Final Report

Grant AOARD-12-4076

Effects of Cognitive Load on Trust

October 2013

Dr. Fang Chen

National ICT Australia (NICTA), Australia

Address: Level5, 13 Garden Street, Eveleigh,
           NSW 2015, Australia
Email: Fang.Chen@nicta.com.au
Phone: +61 9376 2101
Fax: +61 9376 2027

	Form Approved
# Report Documentation Page	OMB No. 0704-0188

Public reporting burden for the 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 Washington Headquarters Services, Directorate for Information Operations and Reports, 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 a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **13 NOV 2013**	2. REPORT TYPE **Final**	3. DATES COVERED **29-08-2012 to 28-08-2013**
4. TITLE AND SUBTITLE **Effects of Cognitive Load on Trust III**		5a. CONTRACT NUMBER **FA23861214076**
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Fang Chen**		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National ICT Australia Limited,Level 5, 13 ,Garden Street ,Eveleigh Australia,AU,2615**		8. PERFORMING ORGANIZATION REPORT NUMBER **N/A**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **AOARD, UNIT 45002, APO, AP, 96338-5002**		10. SPONSOR/MONITOR'S ACRONYM(S) **AOARD**
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) **AOARD-124076**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project focused on the relationship between cognitive load and trust judgments, and the effect of cultural differences in the way trust judgments are made. A literature review is contained in the report, examining differences in cultures. Studies where participants? verbal comments were recorded under different cognitive loads were conducted in Australia. Voice recordings were analyzed for factors such as word choice, pause frequency and duration to find evidence of change in trust behaviors under different cognitive load conditions. Future work is planned to conduct similar studies in the US and Malaysia for comparison to these results.

15. SUBJECT TERMS
Trust, cognitive load

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**74**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Executive Summary

This report presents the research activities undertaken and outcomes achieved by NICTA team as part of the "Effects of Cognitive Load on Trust" project in conjunction with the US AFRL and Sunway University Malaysia. NICTA's role comprised the measurement and assessment of cognitive load through speech and other interaction modalities. The project was focused on the examination of the relationship between cognitive load and trust judgments, and the effect of cultural differences in the way trust judgments are made. The detailed literature review was carried out in this regard, which is detailed in the first section of this report.

The second major milestone in this project was the completion of the planning and design of the user study for all three sites (US, Malaysia and Australia), including coordinating multi-national IRB approval for the experiments. The third milestone involved the implementation of the software for data collection, and the production of supplementary materials to be used in the study. The outcomes for this part of the work are described in the second section of the report. The forth milestone was the execution of the study, which was divided into three data collection phases, one for each site (Australia, US and Malaysia). The data collection summary from all three sites and outcomes achieved is also included in the report.

Finally, detailed data analyses were carried out for the purpose of understanding the relationships between cognitive load, trust (and trustworthiness), and culture. Analyses were conducted to assess the users' experienced cognitive load under various task conditions and analyze the effects of cognitive load on their trust perception. The last section of the report discusses the data analyses conducted and the results of those analyses.

Table of Contents

List of Figures

List of Tables

1. Introduction

Trust is found to be a critical element driving human behavior in both interpersonal and computer-based interactions. It has been realized as one of the most important factors in organizational behavior for all personal and/or computer-supported decision making and task performance (R. Mayer, Davis, & Schoorman, 1995; Schmorrow & Stanney, 2008). Trust is defined as "one party's (or individual's) willingness to accept the vulnerabilities of actions or behavior of the other party based on the expectation that the other will perform the actions important to the trustor" (Dunn & Schweitzer, 2005; R. C. Mayer & Davis, 1999; R. Mayer et al., 1995). Trustworthiness on the other hand is different from trust. Mayer et al. (R. Mayer et al., 1995) found three trustworthiness elements that influence the development of trust in interpersonal situations: ability, benevolence, and integrity. Thus far, only a few studies have looked at how different situational factors influence trust development as reflected in the relative salience of the three trustworthiness indicators. One dominant situational factor that may shape trust perceptions of an information source is culture. Similarly, little is known how cognitive load may affect the different trustworthiness factors during trust development and acquisition.

This 3-year research project was part of a larger international research effort in collaboration with Dr. Lyons and Dr. Stokes (AFRL), and Dr. Lin (Sunway University, Malaysia), with separate proposals submitted through the AFOSR/AOARD programs. A three-part user experiment was designed – one in the US, one in Australia, and one in Malaysia, to investigate the cross-cultural and cognitive load influences on trust. The data collection at all sites has been completed but the data analysis has been completed only for the Australian site. The detailed data analysis of the Australian data and results are available which are being presented in this report.

In the following, project milestones and outcomes are presented followed by a detailed literature review on the aspects of cognitive load and culture and their influence on trust judgments. The experiment and task design, software implementation for the experiment, study execution and data collection are also discussed, followed by various analyses to assess people's experienced cognitive load under various task situations and to examine the effect of cognitive load on their trust perceptions and their overall task performance.

2. Project Milestones and Outcomes

Table 1 presents the milestones and outcomes that were achieved during the course of this project. Despite certain minor changes in the project plan and also an unanticipated replacement of a team member from NICTA, the project completed on time.

Table 1: Project Milestones and Outcomes

Milestone	Deliverable/Outcome	Status
Complete Pre-Pilots (Materials)	• Pilot test the neutrality of the stimulus data to be used in the experiment • Stimulus material in target demographic (Australian) • Make changes to the stimulus material as appropriate to ensure neutrality	Achieved
Experiment Tool Design	• Development of the experimental application to be used • Implement factor manipulations, incl. cognitive load • Implement data collection functionality as part of the design	Achieved
Complete Pre-Pilots (Study Design)	• Conduct pilots on target demographic (6 participants) • Evaluate study design, procedure, physical setup • Assess changes needed at each site	Surpassed
Complete Experimental Study	• Source participants • Run the study • Debrief participants	Achieved
Subjective and Performance Analysis	• Ground truth analysis • Subjective ratings analysis • Performance analysis	Achieved
Linguistic Analysis of Speech Data	• Prepare speech transcriptions and annotations • Run linguistic analyses on text data derived from speech • Report results	Achieved
Signal Analysis of Speech Data	• Collect speech data from other sites • Segment, annotate and label speech data • Build speech models to represent cognitive load levels. • Report results to rest of the team	Achieved
Interactive Data (Mouse movements) Analysis	• Develop features that may be affected by load • Build a parsing tool to extract relevant features • Statistical analysis of results (by load components)	Achieved
Final Year Report	• Produce Final Year report on findings, data summaries and conclusions	Achieved

Milestone	Deliverable/Outcome	Status
Project Management	• Weekly meetings • Team workshops, including conference calls with co-investigators • Year-end final report circulated to AOARD office and all other investigators	Achieved

3. Literature Review

The overall objective of this joint research was to evaluate the relative strength of Mayer's trustworthiness indicators (ability, benevolence, and integrity) in both a collectivistic and individualistic culture. To date, there have been no such studies that have empirically tested this model in different cultures. Additionally, this research also examined the relationship between cognitive load (CL) and trust characterized by the trustworthiness indicators. Another objective was to see whether this CL-trust relationship is affected by cultural factor.

Mayer et al. clearly distinguished trust from other similar constructs like cooperation, confidence, and predictability (R. Mayer et al., 1995). Mayer argues that you can still cooperate with someone even if you don't trust them because of some other external factors, e.g. due to a fear of some kind of punishment. He also argues that prior recognition of risk in case of trust is what differentiates it from confidence. Also, trust and predictability cannot be equated because you cannot trust a party when the party is expected to constantly ignore others and act only for its own interests, just because the party is predictable. Mayer argues that "what is missing from such an approach is the willingness to take a risk in the relationship and to be vulnerable" (R. Mayer et al., 1995).

Mayer also presented characteristics of a trustor and a trustee. The trustee's characteristics of ability, benevolence, and integrity, form the elements of trustworthiness. Benevolence is the perception of a positive orientation of the trustee toward the trustor (R. C. Mayer & Davis, 1999; R. Mayer et al., 1995), e.g. if an employee believes that the supervisor cares about the employee's interests, then the supervisor is considered to have benevolence for the employee. While on the other hand trustor's propensity (or the expectation about the trustworthiness of others or general willingness to trust others) also affects one party's trust in the other party. Mayer proposed that trust is "a function of the trustee's perceived ability, benevolence, and integrity and of the trustor's propensity to trust". Additionally, they also emphasized that the level of trust (as determined by the three trustworthiness factors) will also be affected by the contextual factors such as third parties involved, level of authority of a party in decision making, level of risk perceived, and any alternative options available to trustor.

Some studies argued that the relative effect of the elements of trustworthiness will vary depending on the type of activity to be performed and also by the person's cultural background (Lyons & Stokes, 2010). For example, when a person is required to take some decision or perform some action based on his/her trust judgment about another person or entity then ability or competency of the other person or entity becomes the most important

factor in establishing the trust perception. While on the other hand, when a person is only required to make a judgment about another person or entity without the need to take some decision or perform some personal action based on that judgment, then integrity is regarded as the most crucial factor in the trust perception. This was also confirmed in a study conducted in an organization for a performance appraisal purpose, where employees were found to regard integrity of their top management executives as the most important factor for their trust (R. C. Mayer & Davis, 1999). For the third element of trustworthiness, benevolence, it is expected that it would be the most predictive factor in describing some feelings about a person, e.g. when a person is asked about who they think would be a good friend, benevolence would be the most emphasized factor (Lyons & Stokes, 2010).

Various cultural backgrounds may also affect the level of trust differently. For example, one study expected that people belonging to individualistic cultures (e.g. Western culture) will weigh ability of the trustee more than any other element to judge the trustworthiness of the trustee. On the other hand, people coming from the collectivistic cultures (e.g. Eastern or Asian cultures) will emphasize on trustee's benevolence more than any other factor to determine the level of trust (Lyons & Stokes, 2010). Another study found similar results about the effects that various national and international cultural values can have on people's trust perception (Schoorman, Mayer, & Davis, 2007). The study discussed that people coming from task-oriented cultures establish initial trust more rapidly focusing more on the task completion, as compared to people from relationship-oriented cultures who take more time to establish a relationship before working on the task. With respect to the elements of trustworthiness, the study also discussed that action-oriented and performance-oriented cultures put more value on a person's ability, while collaborative and relationship-oriented cultures emphasize more on the person's benevolence.

3.1.   Cognitive Load and Trust

Cognitive load is a key component of the four-stage model of human information processing (Paas, Tuovinen, Tabbers, & Gerven, 2003). It is the amount of mental demand imposed by the task at hand and has been associated with the limited capacity of working memory and processing of novel information (Baddeley, 2003; Paas et al., 2003). It is clear that, like trust, cognitive load plays an important role in mediating human behavior during collaboration with other humans or automated systems. Several behavioral changes may occur in people's natural work-related task performance due to varying levels of cognitive load induced by various task and non-task related factors. During high cognitive load, they might feel stressed or frustrated with the work situation, which may result in decreased task performance as well as lower level of trust perception for the people responsible for providing such work environment. An understanding of how situations of higher cognitive load affect people's trust perception, therefore, is a critical factor for any organization in order to provide appropriate support to them and hence improve their overall task performance.

Despite their importance, little is known about the relationship between the two – trust and cognitive load – in such contexts. Intuitively we might guess that as cognitive load increases, a person may choose to rely more heavily on colleagues or an automatic system and in the

process display more indicators of an implicit trust of the system. Alternatively, under higher load the person may be reticent to increase their trust and instead adapt their strategy to manage the increased task complexity to avoid increased dependence on others or on an automatic system. Although the authors note that the empirical evidence is limited, Parasuraman & Riley (1997) argue that increased workload is often cited as one of the most important factors in choice to use automation. However, in a comparison of trust in various levels of automation, including manual control, Ruff and colleagues (Ruff, Narayanan, & Draper, 2002) found that as workload increased, subjective reports of trust decreased for automation, but increased for manual control. In an explanation for the equivocal evidence for trust and automation use, Parasuraman and Riley (1997) suggest that complex task domains may prompt different task strategies, such as use of automation during high cognitive load even if trust is low. Thus, under high cognitive load, use of automation and trust (subjectively measured) may not be aligned, as is often assumed. One of the aims of the current study is to clarify the equivocal results of previous research through the examination of the fundamental trust process (i.e., not simply use of automation) under varying levels of cognitive load.

Previous investigation of the effect of cognitive load in user interfaces suggests the entrenchment of established behaviors with increasing load (Ruff et al., 2002). Other research also suggests that with increased cognitive load users revert to older, over learned and simpler types of responses (Oviatt, Coulston, & Lunsford, 2004). When a user has a pre-existing trust of an automated system, the implication is that they will tend to over-trust the system during higher cognitive load (Biros, Daly, & Gunsch, 2004), and a similar result might be expected of another human agent rather than a system. The Affect Infusion Model (Forgas, 1995) suggests that during faster processing, individuals use their affective states as a short-cut to infer their evaluative reactions to a target, such as in judgments of trust. Thus, it is likely that similar reliance on established cultural values and attitudes may rise to the surface during higher cognitive loads. With respect to the elements of trustworthiness, it has been found that action-oriented and performance-oriented cultures put more value on a party's ability, while collaborative and relationship-oriented cultures emphasize more on the party's benevolence.

Based on the theories proposed by Lee and See (2002), and Mayer (1995), an initial conceptual model of trust developed by Dr. Stokes (AFRL) for our research is shown in Figure 1. In this model, trust is an attitude that drives behavior through intention. While trust is affected by the trustworthiness indicators perceived under the influence of various cognitive load levels, at the same time, cognitive load is affecting behavior as a main cognitive constraint using competing resource in the mental space. Due to the error feedback loop of past experience, trust can also be affected dynamically by the fluctuation in cognitive load.
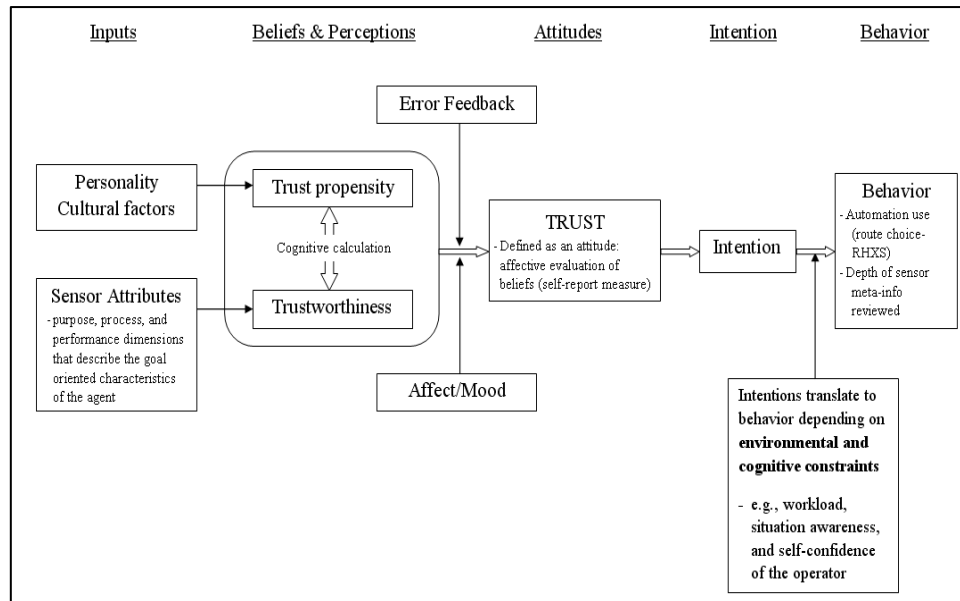
Figure 1: Proposed conceptual model of trust

Based on the literature review conducted by the team, to the best of our knowledge, there are no prior studies that attempted to investigate any relationship between trust and the construct of cognitive load in detail. A few studies have attempted to isolate the effects of high workload and stress on the level of people's trust judgments but their focus was more on the people's trust perception of organizations and of automation systems, e.g. R. C. Mayer & Davis, (1999), and Biros et al., (2004). The former study showed that when good performance is overlooked by an appraisal system in a job environment, employees develop a lack of trust in their employer. On the other hand, when they felt that the appraisal system was fair, their trust for top management increased and they regarded integrity as the most important factor for this trust perception. The study was based on the standard elements of trust and trustworthiness, i.e., the ability, benevolence, integrity, and trustors' propensity as proposed by Mayer et al in 1995 0. In the second example, Biros, Daly, and Gunsch, (2004) presented a study where the objective was to see how people's usage of and dependence on system automation (in other words their trust in system) changes when they experience high task load, especially under information uncertainty situations. It was found that when task load (and hence cognitive load) increases, people continue to rely on the (interaction and decision support) system, even if they have less trust in it. In one study, the researchers found that under high load and critical task situations, people trust the system more when the system behaves in a polite manner and depicts accepted etiquette norms (Raja Parasuraman & Miller, 2004). They also found that people show increased trust in the system they use when the system shows and maintains its reliability and dependability (Raja Parasuraman & Miller, 2004). The study also discussed that as the system's automation becomes more complex (hence causing higher cognitive load), the user's ability and willingness to learn and experience details of the system behavior also decreases and they tend to rely more on the system. The study also discussed that if we can make the system behave similar to the people the users find trustworthy, it will increase their level of trust in the system. The study concluded with the note that high task load along with inappropriate trust level can increase user's mental and cognitive workload (Raja Parasuraman & Miller, 2004).

