Trust and Influence

5 MAR 2012

Joseph Lyons, PhD
Program Manager
AFOSR/RSL
Air Force Research Laboratory
<table>
<thead>
<tr>
<th>1. REPORT DATE</th>
<th>05 MAR 2012</th>
<th>2. REPORT TYPE</th>
<th>3. DATES COVERED</th>
<th>00-00-2012 to 00-00-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. TITLE AND SUBTITLE</td>
<td>Trust and Influence</td>
<td>5a. CONTRACT NUMBER</td>
<td>5b. GRANT NUMBER</td>
<td>5c. PROGRAM ELEMENT NUMBER</td>
</tr>
<tr>
<td>6. AUTHOR(S)</td>
<td></td>
<td>5d. PROJECT NUMBER</td>
<td>5e. TASK NUMBER</td>
<td>5f. WORK UNIT NUMBER</td>
</tr>
<tr>
<td>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</td>
<td>Air Force Research Laboratory, Wright-Patterson AFB, OH, 45433</td>
<td>8. PERFORMING ORGANIZATION REPORT NUMBER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. DISTRIBUTION/AVAILABILITY STATEMENT</td>
<td>Approved for public release; distribution unlimited</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. SUBJECT TERMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. SECURITY CLASSIFICATION OF:</td>
<td>a. REPORT</td>
<td>b. ABSTRACT</td>
<td>c. THIS PAGE</td>
<td>Same as Report (SAR)</td>
</tr>
<tr>
<td></td>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
<td>Same as Report (SAR)</td>
</tr>
</tbody>
</table>
Outline

• Program Trends
• Portfolio Description
• Trust Background & Motivation
• Trust Grants - 2
• Influence Background & Motivation
• Influence Grants - 6
• Recent Transitions
• Collaborations/Synergies
Program Trends

- Trust in Autonomous Systems
- Cross-cultural Trust
- Influence Effects
- Cognitive Mechanisms for Influence
- Computational Methods
- Modeling
NAME: Trust and Influence

BRIEF DESCRIPTION OF PORTFOLIO:
Basic research to explore the science of reliance (i.e., how humans establish, maintain, and repair trust of humans and technological systems) and the science of influence (i.e., understanding how to shape the behavior or attitudes of others).

LIST SUB-AREAS IN PORTFOLIO:
Science of Reliance
• Cross-Cultural Trust – Identify the antecedents of trust in different cultures
• Trust in Autonomous Systems/Autonomy – identify the factors that shape reliance in complex human-machine interactions

Science of Influence
• Understanding the behavioral effects of different influence tactics (air strikes, messaging, developmental activities)
• Understanding the cognitive mechanisms that drive influence effects – identify the avenues of influence for different groups
Trust Background

Trust = willingness of individuals to accept vulnerabilities from the actions of others with little ability to monitor their actions (Mayer et al., 1995)

Assumptions:
• Trust is a human phenomenon
• Trust & trustworthiness are independent (Mayer et al, 1995)
• Trust is relational
  • Humans in cross-cultural interactions
  • Complex human-machine interactions
• Trust is dynamic (Levine et al., 2006)
• Trust is only relevant in the context of risk (Parkhe & Miller, 2000)
• Trust has both affective and analytical underpinnings (McAllister et al., 1995; Lount, 2010; Stokes et al., 2011)
AF Tech Horizon’s 2010
“In the near to mid-term, developing methods for establishing ‘certifiable trust in autonomous systems’ is the single greatest technological barrier that must be overcome to obtain the capability advantages that are achievable by increasing use of autonomous systems” (p. 42)

Operational Challenges:
• Future battle ground - complex human-machine interactions
• Interactions with other cultural groups – where trust will be critical as HUMINT increases in value – partner capacity service core function (Schwartz, 2011)

Science Challenges: Appropriate reliance is really hard!
• Complacency and or under reliance are common pitfalls
• Automation often has unintended consequences (Parasuraman & Riley, 1997)
  • Automation paradox – reliable automation can lead to catastrophic error
• Interpersonal trust models are based on “Western” data/models
• Humans have trust biases (Lyons & Stokes, 2012)
• Little is known about how human trust principles apply to autonomy

Opportunities:
• Identify human-centric trust vulnerabilities before fielding autonomous systems
• Support AFCLC cultural competencies – trust building
Trust Domain/Scope

Interpersonal Trustworthiness
- Ability
- Benevolence
- Integrity

Trust Metrics

Human-Machine Interactions

Cross-Cultural Trust Issues

Human trust

Human Trust Biases/Vulnerabilities

Autonomous Systems & Automation

Human trust
Stokes – Dynamic Trust Model

PI: Charlene Stokes (AFRL) Lab Task
Drs. Lin (Sunway U) and Chen (NICTA)

Objective: Examine contextual factors that influence trust and trustworthiness
• Relationship type, load, culture

Approach: Manipulate trustworthiness and cognitive load in an applicant scenario using US, Malaysian, & Australian subjects

