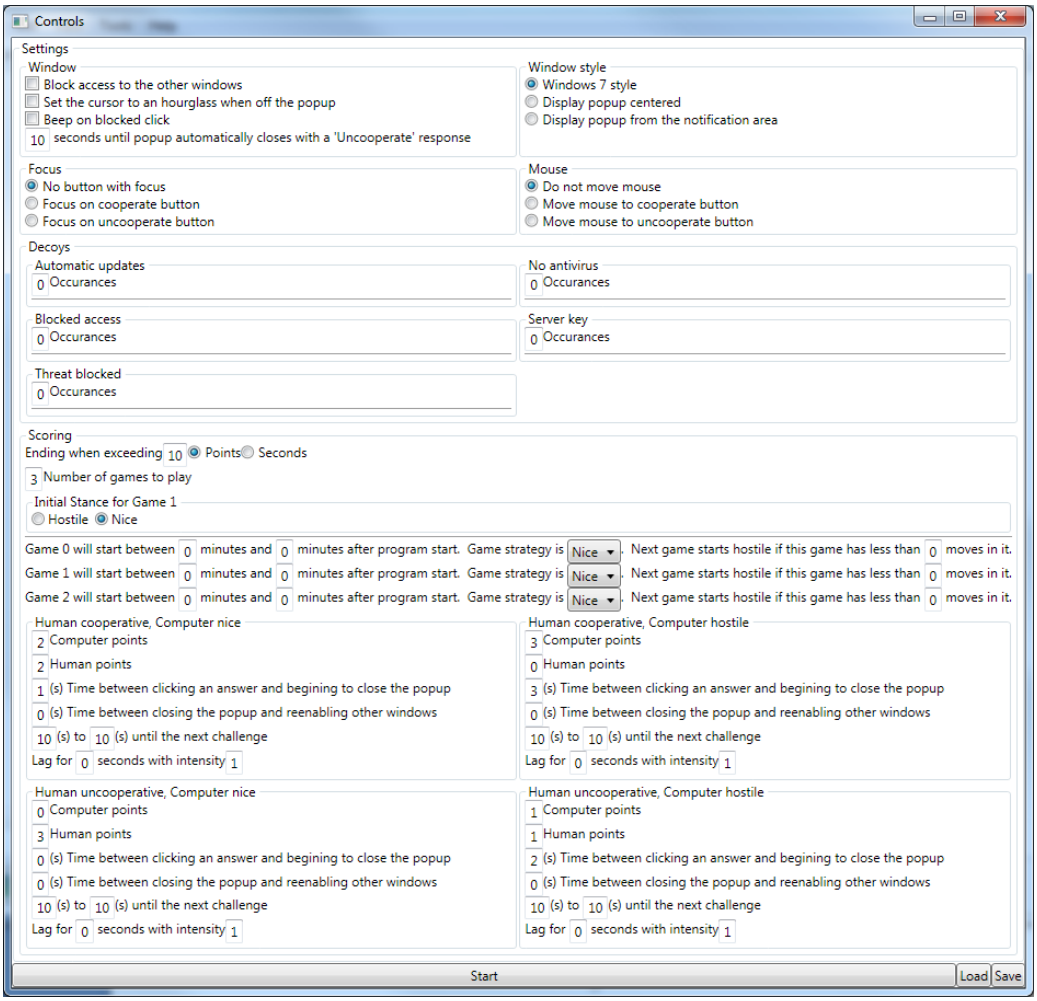



**Figure 3. Modal Window Game Control Panel**

## 3.3 Evaluation

The team conducted three iterations of game pilot tests. Slight adjustments were made to the computer strategies (e.g., when to play nice vs. hostile) and interface after each iteration. In addition, the team conducted a more extensive evaluation with the final software. During the final evaluation, the team introduced a mechanism (Z, Z' windows) for forcing dynamic behaviors from perseverators. This mechanism is described in more detail in the Results section.

### 3.3.1 Participants

During each pilot test iteration, participants completed the study in groups of three to four. Iterations 1 and 2 involved 12 participants each, and Iteration 3 involved nine participants (33 total participants). During the final evaluation, 62 participants completed the study. A professional recruitment agency selected the participants. All had intermediate experience working on computers in an office environment and were proficient in Microsoft Office programs. Participants were told the goal of the study was to measure the effects of poorly performing computers on work productivity. The purpose of this explanation was to motivate users to work quickly and help them understand the amount of work they were given and the intrusiveness of the computer. Pilot test participants earned $100 for two hours of work.

Participants in the final evaluation completed their tasks across two days (with one day between) and received a total of $150.

### 3.3.2 Task

The study task required participants to search for specific entries in a PDF file and add them to a Microsoft Excel file. Participants were informed that they would have the opportunity to earn $20 to $50 per hour, depending on how much work they could complete, and that most participants would earn $100 for the entire session. They were also informed that they needed to complete as much work as possible, and that if they were in the top 25% of productivity level, they would receive a $60 bonus. The facilitator told the participants during each break that they were on the verge of earning the bonus, and that they needed to work hard to maintain their advantage. There was not actually any productivity evaluation; all participants earned $100 in the pilot tests or $150 in the final evaluation.

### 3.3.3 Procedure

Each two-hour study session contained four 30-minute blocks, during which participants experienced two games each (totaling 8 games per session). Each block consisted of 25 minutes of task time (i.e., working on the primary study task concurrent with the game play) and a 5-minute break when participants left their computers. At the beginning of each block, the facilitator re-launched the game application.