Both trust and cognitive load, in their individual capacity, are also known to affect the human behavior including their communication and/or linguistic behavior. For example it has been found that less polite language or communication negatively affects people's trust, while a more polite language improves their trust perception about the communicator (R. Mayer et al., 1995; Raja Parasuraman & Miller, 2004). Other studies found that under high trusting conditions, people repeat each other's linguistic expressions or use similar words, i.e. they show high linguistic mimicry and similarity (Scissors, Gill, Geraghty, & Gergle, 2009; Scissors, Gill, & Gergle, 2008). People also use non-linguistic elements of speech like using more emphasis during their communication (i.e. high pitch and volume) when they trust more (Waber, Williams, Carroll, & Pentland, 2012). Like trust, cognitive load also affects people's linguistic behavior. Several recent studies have shown that high cognitive load causes people to show certain linguistic patterns in communication, especially when they work in a team environment, for example, increased use of pausing, hesitations, and self-corrections (Berthold & Jameson, 1999; Jameson et al., 2009; Khawaja, Ruiz, & Chen, 2008), increased use of negative emotions, more disagreement, decreased use of positive emotions, certain patterns of personal pronouns, and many other linguistic indicators of high cognitive load (Khawaja, Chen, & Marcus, 2010, 2012).

While many of the studies have looked at the trust and cognitive workload separately, only a few have attempted to establish a relationship between the two. Those studies that did try to find a relationship between trust and workload, failed to provide a detailed account of the relationship. Namely, the question of how various levels of cognitive load would affect the trust perception or trustworthiness still remains to be answered. With emerging technologies being available for real-time, automatic, and non-intrusive measurement of cognitive load, this could provide a new dimension into understanding trust judgments in work situations, or during complex problem solving.

4. User Study Design and Materials

Based on our detailed literature review and using the theoretical model of trust presented in Figure 1, we designed and conducted a user study to investigate the relationships between the concepts of trust, cognitive load, and human behavior, especially with respect to decision making. For the purpose we designed a job applicant screening task to be performed by a person under different cognitive load conditions with various trustworthiness elements embedded into the applicant's job profile. The objective was to assess how people's trust perception varied for people having various trustworthiness elements and under various complex workload or cognitive load situations.

4.1. Hypotheses

In order to gain an insight into such relationships, we proposed the following hypotheses concerning the effects of cognitive load and trust:

1. For a fixed level of trustworthiness, increased cognitive load situations are more likely to affect a person's thought process negatively resulting in their certain speech communication and linguistic patterns as compared to low cognitive load situations.

2. For a fixed level of trustworthiness, increasing the task complexity (implicitly cognitive load) will affect both the likelihood of a person to rely more heavily on others and the degree of trust they invest in them.

3. For a fixed level of task complexity, varying the trustworthiness of others will affect both the likelihood of the person to rely more heavily on them and hence the degree of cognitive load they perceive during the task.

4. For a fixed level of task complexity, increasing the trustworthiness of others will result in people's increased efficiency in terms of their decision making and task completion performance. With decreased trustworthiness, decision making efficiency will also decrease.

5. Cultural factors will affect the interdependence of cognitive load and trust, such that cultural biases in trust will be exacerbated under high cognitive load.

6. For a fixed level of trustworthiness, increased cognitive load situations are also likely to affect a person's interaction behavior with the system resulting in certain interaction patterns as compared to low cognitive load situations.

7. Finally, significant correlations will be found between various levels of task complexity (and hence experienced cognitive load) and people's self-reported perception of their cognitive load (subjective ratings).

4.2. *Design: Independent Variable and Repeated Measures*

To produce higher cognitive load tasks, a dual-task paradigm was employed. Subjective ratings of complexity and difficulty were employed after each task set, to ensure that the desired levels of load built into the task design were actually being perceived by the study participants. The full set of all experimental conditions for a given cultural factor can be seen in Table 2.

Table 2: Examination of interdependence between trust and cognitive load: experimental conditions.

		Trust/Trustworthiness	
		Low	High
Cognitive load	Low	Single task, low induced trust	Single task, high induced trust
	High	Dual task, low induced trust	Dual task, high induced trust

In terms of cognitive load measurement, the pair of high-trust tasks allowed a 'control' condition, where we expected to see variation due only to high cognitive load that matches previously observed results (e.g. Shi, Ruiz, Taib, Choi, & Chen, 2007). The task design ensured that multiple methods of cognitive load measurement are available, in particular, recordings of participants' speech and language and logged keystrokes/mouse movement

(behavioral measure) and the aforementioned self-ratings (subjective measure). The study was designed with a 2 (cognitive load, w/n) x 3 (AIB indicators, w/n) x 3 (culture, b/w) mixed design.

All subjects completed both load conditions (low and high load) in a repeated measures design because the implicit load measures are dependent on a baseline to high load comparison. Expected changes in recorded behavioral data features of load will trend one way (e.g. increased pauses in speech during think-aloud of high load tasks).

4.3.   Experimental Platform

The experimental platform simulated a computer-based applicant screening process called the "Human Resources Applicant Selection Tool". Participants were told that they were participating in a user evaluation of a new virtual interview tool being considered for a business application.

Participants assessed potential job candidates and reviewed the applicants' virtual resume which included standard experiential data (i.e., education, previous experience, skills, etc.), interest statements, and referential data provided by previous supervisors. The aim was that the application would have a similar look and feel to that of Facebook or Linkedin. After several design discussions, story boarding and wireframe iterations, the final application was produced and is illustrated in Figures 2-18. Candidate applicants' ability, integrity, and benevolence (AIB trustworthiness indicators) were manipulated through referential data inserted into the tool as well as through narratives provided. Each applicant was described by previous supervisors or co-workers as being high or low on one of the trustworthiness indicators.   Examples of vignette-like descriptions of the trustworthiness indicators from previous research (see R. C. Mayer & Norman, 2004) were adapted for use in the current study. Four applicants were presented for each low or high load session: 1 high benevolence candidate, 1 high ability candidate, 1 high integrity candidate and 1 neutral candidate on all three aspects. For each participant, a total of eight applicants were presented in both the sessions.

For the cognitive load manipulation, a secondary monitoring task was introduced, known as the notification feature, which allowed subjects to receive and "queue" new incoming resumes and applications to be "processed later". This was presented as an additional feature of the tool – and the participants were asked to complete two sessions of the task set, with and without the notification feature. The candidates provided in each session were different instances, but represented the same AIB trustworthiness manipulations such that the entire task was exactly the same except for the notification feature as a dual task that resulted in the high cognitive load under that session.

4.4.   Pre-Screening Survey and Training

The pre-screening survey was implemented using Survey Monkey, over the web. It consisted of 13 questions with 91 items altogether. The full Pre-Screening survey used can be found in the Appendix to this report.

The training video was a 6-minute animated instruction manual on how to complete the tasks. Some still shots of the storyboard of the video have been included in the Appendix also.

4.5.   Task Structure

In each session, the first part involved participants filling out a mood questionnaire to ascertain how they felt at that moment, to be able to check later whether their current emotional state affected their judgment of the candidates. Participants were asked to select their level of emotion for a number of different emotions, along a semantic differential scale, ranging from "Not at all" to "Extremely", using radio buttons. As they were completed, a green tick appeared at the end of the row, denoting the question has been answered.

Initial Survey

Using the following scale, indicate to what extent you feel this way right now, that is, at the present moment.

	Very Slightly or Not at All	A Little	Moderately	Quite a Bit	Extremely
Interested	○	○	○	○	○
Excited	○	○	○	○	○
Enthusiastic	○	○	○	○	○
Alert	○	○	○	○	○
Determined	○	○	○	○	○
Distressed	○	○	○	○	○
Upset	○	○	○	○	○
Irritable	○	○	○	○	○
Nervous	○	○	○	○	○
Jittery	○	○	○	○	○

Figure 1: Initial Mood Survey

Immediately following this, three subtasks were performed on the available candidates.

4.5.1. Subtask One: Screening Candidates

After reading a brief vignette about each candidate applicant, participants were asked to rate their degree of trust for each. The first subtask consisted of answering a series of survey questions about the candidate by clicking on a radio button on a semantic differential scale, as to whether they agree or disagree with a statement being made in relation to this candidate applicant. One by one, they were asked to review the candidates under each of the four tabs on the screen, paying particular attention to the comments included in each section (Education, Experience and Volunteer Experience and Personal Interests, as seen in Fig. 2).

The comment areas were formatted to stand out on the profile page, and the rest of the applicant information was very similar for all candidates (e.g. most had around the same amount of experience and education). Essentially, the differentiators could only be found in the comments from previous employers, supervisors and peers for that candidate – this was where the AIB manipulations were made. Additionally, the applicants were "anonymous" –

they were given code names such as "Applicant 68K" to avoid any bias that may be introduced in the participants upon learning the candidate's name. Icons were used instead of photographs also to avoid any bias relating to the candidate's appearance.

Figure 2: Applicant Vignette including Education, Experience and Personal Interests

As the participant completed the survey questions, a green tick mark appeared at the end of each row, signaling to the participant that that question had been filled in, as seen in Fig. 3. Once all the questions on each page were filled in, and all green ticks appeared on the right hand side of the table did the "Next" button become activated, to allowing the participant to continue moving onto the next page.

Figure 3: Green ticks showing progress through the survey screening questions

4.5.2. Subtask Two: Filling Candidates into Positions

Once the subject completed the review of all four applicants individually, they were asked to bin the applicants in three categories: 1) decision/action oriented - whom the participants would most trust as their boss, 2) relationship oriented - whom the participants would most trust as their friend/colleague, and 3) judgment oriented - whom the participants would most trust to be someone else's supervisor (but not the participant themselves).

Figure 4 shows the second subtask, filling in the positions of supervisor, co-worker and colleague from the available candidates. In order to facilitate this, summaries of the candidates' profiles were presented on a single page, each showing the information in the "comment" areas of the individual profiles in the previous subtask. The profiles were presented in the same order as the tabs in the previous subtask, but arranged on a grid from top to bottom and left to right.

The candidate profiles could be dragged to the positions shown on the right hand column of the window, as the participant chooses. Once the candidate is dragged to a position, the profile summary is greyed out to indicate that profile cannot be chosen for any other position. The participants were also required to fill in some information about how confident they were that they were choosing the right person for the position, along a semantic differential scale of 5 points using a drop down box, from "Extremely confident" to "Very little confidence" in their choice for this position.

Figure 4: Fill Positions for Supervisor, Co-worker and Other's Supervisor

Figure 5: Filling in a candidate for a Supervisor position, with moderate confidence

Once all positions are filled (since there are three positions, only three out of four candidates can be selected), the continue button is activated, as seen in Fig. 6 and the participant is then asked to type in a justification for their choice. A pop up dialogue box is provided for this purpose, and allows the participants to explain their reasons for choosing each candidate for each role, and why they judged them to be the best fit. Once their reasons have been typed in, they can hit the Save button on the dialogue box and move on to the next subtask. This process is illustrated in Fig. 7.

Figure 6: All positions filled, Continue button activated
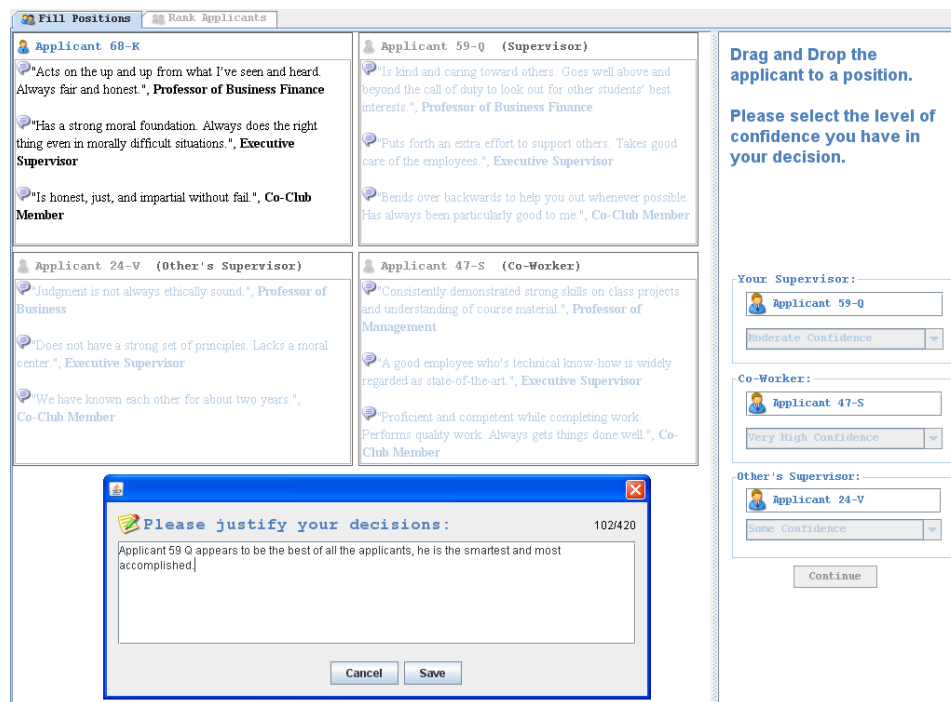
Figure 7: Justification Pop-Up Dialogue Box

4.5.3. Subtask Three : Ranking Candidates

The last subtask involves ranking the candidates in terms of suitability for each position. This means that all four candidates are placed in order from most suitable to least suitable *for each position*. The rankings are made for Supervisor, Co-worker and Other's Supervisor in that order, as can be seen in Fig. 8, 10 and 11. Once all candidate profiles have been ranked, the "Continue" button is activated and the participant can move on from this subtask.