Results: Australian data collected (N=73)
• Load manipulation effective
  High M: 3.625, Low M: 3.037; t(72)=5.201, p<.01
• Trustworthiness dimensions uniquely influence selection
• The unique influence diminish under high cognitive load
  • Dispositional influences stronger

<table>
<thead>
<tr>
<th>LowCL Condition</th>
<th>% of time selected for position of ..</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicant</td>
<td>Supervisor</td>
</tr>
<tr>
<td>Ability</td>
<td>0.250</td>
</tr>
<tr>
<td>Benevolence</td>
<td>0.513</td>
</tr>
<tr>
<td>Integrity</td>
<td>0.238</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HiCL Condition</th>
<th>% of time selected for position of ..</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicant</td>
<td>Supervisor</td>
</tr>
<tr>
<td>Ability</td>
<td>0.250</td>
</tr>
<tr>
<td>Benevolence</td>
<td>0.350</td>
</tr>
<tr>
<td>Integrity</td>
<td>0.400</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.000</td>
</tr>
<tr>
<td>Total</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Atkinson – Role of Benevolence in Trust of Autonomous Systems

PI: David Atkinson (IHMC)

Objective: Experimentally evaluate the Impact of benevolence within a human-machine interaction

• Interpersonal trust and trust in automation models have been treated as orthogonal

Approach:
• Operationalize “benevolence” in an autonomous system scenario
• Evaluate the impact of benevolence using experimental methods

Collaborators:
• Peter Hancock (UCF)
• Deborah Billings (UCF)
• Robert Hoffman (IHMC)
Motivation – Influence

Lt Gen Flynn (Nov 2011)
Need to understand the “Precursors of war” – what are the triggers for attitudes and behavior in different parts of the world

Operational Challenges:
• DoD lacks precision in cultural-based influence – a science base is needed!
• Military Information Support Operations (AF IFO Roadmap, May 2008)
  • Target audience analysis is a top MISO requirements
• Air Force Targeting Center - quantification of behavioral effects for influence

Science Challenges:
• Manipulation of influence tactics not plausible in practice – but maybe in the lab
• Rational actor models may not generalize beyond the laboratory – need field research
• Data mining and modeling tools have outpaced theory in social media research

Opportunities:
• Revolutionize “targeting” within the AF – quantify hidden costs
• Support AFSOC/ACC need for cultural awareness (e.g., AFRICOM)
Influence Domain/Scope

Enabling Factors
- Social Media

Target Audience
- Values/Beliefs
- Behavioral Patterns
- Narrative

Computational Social Science

Military Action
- Kinetic/Non-Kinetic
- Stability Operations
- Trust Building
**Lyall: Terrorism, Governance, and Development**

**PI:** Dr. Jason Lyall (Yale)  
**Lead:** Jacob Shapiro (Princeton)

**Objective:** Evaluate the impact of Kinetic versus Non-Kinetic actions on subsequent violence

**Approach:** Working with CENTCOM, identify unclassified data and used a matching algorithm to compare villages in Afghanistan before and after an event  
- Matched on: geography, pre-existing violence, economic indicators, etc.

**Results:**  
- Kinetic actions led to higher retaliation but the effects were limited to 5km radius

<table>
<thead>
<tr>
<th>Type</th>
<th>15-day</th>
<th>30-day</th>
<th>60-day</th>
<th>90-day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damage</td>
<td>.369*</td>
<td>.542†</td>
<td>1.041*</td>
<td>2.121**</td>
</tr>
<tr>
<td></td>
<td>+9%</td>
<td>+12%</td>
<td>+15%</td>
<td>+19%</td>
</tr>
<tr>
<td>No Damage</td>
<td>.182†</td>
<td>.164</td>
<td>.359</td>
<td>.908</td>
</tr>
<tr>
<td></td>
<td>+6%</td>
<td>+5%</td>
<td>+5%</td>
<td>+7%</td>
</tr>
</tbody>
</table>

**DV:** Changes in mean number of insurgent attacks against ISAF in specified temporal/spatial window  
Based on (Lyall, 2011)

**Impact:** Results transitioned to the Air Force Targeting Center/ACC and briefed to CENTCOM
Matsumoto – Emotions and Intergroup Relations

PI: David Matsumoto (SFSU)  
Co-PI: Mark Frank (SUNY Buffalo)

Objective: Evaluate the role of emotions in predicting violent acts  
• Emotions trigger action tendencies  
• Anger, contempt, disgust

Approach: Examine key leader speeches describing in/out groups prior to acts of violence  
• Code emotions in text

Results: Acts of aggression were preceded by increased anger, contempt, and disgust
ARTIS - MUTUAL INFLUENCE OF MORAL VALUES, MENTAL MODELS AND SOCIAL DYNAMICS ON INTERGROUP CONFLICT

PI: Scott Atran, Rich Davis, Jermy Ginges

**Objective:** To examine how conflict impacts the relationship between religiosity and moral absolutes

- Religious acts may anchor sacred values
- This process could be triggered by threat

