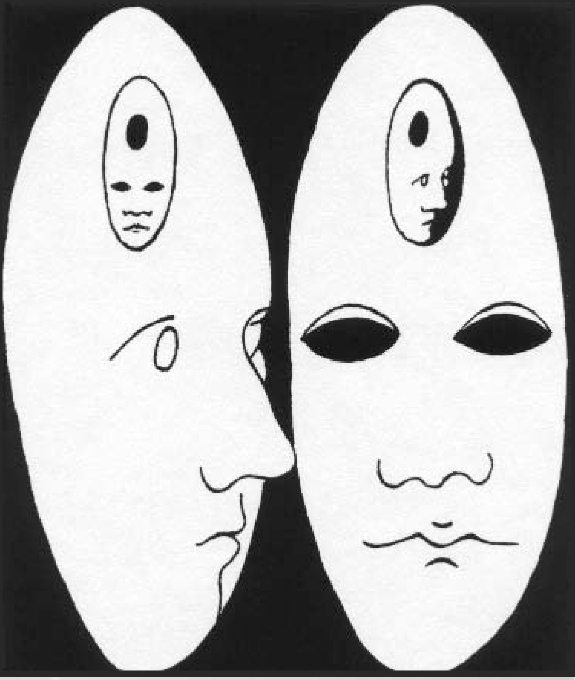
Modeling the Role of Theory of Mind in Social Interaction and Influence



David V. Pynadath
Stacy Marsella

USC

The projects or efforts depicted were or are sponsored by the National Science Foundation, the Air Force Office of Scientific Research and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Motivation

- Contemporary Operating Environment
  - Combatants, non-combatants, NGOs, CNN, etc.
  - Terrorist and insurgency network
  - Socio-economic environment
  - 2nd- and 3rd-order effects of policies, information, ...

- Problem:
  - Analysis, planning, and training all become harder

Modeling and Simulation

- Human-in-the-loop analysis
  - User-centric, not model-centric
  - Facilitate exploration and brainstorming
  - Support critical thinking
- Simulation-based training environments
- Key concerns
  - Provide possible outcomes, *not* single prediction
  - Enable model building by SMEs directly
    - We want to lose our jobs as modelers

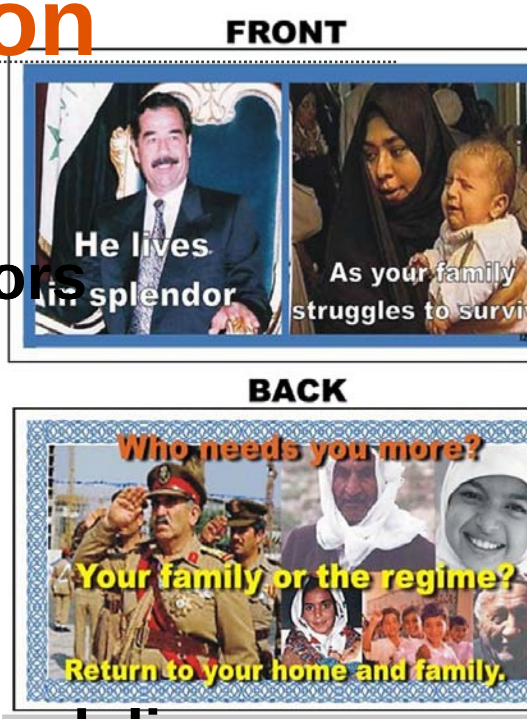
PsychSim

- Framework for social modeling & simulation
  - **Multi**agent **based**
  - Agents represent groups or individuals
  - Each agent models beliefs and generates behavior

- Used in a range of domains
  - Analysis and planning
  - Simulation-based training
  - Basic research on human behavior

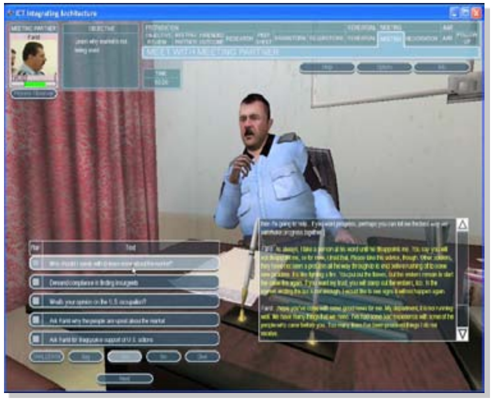
Exploratory Social Simulation

- Funded by OASD/SOLIC
  - Tool for PSYOP analysts and operators
- Follow-on funding by SOCOM
  - Focus on making tool user-friendly
- OSD-ATL/ONR/MITRE
  - Independent evaluation of country modeling
  - Part of Strategic Assessment effort
  - Model developed by MITRE (not us)