Figure 8: Ranking for positions - Supervisor

Figure 9: Ranking for Positions – Supervisor, with initial selections filled

Figure 10: Ranking for the Co-worker/Colleague position

Figure 11: Ranking for the Other's Supervisor position

4.5.4. Task Review

The final section is a Task Review, where some metrics are provided about the number of applicants that have been rated, selected for positions and ranked in the above tasks. The "Future Tasks Queued" section is only used in the high load session, where the Notification feature is activated. The participants' final score will be shown here, namely, the number of "Human Resources" notifications that were spotted and added to the queue.

Figure 12: The Task Review window

At the end of each session, the participants were asked to rate on a Likert scale, with a radio button, how difficult they found the session. We expected that they would rank the session with the Notification feature (and hence causing high cognitive load) to be relatively more difficult than the session without the Notification feature.

Figure 13: Subjective Difficulty Likert Scale

Fifty-percent of subjects completed the low-load session first, while the rest completed the high load session first. An instruction appeared before the participant started the high load session, reminding them of how they were expected to handle incoming notifications, as in Fig. 14. This was a single snapshot recap of the training video, where the whole procedure was shown of spotting and queuing incoming notifications was explained.

Figure 14: Instruction indicating that the Notification feature would be used in this Session

The subtasks to be completed in notifications would appear periodically, for 5 seconds at a time, at the bottom right hand corner of the screen during the entire session. The following format is used:

Figure 15: Notification Item

The incoming notifications would be labeled as belonging to different departments, such as Recruiter, Finance or Human Resources for example, as can be seen in Fig. 15 and Fig. 16. The participants were asked to ignore all notifications, except those addressed for "Human Resources". The participant had 4 seconds to click on the notification item, which would bring up a task list, to be populated with "Human Resources" notifications. Once selected, the item could then be added to this list at the end of the queue, as shown in Fig. 17 and Fig. 18. Participants could also choose to refrain from adding the item onto the list by simply closing the window.

If the participant missed adding "Human Resources" notifications, or incorrectly added notifications from other departments, the border of the notifications would change from green to amber to red, to indicate that the participants should pay more attention to the notifications. When a sufficient number of "Human Resources" notifications are added to the list again, the borders on new notifications will appear green once again.

The number of correct additions overall is tallied throughout the session and a final score is given at the end, in the Task Review.

Figure 16: A Notification appearing for a Recruiter

Figure 17: Adding a Notification for Human Resources to the Task List (queue)

Figure 18: The Notification Item has been added on to the Task List (queue)

4.6.   *Modalities and Data Streams*

A number of modalities and data streams were collected during this experiment. As mentioned before, the experiment was conducted employing dual-task paradigm for higher cognitive load tasks. Subjective ratings of complexity and difficulty were employed after each task set, to ensure that the desired levels of load built into the task design were actually being perceived by the study participants. The Experimental Platform used in the study was developed in-house, that incorporated all data collection, in both versions (high CL and low CL). Following modalities of data were collected:

1.   Survey Responses

Various types of survey responses were collected with respect to candidates' trustworthiness and cultural factors as well as about participants' attitudes and moods:

- **Pre-Screening Survey**
  A pre-screening survey consisting of 13 questions, with a total of 91 multiple choice questions about the participant's attitudes towards their supervisors and peers, honesty, kindness and trustworthiness, as well as some self-identifying ethnicity and personality based questions.

- **Mood Survey**
  This single question survey required participants to rate a series of affective aspects, such as happiness and sadness, according to how intensely the feeling was being experienced at the time.

- **Subjective ratings of mental effort/ task difficulty**
  This single question survey asked participants to rate how difficult the tasks were. It was administered at the end of both the high load and low load sessions.

2.   Behavioral Measures

Different types of behavioral data modalities were also collected during the experiment including speech, text, and interactive data:

- **Speech data: think-aloud protocols**
  Participants were asked to verbalize their thought processes as they work through the three subtasks. Their speech utterances were recorded.

- **Speech Transcriptions**
  The speech data collected was transformed later to speech transcriptions for various task conditions as text data for linguistic analysis purpose.

- **Mouse trajectories**
  These are in the form of (x,y) coordinates, and are sampled with enough resolution to reproduce the entire experiment session. The mouse trajectory data is used to track widget manipulation and log use of the mouse movements or

hovering over specific areas of the application. This data can provide user's behavior in terms of their interaction with the application as well as an indication of their attentional focus.

3.   Performance Measures

Different kinds of performance measures were also captured for the tasks performed during the experiment. These included:

- **Ratings, Filling positions and Rankings**:
  The final responses to the actual subtasks.

- **Time-to-completion**
  Overall time to complete each task and the speech response time to answer each question.

4.7.   *Procedure*

The consent forms were presented to each participant (see Appendix: Consent Forms) a few weeks before the study took place. Participants were assigned randomly to complete a specific experimental condition (high or low cognitive load) first. In the Australian data collection procedure, due to time constraints, participants had to complete a brief background survey online (see Appendix: Pre Screening Survey) in the week following the experiment. However, at the other sites (Malaysia and USA) the participants completed this background survey immediately before the study began.

The experimenter instructed each participant to sit at a specific computer station and to watch the training video, which explained the purpose of the tool (the script can be found in the Appendix). The training content was couched as the examination of a new applicant screening HR tool used to rate job applicants. An experimenter or research assistant was present throughout the entire training and participants could ask questions at any time. No practice trial was provided since the use of the system was demonstrated in the training video.

Once the participants were ready to begin, they started the application and submitted their student identifier to begin. At this point, they completed the mood survey. Depending on which session (high or low load) they were completing first, the Notification feature instruction screen appeared to remind them that they need to queue notifications. The participants were then presented with the first subtask: screening applicants. Additional surveys as described earlier were administered throughout the tasks and at the end of the task session.

Due to time constraints, the participants from Australia, who were all students of INFO3315 course (Human Computer Interaction) at Sydney University, were debriefed the following week during their course tutorial. See Section 4.10 for more details about the participants. To further the participants' learning, an exercise was provided for them in relation to interface

evaluation and methodology of experimentation, which were topics covered in their course (see Appendix: Debriefing and Exercise).

4.8. Pilots Surveys

The cognitive load manipulations to be used in the study were determined through pilot testing for feasibility with the overall study design. Pilot studies were conducted to ensure that the manipulations are effective and that all other applicant information is considered equivalent across load levels and across cultures. Since two rounds of the pilot experiment needed to be completed, the amount of stimulus material needed doubled also. Appropriate language was also needed to be tested for each site to test cultural fit. To address these concerns the following pilot surveys were planned and conducted:

To test the cultural appropriateness of language and bias of trust manipulations, one pilot test was planned at each site (Malaysia, Australia and US) with 50 subjects of a similar background as a test group (e.g. university students). This content pilot, in the form of a web survey, was deployed through SurveyMonkey (can be found in the Appendix). The researchers for each site carefully modified the content to ensure that no bias would be present in or between the candidate profiles that would allow differentiation other than through the AIB indicators – for example, at the Australian and Malaysian sites, the word "college" is seen as less prestigious than the word "university" while in the US, the words are interchangeable. Similarly, some sentences that were to be used as neutral comments could be interpreted as negative comments in Australian culture. For example, if a recommendation for a candidate reads "I have known this person for 2 years", the implication is that there is nothing positive to say about the person except the length of time they have known each other. Other idioms such as use of "above and beyond" and "state of the art" could cause comprehension issues outside of the US.

The US site collected 60 responses to their version of the pilot survey, while the Australian site managed to collect the responses of 30 participants from the University of NSW students in Human Computer Interaction course.

4.9. Pilot Experiments

Various sets of pilot experiments were also undertaken in Sydney and in Dayton, with research assistants. This helped to evolve the experimental software, as well as the UI design and calibrate the secondary task speed and specifications.

It was found that the "think-aloud" protocol added an element of uncertainty in terms of time – as it was found participants would vary in the amount of speech and the depth of explanation they provided. Some participants finished each session in 20-25 minutes, while others took almost 90 minutes. Additionally, during the pilot experiments, it was decided to reduce the number of candidates used in the experiment session from original 6 to 4. This reduction allowed some regulation of the timing of the study, preventing subjects from taking too long and being exhausted at the end of the session. We also provided an extra incentive to the participants of a prize for the participant in every session who was able to most consistently speak throughout the session.

4.10. Participants and Consent Forms

A total of 120 subjects were planned for the user study per site. The study duration was approximately 1 hour. Including pilot studies, a total of 360 subjects were to recruited to participate across the three study locations. Participants could either volunteer to participate (e.g., military personnel), receive course credit at the respective university, and/or receive remuneration in the amount of $20 (or local equivalent) for each 60-minutes of participation, which was the estimated duration of the study.

In the Australian study, 100 students from the INFO3315 course (Human Computer Interaction) at Sydney University, agreed to participate. The consent forms are attached in the Appendix of this document. Of these, 91 completed all sections of the study. The students received course credit in exchange for their participation, as well as snacks and movie ticket prizes for the "top performers" after the session.

To maintain confidentiality, we used student identifiers (numbers and letters) when dealing with the data. Each participant was assigned a subject identifier number. The number will not be attached to any document that included personal identifiers (such as the informed consent form). Some demographic data was requested such as age, sex, class rank, and ethnicity. Only group data was analyzed and all consent forms and data were stored under lock and key when not in use at each site.

5. Operational Processes

5.1. IRB Approvals

NICTA applied for Federal Wide Assurance (FWA) and received word of approval in July 2010 after a process taking several months. NICTA also completed the DoD Addendum following this approval, where NICTA formally acknowledged the requirements to follow IRB directives. An IAIR was also received which indicated that UNSW Ethics Committee would oversee NICTA research at the Australian site. In fact, NICTA's local data collection and user research approval had already been granted by UNSW Ethics Committee (Ref: 08/2010/42) in July 2010, which is still valid until September 2013. This approval was also ratified by the University of Sydney Ethics Committee (Protocol: 14031) in July 2011. All documentation is included in the Appendix.

Finally, cover letters and IRB package for "Trust Management Studies" were submitted to the AFRL IRB including the User Study design and supporting documentation and approved on (Approval 11-21) the 11th June 2011.

5.2. CITI Training

In addition to all these approvals, all Australian team members (Fang Chen, Natalie Ruiz and Asif Khawaja), who were involved in the study with human research participants, regularly completed CITI training and received their completion certificates.

6. Data Collection

6.1. *Schedule*

The overall study and data collection schedule slightly changed due to some unforeseen reasons including technical issues and arrangement of participants, and also due to some team members who left the project. Nevertheless, the Australian data was collected as per original schedule in 2011 and about 100 students from the University of Sydney participated in the user study.

6.1.1. Data Collection Summary: Australia

The Australian group's data collection was completed in October 2011. The collected data underwent preliminary analysis to determine the quality of the data and validation of the protocol. Few minor technical changes in terms of data collection quality were identified in the software tool for other sites to use. The software versions administered at Australia and other sites were not substantially different from each other that would make any confounding problems or other issues. Some statistics of the Australian data collection are as following:

- 91 subjects completed both conditions (high and low CL)
- Approximately 239 survey/response data points per subject
- Speech data: 6.5Gb = 58 hours of speech
- Interactive Behavior: ~96 million data points including mouse trajectories, selection, typing, browsing activity
- Speech Transcriptions: ~80,000 words of manually done transcription data from effective speech (about 50 participants speech transcribed).

6.1.2. Data Collection Summary: Malaysia

Based on the experiences from the Australian data collection, appropriate modifications were made in the Candidate screening application to improve the overall data collection process. Malaysian data was collected in 2012 and about 105 students from the Sunway University participated in the user study. Unfortunately, due to some technical issues with the microphones that resulted in a lot of participants' speech data not recorded correctly, more experiment sessions are being run in Malaysia as of this writing to collect more data equivalent to originally estimated quantity. Here are some statistics for Malaysian data:

- 105 subjects completed both conditions (high and low CL)
- Approximately 275 survey/response data points per subject
- Speech data: 9.5Gb = 88.5 hours of speech
- Interactive Behavior: ~150 million data points including mouse trajectories, etc.

6.1.3. Data Collection Summary: US

As of this writing, the US site is running the experiments and the data collection with roughly 160 students is under progress.

7. Data Analyses

7.1.   Analysis Plan

The hypotheses described in the earlier section of the same title were operationalized as in Table 3.

Table 3: Operationalized Hypotheses of the Study

H1:	Participants from a collectivistic culture (e.g. Malaysia) will rate trust higher when applicants have higher benevolence
H2:	Participants from an individualistic culture (e.g.US, Australia) will rate trust higher when applicants have higher ability
H3:	Participants will bin applicants with higher ability in the Supervisor category, applicants with higher benevolence in the co-worker category, and applicants with higher integrity in the Others' supervisor category.
H4:	The above posited cultural effects will be greater under high cognitive load.
H5:	Overall for any culture, participants' speech features, such as speech rate, pitch, pauses are likely to vary during the high cognitive load tasks vs. low load tasks.
H6:	Linguistic behaviors, such as the negative emotions words, trust and distrust words, personal pronouns and other linguistic and grammatical features are likely to change during the high cognitive load task as compared to the low load tasks.
H7:	Interactive behaviors, such as mouse actions and trajectories are likely to change during the high cognitive load task when compared to the low load task.
H8:	Performance behavior, such as response latency, task completion times are likely to change systematically during high cognitive load tasks vs. low load tasks.
H9:	Subjective ratings in terms of task difficulty are likely to show significant correlations between various levels of task complexity (or cognitive load).

Several analyses were planned to be conducted to test these hypotheses. For the first four hypotheses, the survey data was aggregated based on the pre-established scales used. This data will be analyzed by our US and Malaysian partners for trust and cultural manipulations. In their analyses, they will conduct reliability analyses to ensure that these measures are reliable. Principle component analysis will also be used for the survey and questionnaire answers. The results for these analyses will be published by the US and Malayasian partners in their respective reports separately. For the hypotheses 5 to 9, to test the effects of cognitive load on the participants' behavior, we used various statistical analysis methods (e.g. ANOVA, *t*-tests) depending on the hypothesis being tested. These analyses were conducted only for the Australian dataset as the US and Malaysian dataset were not available at the time of these analyses. Table 4 summarizes various categories of analyses planned for the data.

Table 4: Summary of Analysis Categories Used for the Study

Analysis category	Analysis types in each category	Description
Subjective Ratings Analysis	Task difficulty rating analysis	To validate the experiment design for required task difficulty / mental effort.
Linguistic Analysis of Speech Data	Pause analysis Linguistic behavior analysis Linguistic category analysis	To analyze the speech and linguistic behavioral changes for various cognitive load and trust conditions.

Analysis category	Analysis types in each category	Description
Signal Analysis of Speech Data	Analysis of pitch, tone, speech rate, intensity, energy and other speech signal features.	To analyze the variations in speech signal patterns for different cognitive load conditions.
Interaction Data Analysis	Mouse interaction analysis (mouse actions and trajectories)	To analyze trajectories and mouse activity behavior and their temporal elements.

Performance Analysis	Performance on
	  Time-to-completion
	  (overall and per task),
	  Time-to-respond
	  (response latency)
		The final responses to the actual subtasks.
		To analyze the performance variations under various load and trust conditions.

A number of features of interest were extracted and annotated for these analyses. The following details discuss some of the feature extraction activities carried out on each of the behavioral measures recorded and also some tools that were used to extract those features.

