Advances in Intelligent Systems and Computing 480

Sae Schatz
Mark Hoffman *Editors*

Advances in Cross-Cultural Decision Making

Proceedings of the AHFE 2016 International Conference on Cross-Cultural Decision Making (CCDM), July 27–31, 2016, Walt Disney World®, Florida, USA

Springer

Advances in Intelligent Systems and Computing

Volume 480

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

More information about this series at http://www.springer.com/series/11156

Sae Schatz · Mark Hoffman
Editors

Advances in Cross-Cultural Decision Making

Proceedings of the AHFE 2016 International Conference on Cross-Cultural Decision Making (CCDM), July 27–31, 2016, Walt Disney World®, Florida, USA

Springer

Editors
Sae Schatz
Advanced Distributed Learning (ADL) Initiative
Arlington, VA
USA

Mark Hoffman
Lockheed Martin's Advanced Technology Laboratories
Orlando, FL
USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-41635-9 ISBN 978-3-319-41636-6 (eBook)
DOI 10.1007/978-3-319-41636-6

Library of Congress Control Number: 2016943960

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Advances in Human Factors and Ergonomics 2016

AHFE 2016 Series Editors

Tareq Z. Ahram, Florida, USA
Waldemar Karwowski, Florida, USA

7th International Conference on Applied Human Factors and Ergonomics

Proceedings of the AHFE 2016 International Conference on Cross-Cultural Decision Making, July 27–31, 2016, Walt Disney World®, Florida, USA

Advances in Cross-Cultural Decision Making	*Sae Schatz and Mark Hoffman*
Advances in Applied Digital Human Modeling and Simulation	*Vincent G. Duffy*
Advances in Human Factors and Ergonomics in Healthcare	*Vincent G. Duffy and Nancy Lightner*
Advances in Affective and Pleasurable Design	*WonJoon Chung and Cliff(Sungsoo) Shin*
Advances in Human Aspects of Transportation	*Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico and Andrea Vallicelli*
Advances in Ergonomics In Design	*Francisco Rebelo and Marcelo Soares*
Advances in Ergonomics Modeling, Usability & Special Populations	*Marcelo Soares, Christianne Falcão and Tareq Z. Ahram*
Advances in Social & Occupational Ergonomics	*Richard Goossens*
Advances in Neuroergonomics and Cognitive Engineering	*Kelly S. Hale and Kay M. Stanney*
Advances in Physical Ergonomics and Human Factors	*Ravindra Goonetilleke and Waldemar Karwowski*
Advances in The Ergonomics in Manufacturing: Managing the Enterprise of the Future	*Christopher Schlick and Stefan Trzcielinski*
Advances in Safety Management and Human Factors	*Pedro Arezes*
Advances in Human Factors, Software, and Systems Engineering	*Ben Amaba*
Advances in Human Factors and Sustainable Infrastructure	*Jerzy Charytonowicz*
Advances in The Human Side of Service Engineering	*Tareq Z. Ahram and Waldemar Karwowski*

(continued)

(continued)

Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries	*Sacit Cetiner, Paul Fechtelkotter and Michael Legatt*
Advances in Human Factors in Sports and Outdoor Recreation	*Paul Salmon and Anne-Claire Macquet*
Advances in Human Factors and System Interactions	*Isabel L. Nunes*
Advances in Human Factors, Business Management, Training and Education	*Jussi Kantola, Tibor Barath, Salman Nazir and Terence Andre*
Advances in Human Factors in Robots and Unmanned Systems	*Pamela Savage-Knepshield and Jessie Chen*
Advances in Design for Inclusion	*Giuseppe Di Bucchianico and Pete Kercher*
Advances in Human Factors in Cybersecurity	*Denise Nicholson, Janae Lockett-Reynolds and Katherine Muse*

Preface

The Cross-Cultural Decision Making (CCDM) research focuses on improved decision making across a variety of cultural constructs, including geographical, historical, sociological, organizational, team, and technology interactions. This includes the research of experts and industry practitioners from multidisciplinary backgrounds, including sociology, linguistics, human–computer interaction, human factors engineering, systems engineering, military science, psychology, neuroscience, instructional design, and education, who showcase the latest advances in our understanding of the role of culture on decision making in numerous settings. Improved decision making among members of diverse teams and within organizational systems, and innovative ways to measure and assess that process, comprise the foundation for many projects discussed in these volumes. The influence of culture on decision making is pervasive, as reflected in the diverse disciplines represented by those individuals and entities involved in sociocultural research and engineering. The CCDM collection features papers that discuss emerging concepts, theories, and applications of cross-cultural decision making knowledge. The work described in these chapters reflects dedicated research by a wide range of expert academics and practitioners from around the world.

Each of the chapters of this book was either reviewed or contributed by the members of editorial board. For this, our sincere thanks and appreciation go to the Board members are listed below:

C. Fidopiastis, USA
J. Frank, USA
M. Hail, USA
A. Heaton, USA
D. King, USA
G. Klein, USA
M. Kruger, USA
A. Murata, Japan
S. Numrich, USA
J. Pfautz, USA
P. Picucci, USA
E. Raybourn, USA
E. Reitz, USA
A. Ruvinsky, USA
L. Saner, USA
D. Scapin, France
J. Stodd, UK
R. Taiar, France

We hope that this book, which is the international state-of-the-art in CCDM, will be a valuable source of theoretical and applied knowledge enabling human-centered design of variety of products, services, and systems.

Arlington, TX, USA — Sae Schatz
Cherry Hill, NJ, USA — Mark Hoffman
July 2016

Contents

Part I
Dynamic PMESII Modeling in Complex Environments

Expeditionary Modeling for Population-Centric Operations in Megacities: Some Initial Experiments

Brian Kettler and Jennifer Lautenschlager

Abstract Fueled by globalization and urbanization, megacities are extreme exemplars of complex, population-centric operating environments by virtue of their size, density, complexity, and dynamism. These likely epicenters of future instability make them probable locations of future military operations. As the epitome of complex, population-centric operating environments, megacities and other dense urban areas are of increasing interest to the US military, particularly the Army. Today, course of action analysis for operations in these environments is limited by the availability of analyst expertise, relevant data, and time. This paper presents a future vision of Expeditionary Modeling in which ensembles of diverse computational models are rapidly assembled, tailored, and brought to bear to quantitatively characterize the likely direct and indirect effects of Diplomatic, Information, Military, and Economic (DIME) actions in these environments. This paper describes some preliminary experiments to demonstrate the feasibility of this model-based approach, leveraging extant models for human and urban terrain.

Keywords Megacity · Modeling · Urban terrain · Human terrain · Military planning

1 Introduction

Fueled by globalization and urbanization, megacities are extreme exemplars of complex, population-centric operating environments by virtue of their size, density, complexity, and dynamism. Megacities (cities with populations over ten million) are expected to double in number to over forty cities by 2025. By 2030, cities will

B. Kettler (✉) · J. Lautenschlager
Lockheed Martin Advanced Technology Laboratories,
4301 N. Fairfax Dr., Suite 500, Arlington, VA 22203, USA
e-mail: brian.p.kettler@lmco.com

J. Lautenschlager
e-mail: jennifer.a.lautenschlager@lmco.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_1

have over 60 % of the world's population and 70 % of the world's gross domestic product. These likely epicenters of future instability make them probable locations of future military operations by US and coalition forces on missions such as humanitarian assistance and counterinsurgency. Missions in developing megacities and other dense urban areas such as these must consider the urban operations triad of complex, urban terrain; heterogeneous populations; and fragile infrastructure. These environments are highly dynamic as they are driven by rapid, ongoing population growth; socio-cultural tensions; lack of adequate governance; local and broader political, social, and economic trends; and urban terrain development or damage (e.g., from disasters or conflict).

As the epitome of complex, population-centric operating environments, megacities and other dense urban areas are of increasing interest to the US military, particularly the Army. The Army has identified gaps in its ability to meet the challenges of these environments [1]. This domain presents opportunities to explore how causal knowledge about complex socio-technical environments, embedded in computational models, can provide military planners and operators with new capabilities.

The time available between the decision to intervene in a crisis and the beginning of operations to plan and prepare military operations may be limited to a few weeks or even days [2]. Consequently, only a few friendly (blue) courses of action (COAs) can be developed and evaluated (e.g., wargamed) against those for the enemy (red). Today this is typically done by a group of operators and analysts with few automated tools and often lacking detailed knowledge of the specific operating environment. This also introduces human cognitive biases (such as recency biases).

Computational models can help mitigate these biases and complement human decision making. Such models must span urban and human terrain. The latter include applying human social cultural behavioral (HSCB) models based on social science theory and empirical data. Socio-cultural effects are difficult to predict in general, compared to effects in physical systems. Urban terrain models address phenomena such as land use, traffic and pedestrian behavior. Infrastructure models include those representing physical networks, facilities and services provided (e.g., the power grid, water system, healthcare systems, etc.). Typically these models were created for other purposes such as understanding and managing urban development over time and were not designed with interoperability in mind.

Many challenges exist in building/tailoring, validating, and applying models, especially in an expeditionary setting with limited time, human expertise (regional knowledge, language expertise), and data. Models can take weeks, months, or even years to develop from scratch depending on their scope, availability of data, and the degree of validation required. Data to build, validate, and operate models is often a key limiting factor and may need to be collected in the expeditionary operating environment or mined from open sources (e.g., news, social media, country studies, etc.). Models must be modified as both real world events and areas needing deeper

analysis unfold in the operating environment. When applying models, analysts must be able to understand and trust their results through automated capabilities providing transparency, explanation, and exploration of how the models operates.

Towards this, we have articulated a new vision, "Expeditionary Modeling", which rapidly brings to bear dynamic, heterogeneous ensembles of data, computational models, and subject matter expertise to understand and predict the (in)direct effects of courses of action on the urban and human terrain of dense urban areas. This paper describes our recent work to demonstrate the feasibility of modeling the human and urban terrain for a future conflict scenario set in the megacity of Lagos, Nigeria and to quantitatively evaluate alternative courses of Diplomatic, Information, Military, and Economic (DIME) actions and their Political, Military, Economic, Social, Infrastructure, Information, Physical Terrain, and Time (PMESII-PT) effects using an ensemble of primarily extant models.

The broader Expeditionary Modeling vision is motivated in [3], which presents a wider survey of its challenges and applicable technologies and data sources. This paper focuses more concretely on our initial experiments with these technologies and data sources and presents some quantitative results and lessons learned from our experiments to explore the challenges of rapidly integrating diverse models of varying resolution (e.g., temporal, spatial); provisioning these models from extant, open data sources; mapping different scenario events and courses of action into these models; and making sense of the models' outputs. These tools included Athena, a multi-model social simulation developed by NASA's Jet Propulsion Lab, and CMSim, a socio-culturally realistic simulation of urban terrain and traffic developed by Lockheed Martin. The experiments described in this paper resulted in the following, demonstrated:

- Multi-model COA analysis feasibility was demonstrated by modeling a rich scenario in a complex operating environment with simultaneously simulated courses of action. An existence proof was built using JPL's Athena system, which was able to generate reasonable, quantitative outcomes from interacting direct effects of Red, Blue, and Green courses of action, exceeding the quantitative effects assessment capabilities of a single SME or group of SMEs.
- The feasibility of capturing additional effects of COAs (e.g., physical and subsequent social impacts on the urban terrain/traffic from checkpoints, bombings, etc.) was demonstrated by integrating models of mixed resolution into an ensemble: i.e., Athena at the week–district–group dvslr and CMSim at the road–building–citizen scale.
- The above models were able to be customized to a complex operating environment in several weeks versus months, exploiting open source data.

The remainder of this paper presents our experimental framework (Sect. 2) and summarizes two of our experiments in detail (Sects. 3 and 4).

2 Experimental Framework

The focus of this work is on using computational models to augment today's largely manual processes of analyzing military COA effects. Modeling and simulation has been successfully applied in the past, mostly in force-on-force applications. But promising results from DARPA's Integrated Crisis Early Warning System (ICEWS) program, the OSD/ONR follow-on Worldwide ICEWS program, IARPA's Open Source Indicators and OSD's Human Social Cultural Behavioral Modeling program are recent evidence that models of human terrain can accurately predict outcomes [4–6]. Several social science models take a hybrid approach to analyzing courses of action from a multidimensional (PMESII) perspective including the Conflict Modeling, Planning, and Outcomes Experimentation (COMPOEX) system [7] and National Operational Environmental Model (NOEM) [8].