During the Iteration 1 pilot test, the team investigated four strategies of the Prisoner's Dilemma game:

- Tit for tat (TfT): the computer responded in nice mode when the user was cooperative; the computer responded in hostile mode when the user was uncooperative
- Grim: the computer responded in hostile mode whether the user was cooperative or uncooperative
- Tit for two tats (Tf2T): the computer responded in hostile mode after two uncooperative user moves
- Average: the computer's strategy depended on an average user response

At the completion of the Iteration 1 pilot test, the team re-designed the remaining two pilot tests to deploy only the TfT strategy with the computer beginning in nice mode, switching to hostile if the previous game ended in fewer than eight moves. There were four major reasons for this decision:

- The design results in more distinct tree paths
- The design offers a clear optimum user strategy (ending the game in four moves)
- The strategy offers basic statistical balance (Principle of Indifference, i.e., if users play randomly, they have an equal chance of ending the game with the shortest or longest number of moves)
- Using the other computer strategies confused the participants and required a significantly more complicated analysis to determine whether or not they were behaving randomly

The team applied the same study design to the final evaluation, except that each participant completed two 2-hour study blocks with one day between each block. In all sessions, participants worked simultaneously on their assigned computers. Participants were debriefed about the real purpose of the study at the end of each session (during the final evaluation, participants were debriefed at the end of their second-day session).

# 4 RESULTS AND DISCUSSION

## 4.1 Modeling

Several models were constructed to capture responses across games, and several constructs were created based on the data. The constructs were treated as independent attributes that are elements of an n-dimensional vector space (n = 8, 9, or 10). The attributes were designed to capture aspects of user game play and patterns across many games.

### 4.1.1 Number of Unique Valence Sequences/Attribute 1

This feature captures the number of unique experience pay-off sequences, which indirectly captures game strategy patterns and transforms them into a sequence of unique pay-offs. Since the games are covert, participants have no knowledge of tallied scores, but they sense pay-offs and experiences through the valences determined by their play patterns. This attribute "corrects" for the participants' less-than-perfect knowledge about the games.

### 4.1.2 Z, Z' or Both/Attribute 2

This model captures whether the participants' play pattern triggered a Z' window (which appears for participants with persistent cooperative responses) or a Z window (which appears for participants with persistent uncooperative responses) in normal and inverse mode. Experiencing both types of windows implies that a participant has perseverated in both modes. Experiencing no Z or Z' windows implies that a participant is experimenting or playing randomly.

### 4.1.3 Game Theory-of-Mind Criteria/Attribute 3

The team defined a user win as finishing a game in less than eight moves. The number was derived from the cumulative mass function for games of different lengths. The mean was between seven and eight moves. This attribute examines whether a participant was able to win in normal mode and inverse mode at least once. Z and Z' windows have criteria set to appear consecutively two times if participants continue to perseverate in their responses. By not perseverating (persisting with cooperative or uncooperative moves), participants would only experience one Z or Z' window. This attribute also examines whether users respond to a Z window and finished under the Z and Z' window criteria; it is a measure of whether they changed strategy for the better after a Z or Z' intervention. If participants did change, then we presume they thought about the responses rather than responding as before or playing randomly.

### 4.1.4 Stochastic Index/Attribute 4

This attribute was designed to complement the Z, Z' attribute; it determines how much participants are biasing their games toward optimal or non-optimal play. In normal and inverse mode, the game distributions and the probabilities of game paths appear to follow a normal distribution. Using the beta distribution with our bounds as the shortest and longest game path, and the parameters as the number of cooperative and uncooperative responses, the Cumulative Mass Function (CMF) equals the probability that, given the number of cooperative and

uncooperative responses, a participant is less than or greater than our expected game length median. Therefore, if participants are playing with attentiveness, they will be biased toward the shortest game lengths. If they are playing with less attentiveness, they will be biased toward the longer games (see Figure 4).

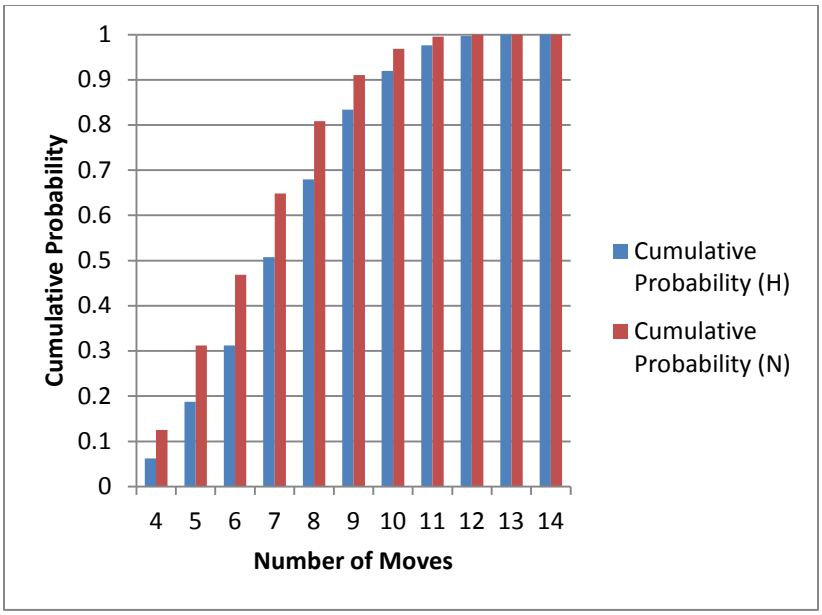



**Figure 4. Stochastic Index**

### 4.1.5 Terminal Mode/Attribute 5

All participants begin play in normal mode. This attribute captures whether their play and strategy allowed them to stay in normal mode, switched and kept them in inverse mode, or allowed them to play through inverse mode and return to normal mode. In a sense, it is a parity bit for the next attribute.