7.1.1. Speech recordings

1. *Data cleaning (e.g. remove cross-talk, noise)*

The speech data was recorded in segments, which corresponded to each of the three subtasks. Since, in Australia, the experiments took place in a classroom laboratory, a number of participants completed the sessions at the same time. Although directed microphone headsets were used, and participants were seated as far away as possible from one another, there was a possibility that cross-talk could have affected the speech recordings. We analyzed the speech recordings and found that most of the recordings were clear enough for the cognitive load classification speech engine we used to process them. Only a few participants' recordings involved had major cross-talk and background noise issues, which were eliminated from the analysis dataset.

2. *Building CL models, training, and testing the data*

We used a speech-based cognitive load classification engine, which we developed as part of previous studies using controlled laboratory experiments (B. Yin, Chen, Ruiz, & Ambikairajah, 2008; Bo Yin, Ruiz, Chen, & Khawaja, 2007) and refined during field trials, to automatically analyze the participants' speech data for features of low and high cognitive load. The engine extracted the predetermined speech signal features and used the speech data to create low load and high load models of speech for each experiment session. Traditionally, the classification accuracy of the system has been very positive with cognitive load classification accuracy levels ranging from 71% to 78% for various datasets in speaker-independent data settings (B. Yin et al., 2008; Bo Yin et al., 2007). Our objective of using the classification system with current study was to verify whether cognitive load can be detected from the acoustic features of speech in this application involving think-out-loud speech, which has never been tested in any of the previous studies. The results are presented in the results section later.

7.1.2. Speech Transcription data

Once the speech data was pre-processed and cleaned, mid-level features such as pause frequencies and lengths were annotated for pausing behavior analysis. The speech data was then manually transcribed and annotated for several linguistic features using a popular speech transcription and annotation tool called ELAN (EUDICO Linguistic Annotator, 2010). Once transcribed and annotated, several linguistic features were extracted from the transcriptions automatically using a text analysis tool called Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). LIWC comes with a built-in dictionary of various psycholinguistic category features comprising 4,500 words. The LIWC dictionary comes with over 80 built-in word categories including negative emotion words, anxiety and anger words, cognitive processes, agreement and disagreement words, as well as grammatical parts of speech like personal pronouns and many others. The LIWC dictionary is customizable and hence two custom word categories relevant to trust analysis – Trust and Distrust words – were added to the dictionary and all trust and distrust words and their synonyms were included in the corresponding categories. The LIWC automatically extracted the linguistic features known to be relevant to trust and cognitive load (see Khawaja et al., 2010, 2012, 2008) from the transcriptions. It extracted these features as percentages of total words spoken in order to deal with participants' verbosity differences. LIWC counts the number of words for a specific linguistic feature by matching the words from the transcription with its dictionary. The average dictionary coverage (the percentage of words captured by the dictionary) for the participants' transcriptions was over 95%. The results of the linguistic analysis are present in the results section. Table 5 lists the LIWC linguistic categories selected from the dictionary for analysis.

Table 5: Selected LIWC Linguistic Categories

Linguistic Category	Example words
Negative emotions	ugly, nasty, bad, fail, sorry
Swear words	damn, shit, fuck
Anger	hate, kill, annoyed
Tentative	maybe, perhaps, guess
Certainty	always, never, absolutely
Achievement	won, done, performed
Trust	trust, believe, sure
Distrust	doubt, disbelieve, suspicious

7.1.3. Mouse interaction features

As part of the study, all participants' mouse interaction with the candidate screening application were also recorded. The recorded data included the type of mouse action performed by the participant, the location of mouse action on the application window, the time the action was performed, the widget the action was performed upon and the session and cognitive load level information. The mouse interaction data was cleaned, parsed, and

analyzed using a statistical analysis tool called R[1]. Various mouse interaction and behavior features automatically extracted from the mouse dataset and analyzed include:

- Time spent moving mouse
- Distance traveled per task/per session
- Categorizing time spent in different screen areas on a per-task basis
- Which areas of the screen were most frequented
- How much time spent on specific widgets, e.g. drop down boxes.

7.1.4. Other performance behaviors

Other application level performance features, such as time-to-completion, time-to-respond both overall and on per task basis were also analyzed for any indication of high load instances within the session or tasks.

7.2.   Analysis Results

Five different types of analyses were conducted for the data collected through the study from the Australian site. The analyses of participants' subjective ratings, their speech data, transcription data, interactive mouse data, and some performance data were performed and the results are presented as following:

7.2.1. Subjective Ratings of Mental Effort

To validate the experiment design for required cognitive load levels the subjective ratings of mental effort or task difficulty were collected from the participants. These were collected at the end of both the high load and low load tasks within each session and were based on a seven-point Likert scale (from 1="Extremely easy" to 7="Extremely difficult"). The analysis of the subjective ratings showed a mean rating of 3.625 for high cognitive load condition and 3.037 for low load condition as shown Figure 19. The pair-wise statistical t-test showed a statistically significant difference between the ratings ($t(72) = 5.201$, $p < 0.001$). This confirmed the effective experiment design that induced the required levels of task difficulty and/or cognitive load levels as expected. The participants found the session with the dual-task more difficult overall than the session without the dual-task. This result is concurrent with other studies based on dual-task paradigm (Leyman, Mirka, Kaber, & Sommerich, 2004; Marcus, Cooper, & Sweller, 1996; Sweller, Merrienboer, & Paas, 1998; Wada, Iwata, & Tano, 2001).

---

[1] The R Project for Statistical Computing (http://www.r-project.org/)
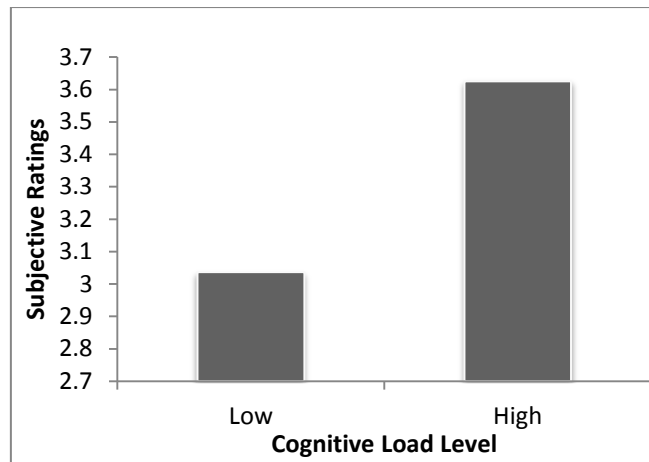
Figure 19: Participants' Subjective Ratings for Task Difficulty ($p < 0.001$)

7.2.2. Speech Data Analysis

As described earlier, the experiment involved two cognitive load levels with one session without dual-task (low load) and the other session with dual-task of handling notifications (high load). Participants' speech under a think-out-loud protocol was captured for both the load levels. We used the speech-based classification engine, as described in section 7.1.1, to see whether we can successfully classify the two load levels from the speech signal features of a think-loud speech. The engine used the participants' recorded speech data to train and test the classification model in ten different repeated folds/rotations for cross validation purposes, in order to eliminate any possible bias due to participants distribution. Within each fold, the engine used 90% of the participants' speech data for model training and remaining 10% for testing and cognitive load classification. No data was duplicated for testing in any fold or rotation.

Although, the classification system has been previously showing an accuracy ranging from 71% to 78% for various speaker-independent data settings, for current speech data, the system showed an overall accuracy of 51%, i.e. only half of the speech data was correctly classified as being low load or high load speech. This could be due to the nature of the speech captured from this study being the 'think-out-loud' speech. Previously the system has been tested with either reading speech or isolated spoken words only.

7.2.3. Performance Analysis

The performance analysis of participants' overall think-out-loud effort, task completion times and times to respond to the survey questions during the tasks was also conducted. When conducting pair-wise t-tests, we found that participants overall spoke proportionally the same amount of effective speech under low load session as compared to high load session ($p = 0.2$). This was also confirmed by a similar average overall word count under both sessions ($p = 0.1$). Interestingly, this shows that for this study involving candidate screening tasks, the dual-task causing high cognitive load does not make any difference to participants' in terms of their effort to think-out-loud their thought process. This result is in conflict to previous findings of similar studies where the amount of speech had either decreased (Kleinman &

Serfaty, 1989) or, in most cases, increased under high cognitive load situations (Katz, Fraser, & Wagner, 1998; Khawaja et al., 2010; Oser, Prince, Morgan, & Simpson, 1991).

Even though the participants' proportional effective speech was same under both load levels, we found that on average, the participants took longer and spent more time to complete tasks in high load session than in low load session ($p < 0.02$). We also found that participants' response latency (time taken from the point they finish reading the question to the point when they start speaking to verbalize their decision thought-process) was higher for high load session as compared to low load session ($p < 0.002$). These results suggest that overall the participants may have taken more time remaining silent or 'silently-thinking' under high load session than under low load condition in order to make their trust judgment decisions. Additionally, even though the overall proportional speech was same under both sessions, interestingly, the participants used longer sentences under high load session, as we found their average number of words per sentence increasing under high load as compared to low load session ($p < 0.001$) and concurrent to our previous findings (Khawaja et al., 2010). These results are illustrated in Figure 20.
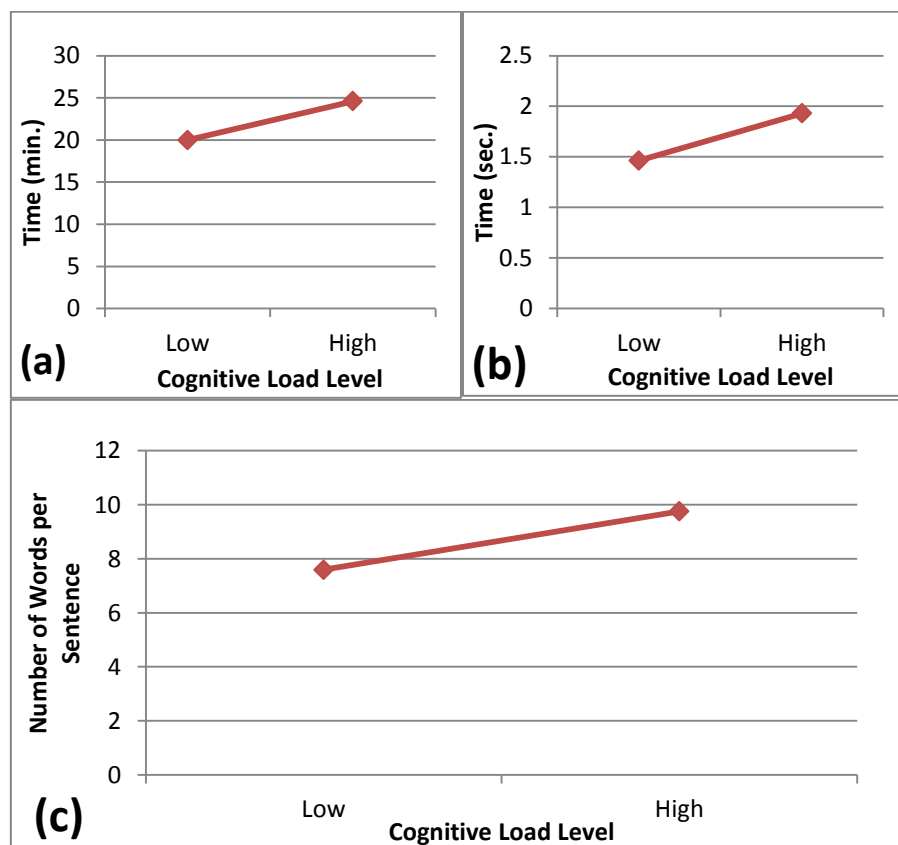
Figure 20: (a) Avg task completion time, (b) Avg Response time, (c) Avg # words per sentence

7.2.4. Linguistic Analysis of Think-Aloud Speech

As described earlier, the participants in the study were asked to verbalize their thought process using a 'think-out-loud' protocol for all the tasks and we recorded their speech and transcribed it using ELAN transcription and annotation tool. Due to the project time constraints, out of 91 participants who completed both sessions, only 55 participants' speech

could be transcribed and analyzed for linguistic behavioral changes for various trust and cognitive load features. The transcriptions were also annotated for various pausing and other mid-level behaviors like hesitations and repetitions. Hence, the linguistic analyses were carried out in three different areas – analysis of pausing behavior, analysis of LIWC-based linguistic categories, and analysis of other speech behavior.

1. *Pause Analysis*

Traditionally in psychology, the pauses during natural speech have been associated to a person's thinking and decision making, i.e. every time a person pauses during the speech, he/she processes currently known information in the working memory to produce the next response (Khawaja et al., 2008). Schilperoord also argued that the more time it takes to produce the response, the more cognitive energy is required to do so (Schilperoord, 2001). Accordingly for the current study, we hypothesized that participants will pause more and longer under high load situation than under low load condition. We manually annotated different pause features mainly for two important factors of pauses – frequency and duration of pauses. The annotations were done in ELAN and covered various pause features as listed in Table 6.

Table 6: Selected Pause Features and their brief Description

Pause Features	Description
Total pause frequency	Total number of pause segments
Freq. of silent pauses	Number of silent pause (voiceless) segments
Freq. of filled pauses	Number of filled pause (voiced) segments, e.g. ahh, um.
Avg. freq. of pauses/min	Average number of pauses per minute (normalized)
Avg. freq. of silent pauses/min	Average number of silent pauses per minute (norm.)
Avg. freq. of filled pauses/min	Average number of filled pauses per minute (norm.)
Total pause duration/length	Average length/duration of overall pauses (in seconds)
Length of silent pauses	Average length of silent pauses (in seconds)
Length of filled pauses	Average length of filled pauses (in seconds)
Percent of total time pausing	Percentage of total time spent in pausing (%)

Pauses inherently originate from breathing activity and are often very brief, so we needed to define a cut-off value for pause lengths (Dechert & Raupach, 1980). Though selected arbitrarily, it usually ranges from 0.25s to 0.3s (Schilperoord, 1996). For our study, we assumed the threshold of 0.3s; any pauses smaller than 0.3s were assumed to be an inherent part of the natural speech and were not used for our analyses. As shown in Table 6, we analyzed the overall pauses as well as silent and filled pauses separately.