The Athena PMESII simulation, developed by the Jet Propulsion Lab for US Army TRADOC, combines over a dozen diverse models (e.g., a six sector economic model, a combat attrition model, and a belief system model) in a tightly integrated fashion to span the PMESII spectrum [9]. Developed since 2009, Athena has proven useful in a number of exercises and real-world operations, chiefly by helping analysts compute the effects of complex courses of action generally determined through wargaming. Athena can take the COAs developed by humans for all actors and replay them in scripted fashion, computing the effects of actions from these COAs given Athena's representation of the target operating environment (OE) and collection of computational models. The calculations that Athena can do surpass the ability of a person or group of people to do without automation, especially across the breadth of Athena's models.

According to its developers, Athena's utility is in illuminating complex, causal dynamics of a situation over time rather than providing accurate point predictions of outcomes. Athena has been used to show quantitatively how a region or country may evolve under the influence of multiple, competing political actors, where actors may be regional powers, insurgents, criminals, or other factions trying to gain political control of "neighborhoods" (modeled provinces or other areas with the operating environment) by collecting support and influence from civilian groups in each neighborhood. Civilian groups include different demographic (social, ethnic, religious, etc.) groups. Actors influence civilian groups via strategies that comprise several dozen tactics that span the spectrum of DIME actions.

Athena's belief system model captures actor and group beliefs on a set of fixed topics and is used to determine the baseline affinity between them based on belief similarity. Each civilian group has a number of variable satisfaction concerns (in aggregate, their 'mood') including their perceptions of quality of life, safety, autonomy, and respect for their cultural beliefs. Simulation events such as conflict, combat, or infrastructure events such as power outages will negatively impact autonomy, safety, and especially quality of life. These effects can spread within and

across neighborhoods. Enough negative events in one neighborhood may adversely impact the support for actors and potentially cause a shift of control in a region.

As with neighborhoods, actors and groups may be modeled at varying levels of granularity. An actor could represent an international organization, country, local government official, insurgent organization, etc. At each simulation tick (week), Athena executes the specified (scripted) tactics for each actor (tactics execute only when the required resources are available and any situation-specific conditions on a tactic are met). Athena updates the corresponding simulation variables.

In our experimental framework, shown in Fig. 1, we applied Athena at a more tactical level than it has generally been used at to capture key relationships at different levels across the PMESII spectrum for a megacity. Experiment 1 (Multi-Model COA Analysis, Sect. 3) both models the OE and computes the effects of DIME actions (Athena's tactics). Athena was augmented (via scripts) with simple disaster models we built for flooding and epidemic spread (not shown).

In Experiment 2 (Mixed Resolution Modeling, Sect. 4), we investigated augmenting Athena with a model, Lockheed Martin's CMSim, of the urban terrain at a fine-grained level of detail (down to roads, buildings, and notional citizens taking trips and causing traffic). Action and events from Athena drove CMSim, which modeled at the minute versus week tick size. Traffic and other outputs of CMSim fed back into Athena. A linking model (not shown) was used to bridge between the two models. Thus Athena was augmented with an external model to compute effects it previously could not.

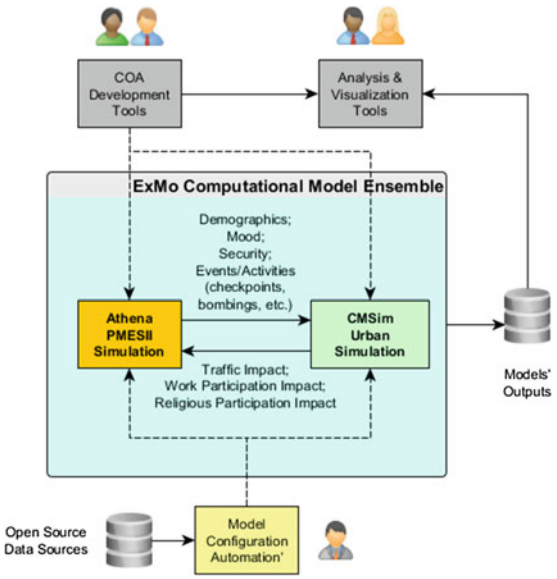


Fig. 1 Expeditionary modeling experimental framework

3 Experiment 1: Multi-model Course of Action (COA) Analysis

The goal of this experiment is to assess and demonstrate the feasibility and utility of a federated multi-model for computing PMESII-PT effects of DIME actions in a complex, population-centric operating environment. For this experiment, we applied Athena to a realistic, future scenario and simulated several COAs we developed. Alternative COAs were compared manually using the output from Athena.

A key task and accomplishment for our experiments was the development of a future, realistic (yet fictional) Lagos, Nigeria scenario experiencing multiple crises over a 12 month period. The scenario began with a flood in eastern Lagos city, to include the Apapa port area and parts of Eti Osa. This led to a cholera outbreak starting in Apapa. The unstable conditions then prompted Boko Haram to make a move into Lagos, where they staged a campaign of suicide bombings and attacks on the police. Response from the Nigerian government was modeled, along with varying US COAs. US forces were present to initially provide foreign humanitarian assistance then later counterinsurgency support.

For Lagos, we modeled the city as eight (large) unique districts (Athena "neighborhoods"), six political actors (including the Nigerian, Lagos State/City, and US Governments; Lagos criminals ("Area Boys" gangs); and Boko Haram), and 24 civilian/ethnic groups (Yoruba Christian, Yoruba Muslim, and "Other" ethnicities for each neighborhood). Major forces included Nigerian military, Lagos police, US military forces, Lagos criminals, and international aid workers.

Blue (friendly), Red (hostile), and Green (neutral) COAs were modeled as scripted actors' tactics, totaling over 300 tactics executed over the scenario. We generated six different blue force COAs, organized in two sets of alternative COAs. Overall scenario generation effort, including data and modeling tasks using Athena, was about 200 h. This included time to familiarize ourselves with Athena. This was counterbalanced by savings in time through reuse of scenario data from an existing similar exercise scenario.

Experiment 1 provided an existence proof for the feasibility of COA analysis and comparison. Interesting effects were observed for DIME actions across the PMESII dimensions. Effects were computed for multiple, simultaneously executing COAs (Red, Blue, and Green). Different effects were observed across different COAs, neighborhoods, and groups. This experiment showed the utility of Athena to compute such effects, well beyond what a single model could do or SME(s) could do manually.

The following sample effects were observed and quantified across our runs:

- political support/control: e.g., variable support for US, Boko Haram, etc.
- military: e.g., causalities, force effectiveness factors, cooperation
- economic: e.g., changing unemployment rate
- social mood (by neighborhood, civilian/ethnic group). These included satisfaction (mood) concerns: quality of life, autonomy, safety (perception), and cultural respect.

- infrastructure: e.g., changing road quality due to funding
- information: e.g., US info ops reduces civilian cooperation with Boko Haram

In analyzing our results, we considered a number of different Athena model outputs found useful in previous studies, focusing at first on the four satisfaction (mood) values specific to each neighborhood and civilian group, along with the mood of neighborhoods and groups. Later, we expanded our analysis to include horizontal and vertical relationships, neighborhood cooperation levels, population levels and flows, civilian casualties, significant events, security values, inter-group support, and several economic values. Anomalies and diverging behavior were visually identified within the graphs. In general, understanding what is significant in a given run and why it happened can be complex. Comparing across runs (e.g., for different COAs) is largely manual and is a ripe area for future automation.

As one example, Fig. 2 shows a COA comparison (COA 5 vs. COA 6) of various mood components by week for all civilian groups in the neighborhoods of Apapa (left), and Central Lagos (right). COA 5 prioritizes infrastructure repairs. COA 6 instead prioritized other concerns such as relief services for the epidemic. COA 6 more rapidly improves quality of life perception (left). US presence improves autonomy, which reflects goodwill toward the Lagos government as it benefits from US improvements to the neighborhoods (right). In Athena, the actor controlling a neighborhood receives credit (blame) when good (bad) things happen in that neighborhood, even if they are not directly responsible for those things.

Figure 3 shows a comparison of different civilian groups within a single neighborhood during one COA. After Boko Haram's arrival in week 6, Yoruba Muslim support increases while support from the others decrease. One reason for this is that, in sharing a common religion, there is greater natural affinity between the Yoruba Muslims and Boko Haram. Another reason is that the insurgent tactics of Boko Haram, a key contributor to the lack of cooperation found in other groups, minimize casualties against Muslims. Finally, an Information Ops campaign in

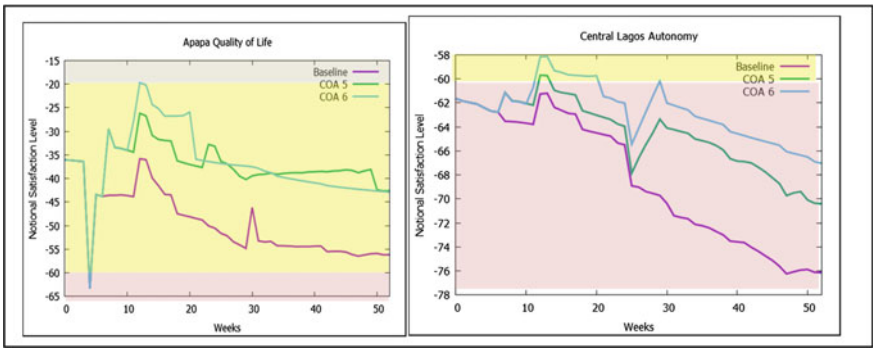


Fig. 2 Sample Experiment 1 outputs showing mood over time by neighborhood (average of all groups) for baseline (no US action) and 2 COAs. A more negative mood score is worse. Both COAs improve mood aspects

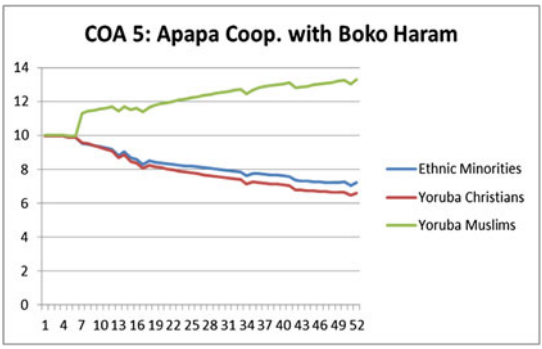


Fig. 3 Cooperation in the Apapa neighborhood between the different civilian groups and Boko Haram. The Yoruba Muslims increase their cooperation following Boko Haram presence around week 6, while the other civilian groups decrease theirs

local mosques, led by Boko Haram, has further increased their support from the Muslim population.

Athena performed capably and provided some interesting results with face validity, as well as a platform for additional experiments (e.g., Experiment 2). A consistent modeling challenge was that some significant actions did not appear to have significant effects in Athena. This was both realistic and problematic. In general, it would take a lot of effort to "move the needle" in a large/dense urban area in terms of deployed forces and other resources. Because Athena does not model physical geography (beyond having "neighborhoods" defined), it uses population size as a proxy for physical area, when the latter is needed to compute an effect. This heuristics probably works well when Athena is used at a more strategic level, but in our scenario the result was some dilution of effect as it assumed our population covered a larger area than it did in reality. While this can be mitigated some by adjusting Athena's coverage (a measure of the fraction of a neighborhood that is affected by some action) parameters, the best answer here may be just to model at a more detailed level with smaller neighborhoods where targeted effects can be localized, and their impact better assessed by the simulator. This would require additional data, time to configure, and run time.

Another challenge is that Athena actors do not exhibit behavior that is not pre-scripted via tactics in a strategy. This is fine in Athena's traditional use where it is used to compute the effects of COAs determined elsewhere (e.g., in a wargame). Conditions can be placed on tactics, and tactics do not execute unless the actor has sufficient resources. Spontaneous tactics (e.g., neighborhood workers holding a strike) cannot occur using just Athena. Mitigations for this were explored in another experiment not addressed in this paper.

A consistent challenge for complex simulations such as Athena is configuring it for a target scenario. This is compounded if the simulation is to be used to support an ongoing operation, in which case the simulation must be updated periodically to synchronize with the state of the real world. In separate experiments, we investigated how Athena (or a similar simulation) could be augmented with automation to help configure it from a variety of data sources. For example, we were able to develop Java code to help extract names and characteristics of actors/groups, typical

tactics, and inter-group dynamics from coded news event data and actor dictionaries [10], dbPedia, and the World Bank's World Development Indicators as part of this effort and directly initialize some scenario inputs in Athena. Additional automation (e.g., salient belief topics and group attitudes towards those) could be extracted from social media. Economic inputs could come from various online indicator sources or even be aggregated from transactional data. Generally some level of user post-processing is still needed to augment or vet the results of such automation. The availability of recent data also poses significant limitation on the automation potential. For example, a previous survey we did showed that data below the Local Government Areas in Nigeria was hard to come by via open sources: e.g., data specific to Lagos City (vs. Lagos State) and its districts.