### 4.1.6 Number of Mode Switches/Attribute 6

Persistent play triggers Z, Z' windows and corresponding mode switches. This attribute indirectly captures perseverant play, which could be optimal or non-optimal (depending on the mode users are in), by counting the number of times a participant switches modes.

### 4.1.7 Number of Combined Valence Sequences/Attribute 7

The cyclic nature of the valence sequences allows a transformation algorithm, applied uniformly on the number of unique valence sequences captured, minimizing the valence sequences to an irreducible form. The number of irreducible sequences is then counted.

### 4.1.8 Valence Differential: Counted Number of Sequences vs. Combined Number/Attribute 8

This attribute quantitatively captures the difference between the number of unique valence sequence patterns and the number of their remaining irreducible forms.

### 4.1.9 Response Time for Last Game Block/Attribute 9

By the end of the experiment, participants had settled into a characteristic response time that reflected the degree to which their decisions had been informed by prior experience.

### 4.1.10 Average Number of Moves/Attribute 10

This attribute classifies participants as fast players, medium players, or slow players. To reach maximum game length, a participant must perseverate with uncooperative responses and one cooperative response. Very few participants reached this number.

## 4.2 Models Tested

Different combinations of these attributes were combined to produce a matrix of similarity scores.

- **Model 1** used the first eight attributes.
- **Model 2** used all the variables in Model 1 + Response Time (RT) (Attribute 9) for the last block.
- **Model 3** used all the variables in Model 2 + average number of moves (Attribute 10).
- **Model 4** was based on a regression model and used only four of the original attributes: terminal mode (Attribute 5), number of mode switches (Attribute 6), number of unique valence sequences (Attribute 1), and number of combined sequences (Attribute 7).

## 4.3 Detection Algorithm

The team analyzed each model's performance for its ability to determine if two data sets came from the same user or a different user. Then, for each model, we generated a matrix of similarity scores reflecting the performance of 40 participants. The 80 x 80 triangular matrix included two sets of results per participant (for Day 1 and Day 2). Our goal was to find a system for comparing each data set and determining whether it was from the same user on a different day or from a different user. The steps in this analysis are described below.

### 4.3.1 Step 1: Multi-dimensional Scaling

A model's similarity matrix was submitted to Multi-Dimensional Scaling (MDS) analysis to generate a matrix of inter-point distances. Ordinal MDS solutions were derived using Kruskal's Stress Formula 1 as a measure of fit. The distances in this MDS space reflect the confusability of each dataset. Figure 5 depicts a two-dimensional MDS space derived for Model 1. Each point corresponds to a single data set. Points near each other in space are likely to be confused as the same user, and those far apart in space can be easily discriminated. (We consider the interpretation of the coordinate axes below.) The coordinates of each point feed into the next stage of the algorithm (e.g., each point has an x,y coordinate to position it in the two-dimensional space).

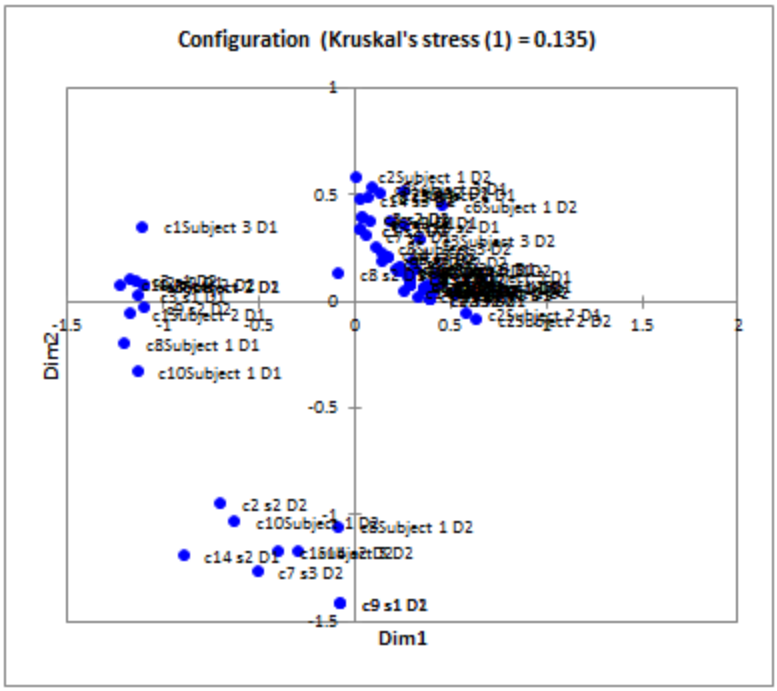



**Figure 5. Two-dimensional MDS Space for Model 1**

### 4.3.2 Step 2: Quantifying Confusability Using Point-wise Distance

Once an MDS space is derived, point-wise distances can be computed using Euclidean distance. The distances can be computed between each data point regardless of the dimensionality (i.e., number of dimensions) of the space.

### 4.3.3 Step 3: Assigning Detection Responses

The point-wise distances are then evaluated against a decision criterion to determine whether they come from the same user or from a different user. The minimum distance is 0, reflecting perfect self-similarity of data points. For the space depicted above, the largest point-wise distance was 2.126. We evaluated decision criteria ranging from .1 to 1 in steps of .1 (e.g., criterion values = .1, .2, .3, … 1 were tested). In other words, when testing a decision criterion equal to .5, any point-wise distance value <.5 received a response of "same" user, while larger values were labeled as a "different" user. This approach allowed us to complete the final step in our analysis, applying signal detection measures to characterize the sensitivity of our model.