We performed paired-sample t-tests and the results confirmed that overall, participants paused significantly more in high load session than in low load session ($p < 0.01$). Detailed investigation showed that both silent and filled pauses showed similar trends ($p < 0.01$). The overall pausing trends are illustrated in Figure 21.
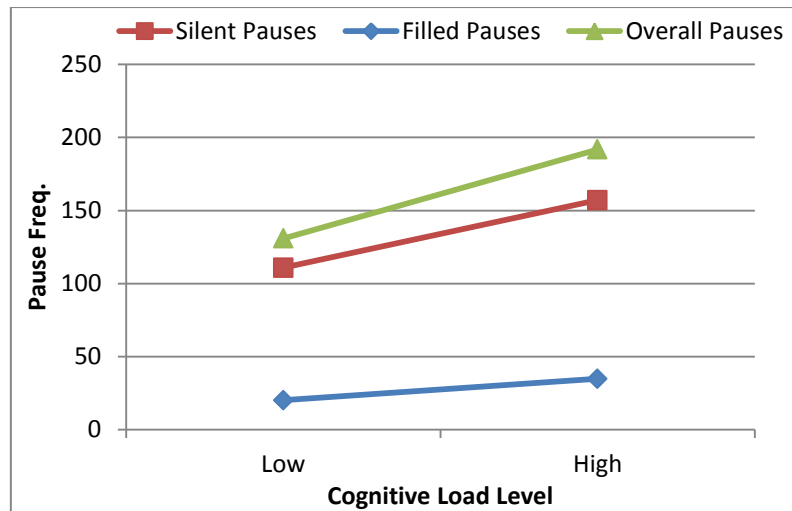
Figure 21: Overall, Silent, and Filled Pause Frequencies ($p < 0.01$)

This pausing analysis was based on total number of pauses used by the participants and did not take the individual differences in speech into account. So to normalize any differences, we calculated average per minute pause frequencies for all participants and compared them for both low and high load sessions. We found that participants still paused significantly more in high load condition than in low load situation ($p < 0.01$). The silent pauses also showed similar significant behavior with more silent pauses under high load session ($p < 0.001$) as before. The filled pauses, although still increasing under high load, did not show a significant difference with normalized frequencies ($p = 0.26$), which tells that the overall increasing pausing difference may have due to the silent pauses only and not the filled pauses. This also informed us that participants used significantly fewer filled pauses (< 5) than silent pauses (~18) overall. These results are illustrated in Figure 22.
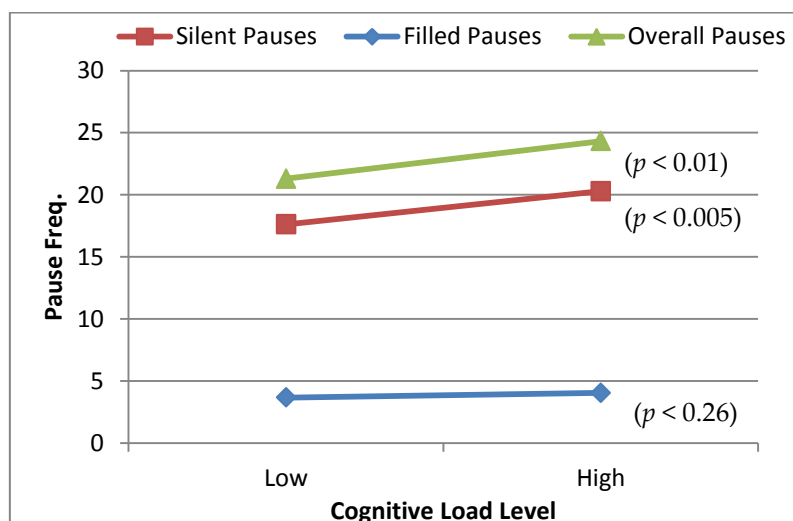
Figure 22: Normalized Pause Frequencies

We also analyzed pause duration/lengths and performed paired-sample t-tests on the overall pausing behavior under both sessions. The results confirmed that overall, participants paused significantly longer under high load condition than in low load condition ($p < 0.0005$).

Detailed investigation of silent and filled pauses showed that participants used longer silent pauses on the average under high load condition than under low load condition ($p < 0.0008$) but there was no significant difference in the pause lengths for filled pauses ($p = 0.1$). The overall pausing duration behavior is illustrated in Figure 23.

Finally, we also calculated for each participant, the overall percentage of total effective speech time the participants spent pausing, in order to normalize any other pausing differences across participants. The results confirmed that overall, participants spent more time pausing under high load session as compared to low load session ($p = 0.0001$), as expected. This is illustrated in Figure 24. These pausing results are also in line with the previous results of performance analysis in section 7.2.3, where the results suggested that participants may have taken more time remaining silent under high load session than under low load session.
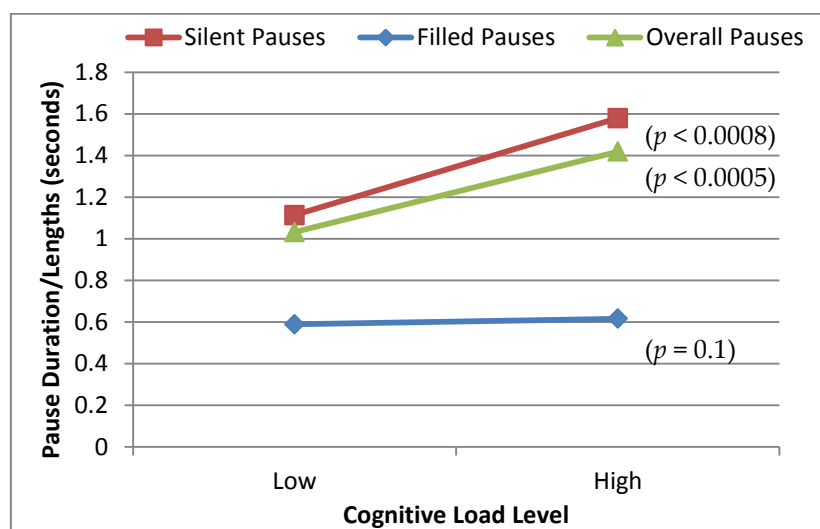
Figure 23: Average Pause Duration/Lengths (in seconds)
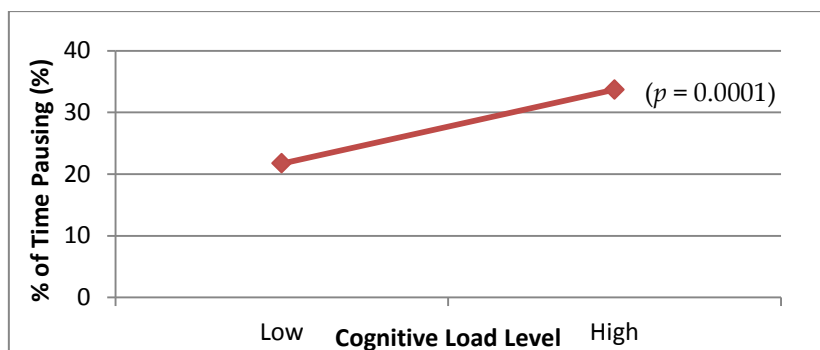
Figure 24: Percentage of Total Speech Time Spent in Pausing (%)

2. *Linguistic Category Analysis*

Apart from pausing features, we also analyzed some selected LIWC linguistic word categories as described in section 7.1.2 and presented in Table 5 earlier. We performed paired-sample *t*-tests for these word categories and found some interesting results with respect to their behavior under two sessions. We found that overall participants used more negative emotion words, more swear or expletive words, and more anger words under the high load session as compared to low load session ($p < 0.05$), as illustrated in Figure 25. These results

reflect the fact that participants feel more anger and frustration when task complexity or cognitive load increases. This is an important finding as high frustration and negative feelings caused by the high cognitive load of the task may drastically impact the person's performance on the task negatively.
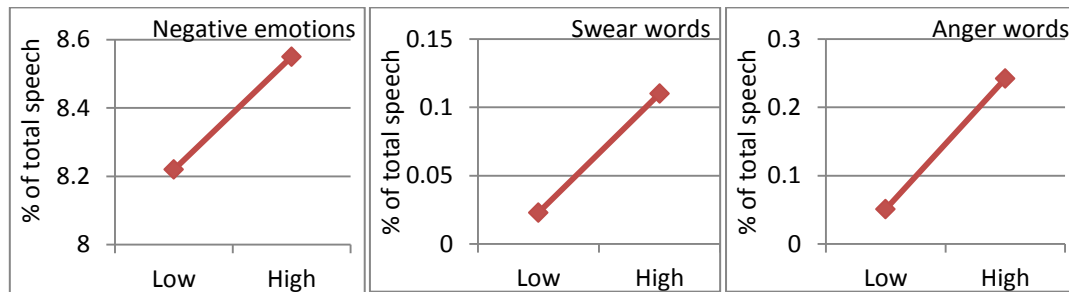
Figure 25: Percentage of Negative emotions, Swear words, and Anger words ($p < 0.05$)

Tentative words like *maybe, perhaps* show a person's doubtfulness about something, while certainty words like *always, sure, absolutely* depict a person's confidence in something. We found for the applicant screening study that the participants used significantly more tentative words and fewer certainty words under high load session than under low load session ($p < 0.05$). We also found that they used fewer achievement words like *won, done, performed* under high load condition then low load condition ($p < 0.05$), which also represents their poor task performance perception. These results are illustrated in Figure 26.
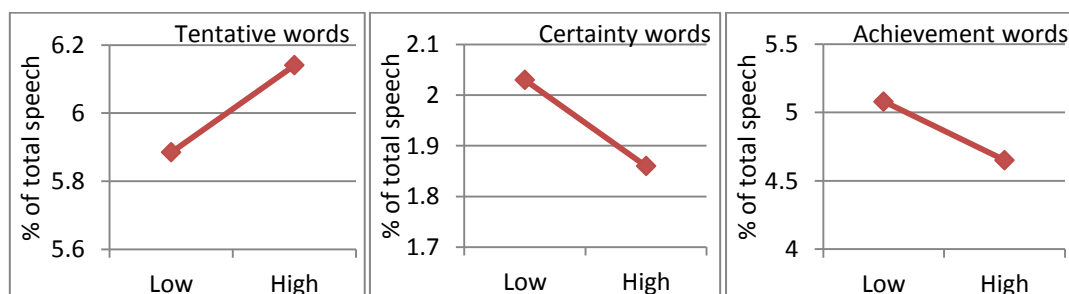
Figure 26: Percentage of Tentative, Certainty, and Achievement words ($p < 0.05$)
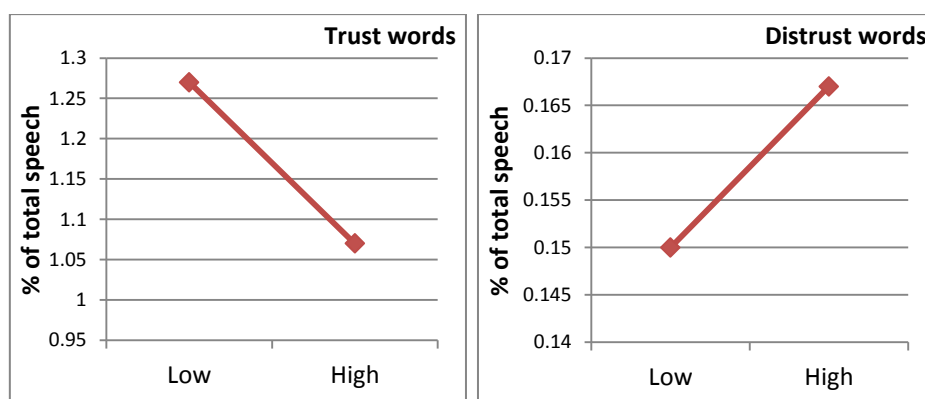
Figure 27: Percentage of Trusting and Distrust words ($p < 0.05$)

Finally and most importantly, from the trust perception viewpoint, we found that participants used significantly less trusting words (like *trust, believe, sure*) and more distrust

words (like *doubt, disbelieve, suspicious*) under high cognitive load session as compared to low cognitive load session ($p < 0.05$), as illustrated in Figure 27. These results tell us that regardless of whether the participants were rating a low or high trustworthy applicant in any session, the high cognitive load caused by the extra dual-task of queuing the incoming new applicants impacted their trust perception negatively and changed their overall trust perception systematically. This finding is important as under highly critical task scenarios (e.g. in military operations, especially where personnel could be interacting with a sophisticated interaction system or a robotic system), a high cognitive load caused by the task or interaction complexity could result in negative or lower trust perception by the person about the system being used, which in turn could possibly affect the person's task performance and/or impact their effective decision making ability.

3. *Other Behavioral Features*

We also looked at some other speech behavioral patterns including the number of times participants hesitate, repeat themselves, make self-corrections, and speak incomplete sentences and analyzed how they change for different cognitive load situations.

Table 7: Other Speech Behavioral Features

Other Behavioral Features	Description
Avg. freq. of hesitations	Number of times they hesitate in their speech
Avg. freq. of repetitions	Number of times they repeat themselves
Avg. freq. of self-corrections	Number of times they correct themselves
Avg. freq. of incomplete sentences	Number of times they speak incomplete sentences

We found that participants on the average hesitated significantly more frequently under high cognitive load session than under low load session ($p < 0.01$). The results also showed that participants repeated themselves significantly more frequently ($p < 0.001$) and left more sentences incomplete ($p < 0.0001$) under high cognitive load than in low load condition. These results are illustrated in Figure 28. In terms of self-corrections, although there was an increasing trend but we did find a significant difference ($p = 0.06$). These results suggest that high cognitive load affects participants' speech communication severely, which may result in some kind of miscommunication between the people working together collaboratively to perform some complex tasks and hence may reduce the overall task performance or in worse cases may even threaten lives in critical scenarios like in war zones.
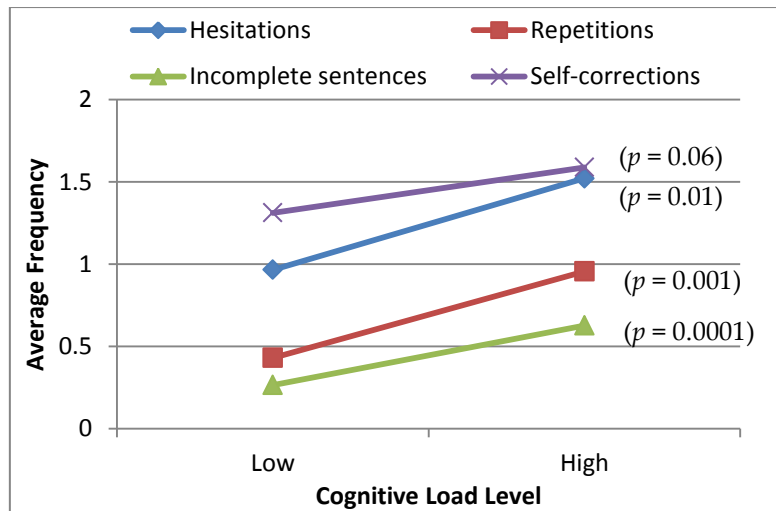
Figure 28: Average Frequency of Hesitations, Repetitions, and Incomplete Sentences

7.2.5. Mouse Interaction Analysis

As discussed earlier, the participants in the study used mouse to interact with the candidate screening application and all of their interaction was recorded. The recorded data included the type of mouse action performed by the participant, the location of mouse action on the application window, the time the action was performed, the widget the action was performed upon and the session and cognitive load level information. All of the 90 participants' mouse interaction data was used for the analyses. We used statistical analysis tool R^2 to clean, parse, and analyze the results. Different mouse interaction behaviors were analyzed as listed in Table 8.