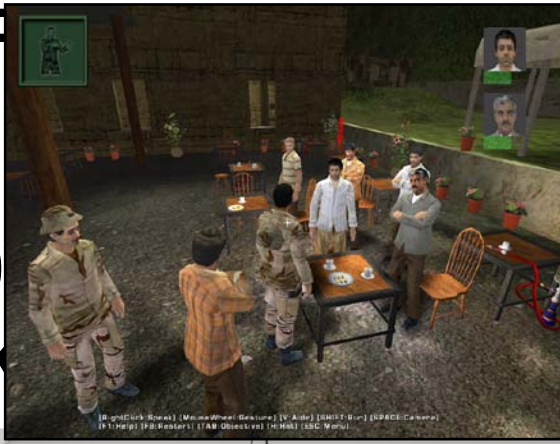

FRONT
He lives in splendor
As your family struggles to survi
BACK
Who needs you more?
Your family or the regime?
Return to your home and family.

Simulation-Based Training

- **BiLAT (Army)**
  - **Negotiation trainer for the military**



- **UrbanSim (Army)**
  - **Urban simulation trainer for stabilization ops**
- **Tactical Language Trainer (DARP**
  - **Foreign language training**



- **RISK (NIMH)**
  - **Teaching young adults to avoid risk behavior**

Range of Theories and Factors

Theories	Appraisal Theory of Emotion, Attachment Theory, Balance Theory, B&L Politeness, Influence Theories, Prospect Theory Personality Theories...
Factors	Trust, Support, Self-deception, Power, Blame, Control, Self-efficacy, Challenge, Threat, Goal congruence, Respect, Positive Face, Negative Face, Reactance, Affinity, Liking