### 4.3.4 Step 4: Compute Detection Performance Measures

Each response produced by the algorithm was classified as follows. True positives or "hits" were assigned to cases where the algorithm responded "same" across Day 1 and Day 2 data sets for a given participant (a maximum of 40 hits were possible). If the detector responded "different" for two data sets that came from the same user, this was classified as a false negative or "miss." False positives (i.e., false alarms) were assigned to responses of "same" when data sets came from different participants. Finally, a response of "different" when a pair of data sets came from different participants was scored as a "correct rejection."

The tallies of hit rate and false alarm rate were converted to the standard signal detection theoretic measure of detection sensitivity, d'. Any given model results in a single level of d'. Adjusting decision-criterion values for a model results in different combinations of hit and false alarm rate. The goal is to find a model that maximizes hit rate while simultaneously minimizing false alarm rate. These values trade off, but for any given sensitivity level (which is determined by the mix of indicators going into the similarity matrix), there is a point where we can maximize hit rate while minimizing false alarms. This relationship can be plotted in an iso-sensitivity curve (commonly called a ROC curve). This function allows us to decide on an acceptable level of hits and false alarms, and a corresponding decision criterion to use to achieve the desired level of detection performance.

### 4.3.5 Detection Algorithm Results

Using the algorithm detailed above, we evaluated an initial set of three models comprising seven, eight, or nine measures of task performance (described above). We evaluated each of these models across a range of MDS dimensionalities. Table 1 reports sensitivity measures (d') for the three detectors based on two-, three-, four-, and five-dimensional MDS solutions. The maximum d' achieved was 1.136, based on Detector 2, five-dimensional using a criterion of .2. Averaged across criterion values, the maximum d' level achieved among models 1-3 was .966 (Detector 1, two-dimensional). A d' value of 1 should result in an overall accuracy level of about 68% correct detection responses (ignoring the distinction between hit and correct rejection rates). The results in Table 1 suggest that there was little improvement moving from two-dimensional to three-dimensional or five-dimensional MDS solutions, and there was no appreciable improvement over Model 1 by Models 2 or 3. The best overall detector (i.e., highest sensitivity across criterion values) was Model 4, which is explained below.

**Table 1. Sensitivity Measures for Three Detectors**

| Sensitivity (d') | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Detector 1 | | | Detector 2 | | | Detector 3 | | Detector 4 |
| criterion | 2D | 3D | 5D | 2D | 3D | 5D | 2D | 3D | 5D | 4D |
| 0.1 | 0.872 | 1.036 | 0.877 | 0.852 | 0.952 | 0.827 | 0.764 | 0.973 | 0.962 | 1.082 |
| 0.2 | 0.829 | 0.869 | 1.058 | 0.783 | 0.924 | 1.136 | 0.757 | 0.982 | 1.008 | 0.990 |
| 0.3 | 1.065 | 0.732 | 0.987 | 0.829 | 0.817 | 0.985 | 0.794 | 0.750 | 0.915 | 1.016 |
| 0.4 | 1.006 | 0.921 | 0.854 | 0.899 | 0.873 | 0.966 | 0.848 | 1.011 | 0.954 | 0.968 |
| 0.5 | 0.912 | 0.952 | 1.003 | 0.862 | 0.957 | 0.933 | 0.902 | 0.949 | 1.070 | 1.061 |
| 0.6 | 0.952 | 0.868 | 0.972 | 0.862 | 0.963 | 1.027 | 0.875 | 0.846 | 1.096 | 1.099 |
| 0.7 | 0.928 | 0.890 | 0.957 | 0.822 | 0.875 | 0.916 | 0.836 | 0.860 | 0.876 | 0.935 |
| 0.8 | 1.040 | 0.773 | 0.865 | 0.915 | 0.773 | 0.728 | 0.932 | 0.869 | 0.779 | 0.853 |
| 0.9 | 1.034 | 0.830 | 0.734 | 0.903 | 0.829 | 0.700 | 1.031 | 0.833 | 0.680 | 0.846 |
| 1 | 1.027 | 0.807 | 0.744 | 0.896 | 0.802 | 0.835 | 0.965 | 0.788 | 0.827 | 0.920 |
| average | 0.966 | 0.868 | 0.905 | 0.862 | 0.877 | 0.905 | 0.870 | 0.886 | 0.917 | 0.977 |

Tables 2-5 display the hit and false alarm rates for each detector across MDS dimensionalities. Again, differences across models and number of dimensions were relatively small, with Model 4 providing the best overall performance and a reasonable tradeoff between hit and false-alarm rates.

**Table 2. Hit and False Alarm Rates for Detector 1**

| Hit & False Alarm Rate - Detector 1 | | | | | | |
|---|---|---|---|---|---|---|
| | 2D | | 3D | | 5D | |
| Criterion | HR | FA | HR | FA | HR | FA |
| 0.1 | 0.425 | 0.144 | 0.300 | 0.059 | 0.200 | 0.043 |
| 0.2 | 0.600 | 0.283 | 0.450 | 0.160 | 0.325 | 0.065 |
| 0.3 | 0.775 | 0.378 | 0.525 | 0.252 | 0.475 | 0.147 |
| 0.4 | 0.825 | 0.472 | 0.675 | 0.320 | 0.525 | 0.214 |
| 0.5 | 0.850 | 0.550 | 0.750 | 0.391 | 0.675 | 0.292 |
| 0.6 | 0.875 | 0.578 | 0.775 | 0.455 | 0.725 | 0.354 |
| 0.7 | 0.875 | 0.588 | 0.825 | 0.518 | 0.775 | 0.420 |
| 0.8 | 0.900 | 0.595 | 0.825 | 0.564 | 0.800 | 0.491 |
| 0.9 | 0.900 | 0.598 | 0.850 | 0.582 | 0.800 | 0.543 |
| 1 | 0.900 | 0.600 | 0.850 | 0.591 | 0.825 | 0.576 |