Table 8: Mouse Interaction Features

Mouse Interaction Features	Description
Distance travelled	Total mouse movements per task per session
Mouse behavior (click, drag, move)	Clicking, dragging, moving behavior per task
Most frequented areas and widgets	Indicators of participants' attentional focus

1. *Mouse Distance Travelled*

We first looked at the mouse interaction in terms of total mouse movements and distance travelled by the participants for all three tasks individually under both low and high cognitive load sessions. The analysis showed that participants' mouse distance travelled was significantly higher for tasks 1 and 3 under high load session as compared to low load session ($p < 0.00$), as illustrated in Figure 29. For task 2, although there was a similar trend for distance travelled but we could not find a significant difference. These results may suggest that participants performed overall more mouse movements when cognitive load was high. This may also be due to the fact that under high cognitive load session, participants were also required to handle the dual-task by adding the new incoming applicants to the applicant queue, which may have required more mouse movements as compared to low load session

---

[2] The R Project for Statistical Computing (http://www.r-project.org/)

where there was no dual-task to handle. Further investigation would be needed to validate the increased distance travelled caused by the high cognitive load.
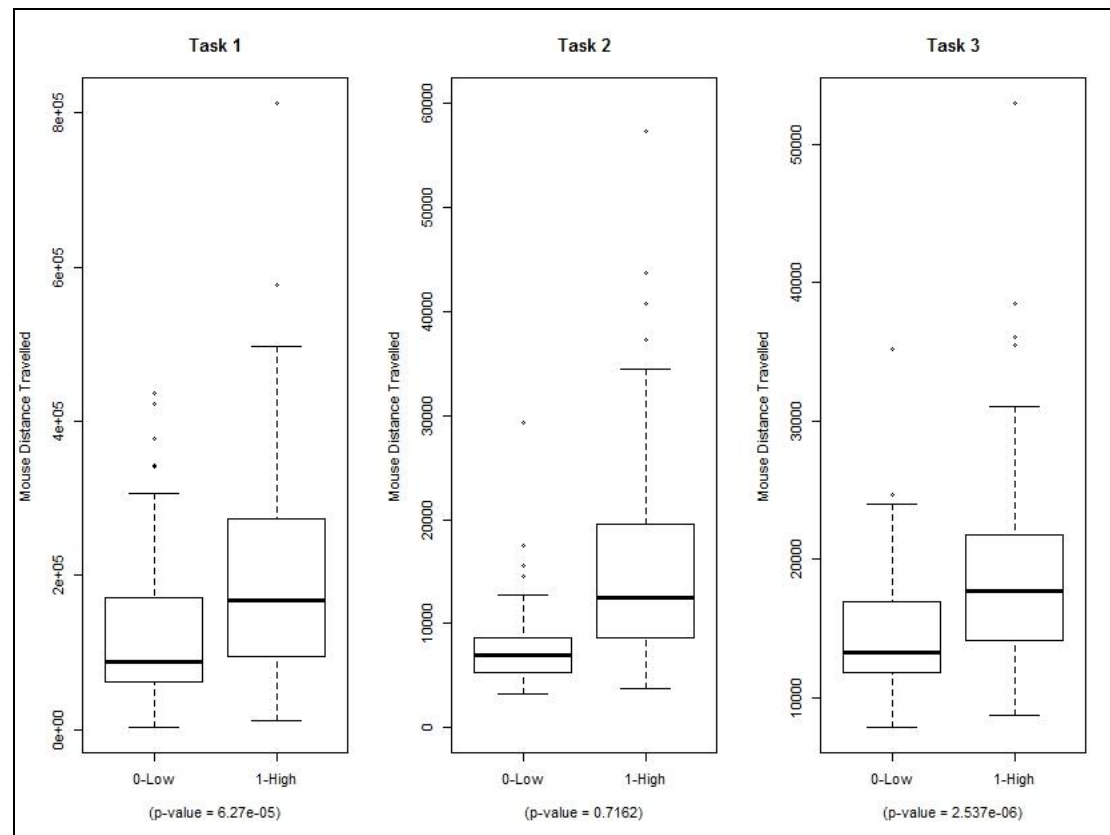
Figure 29: Mouse Distance Travelled for Three Tasks under Low and High Load Sessions

2. *Mouse Events Behavior*

We also analyzed three different mouse action events including mouse clicks, mouse dragging, and mouse move events for each task separately under both low and high load sessions. The results showed that for Task 1 (involving rating), participants performed significantly more mouse clicking and mouse moving actions under high load session than under low load session ($p < 0.00$). For Task 2 (involving positioning), we could not find a significant difference in any of the mouse action events. Finally, for Task 3 (involving ranking), participants performed all mouse events (click, drag, and move) significantly more under high cognitive load session than under low load session (click and move – $p < 0.00$ and drag – $p = 0.01$). These results are illustrated in Figure 30, and show that overall participants' mouse action behavior is affected by the high cognitive load mainly under Task 1 and Task 3; mouse behavior under Task2, like distance travelled analysis, remains unaffected by the cognitive load manipulation.
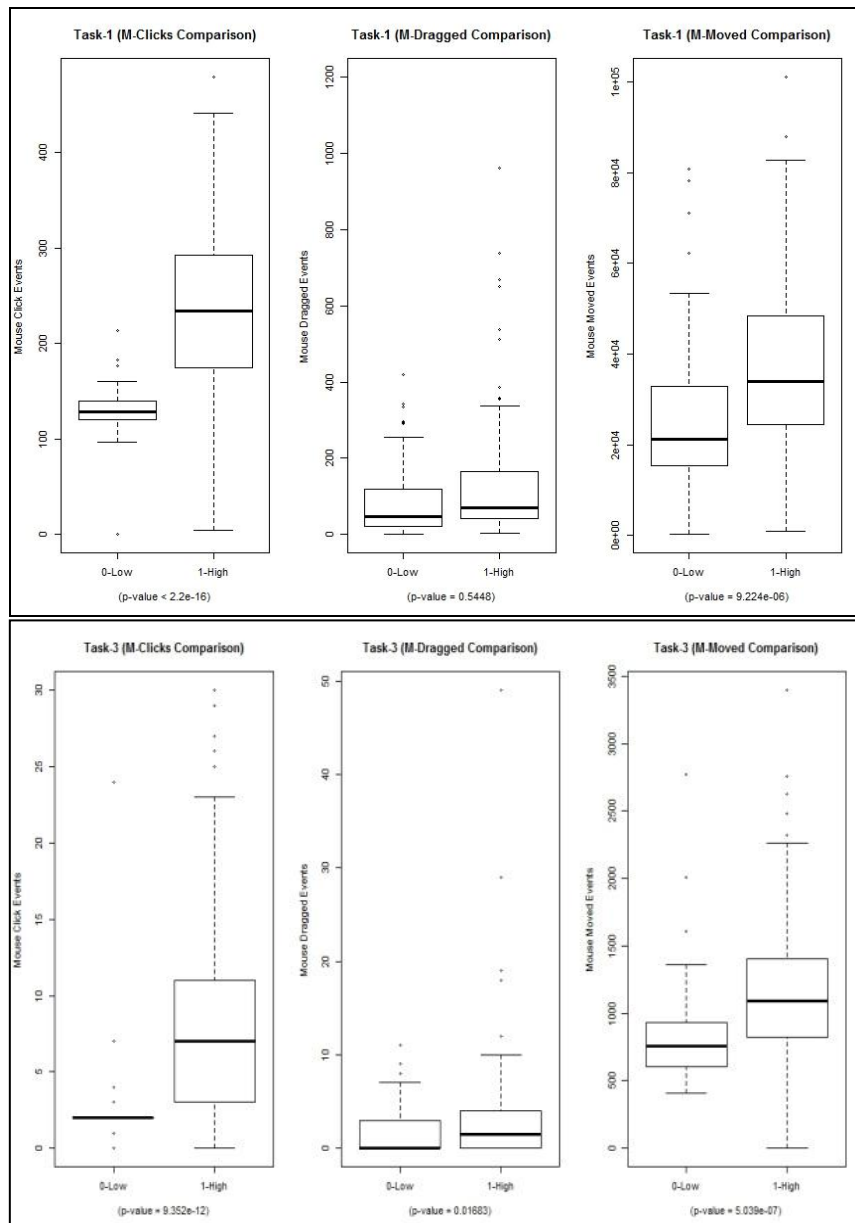
Figure 30: Mouse Actions (click, drag, and move) for Task 1 (top) and Task 3 (below) under Low and High Load Sessions

3.   *Mouse Behavior – Most Frequented Widgets*

For each of the three tasks, there are different kinds of controls or widgets that the participants had to interact with in order to perform the tasks. We performed an analysis to see which areas and widgets of the application the participants interacted with and how. This may not only inform us about the participants' interaction behavior but also about their attentional focus required by different widgets under low and high cognitive load conditions. Task 1 included two important widgets – a survey table to fill in the initial mood survey and a rating panel to rate the applicants for different trustworthiness factors. We found that participants spent significantly more time on these two widgets under high load session as compared to low load session ($p = 0.005$ and $p < 0.00$, respectively), as illustrated in Figure 31.
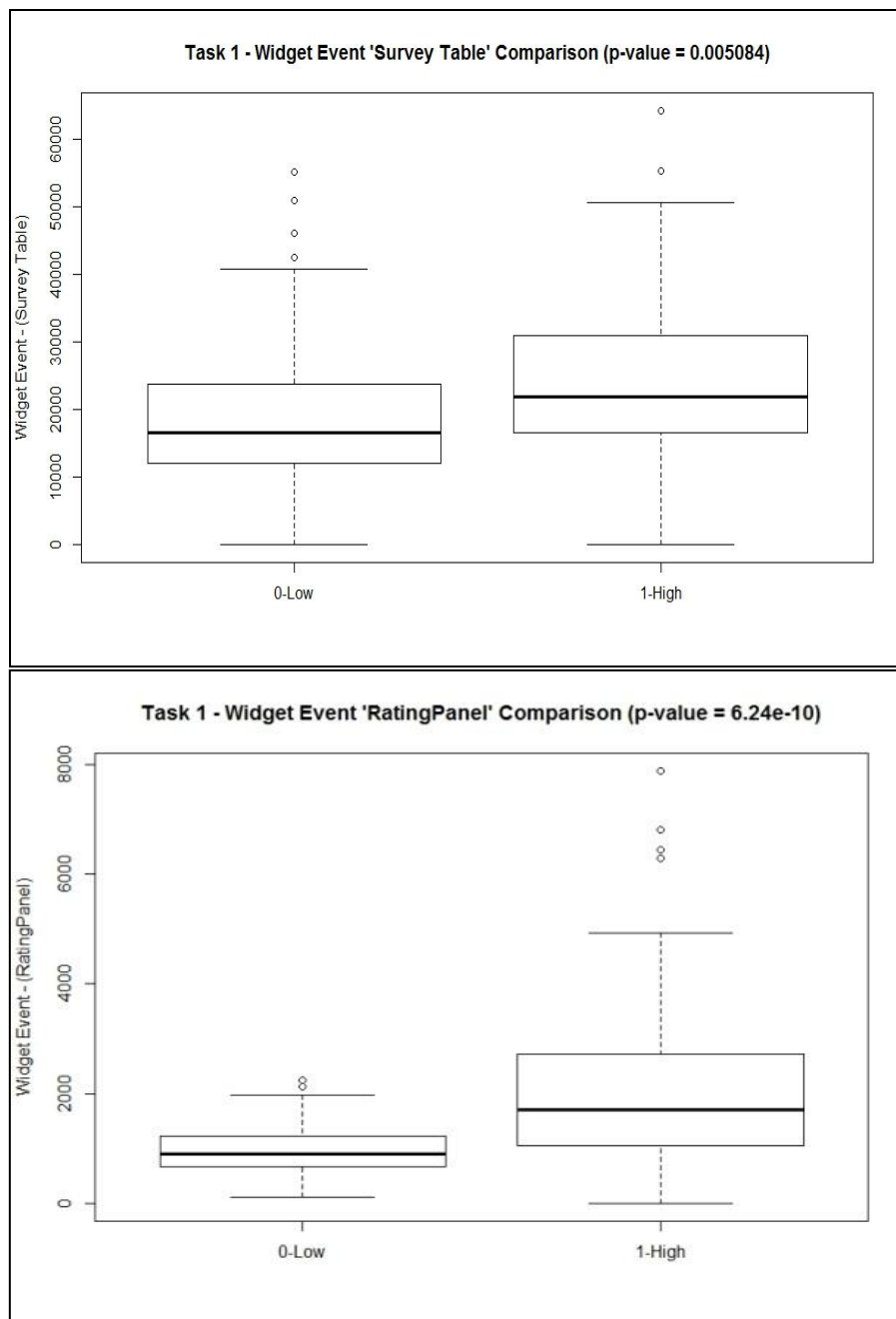
Figure 31: Most Frequented Widgets (Survey table on top and Rating panel below)
under Task 1 for Low vs. High Load Sessions

For Task 2 (involving positioning the applicants for different slots), we found that
participants spent significantly more time on the panel that showed the summary of the four
applicants under high load session as compared to low load session ($p < 0.00$), as illustrated in
Figure 32.
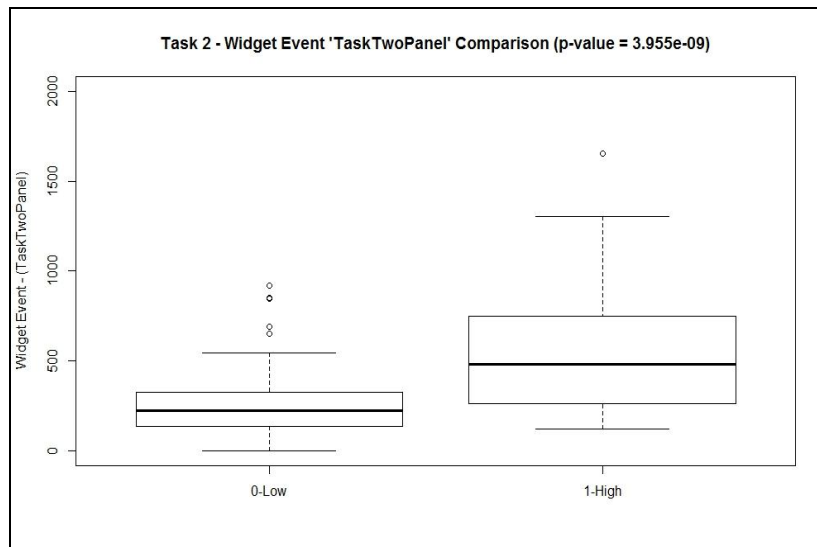
Figure 32: Most Frequented Widget under Task 2 for Low vs. High Load Sessions

Finally, for Task 3 that included two important widgets – an applicant panel that showed the summary of the four applicants and a ranking list that participants were to use to rank the participants for different positions. The results showed that participants spent significantly more time on these two widgets under high load session as compared to low load session ($p < 0.00$ and $p < 0.005$, respectively), as illustrated in Figure 33.
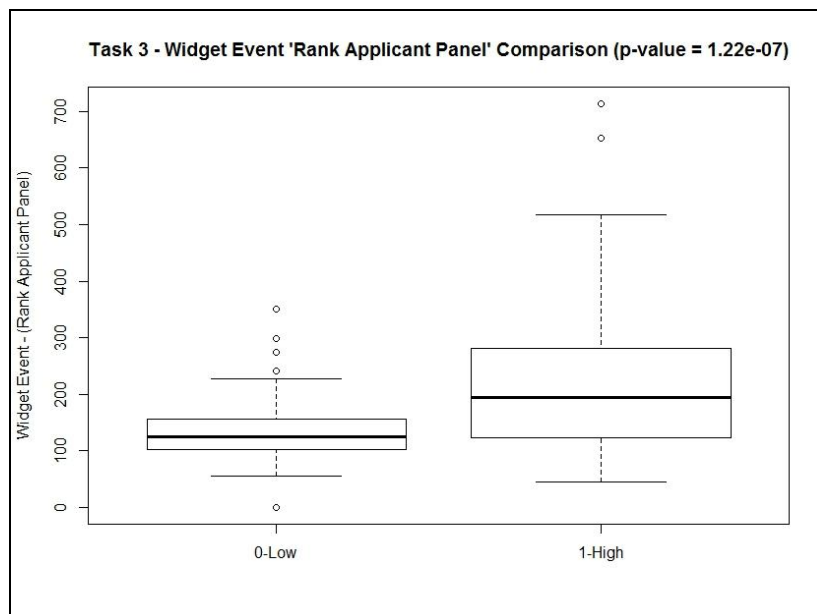
Figure 33: Most Frequented Widgets under Task 3 for Low vs. High Load Sessions
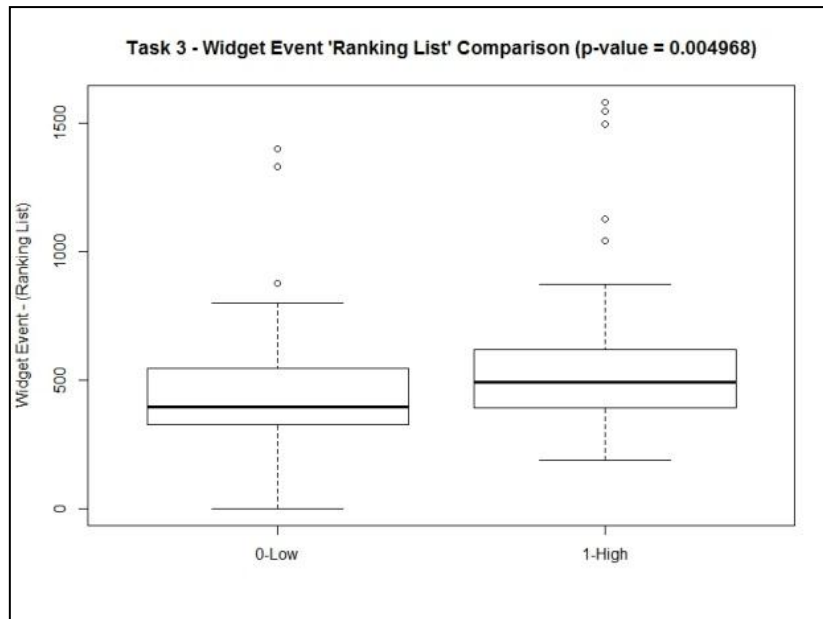
Figure 33: (continued) Most Frequented Widgets under Task 3 for Low vs. High Load Sessions

The mouse interaction behavior analysis is useful as it helps to understand how people's interaction with the system is possibly affected by varying cognitive load conditions. Specifically, it was interesting to find that the participants spent significantly more time on and performed significantly more mouse interaction with the important widgets when they experienced high cognitive load. This may suggest that high cognitive load may have interrupted the participants' constant attention on the widgets required by the task to complete it efficiently and hence they may have ended up spending more time on those widgets. Because low cognitive load session did not involve a dual-task and participants were more able to focus on the task and hence they were able to complete the task much more efficiently without spending much time in frequenting the widgets.