4 Experiment 2: Mixed Resolution Modeling

The primary goal of Experiment 2 was to assess and demonstrate the feasibility and utility of modeling the urban terrain in high resolution to capture *physical* effects of DIME Actions. This complements Athena, which does not model physical distance or terrain. The main mechanism for this was the use of urban terrain simulator, Lockheed Martin's CMSim. The utility of this was evaluated standalone, as well as combined with Athena.

A key component of many urban modeling applications is traffic. Besides urban planning applications, traffic models are of high interest in dense urban areas where maneuver may be impeded during military operations, civilian evacuations, humanitarian relief supply, etc. Traffic can impact workers getting to their place of employment. In Lagos, over 60 % of economic activity is informal transactions such as roadside selling of all manner of goods [11]. Disruption to traffic means disruption of this economic activity and the influx of goods into the city.

CMSim was selected as a good exemplar of an urban terrain simulator (i.e., traffic, trips, etc.). Urban simulation was a focus of US Joint Forces Command's UrbanResolve effort, which has high relevance to modeling megacities [12]. One tool that began under this effort was CultureSim, which represents another type of urban modeling tool and builds on the Joint Semi-Automated Forces (JSAF) system [13]. CultureSim generates socio-cultural realistic agents moving about a city, i.e. individual people taking trips in vehicles or on foot. As trips are generated, the streets fill with vehicular and/or pedestrian traffic, according to the road capacities. Patterns of life can emerge over time from simulated citizens—on foot or by vehicle—taking trips throughout the urban area. The paths of vehicles are modeled, taking into account road connections, capabilities, etc.

CMSim was subsequently developed from CultureSim to be more computationally efficient than CultureSim by sharing data across computational instances at a more aggregate (road vs. entity) level and is an order of magnitude faster [14]. With adequate hardware, the simulation can support hundreds of thousands of simulation entities. CMSim exploits a terrain database, configured for a specific

area that can model road and building characteristics, traffic signals, bad events (e.g., disasters), crowds, etc. This can be partially generated but can require extensive post-processing to fix erroneous assumptions by automated ingest tools about road connectivity, directionality, etc. This experiment also sought to assess and demonstrate feasibility to rapidly configure CMSim for Lagos, exploiting open source data where possible.

CMSim's terrain database was configured for Lagos including modeling eight neighborhoods in our scenario ranging from 500 K to 3.7 M people each. These were model at a ratio of 221:1–1692:1 real to simulation entities (ratio is adjustable depending on fidelity vs. speed tradeoffs). Ten CMSim demographic profiles were created and instantiated for each neighborhood weekly based on Athena's demographic model.[1] Each profile specified a daily itinerary which included taking trips to either fixed or random locations (by type) and at fixed or random times. Road conditions permitting, trips were executed around a time with some probability distribution. Open source data was exploited including OpenStreetMap [15], which provided basic road data. Due to time and data availability constraints, buildings were modeled generically with the right distribution of the right type of building per neighborhood.

This experiment explored mixed resolution modeling through the use of Athena to model more "strategic" behavior of actors/groups weekly at the "neighborhood" (large districts of Lagos, in our scenario) level. CMSim modeled "tactical" behavior of individuals every few minutes at the road/building level. Figure 1 shows this integration. Operationally, Athena was run 1 tick (week) then paused. CMSim was then run to simulate the traffic given the conditions in Athena for that week. Checkpoints and bombings were tactics executed by actors in Athena in specific neighborhoods. The number (and type) of these were fed into CMSim, which determined specific locations and casualties. Flood levels from our simple model were passed to CMSim to determine terrain-based impacts on traffic (i.e., specific road blockages). As these events made areas of the city inaccessible, at least temporarily, citizens were unable to get to their places of work or worship. The resulting labor and worship participation rates, along with a traffic level (as impacted by flooding, checkpoints, and bombings) were fed back to Athena by CMSim, along with the general traffic quality (on a per-neighborhood basis). We used simple linking models to map CMSim outputs to Athena inputs. Then Athena was run for the next week, and the cycle repeated. CMSim was run on different COAs, and we observed differential effects within CMSim. Due to time available, no extensive validation was done save a face validity check that traffic seemed to be behaving realistically for the scenario and Lagos terrain database.

The result was an existence proof for the feasibility and utility of multi-resolution integration. This included integration of models at different

[1]Profiles included: Christian worker/non-worker; Muslim worked/non-worker; and truck driver (modeling the flood of goods around the city). Workers travel to/from fixed locations (i.e., businesses) at relatively fixed times. Religion was used to determine place of worship.

temporal, spatial, and organizational (e.g., group vs. individual) resolution which were successfully integrated and kept in sync through (dis)aggregation. Further, the ability to add complementary functionality to Athena via a loosely-coupled external model was shown. Additional CMSim-Athena integration was envisioned but not done due to time available. The use of multiple models does introduce additional transparency challenges in terms of being able to determine why a particular result occurred. Both Athena and CMSim produce lots of data per tick per run to analyze.

The experiment also showed the utility of adding an urban model. Different *physical* effects were observed for different COAs, including urban "terrain" effects on/from traffic, infrastructure, and (socioculturally-realistic) patterns of life (trips) resulting from actor/group tactics (e.g., checkpoints, bombings, etc.). These effects were more readily apparent at the scale of CMSim versus Athena. In Athena they were observed but overshadowed by other effects (see prior discussion of "moving the needle" in Athena).

An urban terrain model like CMSim was shown to be configurable from open source data. Very detailed terrain databases for CMSim can take many months to build, depending on the level of fidelity sought. For our purposes, building the Lagos terrain database took approximately three person-weeks. While traffic signals and other detail was omitted (due to lack of data and/or time), the fidelity was adequate for our purposes. This bodes well for rapidly getting a capability up and running for a new OE to support a new operation. CMSim could potentially be initially configured from patterns of life data obtained through ISR data exploitation or other means. If we had these normal patterns of life, then CMSim could be configured to emulate them for a particular city as a baseline for subsequent COA analysis.

As a sample result, Fig. 4 shows a comparison of COAs 5 (infrastructure repairs prioritized) and COA 6 (other actions prioritized) within CMSim. Impact on relative traffic quality (1 = highest), work participation, and religious participation are shown by week for the Apapa and East Lagos neighborhoods. Sharp downturn coincides with flooding in each neighborhood. COA 5 results in early improvement to all three metrics.

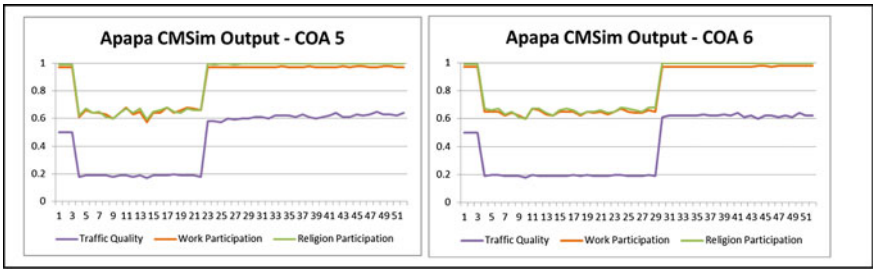


Fig. 4 Experiment 2 sample result: comparison of COAs 5 (infrastructure repairs prioritized) and COA 6 (other actions prioritized) within CMSim. Impact on relative traffic quality (1 = highest), work participation, and religious participation are shown by week

CMSim models had reasonable performance: approximately 45 min/simulated week per scenario neighborhood can scale out with more hardware, making it feasible to run CMSim "inline" at each Athena tick (simulated week). To improve scalability and mitigate data requirements, variable resolution modeling could be used. This would enable modeling some areas in more detail (temporal, spatial, etc.) on demand as needed, assuming data availability. This would be a kind of "model zoom" capability allowing focus on a specific small area in a city.

5 Conclusion

These preliminary experiments have shown the feasibility to take extant models and rapidly (i.e., weeks vs. months) customize those models to a new, complex, population centric operating environment, exploiting open source data. Using these models, we were able to show multi-model COA analysis that indicates potential interesting direct and indirect quantitative effects, which differ across several distinct COAs for a realistic megacity scenario. By integrating a human and urban terrain model at different levels of resolution, the breadth of the resulting model ensemble can better cover the PMESII-PT spectrum of effects. It would be difficult for a group of analysts to come up with all of these effects under time pressures and the dynamicity of the operating environment.

This work has shown the promise of the expeditionary modeling vision, but much remains to be done to fully achieve it. This includes expanding the breadth of actions and effects modeled (perhaps through explicit capture and reuse of context-specific causal knowledge), validating those models both theoretically and empirically, and creating tools to aid in mapping COAs into the inputs of multi-model ensembles and making sense of the models' results. Challenges such as data sparsity, data recency, and the use of models in ongoing operations also need addressing. The operational value would be significant for on-demand, broad, and deeply integrated modeling capabilities that support consideration of greater range of COAs through "what if" exploration for complex operating environments—helping planners to understand and ultimately mitigate undesired direct and indirect effects of their actions.

Acknowledgments This work was sponsored by the Defense Advanced Research Projects Agency via the Office of Naval Research (Contract N00014-14-C-0387). The views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Participants in this effort also included Lockheed Martin Mission Systems and Training; Strategic Analysis Enterprises, Inc.; and Perfecta Federal. US Army TRADOC G27 provided Athena tool support and data.

References

1. US Army SSG Megacities Concept Team: Megacities and the United States Army: Preparing for a Complex and Uncertain Future. Chief of Staff of the Army, Strategic Studies Group (2014)
2. RAND Corporation: Standing Up a More Capability Joint Task Force Headquarters: Research Brief. RAND Arroyo Center (2011)
3. Kettler, B.P., Hingst, R.G., Hoffman, M.A.: Expeditionary modeling for megacities and other dense urban areas. In: Cohn, J.V., Freeman, H., Schatz, S.L., Combs, D.J. (eds) Modeling Sociocultural Influences on Decision Making: Understanding Conflict, Enabling Stability. CRC Press (2016, in press)
4. Kettler, B., Hoffman, M.: Lessons learned in instability modeling, forecasting, and mitigation from the DARPA integrated crisis early warning system (ICEWS) program. In: 2nd International Conference on Cross-Cultural Decision Making, San Francisco (2012)
5. Intelligence Advanced Research Projects Activity: Open Source Indicators Program. http://www.iarpa.gov/index.php/research-programs/osi
6. Boiney, J., Foster, D.: Progress and Promise: Research and Engineering for Human Sociocultural Behavior. The MITRE Corporation (2013)
7. Waltz, E.: Situation analysis and collaborative planning for complex operations. In: 13th International Command and Control Research Symposium (2008)
8. Salerno, J.J, Romano, B., Geiler, W.: The National Operational Environment Model (NOEM). In: Modeling and Simulation for Defense Systems and Applications, Orlando, Florida (2011)
9. Chamberlain, R.G., Duquette, W.H.: Athena in 2013 and Beyond. Jet Propulsion Lab Publication, Pasadena (2013)
10. Lockheed Martin Corporation: Integrated Crisis Early Warning System (ICEWS) Coded News Event Data. Harvard Dataverse. https://dataverse.harvard.edu/dataverse/icews
11. Packer, G.: The Megacity. The New Yorker (2006)
12. Adamson, W.G.: Toward better readiness. US Army War Coll. Q. Parameters **45**(1), 45–54 (2015)
13. Moyer, D.: Obstacles to producing realistic traffic in a distributed environment. In: Spring Simulation Interoperability Workshop, Huntsville, Alabama (2006)
14. Moyer, D., Macannuco, D.: A mixed resolution simulation for highly scalable urban and regional populace modeling. In: Fall Simulation Interoperability Workshop (2007)
15. OpenStreetMap Foundation. https://www.openstreetmap.org

Time-Series Analysis of Blog and Metaphor Dynamics for Event Detection

Brian J. Goode, Juan Ignacio M. Reyes, Daniela R. Pardo-Yepez, Gabriel L. Canale, Richard M. Tong, David Mares, Michael Roan and Naren Ramakrishnan

Abstract Open source indicators (OSI) like social media are useful for detecting and forecasting the onset and progression of political events and mass movements such as elections and civil unrest. Recent work has led us to analyze metaphor usage in Latin American blogs to model such events. In addition to being rich in metaphorical usage, these data sources are heterogeneous with respect to their time-series behavior in terms of publication frequency and metaphor occurrence that make relative comparisons across sources difficult. We hypothesize that understanding these non-normal behaviors is a compulsory step toward improving

B.J. Goode (✉) · N. Ramakrishnan
Discovery Analytics Center, Virginia Tech, 900 N. Glebe Rd., Arlington, VA, USA
e-mail: bjgoode@vt.edu

N. Ramakrishnan
e-mail: naren@vt.edu

J.I.M. Reyes · D.R. Pardo-Yepez · G.L. Canale · D. Mares
Center for Iberian and Latin American Studies (CILAS), UC San Diego,
Gilman Dr., La Jolla, CA 9500, USA
e-mail: jireyes@ucsd.edu

D.R. Pardo-Yepez
e-mail: dpardoye@ucsd.edu

G.L. Canale
e-mail: gcanale@ucsd.edu

D. Mares
e-mail: dmares@ucsd.edu

R.M. Tong
Tarragon Consulting Corporation, 1563 Solano Avenue, Berkeley, CA #350, USA
e-mail: rtong@tgncorp.com

M. Roan
Department of Mechanical Engineering, Virginia Tech, 445 Goodwin Hall,
Blacksburg, VA, USA
e-mail: mroan@vt.edu

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_2

analysis and forecasting ability. In this work, we discuss our blog data set in detail, and dissect the data along several key characteristics such as blog publication frequency, length, and metaphor usage. In particular, we focus on occurrence clustering: modeling variations in the incidence of both metaphors and blogs over time. We describe these variations in terms of the shape parameters of distributions estimated using maximum likelihood methods. We conclude that although there may be no "characteristic" behavior in the heterogeneity of the sources, we can form groups of blogs with similar behaviors to improve detection ability.