**Table 3. Hit and False Alarm Rates for Detector 2**

| Hit & False Alarm Rate - Detector 2 | | | | | | |
|---|---|---|---|---|---|---|
| | 2D | | 3D | | 5D | |
| Criterion | HR | FA | HR | FA | HR | FA |
| 0.1 | 0.375 | 0.121 | 0.250 | 0.052 | 0.175 | 0.039 |
| 0.2 | 0.550 | 0.256 | 0.450 | 0.147 | 0.350 | 0.064 |
| 0.3 | 0.675 | 0.354 | 0.550 | 0.245 | 0.475 | 0.148 |
| 0.4 | 0.775 | 0.443 | 0.650 | 0.313 | 0.575 | 0.219 |
| 0.5 | 0.825 | 0.529 | 0.750 | 0.389 | 0.650 | 0.292 |
| 0.6 | 0.850 | 0.569 | 0.800 | 0.452 | 0.750 | 0.362 |
| 0.7 | 0.850 | 0.585 | 0.825 | 0.524 | 0.775 | 0.436 |
| 0.8 | 0.875 | 0.593 | 0.825 | 0.564 | 0.775 | 0.511 |
| 0.9 | 0.875 | 0.598 | 0.850 | 0.582 | 0.800 | 0.556 |
| 1 | 0.875 | 0.600 | 0.850 | 0.593 | 0.850 | 0.580 |

**Table 4. Hit and False Alarm Rates for Detector 3**

| Hit & False Alarm Rate - Detector 3 | | | | | | |
|---|---|---|---|---|---|---|
| | 2D | | 3D | | 5D | |
| criterion | HR | FA | HR | FA | HR | FA |
| 0.1 | 0.325 | 0.112 | 0.250 | 0.050 | 0.200 | 0.036 |
| 0.2 | 0.525 | 0.244 | 0.475 | 0.148 | 0.325 | 0.072 |
| 0.3 | 0.650 | 0.342 | 0.525 | 0.246 | 0.450 | 0.149 |
| 0.4 | 0.750 | 0.431 | 0.700 | 0.313 | 0.575 | 0.222 |
| 0.5 | 0.825 | 0.513 | 0.750 | 0.392 | 0.700 | 0.293 |
| 0.6 | 0.850 | 0.564 | 0.775 | 0.464 | 0.775 | 0.367 |
| 0.7 | 0.850 | 0.579 | 0.825 | 0.530 | 0.775 | 0.452 |
| 0.8 | 0.875 | 0.586 | 0.850 | 0.567 | 0.800 | 0.525 |
| 0.9 | 0.900 | 0.599 | 0.850 | 0.581 | 0.800 | 0.564 |
| 1 | 0.900 | 0.624 | 0.850 | 0.598 | 0.850 | 0.583 |

**Table 5. Hit and False Alarm Rates for Detector 4**

| Hit & False Alarm Rate - Detector 4 | | |
|---|---|---|
| | 4D | |
| criterion | HR | FA |
| 0.1 | 0.375 | 0.081 |
| 0.2 | 0.475 | 0.146 |
| 0.3 | 0.575 | 0.204 |
| 0.4 | 0.700 | 0.329 |
| 0.5 | 0.775 | 0.380 |
| 0.6 | 0.825 | 0.435 |
| 0.7 | 0.825 | 0.500 |
| 0.8 | 0.825 | 0.533 |
| 0.9 | 0.850 | 0.576 |
| 1 | 0.875 | 0.591 |

## 4.4 Interpretation of Measures and an Additional Model

To supply some insight for interpreting the contribution of each unique predictor in a model, we conducted a series of regression analyses. Regressing each individual indicator onto the MDS coordinates for a given model indicates the degree to which each is responsible for the configuration of point-wise distances.

The results indicated most model predictors contributed statistically reliably to many of the model configurations. For example, Dimension 1 values from the Model 1 two-dimensional MDS configuration were reliably predicted by the number of unique valence sequences, the stochastic index, the number of mode changes (and mode finished on), and the number of combined valences (all $p<.05$). The strongest predictor (largest coefficient) of values on this dimension was the number of mode changes experienced. For Dimension 2, every model input

variable made a significant contribution to performance (all coefficients $p<.05$). The largest predictor coefficients corresponded to number of mode switches, and the number of unique valence sequences.

For the three-dimensional version of this model, we found virtually the same results pattern. All variables were significant predictors of Dimension 1 values, with number of mode switches the largest contributor. Dimension 2 was predicted strongly by the number of unique valence sequences. For Dimension 2, only the stochastic index and number of combined valences failed to reach significance. Dimension 3 again was predicted most strongly by a combination of mode and valence sequence measures. However, the stochastic index did predict strongly in the five-dimensional version of this model. The response time measure also was a significant contributor to performance in the Model 2 three- and five-dimensional instantiations.

The patterns observed in the regression analyses were corroborated by factor analyses conducted on the same data. Factor loadings from number of mode changes and unique and combined valence sequence measures consistently mapped onto the dimensions as described above. The measure of indicator overlap, though, was relatively low across models (the largest observed value of Cronbach's alpha was .636 in Model 1-3D). This result suggests that many of the predictors made unique contributions to model performance.

Based on these exploratory analyses into the primary contributors to our detection results, we tested an additional model whose similarity values were derived solely from the four indicators suggested most strongly by the regression analyses. Detector 4 included similarity values based on the two mode measures and the two valence measures. If our exploratory analyses were reliable, then we should expect these indicators to be sufficient for detection performance commensurate with that reported above.