8. Conclusion

In this report we presented the research activities undertaken and outcomes achieved by NICTA as part of the "Effects of Cognitive Load on Trust" project in conjunction with the US AFRL and Sunway University Malaysia. NICTA's role was to conduct the measurement and assessment of cognitive load through speech and other interaction modalities, which has been fulfilled successfully. We presented a detailed literature review of the relationships between cognitive load and trust judgments, and the effect of cultural differences in the way trust judgments are made.

We also discussed in the report the detailed planning and design of the user study for all three project sites (US, Malaysia and Australia), including coordinating multinational IRB approval procedures. The detailed implementation of the experimental software for data collection, and the production of supplementary materials used in the study were also described. The outcomes for this part of the work were described in the second section of the report. The study execution, which was divided into three data collection phases, one for each site (Australia, US and Malaysia) along with their data collection summary and outcomes was also presented in the report.

Finally, a detailed data analysis and results were presented and discussed with regard to understand the relationships between cognitive load and trust. Five different types of analyses were conducted for the data collected through the study from the Australian site. These analyses included subjective ratings analysis, speech signal data analysis, performance data analysis, linguistic analysis, and interactive mouse data analysis. The results showed participants' varying behavioral indicators under different levels of cognitive load.

9. References

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Journal of Neuroscience*, *4*, 829–839.

Berthold, A., & Jameson, A. (1999). Interpreting Symptoms of Cognitive Load in Speech Input. In *Seventh International Conference on User Modeling (UM99)*. Springer Wien New York.

Biros, D. P., Daly, M., & Gunsch, G. (2004). The Influence of Task Load and Automation Trust on Deception Detection. *Group Decision and Negotiation*, *13*, 173–189.

Dechert, H. W., & Raupach, M. (1980). *Towards a Cross-Linguistic Assessment of Speech Production*. Frankfur: Lang.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and Believing: The Influence of Emotion on Trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748.

ELAN: (EUDICO Linguistic Annotator). (2010, April). Max Planck Institute for Psycholinguistics, The Netherlands. Retrieved from http://www.lat-mpi.eu/tools/elan/;

Forgas, J. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*, 39–66.

Jameson, A., Kiefer, J., Müller, C., Großmann-Hutter, B., Wittig, F., & Rummer, R. (2009). Assessment of a User's Time Pressure and Cognitive Load on the Basis of Features of Speech. In *Resource-Adaptive Cognitive Processes* (p. 171). Springer Berlin Heidelberg.

Katz, C., Fraser, E. B., & Wagner, T. L. (1998). Rotary-Wing Crew Communication Patterns Across Workload Levels. In *RTO HFM Symposium on "Current Aeromedical Issues in Rotary Wing Operations."* (p. pp–14.1–14.3.). San Diego, USA.

Khawaja, M. A., Chen, F., & Marcus, N. (2010). Using Language Complexity to Measure Cognitive Load for Adaptive Interaction Design. In *Proceedings of International Conference on Intelligent User Interfaces (IUI 2010)* (pp. 333–336). Hong Kong, China: ACM.

Khawaja, M. A., Chen, F., & Marcus, N. (2012). Analysis of Collaborative Communication for Linguistic Cues of Cognitive Load. *International Journal of Human Factors and Ergonomic Society*, *54*(4), 518–529. doi:http://dx.doi.org/10.1177/0018720811431258

Khawaja, M. A., Ruiz, N., & Chen, F. (2008). Think before you talk: An empirical study of Relationship between Speech pauses and Cognitive load. In *Australasian Computer-Human Interaction Conference (OzCHI'08)* (pp. 335–338). Cairns, Australia.

Kleinman, D. L., & Serfaty, D. (1989). Team performance assessment in distributed decision making. In R. Gibson, J. Kincaid, & B. Goldiez (Eds.), (pp. 22–27). Presented at the Interactive Networked Simulation for Training Conference.

Lee, J., & See, K. (2002). *Trust in computer technology and the implications for design and evaluation* (No. FS-02-02).

Leyman, E., Mirka, G., Kaber, D., & Sommerich, C. (2004). Cervicobrachial muscle response to cognitive load in a dual task scenario. *Ergonomics*, *47*(6), 625–645.

Lyons, J. B., & Stokes, C. K. (2010). Exploring a dynamic model of trust management. *A white paper by Air Force Research Laboratory*.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding Instructions. *Journal of Educational Psychology, 88*, 49–63.

Mayer, R. C., & Davis, J. H. (1999). The Effect of Performance Appraisal System on Trust for Management: A Field Quasi-Experiment. *Journal of Applied Psychology*, *84*(1), 123–136.

Mayer, R. C., & Norman, P. M. (2004). Exploring Attributes of Trustworthiness: A Classroom Exercise. *Journal of Management Education*, *28*(2), 224–249.

Mayer, R., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709–734.

Oser, R., Prince, C., Morgan, B., & Simpson, S. (1991). *An Analysis of Aircrew Communication Patterns and Content* (No. 90-009). Orlando, FL: Naval Training Systems Centre, Human Factors Division. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA246618

Oviatt, S., Coulston, R., & Lunsford, R. (2004). When do we interact multimodally?: Cognitive load and multimodal communication patterns. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces* (pp. 129–136). New York, NY, USA: ACM.

Paas, F., Tuovinen, J. E., Tabbers, H., & Gerven, P. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63–71.

Parasuraman, R, & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253.

Parasuraman, Raja, & Miller, C. A. (2004). Trust and Etiquette in High-Criticality Automated Systems. *Communications of the ACM*, *47*(4), 51–55.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007. *www.liwc.net; last accessed: April 2012*. doi:www.liwc.net; last accessed: April 2012

Ruff, H., Narayanan, S., & Draper, M. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators & Virtual Environments*, *11*(4), 335–351.

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Amsterdam / Atlanta: Rodopi.

Schilperoord, J. (2001). On the Cognitive Status of Pauses in Discourse Production. In *Contemporary tools and techniques for studying writing*. London: Kluwer Academic Publishers.

Schmorrow, D., & Stanney, K. (2008). *Augmented cognition: A practitioner's guide*. Santa Monica, CA, USA: Human Factors and Ergonomics Society Press.

Schoorman, F. D., Mayer, R., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. *The Academy of Management Review*, *32*(2), 344–354.

Scissors, L. E., Gill, A. J., Geraghty, K., & Gergle, D. (2009). In CMC we trust: the role of similarity. In *Proceedings of the 27th international conference on Human factors in computing systems* (pp. 527–536).

Scissors, L. E., Gill, A. J., & Gergle, D. (2008). Linguistic mimicry and trust in text-based CMC. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 277–280).

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic Skin Response (GSR) as an Index of Cognitive Load. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'07)* (pp. 2651–2656). San Jose.

Sweller, J., Merrienboer, J., & Paas, F. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296.

Waber, B., Williams, M., Carroll, J., & Pentland, A. (2012). A Voice is Worth a Thousand Words: The Implications of the Micro-coding of Social Signals in Speech for Trust Research. In F. Lyon, G. Möllering, & M. Saunders (Eds.), *Handbook of Research Methods on Trust* (pp. 249–258). Edward Elgar, UK.

Wada, F., Iwata, M., & Tano, S. (2001). Information Presentation based on Edtimation of Human Multimodal Cognitive Load. *Proceedings of IFSA World Congress and 20th NAFIPS International Conference*, *5*, 2924–2929.

Yin, B., Chen, F., Ruiz, N., & Ambikairajah, E. (2008). Speech-based Cognitive Load Monitoring System. In *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'08)* (pp. 2041–2044). IEEE Press.

Yin, Bo, Ruiz, N., Chen, F., & Khawaja, M. A. (2007). Automatic Cognitive Load Detection from Speech Features. In *Australasian Computer-Human Interaction Conference (OzCHI'07)* (pp. 249–255). Adelaide, Australia.

10. Appendices

1. Consent Form

UNSW

NICTA

UNSW Approval No (08/2010/42)
University of Sydney HREC Protocol No (14031)

THE UNIVERSITY OF NEW SOUTH WALES (UNSW) AND NATIONAL ICT AUSTRALIA
(NICTA)

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM

Applicant Screening Tool

[Participant selection and purpose of study]
You *(i.e. the research participant)* are invited to participate in a study of a new "Applicant Screening Tool". We *(i.e. the investigators)* hope to learn to investigate how people rate applicants using a newly developed software tool. You were selected as a possible participant in this study because you are sufficiently removed from this area of expertise. A total of approximately 360 participants will be enrolled over several phases of the study.

[Description of study and risks]
If you decide to participate, we will ask you to interact with a computer simulation and answer questions using a computer. During this task, you will be asked to review several pieces of information for numerous applicants and rate their suitability for different roles. You will also be required to complete a few questionnaires related to the task. We will also be collecting speech data, mouse trajectories and keyboard strokes, but this should not interfere with your ability to complete the task.
We will also train you such that you are confident in completing the tasks. The training will happen at the beginning of the session and give you an opportunity to familiarise yourself with the system's functionality and the tasks to be completed.
Your participation will not involve risks or discomforts greater than you would encounter when using a computer.

[Confidentiality and disclosure of information]
Data will be collected from participants in the United States of America, Australia, and Malaysia. Data will be shared with the investigating researchers at each of the three data collection sites. Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission, except as required by law. If you give us your permission by signing this document, we plan to publish the results in academic journals and conference proceedings, as well as to build models on features derived from the recorded behavioural signals. These models cannot be reversed to recover any of the original signals, and the models may be used for our own evaluations, as well as be used offsite by outside third parties that we would grant the permission. In any publication of results and distribution of models or related software, information will be suppressed in such a way that you cannot be identified.

[Recompense to participants]
You will be given course credit as part of the user study package at the completion of the study.

Complaints may be directed to the Ethics Secretariat, The University of New South Wales, SYDNEY 2052 AUSTRALIA (phone 9385 4234, fax 9385 6648, email ethics.sec@unsw.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

[Your consent]
Your decision whether or not to participate will not prejudice your future relations with The University of Sydney or The University of New South Wales or National ICT Australia. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.
If you have any questions, please feel free to ask us. If you have any additional questions later, Dr. Fang Chen on (02) 9376 2101 will be happy to answer them. You will be given a copy of this form to keep.

THE UNIVERSITY OF NEW SOUTH WALES AND NATIONAL ICT AUSTRALIA

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM
(continued)

Applicant Screening Tool

You are making a decision whether or not to participate. Your signature indicates that, having read the Participant Information Statement, you have decided to take part in the study.

....
Signature of Research Participant

....
Signature of Witness

....
 (Please PRINT name)

....
(Please PRINT name)

....
Date

....
Nature of Witness

....
Signature(s) of Investigator(s)

....
Please PRINT Name

REVOCATION OF CONSENT
Applicant Screening Tool

I hereby wish to **WITHDRAW** my consent to participate in the research proposal described above and understand that such withdrawal **WILL NOT** jeopardise any treatment or my relationship with The University of New South Wales and National ICT Australia Ltd.

....
Signature

....
Date

....
Please PRINT Name

The section for Revocation of Consent should be forwarded to Dr. Fang Chen, Locked Bag 9013, Alexandria, NSW 1435, Australia.

2. *Protocol for Training Applicant Screening Tool*

Over the next 60 minutes you will be asked to complete a series of tasks.

This information will be used to evaluate a new applicant screening tool, which will be used by human resource departments and potential employers.

The design of the tool was modelled after social media websites to standardize the hiring process for positions with large number of applicants.

The tool provides employers an applicant's virtual resume and assist employers in selecting the best fitting applicants for various positions.

The tool includes several different features that may be added to the final product.

You will be asked to complete survey questions and we will record your verbal responses to better evaluate the tool and different features provided.

Please describe the steps you take and decisions you make verbally. Please also verbalize any emotional responses you have to the applicants.

A demonstration of this will be provided in a training video.

Following the training video, you will be presented with two versions of the tool with varying features. The training exercise will begin now, which will outline these features.

Script Task 1: Rating Applicants

This screen shot shows an overview of each applicant's virtual resume.

Each resume provides limited information of the applicant.

At the top of the screen, moving from left to right, you will see a tab for each applicant. There are three categories of information for each applicant:

education,

experience, and

volunteer experience and personal interests.

Within each of these categories is a statement about the applicant.

As part of the audio recording, please read aloud each recommendation.

For example:

After reviewing the profile and reading aloud each recommendation, you must complete the questions listed at the bottom of the screen.

You will complete this for each applicant. Use this task to familiarize you with the applicants. After completing the questions for each of the four applicants, you will now move onto selecting the applicant for the best fitting position.

Task 2 : Selecting Applicants into Position

This screen shot shows a summary of the information presented in the four applicant resumes.

Use this summary page to select the applicant who is most appropriate for the position listed on the right side of the screen.

This includes:

a supervisor within your department,

a co-worker within your department,

and a supervisor for a different department.

You will not be able to select the same applicant for more than one position.

To place an applicant in a position drag and drop the resume into the appropriate slot.

Please note that you can change your response by dragging a different applicant in a previously filled slot.

Once you have selected an applicant for a position, please rate how confident you are in the decision made, which ranges from very high confidence to very low confidence.

Please remember to verbalize your actions, feelings, and thoughts throughout your decision process.