Keywords Open source indicators · Metaphor · Temporal clustering

1 Introduction and Related Work

Open source indicators are a useful tool for identifying precursor signals in social media for large scale events. Past work has focused on forecasting protest events and disease outbreaks [1], but the methods employed for mining social media apply to a number of useful event forecasting scenarios. The open source indicators used for forecasting are numerous and include well-known outlets such as Facebook, Twitter, and RSS feeds of news sources and blogs. Success in finding documents within these sources that are of interest is typically subject to various algorithms using keywords [2, 3]. However, our recent work has suggested that more complex keys, such as linguistic metaphors, can be used to identify signals of interest.

The focus of this paper is to understand blog and metaphor content behavior along three dimensions in order to generate signals for event detection. Prior research suggests that blog behaviors can be classified as highly clustered (i.e., "bursty") in terms of temporal rates of appearance and diffusion across a blog network [4]. Such behavior may be intrinsic to human prioritizing behaviors [5, 6], and manifest in blog and metaphor usage.

In this paper, we discuss the methodology of recent work wherein blogs in Latin America are filtered for political metaphors to create signals for event detection. In particular, we address the issue of combining heterogeneous blog sources with sparse key metaphor signals. There are many different methods for addressing this problem, and the approach we take here is one of analyzing occurrence clustering by not assuming exponential rates of blog publication and metaphor appearance using a discrete Weibull distribution. Similarly, we analyze word counts (blog lengths) through the lens of a log-normal expectation. The conclusions we draw are that parsing similar blogs along these dimensions, with an eye toward complex system analysis, results in a set of signals that can be combined to yield an event signal without there being a "characteristic" blog to normalize. We conclude the paper with an example of how this information is aggregated to produce an event signal.

2 Characteristics and Analysis of Blogs and Metaphors

Our data set consists of blog documents that were extracted from four Latin American countries for this study: Argentina (89 blogs), Ecuador (59 blogs), Mexico (97 blogs), and Venezuela (82 blogs). The number of documents in each blog ranges from less than 100 to more than 10,000. In total, 589,089 documents are analyzed. We study three properties of these blog documents grouped by blog source: length (word count), publication frequency, and frequency of political metaphor usage. The distributions that we present of these features reveal interesting properties of the blog sources and political metaphor content. We use maximum likelihood estimation to model the word count as a log-normal distribution. Publication frequency and frequency of metaphorical usage features tend to follow a discrete Weibull distribution [7] and show evidence of varying degrees of temporal clustering. The shape parameter of the discrete Weibull distribution is a measure of related blog behavior for temporal clustering.

Word Counts. Word count distributions are formed for each blog in each country. There are two main types of word count distributions that emerge from the data. The first, shown in Fig. 1a, is one where the expected document length increases with length up to the peak of the distribution, and then starts to decay. One way to think of the distribution is not just a measure of the length of the document viewed as a whole. Rather, the word count distribution is the likelihood that if words were being displayed to an observer, what is the likelihood that a thought will be complete in some finite number of words? In the first case in Fig. 1a, the documents composing this blog series typically require a lead-up to complete the document. This assumes that these documents are full text and prose and not just lists. These types of documents are documents corresponding to authors that are more verbose in their prose. The second type of distribution, shown in Fig. 1b, is a very heavy tailed distribution where the likelihood of a continued document length decreases with document length. These are more "twitter" style blogs. Assuming full text and prose, then these documents tend to get their message out earlier than later in the document. Such documents can correspond to comments about videos or take the format of alerts where one tries to make the point as soon as possible.

The distributions here do not reveal anything about the content of the blogs – that would be a representation error due to speculation. However, if we assume that the documents are written for a purpose, then the distributions do reveal information about the relative nature of where the purposeful information appears in the document. A closer inspection of the distributions (a) and (b) in Fig. 1 show that they are not very different in terms of frequency of length leading toward the right tail. In fact, the Fig. 1b tail appears to be slighyly longer. However, the short documents in Fig. 1b suggests that there is a prevalence for short and concise documents given the increased probability of occurrence.

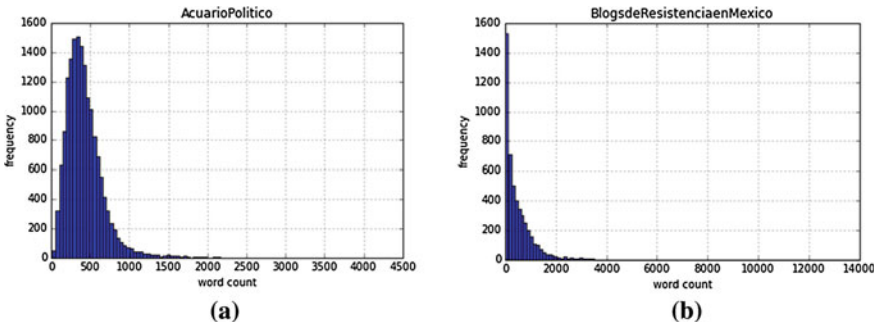


Fig. 1 Distribution of blog length (by word count) for blogs "Acuario Politico" and "Blogs de Resistencia en Mexico". Cast as a log-normal distribution, this represents the number of words until a document ends. In this particular case, such a distribution describes the probability that a "thought" will end given the amount of words at present in the document

For each country, a maximum likelihood estimate (MLE)[1] of the log-normal distribution was fit to the documents in each blog. Box plots of the log-mean and log-standard deviation are shown in Fig. 2 for each country. For each country, the quartiles for both log-mean and log-standard deviation tend to be comparable on the log scale. The majority of the mass of the distributions tend to be centered around a document length of 400 with a confidence interval of spanning [500, 800] words (Cox method). On the log scale, the differences in the mean quartiles can be quite substantial. Argentina and Ecuador, for instance, show more diversity in the expected blog length values than Mexico and Venezuela. The larger the spread of the quartiles indicates more diversity in the lengths of blogs, and subsequently, the prevalence for how long information is communicated in a particular blog.

Publication Frequency. As discussed above, the temporal behavior of blogs is subject to clustering (i.e., "bursty") behavior along several dimensions including temporal publication frequency and references to other blogs. We hypothesize that the temporal clustering (publication frequency) behavior of the blogs is indicative of the type or quality of the content of the documents published. An example is shown in Fig. 3 where two very different publication behaviors are seen. In Fig. 3a the blog "Carpe Diem" is seen to exhibit numerous short-order occurrences with fewer long-term occurrences decreasing in probability as the temporal duration increases. This trend is indicative of temporal clustering, and increases in publication activity could serve as an event indicator. Many blogs follow this format. There are also cases seen in Fig. 3b where there is a definitive publication frequency. For this blog, publication appears to occur regularly around every two weeks with few deviations from the trend. The issue with this sort of blog is that

[1]MLE for all sample distributions analyzed with Q-Q plots. Results hold well for most distributions. Distributions with larger tails show deviations in the higher quartiles, but with no particular trend. Errors in tail estimation are attributed to non-aggregate and micro-level events beyond the scope of this paper.

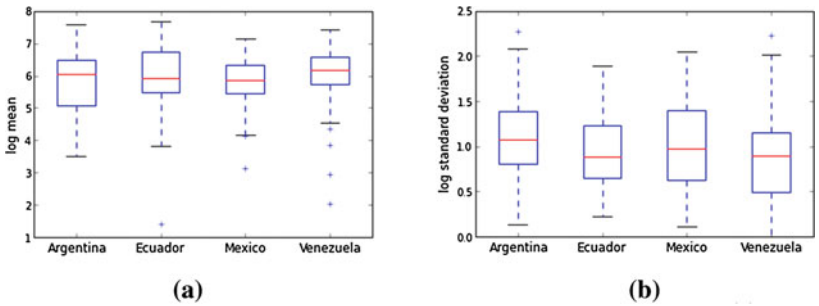


Fig. 2 Box plot of length log-mean and log-standard deviation parameters by country

there is an increased chance that the text of the recorded date does not coincide well with actual events. In this case, event detection will have to take into account the temporal references in language present in the document. However, it may also be indicative of a more reputable news source that publishes regularly, and the content may have higher quality content. Removing speculation as to the cause, simply aggregating blogs of different temporal clustering characteristics may lead to inconsistencies with respect to both temporal alignment and content type.

We propose that the temporal clustering in our data can be measured by the shape parameter, β, of the discrete Weibull distribution. Specifically, if we measure deviation from an exponential temporal duration of publication, then $\beta = 1$ indicates exponentially decreasing occurrence rates with time. $\beta < 1$ indicates temporal clustering behavior like that seen in Fig. 3a, and $\beta > 1$ indicates behavior of increasing likelihood of publication with greater temporal duration. The box plots shown in Fig. 4 show how the proportion of blog temporal clustering in our dataset in every country. For Argentina, Ecuador, and Venezuela three quartiles exhibit

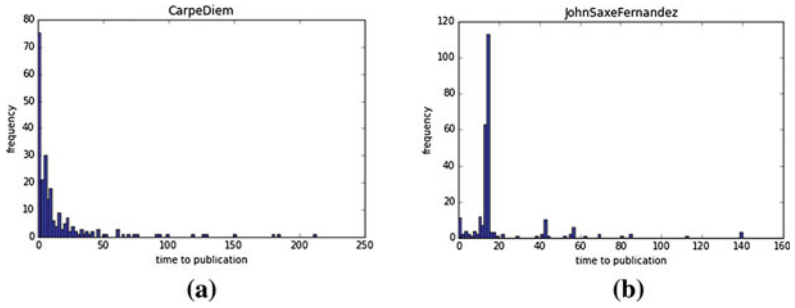


Fig. 3 Distributions of publication frequency. The distribution on the *left* **a** shows temporal clustering where the distribution on the *right* **b** shows a regular bi-weekly publication rate. When combining sources, different clustering behaviors in aggregate can change the perceived dynamics of the underlying content. Grouping by similar clustering behaviors can help mitigate effects of different time stamps

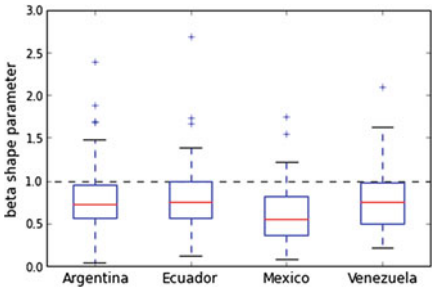
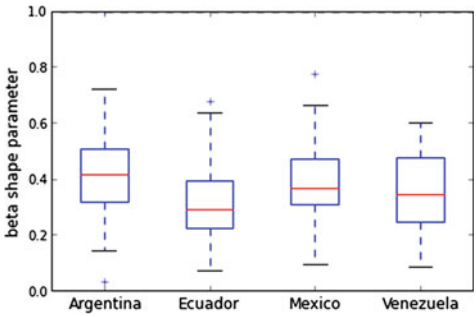


Fig. 4 Box plot of distribution of publication frequency. The *dashed line* represents a uniform rate of publication. All of the datasets in each country tend to exhibit mostly clustered publication rates, with one quartile of the blogs appearing to be a more uniform publication rate. However, the political blogs selected in Mexico tend to display more temporal clustering than the rest of the dataset

temporal clustering where one exhibits less temporal clustering. Our Mexico blog dataset exhibits more temporal clustering than the other countries.