**Approach:** Surveyed Palestinians on attitudes toward Israeli-Palestinian issues

- Recognition of Israel as Jewish State, right of return for refugees, sovereignty, etc.
- Assessed: religiosity, perceived conflict, and moral absolutism

**Results:** Conflict intensifies commitment of religious devotees to sacred values

**Impact:** Prior research on sacred values has been presented to Congress, DARPA, House of Lords, Dept of State – and published in top-tier journals (e.g., Science)

Figure 1. Likelihood Of Being a Moral Absolutist (%) Depending On Religiosity And Perceived Threat to the Palestinian People (+/- 1 SD).
Objective: To assess the cognitive processing of sacred decisions.

Approach: Used fMRI, exposed subjects to statements varying in sacredness.
- "I’m a dog person; I believe in God"
- Offered subjects money to sign a document contradicting their value.

Results: Statements identified as sacred evidenced rule-based activity whereas bid statements showed utilitarian activity.

Impact: Transitioned to DARPA.
Axelrod – Case-based Influence

PI: Robert Axelrod (Michigan)

Objective: Examine the use of analogies and exemplars within different cultural groups
• Examine cultural sensemaking/narrative – historical references
• Salience, similarity, compellingness

Approach: code newspapers after significant events to examine use of historical cases
• 9/11; Arab Spring; Osama Letters

Results:
• Wide variety of historical analogies used
• Varies by geography
  • Cases more salient when more proximal

Following 9/11

- New York Times - Opinion Pieces
- New York Times - Front Page

- Mentions
- Cases
- Deny
- Avoid
- US involved
- Terrorist attack
- Middle East
- Historical Analogy
- post WWII
Axelrod – Case-based Influence

Demonstrations in Egypt 2011

-mentions
-cases
-deny
-avoid
-US involved
-Revolution/demonstration
-Middle East
-Historical Analogy
-post WWII

Osama Letters/Speeches

-mentions
-cases
-deny
-avoid
-US involved
-Terrorist attack
-Middle East
-Historical Analogy
-post WWII

DISTRIBUTION A: Approved for public release; distribution is unlimited.
PI: Kristina Lerman (USC)
MURI Lead: Jeff Brantingham (UCLA)

Objective: Examine the utility of social network analysis metrics for understanding influence

Approach: (sample)
• Examine utility social networks using DIGG
• Influence = re-posting
• Alpha Centrality vs. Page Rank

Results:
• Alpha Centrality better predictor of influence than Page rank for more local network structures
PI: Jeff Brantingham (UCLA) & team!

Objective: Use mathematical models and statistics to forecast patterns of violence in LA gangs

Approach:
• Apply statistical models to gang activity

Results:
• Model indentified dyadic rivalry as most violent – supported by actual crime data

Impact:
• Used by Predictive Police Unit
• Used by 711 HPW Forecasting CTC
Recent Transitions

• AF Minerva Princeton Project:
  – Direct transition to COIN efforts for ISAF, heavily cited in the 2011 Foreign Relations Committee Report on Aid in Afghanistan
  – Briefings to Gen McChrystal, Chairman of the Joint Chief of Staff, ISAF, DARPA
  – Lyall – airpower study transitioned to the Air Combat Command/Air Force Targeting Center
• UCLA MURI used by predictive policing unit in CA (featured on ABC/NBC news)
  – Transitioning to 711 HPW Forecasting CTC
• Yaneer Bar-Yam – Statistical models of geographic/ethnic boundaries & link between food costs and riots presented to Strategic Multi-layer Assessment community – Warfighters in the MISO domain
  • Cited in Top 10 Science Discoveries of 2011 by Wired Magazine!
• Sandy Pentland – computational methods applied in the DARPA Nexus 7 initiative
• Sacred values work has transitioned to the Navy and DARPA impacting national policies -- U.S. Senate, U.S. State Dept., House of Lords, DARPA (Featured in Time magazine article on Gaddafi)
• Greg Burns - Neurobiology of Violence leveraged by DARPA
Collaborations/Synergies

Trust
ARL: Trust work related to interfaces, robotics, networks
ARI: Trust in networked teams
NICTA: Funding technology development, some work on trust and cognitive load
NAVAIR: Trust & culture interests
IARPA: Funding large trust initiative on physiology of trust
ONR: Machine Ethics
AFRL: Trust is a core research area – close collaboration with 6.1 and 6.2

Influence
ONR: Close collaboration, co-funding
ARO: Training, mission rehearsal, face-to-face negotiation/interaction, etc. Focus on near-term: “something for the soldier”
DARPA: Interest in culture but more focused on neuroscience and training
OSD HSCB: Modeling, ops analysis, training (mainly 6.2-6.3)
Dept. of State: Social Media interests, possible co-funding
Air University Culture & Language Center: focus is mainly on training related to language and culture
Defense Equal Opportunity Management Institute: focused on cultural training