For example, you may say: <>
When you click on the continue button, a pop-up will ask you to justify your decisions.
Type in your justifications and click save.
Please provide a specific and brief explanation for your decision.

Task 3: Ranking Applicants into Position
In the next task, you will need to rank the participants from best to worst for the three positions.
Do this by dragging the applicants to each ranking option.
Again, you can change the order of the ranked applicants at any point before clicking the "continue" button by dragging a different applicant into the ranking slot.
When completing the ranking be sure to verbalize your actions, feelings, and thoughts.
For example, you may say: <>

Task Summary and Scores
Once you have finished these three tasks, you will be presented a review of your performance.
This review covers the number of applicants rated, selected, and ranked.
You are also provided feedback on your accuracy of queuing future applicants.

You will complete two full sessions of these tasks.
You will use the full version of the tool in each session, but only one session will include the queuing and notification feature of the tool. You will be told which version you will be completing before you begin.

Ok, lets' begin!

The Notification Feature
You will evaluate two versions of the tool. One of them will have a special notification feature.
You will notice a pop-up in the bottom right of the screen as you are using the tool.
This feature indicates that new resumes have arrived and you will need to add or queue future applicants.
If a resume is listed as relevant for "Human Resources," you will need to add the resume to the queue by clicking on the box and then clicking "ADD."
If the resume is listed as relevant for a department other than "Human Resources," such as "Finance," do not add the resume to the queue.
You can exit the queue box without adding an applicant by clicking the "X" at the top of the box.
If you incorrectly add an applicant to the queue, the applicant will be highlighted red to provide you immediate feedback of your error.
You will only have 4 seconds to decide if an applicant needs to be added to the queue before this information disappears.
You will be scored on the number of applicants correctly added to the queue so be sure to carefully attend to this information when the pop-up appears.

3. *Pilot Survey*

Thank you for participating in our survey. The survey will take between 15 and 20 minutes to complete. The survey is being used to develop and validate questions for a study on recommendations for job applicants. Participation is voluntary. Data will only be used for

study development purposes and no personal identification information will be recorded.

At the end of the survey, you will be asked for your student number. This is so we can organise to get your Event movie voucher to you in appreciation for your time, and your responses will remain anonymous.

If you have any questions or concerns, please contact Dr. Kevin Eschleman (kevin.eschleman@wpafb.af.mil). Thank you again for your participation.

1. Please choose the answer that best describes you.

○   Please choose the answer that best describes you.

Native Australian: You were born in Australia and have lived here most of your life and identify as Australian.

○   **Near-native Australian:** You were NOT born in Australia, but you have lived here for the majority of your life and identify as Australian.

○   **Non-native:** You were NOT born here and you do NOT identify yourself as Australian.

2. What is the ethnicity you most closely identify with?

Rating Applicant Recommendations

In the following section, you are given a list of recommendations that will be used to describe potential employees. The recommendations are provided by coworkers, supervisors, teachers, and friends. Please identify the personal characteristic you think the recommendation describes. Your options include:

High Ability: The recommendation reflects the competencies and skills relevant to being a successful employee.

High Benevolence: The recommendation reflects that the applicant considers other people's interests before making decisions or acting.

High Integrity: The recommendation reflects that the applicant adheres to a set of principles and values that are considered positive.

Neutral: The recommendation does not strongly reflect any specifc personal characteristic.

Other: Mark "Other" if you believe the recommendation describes a personal characteristic that was not provided as an option.

3. This person has a strong moral foundation.

◉   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify) [   ]

4. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

5. This person is inconsistent in performance and completion of tasks.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify) [   ]

6. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

7. A good employee whose technical know-how is widely regarded as state-of-the-art.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

8. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

9. We have known each other for about two years.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

10. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

11. This person has a questionable value system.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

12. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

13. We have similar interests.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

14. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

15. This person's quality of work is not adequate.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

16. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

17. This person always does the right thing even in morally difficult situations.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

18. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

19. This person needs to learn to consult with other team members before acting.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

20. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

21. I am surprised we have not interacted more.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

22. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

23. This person is not honest nor upfront about intentions.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

24. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

25. This person is currently an employee at our organization.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

26. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

27. Questionable ability to perform essential job tasks.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify) [   ]

28. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

29. Please mark integrity.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify) [   ]

30. This person enrolled in several of my courses while in school.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify) [   ]

31. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

32. This person does not meet expectations of performance in all areas.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify) [   ]

33. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

34. This person always gets things done well.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify) [   ]

35. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

36. This person has a tendency to ignore others.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify) [   ]

37. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

38. We have not had an opportunity to interact much.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

39. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

40. The work devoted to planning and organizing is always careful and complete.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

41. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

42. This person needs to improve on basic work skills.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

43. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

44. This person has been an active member of the club for several years.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

45. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

46. This person is one of my most fair and impartial employees.

○   This person is one of my most fair and   ○   Integrity
impartial employees.   Ability
                                                              ○   Neutral
○   Benevolence

Other (please specify)

47. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

48. This person does not demonstrate a commitment to helping others.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)|_____

49. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

50. This person consistently demonstrated strong skills on class projects and understanding of course material.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)|_____

51. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

52. The interests of the team and coworkers were always placed before this person's self-interests.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)|_____

53. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

54. This person is not always truthful and honest.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)|_____

55. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

56. This person is proficient and compentent while completing work.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)|_____

57. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

58. This person has a strong character.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

59. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

60. This person is always excited about the club.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

61. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

62. This person lacks consistency in values and principles.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

63. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

64. This person does not put in an extra effort to help others.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

65. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

66. This person performs quality work.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

67. Based on your above response...

1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

68. This person has strong values that are well respected.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

69. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

70. This person is honest, just, and impartial without fail.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

71. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

72. This person consistently completed tasks at a high level.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

73. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

74. This person does everything possible to help out whenever possible.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

75. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

76. Has always been particularly good to me.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

77. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

78. This person takes good care of the employees.

○ Ability   ○ Integrity

○ Benevolence   ○ Neutral

Other (please specify)

79. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

80. I was always confident the work would be completed efficiently and competently.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

81. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

82. This person is kind and caring toward others.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

83. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

84. This person goes well above and beyond the call of duty to look out for other students' best interests.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

85. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

86. This person puts forth an extra effort to support others.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

87. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

88. This person has been part of our team for several years.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

89. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

90. This person is a valuable employee who performed at a high level.

  ○   Ability   ○   Integrity

  ○   Benevolence   ○   Neutral

Other (please specify)

91. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

92. This person is respectful of other people and their opinions.

  ○   Ability   ○   Integrity

  ○   Benevolence   ○   Neutral

Other (please specify)

93. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

94. This person lacks a moral center.

  ○   Ability   ○   Integrity

  ○   Benevolence   ○   Neutral

Other (please specify)

95. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

96. This person needs to become more compassionate and caring toward others.

  ○   Ability   ○   Integrity

  ○   Benevolence   ○   Neutral

Other (please specify)

97. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

98. This person's judgement is not always ethically sound.

  ○   Ability   ○   Integrity

  ○   Benevolence   ○   Neutral

Other (please specify)

99. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

100. This person is very supportive of others in the workplace.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

101. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

102. Please mark benevolence.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

103. This person acts on the up and up from what I've seen and heard.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

104. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

105. This person is always fair and honest.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

106. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

107. This person does not tolerate or listen to others.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

108. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

109. This person does not have a strong set of principles.

○   Ability   ○   Integrity

○   Benevolence   ○   Neutral

Other (please specify)

110. Based on your above response...
1 - Very Low   2 - Low   3 - Moderate   4 - High   5 - Very High

4. Pre - Screening Survey

You have been offered the opportunity to participate in a cross-cultural decision making study using the "Applicant Screen Tool." The purpose of this research is to investigate how people rate applicants using a newly developed software tool. Please complete the following survey in preparation for the experiment. If you have questions or concerns regarding the survey, please contact Dr. Natalie Ruiz.

Dr. Natalie Ruiz
Natalie.Ruiz@nicta.com.au
T +61 2 9376 2160

1. Please provide your student ID (unikey). This information will be used to link your survey responses to your participation in the study.
2. The following statements describe beliefs about the world. Please indicate your agreement with each statement.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

An individual who is currently honest will stay honest in the future.
Any phenomenon has numerous numbers of causes, although some of the causes are not known.
Everything in the universe is somehow related to each other.
We should consider the situation a person is faced with, as well as his/her personality, in order to understand one's behavior.
Nothing is unrelated.
A person who is currently living a successful life will continue to stay successful.
Any phenomenon entails a numerous number of consequences, although some of them may not be known.
Even a small change in any element of the universe can lead to significant alterations in other elements.
The whole is greater than the sum of its parts.
Future events are predictable based on present situations.
When disagreement exists among people, they should search for ways to compromise and embrace everyone's opinions.
Everything in the world is intertwined in a causal relationship.
It is more desirable to take the middle ground than go to extremes.
Current situations can change at any time.
The whole, rather than its parts, should be considered in order to understand a phenomenon.
It is not possible to understand the parts without considering the whole picture.
If an event is moving toward a certain direction, it will continue to move toward that direction.
Every phenomenon in the world moves in predictable directions.
It is more important to pay attention to the whole than its parts.
It is more important to pay attention to the whole context rather than the details.
It is desirable to be in harmony, rather than in discord, with others of different opinions than one's own.
Choosing a middle ground in an argument should be avoided.
It is important to find a point of compromise than to debate who is right/wrong, when one's opinions conflict with other's opinions.

We should avoid going to extremes.

The following statements describe individual characteristics. Please indicate your agreement with each statement.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

I would prefer complex to simple problems.
I like to have the responsibility of handling a situation that requires a lot of thinking.
Thinking is not my idea of fun.
I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.
I find satisfaction in deliberating hard and for long hours.
I only think as hard as I have to.
I prefer to think about small, daily projects to long-term ones.
I like tasks that require little thought once I've heard them.
The idea of relying on thought to make my way to the top appeals to me.
I really enjoy a task that involves coming up with new solutions to problems.
Learning new ways to think doesn't excite me very much.
I prefer my life to be filled with puzzles that I must solve.
The notion of thinking abstractly is appealing to me.
I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
I feel relief rather than satisfaction after completing a task that required a lot of mental effort.
It's enough for me that something gets the job done; I don't care how or why it works.
I usually end up deliberating about issues even when they do not affect me personally.
Imagine yourself in a job setting and rate your agreement with the following items.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

It is important to have a good working relationship with your direct superior.
It is important to be consulted by your direct superior in his/her decisions.
A subordinate should not be afraid to express disagreement with his/her superior.
A structure with a subordinate having two bosses should be avoided.
People at lower levels in the organization should carry out the requests of people at higher levels without questions.
People at higher levels in organizations have a responsibility to make important decision for people below them.
Once a manager makes a decision, people working for the company should not question it.
In work-related matters, managers have a right to expect obedience from their subordinates.
An organization's rules should not be broken, not even when the employee thinks it is in the company's best interest.
Managers should make most decisions without consulting subordinates.
It is frequently necessary for a manager to use authority and power when dealing with subordinates.
Managers should seldom ask for the opinions of employees.
Employees should not disagree with management's decisions.

Managers should not delegate important tasks to employees.
Most organizations would be better off if conflict could be eliminated.
One can be a good manager without having precise answers to most of the questions that
subordinates may raise about their work.
In order to have efficient work relationships, it is often necessary to bypass the hierarchical lines.
I am uneasy in situations in which there are no clear rules or guidelines.
Conflicts with our opponents are best resolved by both parties compromising a bit.

The following statements describe other people. Please indicate your agreement with each
statement.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

Most people are basically honest.
Most people are trustworthy.
Most people are basically good and kind.
Please identify how much the following statements describe you in general.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

I am the life of the party.
I sympathize with others' feelings.
I get chores done right away.
I have frequent mood swings.
I have a vivid imagination.
I don't talk a lot.
I am not interested in other people's problems.
I often forget to put things back in their proper place.
I am relaxed most of the time.
I am not interested in abstract ideas.
I talk to a lot of different people at parties.
I feel others' emotions.
I like order.
I get upset easily.
I have difficulty understanding abstract ideas.
I keep in the background.
I am not really interested in others.
I make a mess of things.
I seldom feel blue.
I do not have a good imagination.

7. Please identify your gender/sex.

☐ Male

☐ Female

8. Please identify your school/military status.

☐ College Freshman

☐ College Sophomore

☐ College Junior

☐ College Senior

☐ Military

Other (please specify) [   ]

9. Please select the response that best describes your ethnicity.

☐ Non-native

☐ Near-native

☐ Native

Self-identified ethnicity: [   ]

10. What is your age in YEARS?

11. What is the hand you primarily use (handedness)?

○ Right

○ Left

○ Both (ambidextrous)

12. Do you wear glasses or contact lenses to read?

○ Yes

○ No

13. Have you ever been involved in hiring or human resources?

○ Yes

○ No

5. Debriefing and Student Exercise

INFO3315 - Experiment 2 In-Class Exercise
HR Applicant Screening Tool

<u>Debriefing</u>
Trust is a critical variable in military operations, be it trust in leadership and team members or between opposing sides at the negotiation table. The purpose of this study is to examine the trust process across cultures and establish a fundamental understanding of how trust operates. The three data collection venues include: the U.S. (Wright State University), Malaysia (Sunway University), and Australia (NICTA)
The process of a trust judgement has both affective and cognitive components.

The HR applicant screening tool was a "cover story" in order to get samples of people making trust judgements based on 3 aspects: Benevolence, Integrity and Ability. The applicant's details were manipulated along these 3 variables. In the experiment you participated in (the "Australian" group), the cognitive load was also manipulated using a "dual task" design (The Notification Feature). This was an attempt to disrupt the trust judgement by overloading the "cognitive" processing and forcing subjects to use more "affective" processes to make the judgements. This would either change the applicant you trust the most, or not, and would give us a better idea on how people make trust judgements.

<u>Methodology</u>

1. In this testing experiment, all participants completed both versions of the software, with and without the notification feature.
a) What is this kind of design called?
b) What are the expected benefits of this design type?
c) A rank order effect may be observed in this experiment design. What is it and how can it be counteracted?

2. As a tester, you were asked to think-aloud while completing this experiment. What kind of information can we get from thinking-aloud protocols?

<u>UI Design and Implementation</u>

3. The tabbed pane widget used to display the applicant profiles and question panels cannot be used for navigation (to switch from one tab to another). Instead, the user is prompted to click "Continue" to progress within the task, and from profile to profile.

a) Cite at least one benefit of using this specific design combination.

b) What is its major drawback?

4. The application designers wanted to highlight some specific sections of the applicants' profiles. Cite 2 ways in which they tried to achieve this.

5. What other interaction patterns could have been used instead of radio buttons to record the answers to each review question?

Task 2: Fill positions and Task 3: Rank applicants

6. The layout of this panel changed from the previous task panels. Some interface features changed and some others stayed the same. Why do you think that was done?

7. The drag and drop mechanism allowed easy placement of the applicants into the ranking slots. Sketch a wireframe of an alternative mechanism that could be used to achieve the same goal of ranking the applicants.

8. The dialogue box to "Justify your decisions" popped up once your selections were made, which was not a good way to solicit feedback about the rankings. Give 2 ways in which it could have been done better.

Notification Feature

9. The notification feature allows incoming applicants to be sorted as they come in, at the same time as reviewing applicants. This avoids a build up of new applicants that need to be sorted through later.

a) What interface and interaction aspects of this feature worked well? (Name one)

b) What are the limitations of this implementation of the feature? (Name one)

c) How can these be overcome?

6. *Training Video Storyboard*

74