Lessons Learned From Conducting the SCAN Game

Dr. Jay Nunamaker
UA Regents' Professor of MIS
Director, Center for the Management of Information, University of Arizona

Overview

1. **Create a more realistic scenario: control and realism**
2. **Minimize facilitator presence in the process**
3. **Improve experiment setup to enhance audio and video quality**
4. **Synchronize data sources as they are collected**
5. **Evaluate the feasibility of a smaller experiment**
6. **Evaluate additional speech recognition services for game transcription**
7. **Ask more debriefing questions to investigate game strategies**

What we would do next time?

1. Create a more realistic scenario: control and realism

- Design another experiment scenario closer to the military application contexts. Game scenarios are helpful to obtain group truth and reduce sanctioned lies but far from the military scenarios. We've learned a lot from playing and analyzing the games
- Interview army personnel to get a better idea of what they do when they go to the field

What we would do next time?

2. Minimize facilitator presence in the process

- Adopt self-facilitation
- Use pre-recorded audios to automate facilitation
- Facilitator visualization / editing of game events with round designations
- Move facilitators to another room

What we would do next time?

3. Improve experiment setup to enhance audio and video quality

- Use **lavalier microphones** to capture audio with good quality. Problem: transcripts from Watson are disappointing
- Add **illumination** for each participant from Microsoft Surface or desktop. Problem: overhead lighting and / or background lighting cast shadows over participant faces
- Evaluate the use of Surface tablets and separate video encoding from experiment devices to reduce strain on devices

What we would do next time?

4. Synchronize data sources as they are collected

- Use synchronized cameras controlled by a central experiment management system
- Align and concatenate individual players' videos before identifying speakers
- Use or adjust a video codec so that recorded audio does not drift from video
- Record key timestamps to facilitate synchronization and segmentation
- Use both audio dings and a colored strobe flash at important timestamps (e.g., start of rounds and start of surveys)
- Integrate timestamping features with encoding devices

What we would do next time?

5. Evaluate the feasibility of a smaller experiment

- Prepare for lack of attendance
- Maintain a waitlist of participants to backup when not enough participants arrive
- Emphasize the importance of attending scheduled sessions when recruiting participants and send multiple reminders before the session
- Consider using smaller teams (e.g., 6 participants)
- Current deception research focuses on dyads, and dominance research in groups usually has a team of four
- Will it be easier for villagers to figure out the spies if the teams are smaller?
- The minimum number of players in the original Resistance game is 5
- Observation: best data is from games played by 8 experienced players

What we would do next time?

6. Evaluate additional speech recognition services for game transcription

- Differentiate deceptive strategies and deception utterances
- In the 40 games tagged by UCSB, 11.54% of the turns-at-talk by spies are lies / misleading statements
- Differentiate various forms of deception
- 27.7% of the deceptive turns-at-talk by spies are outright lies, and 72.3% are misleading statements (also considered to be lies)

What we would do next time?

7. Ask more debriefing questions to investigate game strategies

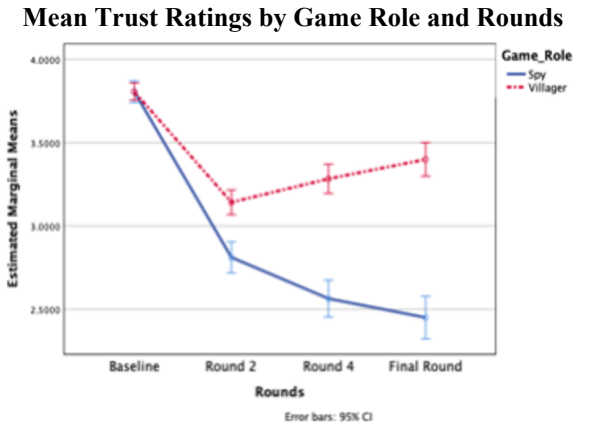
- Leave time to **discuss** debriefing strategy
- Ask in depth questions about their strategies

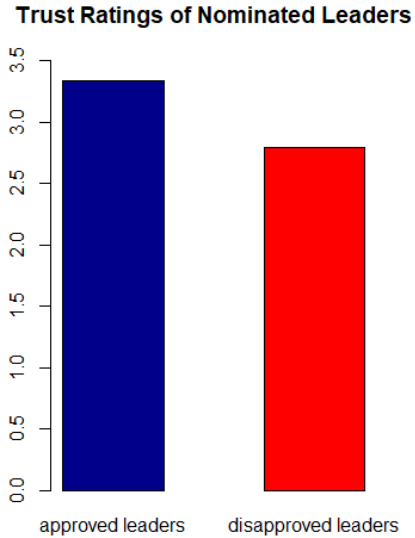
Initial Findings from the Trust Analysis