Key Challenge

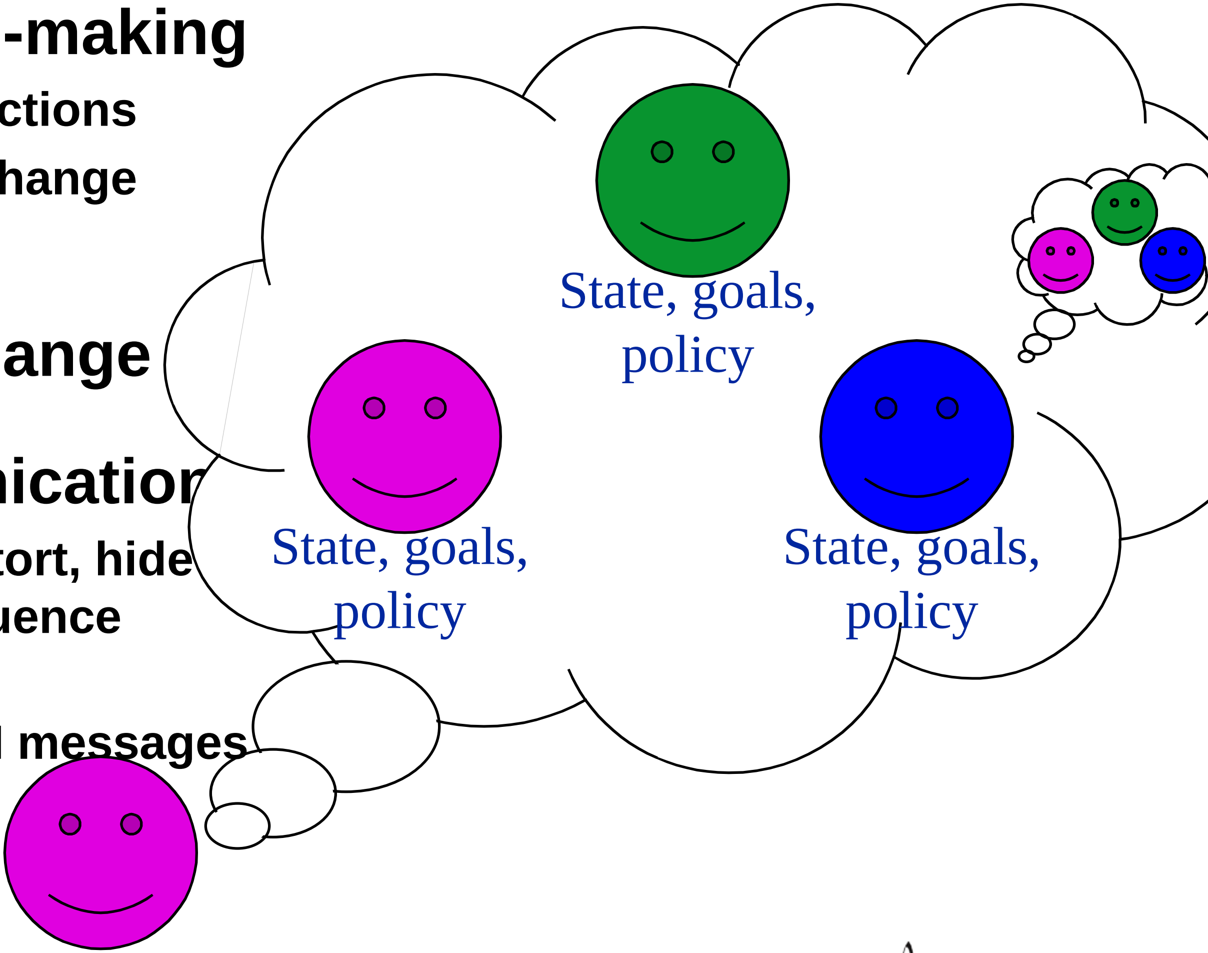
- *Goal:* Expressive simulation framework
  - *Theoretical:* Capable of modeling these factors
  - *Practical:* Useful in a range of domains/applications
- *Problem:* How to make them user friendly?
  - Must be easy to author and calibrate
  - Must be easy to understand and explain

Solution

- **Capable, but constrained, architecture**
  - ***Theory of Mind (ToM)***
    - **Agents have subjective perspectives about others**
  - ***Decision Theory / Subjective Expected Utility***
    - **Agents pursue their own goals**

Theory of Mind

- **Informs decision-making**
  - **Predict others' reactions**
  - **Select actions to change others' beliefs**
- **Informs belief change**
- **Informs communication**
  - **Communicate, distort, hide information to influence others**
  - **Communicate ToM messages**



Decision Theory

- Maximum Expected Utility
  - Agents choose behavior to maximize utility
  - Bounded rationality
  - Domain-independent algorithms

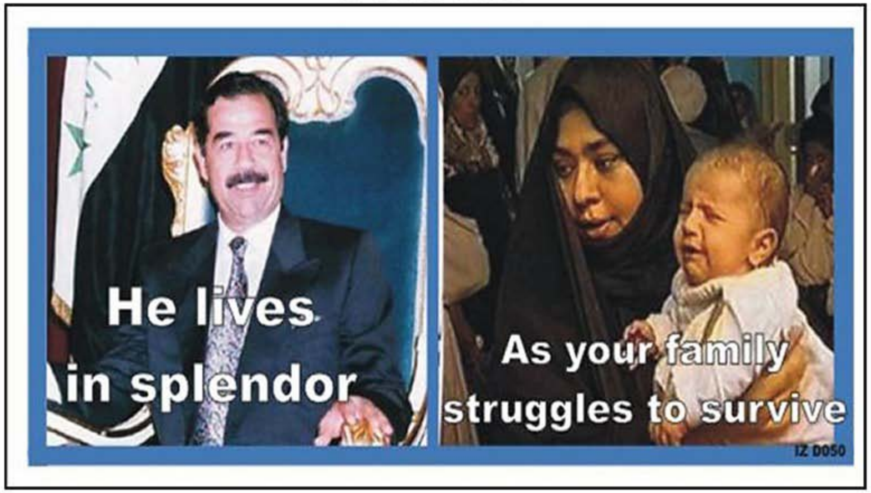
- Quantitative models are sensitive to degrees
  - Tradeoffs among conflicting goals
  - Risk attitudes when deciding under uncertainty