Metaphor Usage. In each blog document, linguistic metaphors were discovered using a metaphor extraction algorithm [8]. The metaphors are mapped to a set of political target and source concepts (see Methods section). Instead of key words, these concept mentions serve as the basis for forming the event signal. The frequency of occurrence of specific source and target concepts in the blog documents is recorded over time. Similar to publication frequency, if linguistic metaphors appear regularly in particular blogs, then it may be more difficult to distinguish the signal from the noise, or ordinary use of language. For this, we use MLE to estimate the shape parameter of the discrete Weibull distribution for metaphor time series data, for each blog. The resulting box-plots separated by country are shown in Fig. 5. These results show that all of the blogs tend to exhibit a degree of temporal clustering with respect to metaphor usage, with some blogs being very cluster-centric in their use of political metaphor. Across countries, similar results are seen, with Ecuador exhibiting less clustering and Venezuela having a larger range at the center of the shape parameter distribution.

Fig. 5 Box plot of shape parameter distributions for metaphor time series data



These results are encouraging in that metaphor usage within blogs appears to be clustered and more likely to surround an event of interest. However, this does not preclude that metaphors are not persistent throughout a population and only referenced on occasions of interest. To the contrary, some metaphors like those that compare elections to a game tend to be quite persistent when aggregated across our whole blog corpus. To compensate for this, we use the results of publication frequency and blog length to aggregate across groups of similar blogs to effectively normalize for different blog dynamics and reduce the temporal complexity present within each aggregated unit of blogs.

3 Event Detection

We now provide an example of a use case where we cluster metaphor time series based on correlated metaphor behavior and temporal clustering of blog publication frequency. For this example, we use our Argentinian dataset and plot these metrics on the set of axes shown in Fig. 6. Information about the blog dynamics are shown on the axes themselves. The y-axis is the occurrence measure ($k = \beta$) for the blog publication frequency, and the frequency of document publication are shown on the x-axis. Each marker represents a subset of the Argentinean blogs displaying similar trends in specific metaphor activity based on correlated metaphor time series. The opaqueness of the marker shows the metaphor time series trending strength of

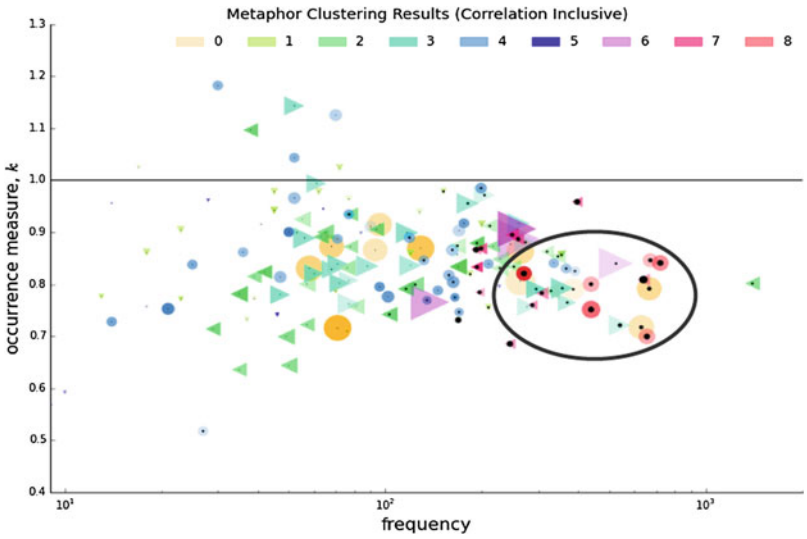


Fig. 6 Plots of correlated groups of blogs and metaphors in terms of frequency of appearance and occurrence measure. We look at groups of blogs and metaphors that have low occurrence measures (high clustering) and high frequency of appearance

groups of metaphors separated by co-clustering on signal correlation with a 30-day rolling Hamming window. The size of the marker indicates the number of similar blogs, and the size of the black marker represents the number of similar metaphor clusters within each subset of blogs. The end result of this procedure are groups 1–8 that represent aggregated collections of *both* blogs and metaphors displaying similar behaviors as defined in the previous sections.

For this example, we target the lower right region of the plot which contains blogs that show large numbers of entries with a propensity to show temporal clustering (low occurrence measure). We also target highly correlated metaphors which give more signal weight to the temporally clustered metaphor time series. Of the available groups, Group 8 is well represented in this region of the chart. This group is represented by the following metaphor source-target concept pairs: GOVERNMENT:{MACHINE, MOVEMENT, STRUGGLE} and WEALTH: {MOVEMENT, RESOURCE}. This indicates that there were instances of linguistic metaphors referencing concepts that GOVERNMENT is like a MACHINE or WEALTH is like a RESOURCE.

We plot the aggregated time series for Group 8 in Fig. 7. We investigate the signal for peaks in metaphor usage around particular dates. For Group 8, the concept pair of GOVERNMENT and MACHINE shows several peaks in the years spanning

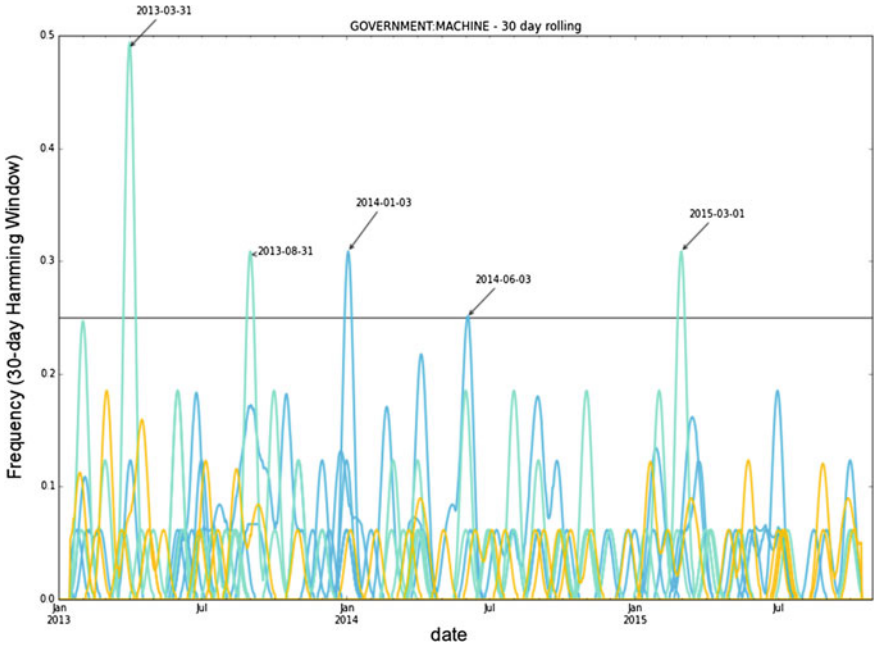


Fig. 7 The time series are shown for the GOVERMENT:MACHINE metaphor concept. The time series displayed here are all from blog and metaphor combinations that show similar temporal clustering, document length, and correlated metaphor usage

2013–2015. The highest peak, which occurs on 3/31/2013 on a centered monthly rolling average corresponds to the dates following when Pope Francis, from Argentina, was elected Pope by the Catholic Church. Also prevalent during this time was the Kirchner money laundering allegations appearing on an Argentinean news show [9]. The documents composing these time series make references to both Pope Francis and government finance, through the lens of political discontent.

4 Conclusions

Blog sources can be widely varied in terms of characteristics governing document length, publication frequency, and content (e.g., metaphor usage). For event detection, we use the key linguistic metaphor concepts to form time series for event detection. Our analysis of blog dynamics from our set of Latin American blogs shows that the document length tends to follow a log-normal distribution. This is indicative of the preference for communicating ideas, either in short bursts or more lengthy discussions. The publication frequency of the blogs were measured using maximum likelihood parameter estimation from a discrete Weibull distribution. The shape parameter of this distribution serves as a measure of temporal clustering of document publications. Blogs with high clustering behavior are more likely to coincide with events of interest than those with a constant publication rate. Our content analysis revealed that no one blog was likely to have persistent metaphor usage, contributing to the strength of the time series signal. Finally, we demonstrate a grouping relation among both blogs and metaphor content in order to generate a signal for event detection. For this, we targeted blogs with high temporal clustering, large document lengths, and similar trends in metaphors. We demonstrate that we can extract peaks by aggregating blogs of similar behaviors.

5 Methods

The blogs in this study were read and selected for political slants and references to political topics. Blog searches were conducted either using a directory (e.g., http://blogsdemexico.com.mx) or through Google keyword searches. Political blogs were identified by references to political entities, events, and people (e.g. “Maduro blogspot”). Also included in the corpus are blogs discussing issues such as the economy. Our blog dataset is biased toward political blogs, and specifically those with high metaphor content. Over all countries included in the analysis, the mean linguistic metaphor to document ratio was 0.49, meaning 1 out of every 2 documents on average contains a metaphor reference. All blogs were extracted using an automated tool to search, pull each document present within the blog structure, and identify meta information such as publication date, author, and title. Blogs were excluded if they were coded using JS Widgets.

The metaphors referenced in the text were identified using software, called the Metaphor Detection System (MDS), developed on the IARPA Metaphor program.[2] The MDS detects linguistic metaphors in Spanish, and has a detection F-score of 0.74 (precision 0.82, recall 0.68). Each linguistic metaphor is mapped to a target concept: BUREAUCRACY, DEMOCRACY, ELECTIONS, GOVERNMENT, POVERTY, TAXATION, and WEALTH. Each of the target concepts are linked to a source concept describing the target as mapped in the linguistic metaphor generating a source-target concept pair. The frequency of appearance of these concept pairs over time constitutes the time-series. These pairs are established a priori in the MDS architecture and can serve as a bias.

All MLE parameter estimates for distributions were found using R statistical software. All co-clustering operations were conducted using the scikit-learn package in Python. Co-clustering was performed over 20 randomized trials to determine the optimal number of clusters. The metaphor time series used to generate the correlation matrix were filtered with a central Hamming window of length 30 days to smooth out sparse signal irregularities.

Acknowledgments Supported by the Intelligence Advanced Research Projects Activity (IARPA) via DoI/NBC contract number D12PC000337, the US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

References

1. Ramakrishnan, N., et al.: 'Beating the news' with EMBERS: forecasting civil un-rest using open source indicators. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1799–1808. KDD '14, ACM, New York, NY, USA (2014)
2. Muthiah, S., et al.: Planned protest modeling in news and social media. Innovative Appl. Artif. Intell. (2015)
3. Chakraborty, P., et al.: Forecasting a moving target: ensemble models for ili case count predictions. In: Proceedings of the 2014 SIAM International Conference on Data Mining, pp. 262–270 (2014)
4. Goetz, M., Leskovec, J., McGlohon, M., Faloutsos, C.: In: International AAAI Conference on Web and Social Media (2009)
5. Barabasi, A.: The origin of bursts and heavy tails in human dynamics. Nature **435**, 207–211 (2005)
6. Zipf, G.K.: Human Behavior and the Principle of Least Effort. Addison-Wesley Press Inc, Cambridge (1949)
7. Nakagawa, T., Osaki, S.: The discrete Weibull distribution. *IEEE Trans. Reliab.* **R-24**(5), 300–301 (1975)

[2]See: http://www.iarpa.gov/index.php/research-programs/metaphor.

8. Mohler, M., Tomlinson, M., Rink, B.: Cross-lingual semantic generalization for the detection of metaphor. Int. J. Comput. Linguist. Appl. **6**(2), 115–136 (2015)
9. Pérez, S., Turner, T.: In Argentina, mix of money and politics stirs intrigue around kirchner. Wall Street J. July 28, 2014. [online] Accessed March 9 2016

Using Stories to Deepen Shared Human-Computer Understanding of PMESII Systems

Robert E. Wray, Jeremiah T. Folsom-Kovarik, Dylan Schmorrow, Randolph M. Jones and Robert Marinier

Abstract Humans tend to represent and to understand the world in terms of stories, while computer reasoning tends to require formal, mathematical representations. This paper describes a research prototype that enables computers to parse human stories and use collections of those stories to inform causal modeling of political, military, economic, social, infrastructure, and information (PMESII) systems. We introduce the need for causal modeling, the approach we have taken in implementing an initial proof-of-concept and the results from pilot testing of the software that illustrates functional capabilities and opportunities for deepening story-based computer interpretation of stories.