Indeed, Model 4 performed better than the other detection models. An MDS dimensionality of four was selected based on a clear elbow at this point in the stress plot for dimensionalities ranging from two to five. The model fit the data extremely well (Kruskal's stress = .022; $R^2$ = .969). Table 5 above indicates that a sensitivity value (d') of .977 resulted for this model (this value is independent of the criterion value). Figure 6 below plots the hit and false alarm rates resulting from a range of criteria using this detection model. The figure also depicts a theoretical ROC curve for the observed sensitivity level (i.e., d'). To make use of this detection model, one must simply decide what level of hit rate is necessary and what false alarm rate is acceptable. Detector 4 is capable of hit rates as high as 88% and was the best performing model we tested.

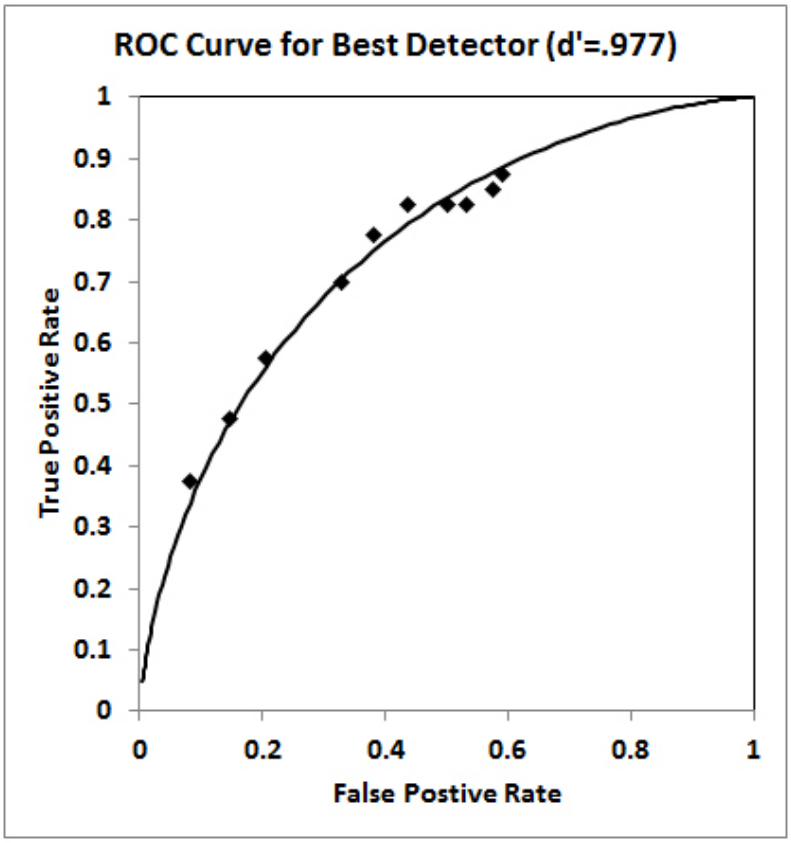



**Figure 6. Hit and False Alarm Rates Resulting from Model 4**

The results illustrate the covert cognitive game method can meet the 80% correct criterion for four models. The question of false positives requires further consideration.

Hit: The detector authenticates an authorized user (true positive).

Miss: The detector fails to authenticate an authorized user.

False Authentication (or alarm or positive): The detector authenticates an unauthorized user (false positive).

True Rejection: The detector fails to authenticate an unauthorized user.

A:= an authorized user

~A:= an impostor (not authorized)

D:= a detection, an authentication

~D: A non-authentication. The detector rejects the user.

All games lead to a D or ~D.

P(D): The probability that the detector will authenticate. This includes Hits and False Authorizations

P(~D) The probability that the detector will *__not__* authorize a user. This includes True Rejections and Misses (false rejections)

P(D|A) := Hit rate

P(D|~A) := False authentication rate

P(A|D) =[ P(D|A)*P(A)]/P(D)  The Bayesian Posterior

The ROC curve was constructed from data that compared every user against all others. It is a worst-case scenario for false alarms because there are only 40 same-user comparisons, whereas there are 40*(39)/2 (= 780) different-user comparisons. The detector captures the degree of within-subject similarity (hit rate) and across-subject (false alarm rate) similarity.

However, the value of a false alarm rate depends greatly on the concept of operations (CONOP). For example, suppose that the hit rate and false alarm rate obtained from our data predict the performance of our detector against 1,000 authorized users, who each make 10 authentication attempts per day. Now suppose one of those 1,000 also tries to impersonate another user three different times during the same day. Setting our criterion to 0.825 hit rate, 0.467 false alarm rate, out of the 10,003 trials, we get 8,250 hits (correct authorizations) and therefore 1,750 misses, defined as our detector rejecting an authorized user). We also get 1.86 correct rejections but 1.14 false authentications. That is, about half the time, the impostor gets through. Lowering the criterion to a hit rate of 0.775 can cut the number of false positives to 1.14, but this reduction comes at the cost of 7,750 hits and 2,250 misses. An authorized user is then failing to be recognized 22.5% of the time. At a hit rate of 0.700, the number of false alarms drops below 1, which means that an impostor would likely fail with only three attempts. Under a different CONOP, with more attempts at impersonation (30 rather than three), the impostor succeeds about one-third of the time. If the impostor is allowed 100 attempts at impersonation, the false alarm rate needs to drop to 0.01 to ensure that an impostor is excluded. This trade-off is inherent to detectors and cannot be ignored.  For the Phase 1 project, the 0.800 hit rate was easily achieved at the cost of higher-than-desirable false alarm rate.