Why Study Perceived Trustworthiness

1. Perceived trustworthiness (along with dominance and nervousness) could be an indirect (proxy) measure for **deception**
   - Spies were trusted less than villagers
   - Trust ratings declined over the course of the game
   - Trust ratings of villagers remained higher and showed an upswing over time

2. Perceived trustworthiness could influence **decision-making** (e.g., votes on nominated leaders)
   - Nominated leaders who were approved are more trusted than those who were disapproved





Behavioral Predictors of Trustworthiness

Dependent Variable: Trustworthiness Score		Model
Control Variables	Gender (Male = 1)	
	Game Experience	**0.172(0.092)** *
	Native English Speaker	
	Game Status	**0.049(0.028)** *
Vocalic Features	TaT duration	**0.016(0.008)** **
	F_0-mean	
	F_0-Sd	
	Loudness-mean	
	Loudness-Sd	
	HNR-mean	
	HNR-Sd	
	Jitter-mean	
	Jitter-Sd	
	Shimmer-mean	
	Shimmer-Sd	
Main Effects	T3 (After Round 2)	
	T2 (Round 1 and 2)	**-0.500(0.126)** ***
	Game Role (Spy = 1)	
Interactions	Game Role * T3	**-1.360(0.181)** ***
	Game Role * T2	**-0.478(0.188)** *

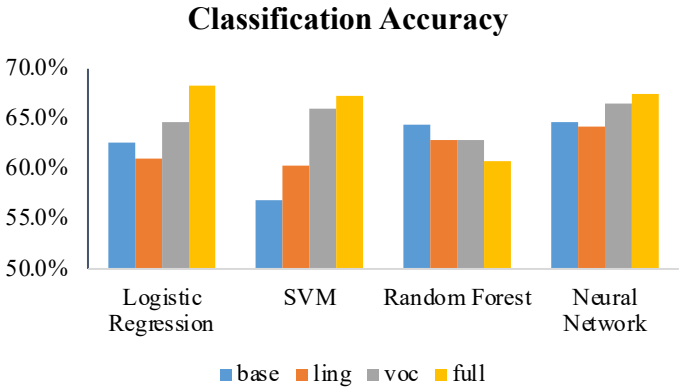
*Note: *p<0.1; **p<0.05; ***p<0.01, non-significant coefficients omitted*

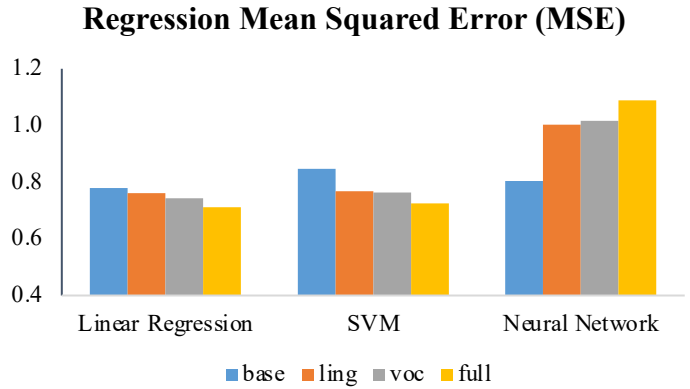
Dependent Variable: Trustworthiness Score		Model
Control Variables	Gender (Male = 1)	
	Game Experience	
	Native English Speaker	
	Game Status	**0.077***(0.028)**
Linguistic Features	Number of Words	**0.084****(0.042)**
	Positivity	
	Negativity	
	Hedge Ratio	
	Disfluency Ratio	
	First Person Ratio	
	Second Person Ratio	
	ARI Readability	**0.026***(0.015)**
Main Effects	T3 (After Round 2)	
	T2 (Round 1 and 2)	**-0.608***(0.126)**
	Game Role (Spy = 1)	
Interactions	Game Role * T3	**-1.449***(0.173)**
	Game Role * T2	**-0.430****(0.179)**

*Note: *p<0.1; **p<0.05; ***p<0.01, non-significant coefficients omitted*

- **In adversarial group settings, cues of perceived trustworthiness** include longer turn-at-talk duration, more words, higher comprehensibility (ARI Readability score)
- **Deceivers became less trusted as the game progressed**

Predicting Perceived Trustworthiness (Research in Progress)





Preliminary Findings

- Using Full dataset (linguistics and vocalics added to base) performs the best in most models
- Using Vocalics performs better than Linguistics in most models

Next Step

- Analyze full sample of trust data from behavioral predictors