Keywords PMESII modeling · Computational narrative · Human-systems integration

1 Introduction

A gap, sometimes a gulf, typically exists between human understanding and computer understanding of a current situation, a past event, or a simulation. The gap leads to misunderstandings, mistakes, and missed opportunities. We contend that

R.E. Wray (✉) · J.T. Folsom-Kovarik · D. Schmorrow · R.M. Jones · R. Marinier
Soar Technology Inc., 3600 Green Court Suite 600, Ann Arbor, MI 48105, USA
e-mail: wray@soartech.com

J.T. Folsom-Kovarik
e-mail: jeremiah@soartech.com

D. Schmorrow
e-mail: dylan.schmorrow@soartech.com

R.M. Jones
e-mail: rjones@soartech.com

R. Marinier
e-mail: bob.marinier@soartech.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_3

this gap derives from differences in representation. Humans tend to represent and understand their world in terms of stories [1, 2], while computer reasoning tends to require formal, mathematical representations.

We are exploring ways in which computers can parse human stories and use collections of stories as a form of causal modeling. The approach, *story-based causal inference* (SBCI) includes both causal inference and analogy to comparable stories. The goal of SCBI is to enable computers to reason about and extend human stories. We hypothesize that bridging this representational gap will allow computers to help humans in new ways. For example, they can assess humans' plans and raise alerts about potential problems, identify potential problems or issues gleaned from reviewing stories about past events, and can support human-computer teaming without introducing technical burden on the human users.

This paper describes an initial proof-of-concept implementation of SBCI. We focus on the domain of military task force planning that requires consideration of political, military, economic, social, infrastructure and information (PMESII) systems. PMESII domains are especially apt for the challenge of story-based reasoning because complex causal interactions within and between these systems are difficult for even experts to understand and predict. In a planning context, where the planners may not have deep expertise in either PMESII systems or direct knowledge and experience in the area of deployment, computer-based support to identify potential interactions and implications may provide immediate impact in helping form better plans and more robustly identify points of potential failure and contingencies. In this PMESII context, we introduce the SBCI prototype, illustrate with some specific examples of the capability it provides, and summarize an initial software test of the prototype.

2 Illustrative Challenge: PMESII Modeling for Planning

Military organizations have invested significant research and resources in the development of processes to make the planning process more routine and less fragile to the limited expertise of individuals within a (multi-echelon) planning team. However, missions and requirements grow more complex and analytical methods are sometimes insufficient for today's complex missions. This is even more evident when a deployment must be planned in a short amount of time, which both limits what options can be considered and the feasibility of bringing in (many) experts to help widen and deepen planners' understanding of the mission in the context of its particular plans.

Short-term planning horizons, limited intelligence, and limited expertise compound the dangers of necessary assumptions in the planning process. Unacknowledged and incorrect assumptions lead to surprises, missed opportunities, and failures to consider second- and third-order effects of plans. Further, one of the most difficult tasks in the planning process is to generate and assess future states, to make meaningful forecasts, even when one's assumptions are inexact and partially

informed. Even experts with years of experience often are no better than novices in their ability to assess potential future states [3]. Thus, a core technical challenge of computer-supported planning is to enhance a planner's ability to surface and test assumptions, to help identify important and/or likely possible futures, and to understand how one's assumptions interacts with those possible futures.

Consider, as a specific example, mission planning in a Joint Task Force (JTF). Such task forces are typically faced with complex, novel, and urgent missions. This combination of factors makes planning more difficult and does not provide time for in-depth study and analysis of all factors that may impact the mission.

To illustrate, a JTF might be formed for the purpose of providing disaster relief immediately after a foreign natural disaster such as a typhoon striking a large island city in the Pacific. Such a problem requires the planning team consider many different factors such as the political, economic, military, social, infrastructure, and informational (PMESII) [4] systems within the society. Although good models of these factors may be present prior to the crisis, they most likely will not have been updated to reflect the impact of the crisis on these PMESII factors (washed out roads, disjointed political coordination due to downed communications, etc.). Additionally the presence of an outside military force, even in a humanitarian deployment, can drastically impact any prior models and expectations.

The prototype we describe is exploring how we can capture and use stories about an immediate disaster-relief situation and stories of similar past situations could let a human describe the disaster and their plan for relief, while a computer reasons about the human's stories and responds to them in order to enhance the human's understanding and plans. A desired outcome of this interaction is that the computer system would identify potential gaps, misconceptions, or unstated assumptions in the user's story. By acting on recognized gaps, the system could then assess the humans' understanding and plans, alerting them to important assumptions or opportunities overlooked in the time-stressed planning process.

3 Story Graphs

Human understanding of the world is not simply symbolic but is also grounded in stories. Human experience is subjectively narrative with its cognitive basis in the episodic memory that links memories of events in the context of a linear timeline [5]. Humans understand which plan actions and outcomes are plausible using this narrative experience. As above, we hypothesize that stories can provide a bridge between computers and a human planner.

Computer representations of stories exist [6, 7] and we are drawing on such prior work to inform our representation of stories. At a high-level, a *story* in our conception (Fig. 1) is a sequence of statements about world states and state-changing events, which are represented in a story graph. The story graph allows unification of the representation of knowledge and data from a multitude of sources such as past experience, real-time mission data, computer simulations, and expert knowledge.

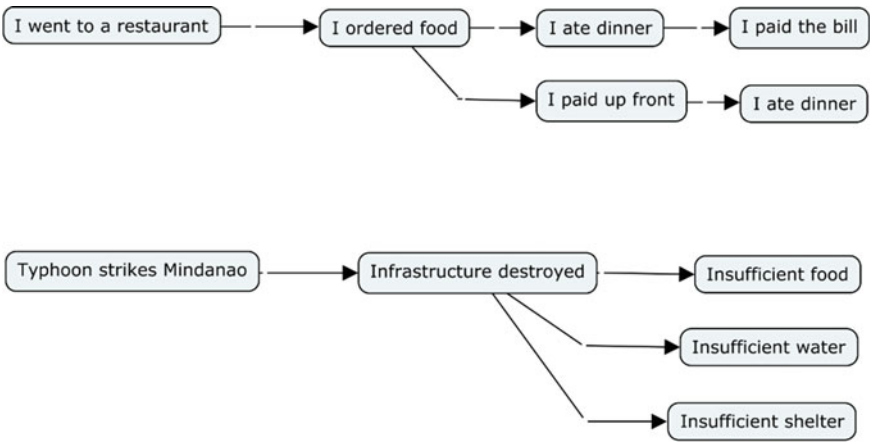


Fig. 1 Examples of simple stories

Nodes in the story graph describe events in the sequence. Links express the concept that a preceding event can cause (or contribute to the cause of) the subsequent event. At this point in our research, we are imposing minimal requirements on the semantics of nodes and links to provide greater flexibility in exploring how to use these stories for computer-based inference. For example, we expect to adopt in the future a richer vocabulary of linkages, so that graphs can express AND, exclusive OR, relative contributions, and so forth. Other researchers have described formal definitions of the kinds of causal links that can appear in stories [8]. However, there is no agreed standard so we took a least-commitment approach in defining causal linkages in this initial research.

Conceptually, stories are recursive so that a high-level node ("I ordered food") could be composed into a more detailed representation of the event ("I asked about the specials", "The server said…"). Eventually, nodes (and thus stories) are grounded in semantic knowledge of the world that makes each described event and state have meaning in the whole [9]. These story graphs can relate what could plausibly happen next, as opposed to what must deterministically happen. The graph structure allows for branching and merging, representing alternative linear paths, of which more than one may be true.

4 Story-Based Causal Induction

Figure 2 illustrates, at an architecture level, how the prototype system we have developed uses story graphs to generate suggestions for consideration during planning. Although our research program encompasses methods for extracting computer-based stories from other sources, such as natural language (stories from a

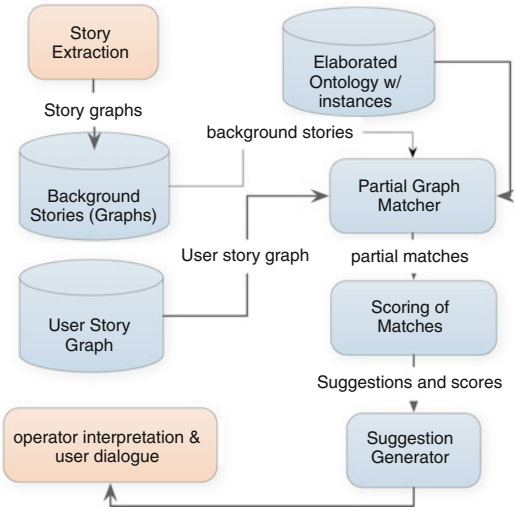


Fig. 2 Conceptual view of analogical story matching

newspaper), sensors, non-text sources (e.g., maps) and from users directly, detailed functional representations of these systems are elided in this view (orange boxes).

The primary focus in this paper is to describe how we can use past stories to inform or enrich a developing plan (where a partial plan is an "in progress" user story). We term this process story-based causal inference (SBCI), which is represented in the figure with the blue boxes. Story-based causal inference allows the computer to reason about, connect, and integrate story graphs. By finding relevant knowledge in a story graph knowledge base and combining them with the story of the current situation, the computer system can help human users surface unstated assumptions and identify unexpected outcomes, improving their foresight and mission effectiveness.

SCBI employs a common story graph representation, whether individual stories derive from users or background stories. They can be elaborated with semantic knowledge from an ontology. User stories are matched to other stories via a partial matching process. The SCBI prototype uses partial matching for two complementary reasoning methods, analogical matching and causal inference, which we describe further below. Resulting partial matches are then scored and the highest scoring matches are translated from story graphs to a more human-understandable text representation for presentation to users.

4.1 Analogical Matching

Analogy [10, 11] is apt for finding possibly relevant information and incorporating it into a story graph to elaborate or extend user stories. Analogy is also familiar to military decision makers who are expected to create mission analyses that draw on

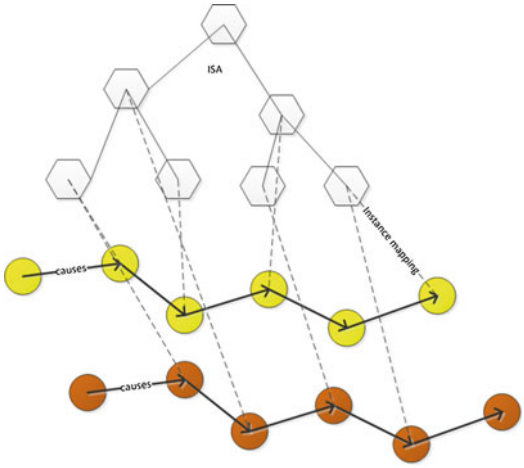


Fig. 3 Conceptual view of analogical story matching

knowledge from past events with similar characteristics or in the same geographical region. We drew on cognitively inspired planning algorithms that carry out goal-based graph search. They incorporate various heuristics designed to overcome worst-case computational complexity of graph matching by leveraging domain knowledge and machine learning [12]. We drew on elements of this prior work to create a prototype that could draw analogical connections between different but similar story graphs.

The computational problem is represented conceptually in Fig. 3. A user story (orange) and a prior story (yellow) have similar structure. Representational elements within a node are mapped to a common ontology (hexagons) as instances. For example, in the prior story illustrated in Fig. 4, the first node would be mapped to the political entity *Haiti* in the ontology and the natural disaster *Earthquake*. These mappings to the ontology then enable additional connections via ontological relationships. The ontological representation of Haiti might link further to its political and military organization, its economic structures, and so on.

Matching of nodes in stories can be exact or partial. Exact matches correspond to finding a previous occurrence of the same conditions in a past story. For example, a common node appearing in many disaster relief stories might be *lack of potable water*. If the same node appears in the user story, the system could immediately match it to many past stories. In Fig. 3, the second node in both the user story and

Fig. 4 An example historical story used in analogical mapping

the past story are mapped to the same node in the ontology, representing an exact match.

Most graph matches are not exact. Partial graph matching uses the ontology to help the analogical matching process identify possibly relevant or interesting matches. With these augmentations, the prototype can carry out different types of analogy:

1. **Generalization** makes use of the ontology links to attempt matches between nodes that talk about different objects, but the objects share a common ancestor in the ontology (consider the third nodes in the user and past story). For example, a generalization of the current situation, which contains a story node medical service degraded, might infer that the node should match to a past story with the node *police service degraded* because both objects are subclasses of services.
2. **Transformational analogy** [13] refers to possible analogies between relations that the system can hypothesize might exist and uses graph structure rather than a match on node contents. For example, if a node contains the relation *is-present-in(disease, region)* then the system might recognize a structural alignment with another node that contains *occurred-in(disaster, region)*. Structural similarity can be enhanced if additional nodes and edges match across the two graphs.