PsychSim: Architectural Claim

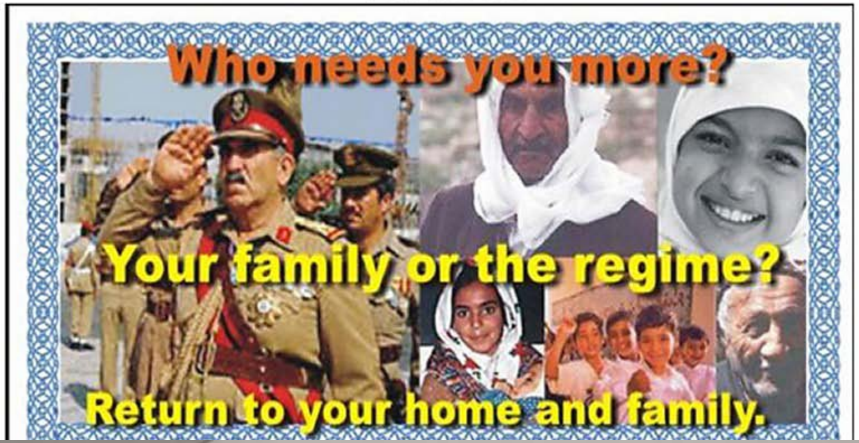
- ToM & DT is sufficient for modeling key factors
  - Factors may be derived from existing base components
- Advantage: Simplifies models
  - New phenomena derive from already authored parameters
  - As opposed to authoring new content for each new module
- Advantage: No additional integration
  - New phenomena operate in same framework as existing ones
  - As opposed to explicit management of module interactions
  - Therefore, existing algorithms apply
- Advantage: Framework is extensible

Modeling Influence

FRONT



BACK



- Theory of mind
  - what do soldiers think of:
    - Saddam
    - themselves
- Decision theory
  - Saddam cares about
    - his own welfare, vs.
    - the Iraqi people's welfare
  - the soldier cares about
    - the regime, vs.
    - his family's welfare

Modeling Influence: Consistency

- Is message consistent with what I've seen?
  - Also: Consistency with norms, cherished beliefs; with subgroup (In/Out group, consensus)
- If message is true, does past behavior make more sense?
  - "Makes more sense" = "has higher utility to actor"
- Saddam cares more about himself?
  - Consistent with any observed "selfishness"
  - But inconsistent with any observed philanthropy

Modeling Influence: Self-Interest

- Is message good news for me?
  - Wishful thinking, self-deception, motivated inference
- If message is true, am I better off?
  - “better off” = “higher utility to me”
- Example message is good news?
  - Saddam being a selfish leader = lower utility
  - My family struggling to survive = lower utility

Modeling Influence: Sender's Interest

- Does sender benefit have ulterior motive?
  - If so, I am less likely to believe it
- If I believe message, is sender better off?
  - "better off" = "higher utility to sender"
- Does coalition have ulterior motive?
  - If I return to my family, Iraqi army is weakened
  - Thus, coalition is more likely to achieve its goal

Modeling Influence: Bias factors

- Do I like sender of message?
  - Has sender’s behavior benefited me in the past?
  - “benefited me” = “increased my utility”

- Do I trust sender of message?
  - Has sender been truthful in the past?
  - “truthful” = “sent messages that I believe to be true”

Research on other factors

- Trust & Cross-organization Info Sharing
    - USC Marshall School of Business, funded by Lockheed Martin
- Self-deception / handling EU paradoxes
    - Ito, Pynadath & Marsella (IVA08, AAMAS09)
- Emotion (appraisal theory)
    - Si, Marsella, Pynadath (IVA08)
- Stereotype formation
    - Pynadath & Marsella (AAAI07)
- Influence Theory and Message Acceptance
    - Marsella, Pynadath, Read (ICCM04); Pynadath, Marsella (IJCAI05)
- Attachment Theory

Summary

- ToM and DT have proven sufficient so far
  - PsychSim currently realizes a range of factors
  - Uses information already present in behavior model
  - Obviously not yet exhaustive

- Exploratory Social Simulation
  - To aid experts in analyzing complex social situations
  - To support training for soldiers, analysts, etc.