Calculating the Bayesian posterior results in the probability that, if the detector authenticates a user, the user is in fact authorized. For the actual detector, it is well over 99.7%. The high value reflects the CONOP in that it is the ratio of correct detections divided by total detections. For a different CONOP with more attempts at impersonation and with more false alarms, the Bayesian posterior is lower.

The Bayesian posterior for the current detector is still high under the second CONOP because it is the ratio of true authorizations to all authorizations for this scenario (30 attempts at getting in by an imposter and 10,030 total authorizations). With different assumptions, the Bayesian posterior can go from near 0 (if everyone is an impostor) to near 1 (there are no impostors). For future work, two key data are needed: 1) the ratio of attempted impersonations to actual user attempted authorizations; 2) a criterion number of attempts an imposter can make.

A simulated detector with a hit rate of 0.8 and a false alarm rate of .01 yields 8,000 hits (2,000 misses) but only 0.3 false alarms. Therefore, an imposter needs about 100 attempts before he or she succeeds in being authorized, and a user fails to be authorized about 20% of the time. With a better detector (.01 false alarm rate, .85 hit rate, the user fails 15% of the time and the imposter still needs approximately 100 attempts to be authorized).

## 5 CONCLUSIONS

The SwRI team developed a game theoretical approach to authenticating users. After a series of pilot experiments, we conducted a larger, two-day experiment. None of the participants detected the game, although many reported strategizing when they chose how they would respond to the game interruptions. After analyzing the data, we identified 10 attributes that characterized all participants. These attributes were compared in a similarity matrix and then reduced through MDS techniques. Finally, a four-dimensional subspace of the variables was identified and used to construct a ROC curve. The required 80% hit rate was exceeded. The false alarm rate was higher than desired; however, evidence indicates that this can be reduced through further research.

## 6 RECOMMENDATIONS

In future, a CONOP is needed to evaluate the quality of a detector. There are accepted values of quality; however, the final effects are best understood if two constraints are established: 1) the ratio of authentication attempts by legitimate users to attempts at impersonation and 2) the number of authentication failures allowed before a user is locked out. These two parameters can bound both the number of days an authorized user can go without failure and the number of attempts and impersonator needs to be confident of success.

## 7 REFERENCES

Binmore, K. (2007). *Game Theory: A Very Short Introduction*. New York: Oxford University Press.

Micko, H.C. (2000). *Experimental Matrix Games*. http://www.mathpsyc.uni-bonn.de/doc/micko/content.htm

# 8 APPENDIX A ANALYSIS SURVEY QUESTIONS

## Trait Questions

The following questions were asked via face-to-face interviews.

1) When software needs to be installed on your computer, how often are you the one who does the installation? (scale: 1 = never; 5 = always)
2) You open a piece of software and a dialog window opens prompting you to download the latest update for the software. When would you typically comply with this suggestion? (options: never, after a few days, after a few minutes, immediately, I'd wait for a reminder)
3) You decide to allow your computer to download and install the latest version of a software program. During the process, you are prompted to enter the administrator password/permissions. How likely are you to proceed? (scale: 1 = never; 5 = always)
4) A dialog box appears on your screen telling you that a familiar software program (e.g., MS Word) wants to access the internet. How likely are you to give permission? (scale: 1 = very unlikely; 5 = very likely)
5) A dialog box appears on your screen telling you that a software program that you do not recognize or use regularly wants to access the internet. How likely are you to give permission? (scale: 1 = very unlikely; 5 = very likely)
6) If you discover that your work computer is no longer connected to the internet, what would you do?
7) How likely are you to install or check the status of device drivers? (scale: 1 = very unlikely; 5 = very likely)
8) How often do you connect or insert the following media into your work computer during a typical day? (scale: 1 = very rare; 5 = very often)
    a. Flash drives
    b. Memory cards
    c. External drives
    d. CDs or DVDs
9) Which of the following do you use?
    a. Mouse? Roller ball? Trackpad?
    b. How often do you use the right-click functionality on these? (scale: 1 = never; 5 = frequently)
10) Do you set the wallpaper on your computer?
11) Do you modify the screensaver on your computer?
12) Do you adjust the cursor blink rate on your computer?
13) Do you adjust the keyboard repeat delay (time to hold key before repeat)?
14) Do you adjust the keyboard repeat rate on your computer?
15) Do you adjust the mouse pointer speed on your computer?
16) Do you set the mouse double-click speed on your computer?
17) Do you set the mouse button configuration on your computer?
18) Do you add shortcuts to the desktop?
19) Do you add shortcuts to the start menu?
20) Do you add shortcuts to the Task Bar?