Completing the analogy requires inference from a matched past story to the current situation. In the prototype, we remap objects within nodes from the current situation story to the matched graph. For example, if the current situation contains *disease is present in Affected Region* and the analogy is to *disease in a region degrades medical services*, then the remapped inference produces *this medical service in Affected Region will be degraded*. This subgraph appears as a new node dependent on the original node with a linking edge. In the prototype, all inferences led to adding nodes in a story. The new node also becomes available for iterative application of the matching algorithm to find more suggestions that follow from an inference.

4.2 Causal Inference

In some cases, users of a system such we are developing may want to simply assert that some causal relationship exists, rather than needing to derive or discover one thru analogy. The example in Fig. 5 presents an inference rule that says that disease conditions are more likely to increase in effect in a region when the medical service in that area has been degraded. The nodes preceded by "=" serve as variables, so that the rule is applied to a particular medical service, region, and disease. Such a rule could be used, for example, to anticipate specific disease outbreaks following degradation of medical services in an area after a typhoon.

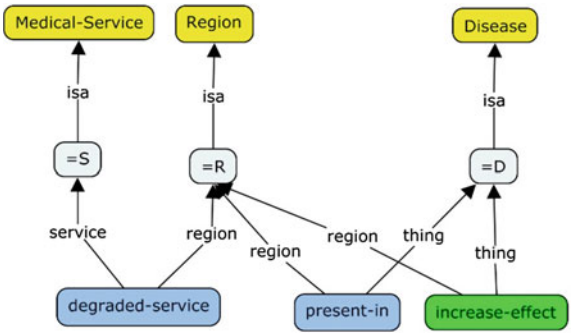


Fig. 5 An example graph-based inference rule

From an information content perspective, these rules capture and express causal properties that may not be evident in individual stories. They can fill in gaps and help the system be more efficient in finding relationships and connections. Computationally, they are expressed in the prototype in the same story graph representation that is used to capture user stories and past histories. The ontological mapping process outlined earlier allows the variables defined in the rule to be bound to specific elements from user stories.

This design choice has one immediate benefit and one potential long-term benefit. The immediate benefit is that the prototype can use the analogical mapping process presented earlier to apply these causal rules. The same underlying computational process is used for both analogy and causal inference. The long-term benefit of this uniformity in representation is that it can support the gradual transformation of analogical connections between stories to causal rules, enabling a kind of causal learning. For example, after the observation of many typhoons, hurricanes, earthquakes and other natural disasters and the outbreak of different diseases following these disasters, the type of system we are developing could likely learn the kind of rule represented in Fig. 5.

4.3 Match Scoring and User Feedback

Many possible matches could be generated by the partial mapping outlined above. The prototype includes a scoring function that attempts to estimate the logical sufficiency and logical necessity of individual graph components with the goal of assessing the "aptness" of the potential inference. For example, if a subgraph of a user story is also present in a past story, then the prototype estimates the logical sufficiency (LS) of the nodes in the augmented subgraph. If a subgraph of the user story is missing from a candidate past story, then the system estimates the logical necessity (LN) of the missing nodes. In other words, each matching node improves the match score and each missing node decreases the match score. The scoring

function was tuned manually for the software test event. Long-term, it would be preferable for the system to learn the relative weights of necessity and sufficiency matches.

Scoring of candidate matches are currently used to rank the candidates for presentation to the end user. The same scores could be used to filter and prioritize the alerts that result from story additions. However, the current scoring only takes into account the "goodness" of the match. We expect that users will also be interested in seeing story elaborations that are relevant or important for their plan. This expectation was confirmed in the system test event: match scores by themselves did not contain all the information needed to determine which story additions should result in alerts. There is also a need to estimate the likelihood and the impact of each story addition in order to highlight alerts that make a difference and to justify the suggestion to an end user. Story suggestions should be relevant to the users plan. To-date, we have focused on two types of suggestions: surfacing an unstated assumption and identifying an unexpected possible outcome of the situation.

1. **Surfacing an unstated assumption** is represented by the insertion of a story node that has a directed edge pointing into an existing node. For example, a user might create a story graph about disaster relief that contains the two nodes *lack of potable water* → *water delivery*. By analogy to previous stories, the system might have access to multiple nodes that in the past have been linked to *deliver water* (e.g., *truck convoy*, *set up desalination*). The system could then suggest any of these as possible assumptions that might be underlying the user's story. If more information is available from the rest of the user story or from any mission parameter inputs, the system could use the information to narrow the choices. For example, a desalination plant resource might suggest that the unstated assumption in water delivery was that the desalination plant is available.
2. **Identifying unexpected outcomes** is represented by the suggestion of a story node that only has edges pointing into it from the established story. The new node extends the story graph as a whole. Imagine that a past story contains knowledge about the available desalination plant—its capacity is only designed to support the personnel embarked on its own ship, so supporting more people in the stricken population will lead to reduced water available to the rescuers. (Such knowledge could come from a narrative when the desalination plant was overburdened, or more simply from general expert knowledge about the tool.) If the system suggests a new node with the content *not enough capacity*, the new node is a result of the existing node *set up desalination* and is thus an unexpected outcome. The combination of the two nodes together, a surfaced assumption and an unexpected outcome that invalidates the assumption, together show how the prototype could raise a salient alert to the end user.

5 Prototype Test and Results

We conducted a small-scale software test of story matching and inference. Five testers (subject matter experts with experience in military planning) interacted with the system individually over the course of two hours. In the first hour, they created stories that reflected their analyses of a disaster relief scenario.

Operators from the research team translated these stories directly into story graphs because our goal was to test the functionality of SBCI, rather than the adequacy of user interfaces for capturing user stories. The stories the testers created were then used to test SBCI in the second hour. We used a background corpus of natural disaster events (e.g., Hurricane Katrina) and location-specific data (similar to *CIA World Factbook*), encoded into story graphs (2,575 total story nodes connected to an ontology of about 9,000 classes). We then applied SBCI to the planner's stories and this background knowledge to attempt to generate examples of unsurfaced assumptions and unexpected outcomes in the testers original stories.

Results of the system test showed that the testers were able to create story graphs describing their analysis. The captured graphs contained on average 105 nodes (67–133). The system then offered an average of 16 story suggestions (11–36) for each tester. The testers accepted on average 6 suggestions (2–11). Table 1 summarizes a few examples of suggestions generated by the prototype that were accepted by the testers and added to their stories.

Even when they did not accept a change, the testers sometimes did make other changes that were related, so the suggestion may have served as a reminder. We also observed that testers applied some interpretation to accepted suggestions. For example, the elaboration in Table 1 about the potential for violence resulted in an

Table 1 Examples of story elaborations produced by SBCI in the software test event

Type	Original plan	Elaboration
Unstated assumption	Restore unlading capability and bring in supplies and building materials thru the seaport	Storm effects, in addition to lading deficits, may limit navigational aiding and docking capacity
Unstated assumption	NGO can facilitate delivery of medical services in the absence/degradation of local medical services	NGO offers technical expertise and capacity in other areas, including water supply and communication with population
Unstated Assumption	Establish communications with provincial government	Local government includes an autonomous region with a distinct governance structure
Unexpected outcome	Establish food distribution points in rural areas	Local government may actively hinder food distribution to areas that are "troublesome"
Unexpected outcome	Establish food distribution points in urban areas	Fighting and violence sometimes occur at food distribution points in urban areas following disasters

elaboration of the planner's story to assess security needs rather than to expect violence.

One of the limitations of the test was the relatively small amount of time allowed for a large planning problem. As a result, many suggestions may likely have been evident to planners given more time. Although they were free to draw on their own knowledge, the planners did have available (during both hours of the event) all the source material that was encoded as background stories; that is, all the data that the system used to make its suggestions was available to the testers. This represents the practical constraint that sometimes available knowledge cannot be retrieved or applied in time-constrained contexts, in both human and computational systems [14].

The test event also highlighted several complementary use cases for story-based causal inference. For example, in a JTF mission analysis and planning scenario such as we tested, it might be necessary for users from multiple echelons (ranks, roles, and expertise) to contribute to a single story. This use case highlights the need for a hierarchical story graph, which would enable different users to review and fill in the story with different levels of detail. As a second use case, which we did not focus on in our prototype but test users expressed a desire for, it should be possible for the story representation to identify a "hinge" (crucial decision point) or pattern in mission data. For example the story should be able to identify that one possible branch is likely to lead to the same completion under a number of initial conditions, while another branch has good possible outcomes but also several negative outcomes are possible.

6 Conclusions

Our long-term goal is to explore the impact of using stories to bridge the gap between human and computer understanding. We hypothesize mutually shared representations will improve human-computer teaming and mitigate the common and sometimes disastrous mistakes that arise from misunderstanding or misapplication of information by computers or humans by the other.

The story-based causal inference prototype offers positive but preliminary evidence that a shared, story-based representation can be used constructively and collaboratively by humans and computers. In the context of the military planning example, it allowed unstated assumptions to be surfaced and unexpected outcomes or implications to be identified during planning. It drew on data about local PMESII systems and prior stories of natural disasters to help deepen and extend testers' available knowledge of their plans, such as the need to consider two parallel governmental authorities in one region, or the socio-economic factors influencing food distribution in different areas. These suggestions would not transform a typical

military planner into a PMESII expert, but they would aid planners unfamiliar with the intricacies of PMESII systems to better factor PMESII considerations into plans.

Although promising, additional work is needed to make SBCI practical for the military planning use cases described in the paper. We see three core areas of future research and development:

1. **Scalability**: SBCI employs heuristics to limit the impact of the costs of general graph matching, but its actual scalability to real-world problems is not yet established. We expect gains in scalability will come from a richer vocabulary of causal linkages. Different kinds of links and indices to them will help the partial matching process target subgraphs within the large corpus of background stories more efficiently. However, this direction is in tension with the relative simplicity and understandability of the linkages to potential users. Thus, the goal is to find a representation that remains readily understandable, while providing sufficient discriminating power for matching.
2. **Semantic Integration**: The prototype used a standard ontology and some general mapping rules to connect elements within story nodes to the ontology. For the future, a more sophisticated mapping is needed, along with both larger and more domain-specific ontologies. Better mappings will enable improved partial matching and more targeted match scoring. For example, analogical match could be made more precise if the matching process could traverse the ontology along with the story in deciding a potential match. In the prototype, we used only ISA mappings. For match scoring, a more domain-specific ontology would facilitate scoring based on relevance in addition to semantic and structural similarity.
3. **Learning and Generalization**: The current prototype does not improve its performance with experience. Several different kinds of learning would improve overall performance. Most importantly, as discussed above, a process that allows the system to learn new causal inference rules would both help improve capability and scalability because causal rules rely more on exact match than partial match.

Acknowledgments This work is supported by the Defense Advanced Research Projects Agency project HR0011-15-C-0067. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Department of Defense or the Defense Advanced Research Projects Agency. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. We would like to thank collaborators and sponsors at DARPA who provided insights and operational perspectives in the development of story-based causal inference and Mike Lerario and Kevin McEnery for their insights and perspectives on the challenges of military planning.

References

1. Kahneman, D., Tversky., A.: The simulation heuristic. In: Kahneman, D., Slovic, P., Tversky, A. (eds.) Judgment under Uncertainty: Heuristics and Biases, pp. 201–208. Cambridge University Press, Cambridge UK (1982)
2. Fisher, W.R.: Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action. University of South Carolina Press, Columbia, SC (1989)
3. Tetlock, P.E., Gardner, D.: Superforecasting: The Art and Science of Prediction. Crown Publishers, New York (2015)
4. Zacharias, G.L., MacMillan, J., Susan B. Van Hemel (eds.): Behavioral Modeling and Simulation: From Individuals to Societies. National Academies Press, Washington, DC (2008)
5. Tulving, E.: Elements of Episodic Memory. Oxford University Press, Oxford (1985)
6. Schank, R., Abelson, R.: Scripts, Plans, Goals, and Understanding. Erlbaum, Hillsdale, NJ (1977)
7. Richards, W., Finlayson, M.A., Winston, P.H.: Advancing Computational Models of Narrative. MIT CSAIL (2009)
8. Shapiro, B.P., van den Broek, P., Fletcher, C.R.: Using Story Based Causal Diagrams to Analyze Disagreements About Complex Events. Discourse Processes (1995)
9. Verheij, B.: Virtual Arguments: On the Design of Argument Assistants for Lawyers and Other Arguers. TMC Asser Press, The Hague, The Netherlands (2005)
10. Hofstadter, D.R.: Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. Basic books (2008)
11. Gentner, D., Holyoak, K., Kokinov, B.: Analogy: Perspectives from Cognitive Science. MIT Press, Cambridge, MA (2000)
12. Jones, R.M.: Problem solving via analogical retrieval and analogical search control. In: Foundations of Knowledge Acquisition, pp. 227–262. Springer (1993)
13. Kuchibatla, V., Muñoz-Avila, H.: An analysis on transformational analogy: general framework and complexity. In: Advances in Case-Based Reasoning, pp. 458–473. Springer (2006)
14. Newell, A.: The knowledge level. Artif. Intell. **18**, 82–127 (1982)

Modeling Causal Relationships in Sociocultural Systems Using Ensemble Methods

Amy Sliva, Scott Neal Reilly, David Blumstein, Steve Hookway and John Chamberlain

Abstract Analyzing the political, military, economic, social, information, and infrastructure (PMESII) effects in a sociocultural system requires models that capture the causal and predictive dynamics. However, given the complexity of PMESII factors and the diversity of available data sources, accurately modeling causal relationships requires incorporating multiple domains of study and a variety of analytic methods. In this paper, we present an ensemble approach to modeling causal relationships of sociocultural systems, applying insights from machine learning where ensembles consistently outperform individual approaches. We describe three different types of ensemble models and combinations and explore the application of this approach in experiments using both synthetic and real-world datasets.