21) Do you use the function keys?
22) Do you adjust the volume setting?
23) Do you ever plug in headphones?
24) Do you adjust the screen brightness?
25) How often do you use Microsoft Word? (options: daily, a few times a week, a few times a month, very rarely, never)
26) Rate your experience level using Word? (scale: 1 = beginner; 5 = advanced)
27) In Word, do you use the document review options (e.g., Track Changes)?
28) In Word, do you use the outline view to organize documents and display different levels?
29) In Word, do you use the Hidden Text option?
30) In Word, how likely would you be to use the Options menu to modify which buttons appear on the screen? (scale: 1 = very unlikely; 5 = very likely)
31) In Word, do you customize the Quick Access bar located at top left of the window?
32) How often do you use Microsoft Excel? (options: daily, a few times a week, a few times a month, very rarely, never)
33) Rate your experience level using Excel (scale: 1 = beginner; 5 = advanced)
34) Do you use keyboard shortcuts in Excel? (scale: 1 = never; 5 = frequently)
35) In Excel, do you create or modify Macros? (scale: 1 = never; 5 = frequently)
36) In Excel, do you use filtering functions? (scale: 1 = never; 5 = frequently)
37) In Excel, do you use pivot tables? (scale: 1 = never; 5 = frequently)
38) In Excel, do you use add-ins via the Tools menu? (scale: 1 = never; 5 = frequently)
39) In your internet browser, do you use options like setting your homepage?
40) In your internet browser, do you use the option to clear history?
41) In your internet browser, do you use options like managing add-ons?
42) In your internet browser, do you make changes to security settings?
43) If your browser prompts you to set a default search provider or use an accelerator, do you do this?
44) How often do you use Windows Explorer to access and organize files and folders? (scale: 1 = never; 5 = frequently)
45) How often do you use the Windows Control Panel to make changes like adding/removing applications, adding devices, etc? (scale: 1 = never; 5 = frequently)
46) How often do you use Windows performance monitoring tools via Windows Task Manager? (scale: 1 = never; 5 = frequently)
47) How much time do you spend per day using the keyboard to type up documents? (options in hours: < 1, between 1 & 2, between 2 & 3, more than 3)
48) How much time each day do you spend searching for information via the web? (options in hours: < 1, between 1 & 2, between 2 & 3, more than 3)
49) How many windows do you tend to leave open on your screen at once?
    a. Usually one or two
    b. Usually more than three

50) How many windows do you tend to leave open, but minimized?
    a. Usually one or two
    b. Usually more than three
51) Which of the following are you most likely to do while browsing the web? (options: open additional tabs, open additional overlapping browser windows, use a single window)
52) How much time would you say you spend focusing attention on your computer screen each day? (options in hours: < 1, between 1 & 2, between 2 & 3, more than 3)
53) What are the top five things you do on your computer during a typical day at work?
54) Rate your level of agreement with the following statements:
    a. I tend to work with a small set of applications for long periods of time (e.g., Word, Excel, PPT). (scale: 1 = disagree; 5 = completely agree)
    b. I interact directly with my computer only periodically throughout the day in short bursts (e.g., to respond to e-mail, to look up information online, occasionally write a note or edit documents).(scale: 1 = disagree; 5 = completely agree)
55) How many monitors do you use on your computer?
56) Do you feel that you have enough screen space for all the applications you have open at a time?
57) While on your computer, how often do you feel like you are rushing to get your work done? (scale: 1 = never; 5 = regularly)
58) If you encounter a printing problem while using an application such as Word (and the problem is not the printer itself), which of the following are you most likely to do?
    a. Try to resolve it myself
    b. Ask someone how to solve it (other than tech support)
    c. Wait for tech support
59) You open Word and discover that the page view is different from how you normally like it. How do you typically respond?
    a. Try to change the settings myself
    b. Ask someone how to make changes
    c. Ask tech support
    d. Live with it the way it is
60) An application on your computer is not working properly or as you expect. How would you tend to respond?
    a. Troubleshoot the problem myself
    b. Ask someone else to look into it
61) How often do you use keyboard shortcuts to perform the following tasks? (scale: 1 = never; 5 = frequently)
    a. Copy/cut/paste
    b. Save a file
    c. Open a file

d. Switch between windows (alt+tab)
e. Other (please specify)

62) When performing the following actions inside an application, which are you more likely to do (excluding using keyboard shortcuts): click on an icon or use a drop-down navigation menu?
    a. Saving a file
    b. Opening a file
    c. Printing a file
    d. Other (please specify)

63) What are some of the most frustrating things you experience while using your computer at work?

64) Are there any instances where your computer does something that makes you happy or relieved?

## Daily Diary Questions

The following questions were asked via a web-based survey twice per day for one week.

1) How many separate applications did you use during the last 3-4 hours? (options: 0-2, 3-4, 5-6, more than 6)
2) What was the maximum number of applications you had running simultaneously? (options: 0-2, 3-4, 5-6, more than 6)
3) How many applications did you have open for more than 5 minutes during this period? (options: 0-2, 3-4, 5-6, more than 6)
4) Which applications were open for more than 5 minutes?
5) Please estimate how much time you spent focusing your attention on each of these.
6) During this period, how many applications did you open and use for less than 5 minutes before closing? (options: 0-2, 3-4, 5-6, more than 6)
7) Which applications were open for less than 5 minutes?
8) In the applications that you opened, were there any warnings generated by the application that you had to respond to?
9) If yes, please describe.
10) Were there any Windows Operating System warnings you had to respond to?
11) If yes, please describe.
12) Did you get any prompts to update software?
13) If so, how did you respond? (options: closed dialog box, left dialog box open but ignored, chose "no," chose "remind me later," chose "ok" to update)
14) Did any applications make recommendations (e.g., go here for help)?
15) If so, please describe.
16) Were there any messages displayed that were difficult or impossible to understand?
17) Were there excessive scrolling requirements?
18) Did you notice any unusually long loading times?
19) Did you receive any administrative permission prompts?
20) Did your computer cause any confusing or annoying events during the last 3-4 hour period?
21) If yes, please describe.

# 9 LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

| | |
|---|---|
| AFRL | Air Force Research Laboratory |
| CMF | Cumulative Mass Function |
| CONOP | Concept of Operation |
| DARPA | Defense Advanced Research Projects Agency |
| DoD | Department of Defense |
| LLC | Limited Liability Company |
| MDS | Multi-Dimensional Scaling |
| ROC | Receiving Operating Characteristic |
| SwRI | Southwest Research Institute |
| TfT | Tit for Tat |