Keywords Dynamic PMESII modeling · Sociocultural systems · Prediction · Causality · Ensemble learning · Machine learning

1 Introduction

When planning operations or designing policy interventions, military decision-makers and policy experts must have an understanding of the world around them, particularly the political, military, economic, social, information, and infrastructure

A. Sliva (✉) · S.N. Reilly · D. Blumstein · S. Hookway · J. Chamberlain
Charles River Analytics, 625 Mount Auburn Street, Cambridge, MA 02138, USA
e-mail: asliva@cra.com

S.N. Reilly
e-mail: snealreilly@cra.com

D. Blumstein
e-mail: dblumstein@cra.com

S. Hookway
e-mail: shookway@cra.com

J. Chamberlain
e-mail: jchamberlain@cra.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_4

(PMESII) implications on and effects of these policies. Analyzing these effects requires models that accurately capture the causal and predictive dynamics of a sociocultural system, enabling policy makers to make informed decisions as to the likely results of proposed interventions. Unfortunately, it is notoriously difficult to analyze causality rigorously, more so in complex sociocultural systems. Controlled experiments are rarely possible and capturing rich and complex causal and predictive relationships from available data is extremely challenging.

Making practical use of causal concepts challenges researchers in a wide range of fields, including philosophy, economics, statistics, artificial intelligence, cognitive science, physics, and social sciences. As a result, researchers have developed an equally wide range of approaches to recognize, represent, and reason with causal relationships to understand the past and predict the future. For example, Granger causality [1, 2] and convergent cross mapping [3] can detect possible causality from time series data, probabilistic graphical models provide tools for describing and reasoning about causality and uncertainty [4, 5], and recent advancements in artificial intelligence provide approaches for determining the directionality of causal linkages from observational data [6, 7]. Consequently, the diversity of factors and data sources that comprise sociocultural systems strongly suggests that modeling causal relationships of PMESII factors cannot be successful without incorporating aspects from multiple domains of study and insights from a variety of analytic methods.

In this paper, we present an ensemble modeling approach to understanding causality for PMESII modeling, treating the variety of causal analysis methods and data sources as an asset rather than a challenge. To mitigate the impacts of PMESII modeling complexity, we leverage computational data mining and data analysis techniques to accelerate the analysis and modeling process, recommend causal hypotheses that might be unanticipated by human experts, and help humans validate their theories against available data. Our approach applies insights from machine learning, where similar types of ensembles have been consistently able to outperform individual learning approaches, to (1) employ different analytics as appropriate to disparate data sources; (2) apply distinct analytic approaches to address discrete aspects of the sociocultural system; and (3) combine the results into a single coherent analysis.

The remainder of this paper is organized as follows. In the next section, we discuss different types of automated and semi-automated methods for identifying and modeling causal relationships from various types of sociocultural data. Section 3 presents our approach to ensemble causal models. In Sect. 4, we illustrate our approach by applying ensemble causal models to both synthetic and real-world data sets. First, we developed an ensemble combination of additive noise models to determine the causality from observational, non-temporal data. Second, we constructed an ensemble of diverse causal analysis techniques for time-series data that exploited metadata to improve performance. In both of these cases, the ensemble outperformed the individual approaches and produced a more accurate model of causal relationships. Third, we used a human-in-the-loop process to develop an

ensemble model for extracting causal relationships from real-world data about Anbar province in Iraq to form a more complete picture of the causal dynamics and PMESII effects in the region. Finally, in Sect. 5 we discuss the implications of this research and directions for future work.

2 Modeling Causality in Sociocultural Systems

Causality is extremely difficult to analyze, in part because there is no agreed upon definition of what makes a relationship "causal" or "predictive." Goldthorpe [8] identifies three possible definitions of causality that have been prominent in social science research: (1) causality as robust dependence, where causation implies association or correlation with strong predictive power; (2) causation as consequential manipulation where different consequences are observed when the causal variable is manipulated or varied; and (3) causality as a generative process, where the relationship is determined by some underlying social process or mechanism, which itself may be unobservable. Each of these definitions can provide insight into causal relationships in sociocultural models of PMESII factors, and their usefulness depends on the current decision-making context and the available information. For example, if a Commander must make an immediate decision for a transportation route, a causal model of the first type that predicts IED placement from the relative locations of roads, infrastructure, known weapons caches, and previous attacks, may be sufficient for route planning. However, for larger operations, such as reconstruction or humanitarian aid, a more detailed understanding of the causal relationship between local ethnic demographics and how this influences which water sources a household uses and how road closures or construction may impact their ability to acquire water, may be necessary. In this section, we review several automated methods for analyzing causal/predictive relationships and extracting them from data (or validating hypotheses against observational data). These techniques are drawn from a variety of social, natural, and computational domains and include a range of interpretations on the nature and definition of causality.

Many models adhere to the first definition of causality and describe relationships that are highly correlated with a strong predictive component. For example, because causality has an inherently temporal component—the effect does not precede the cause—techniques that exploit this can help analyze and extract causal models from time-series data. The most basic model of predictive power is simple correlation, such as the classic Pearson correlation measure; by adding a temporal offset to the correlation to mimic the sequencing of cause and effect, Pearson can successfully capture many causal/predictive relationships in time-series data. Granger causality [1, 2], a technique familiar to many social science researchers, takes strong predictive correlation further. Granger causality was originally introduced for time series analysis in economics and assumes a temporal order of cause and effect and that the causal variables provides information about the effect that can otherwise be unavailable.

Dynamic time warping is a similar approach that was developed by the authors for reasoning over arbitrary time series data [9, 10]. It was originally created for gait recognition where temporal offsets may be inconsistent over time (similar to how the distance between steps may change depending on whether a person is walking or running, an increase in crime might happen anywhere from six to 12 months after an uptick in unemployment). Convergent cross mapping (CCM) [3], a recent advance in biological studies, can identify even deeper, feedback relationships among time series, enabling scientists to model cyclic causal relationships, such as the feedback relationship between poverty and conflict known as the conflict trap [9, 11, 12]. These approaches can be very useful for identifying causal/predictive relationships in time-series datasets, for example validating models relating drops in unemployment with successive increases in criminal activity. However, we have found that this temporal data only represents a small proportion of the available sociocultural data sets.

Capturing Goldthorpe's third definition, relatively recent advances in the uncertainty in artificial intelligence (UAI) community seek to model the structure of some underlying process that is responsible for generating observed outcomes, capturing the direct and indirect causal relationships that exist between PMESII factors. For example, rather than simply indicating there is a statistically strong relationship between unemployment and crime data, these approaches will explicitly model the causal chain from lack of jobs in the legitimate economy to unemployment to economic hardship to criminal behavior, capturing the entire process that links the observed rise in unemployment with the subsequent rise in crime. Unlike the causal/predictive methods described above, the observational data used to construct and validate these models can be non-temporal, such as survey data (e.g., household development indicators surveys) or cross-sectional records (e.g., election outcomes for different regions) often encountered when trying to understand various PMESII characteristics.

Graphical probabilistic models, such as Bayesian networks [5], have provided the foundation for work in causal analysis since Pearl's *Causality: Models, Reasoning, and Inference* [4]. These models have proven useful as formal tools for describing and reasoning about causality for many fields of inquiry (e.g., epidemiology, policy, and social sciences). Graphical models mimic naturalistic ways of thinking about causality by explicitly representing the *structure* of causal relationships. Figure 1 shows an example of a graphical model illustrating the causal links between economic factors and increases in criminal activity. With their basis

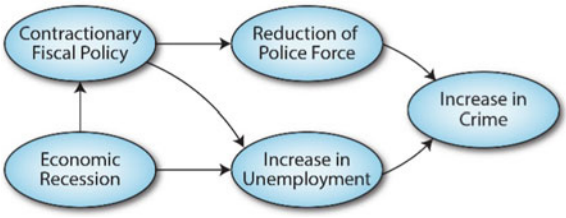


Fig. 1 Graphical causal model illustrating the causal chain from economic recession to increases in crime

in formal probability theory, these models also facilitate rigorous mathematical reasoning and inference; however, the intuitive graphical structure means that it is often straightforward to reason visually over these graphs and make complex assessments about causal interactions in real-world situations quite easily and naturally, often without needing to resort to algebraic analysis. The graphical model structure can either be provided by human experts, or through automated structure learning techniques, such as heuristic search or constraint satisfaction (e.g., inductive causation (IC) [13] or the PC algorithm [14]). However, these automated structure learning methods depend on large quantities of data from all variables in the system, which in many cases may be unavailable or researchers have not identified all of the necessary variables present in the system. Mooij et al. [6] have developed a technique for determining the directionality of a causal relationship between only two variables that does not rely on any additional structural knowledge, enabling structured causal analysis from more limited observational data. For example, if we know that there is some causal relationship between crime and unemployment, we can determine the directionality of this link simply from observational data about these two variables, rather than requiring knowledge of other factors in the graph. This approach assumes that there is noise in any causal relationship and that by evaluating the effects of some standard noise models on causal relationships, it is possible to tease apart cause and effect in some situations. This has proven effective in many practical cases (see Sect. 4).

The wide range of methods described above—by no means an exhaustive list—make clear that is no single causality analysis technique that addresses all of the challenges and complexities of modeling PMESII factors in sociocultural systems. Rather, social scientists have at their disposal a suite of causal/predictive modeling approaches when dealing with observational data, each with strengths and weaknesses. For example, time-series analysis techniques can create useful models, but may draw superficial or over-generalized conclusions; probabilistic graphical models provide intuitive views of causal relationships, but require a priori identification of the important variables; and additive noise models can recognize causal structure between two variables, but cannot account for confounding relationships. Having these analytic techniques available allows scientists to develop more robust models of causal relationships in complex systems. However, we believe the real strength of this analysis suite lies not in finding the perfect technique for a particular system, but in combining these approaches in unique ways to produce a more comprehensive model [9, 10].

3 Ensemble Methods for Causal Analysis

Ensemble reasoning has proven extremely useful in the computational modeling and the machine learning communities to provide capabilities beyond those provide by any individual technique. For instance, the DARPA ICEWS program was successful at predicting nation-state instability by combining a variety of

independent predictive models and combing the results into a single prediction that was more accurate than any individual model provided [15]. Also, when using machine learning to, for instance, learn a classifier, it is possible to create ensembles that combine various learning techniques (e.g., support vector machines, neural networks, naïve Bayes classifiers, combinations of decision trees)—each with its own strengths and weaknesses—to provide better results than any individual approach [16]. We extend this concept to the domain of causal reasoning to create multiple causality-analysis techniques, each with its own strengths and weaknesses, that more thoroughly and accurately model causality in real-world systems.

3.1 Constructing Ensembles

There are many possible structures for combining multiple causal reasoning methods into ensembles and for joining multiple ensembles into larger ensembles. We call the building blocks of our approach data ensembles, chain ensembles, technique ensembles, and nested ensembles. Figure 2 shows these graphically.

Data ensembles. There are qualitatively different types of causal relationships to reason about (e.g., causality when there are hidden variables, non-deterministic causality) and qualitatively different types of available data to learn from (e.g., continuous, categorical, noisy). Just as different machine learning techniques can learn different kinds of things from different kinds of data, we need different causal analysis techniques to extract causal models based on the nature of the causal relationship present and the types of available data. For instance, Granger causality works on temporal data; PC works on non-temporal data.

Chain ensembles. Most causal analysis techniques are very specific about what kind of knowledge they extract, so they need to be composed with algorithms that perform other types of reasoning to provide the necessary analysis. For instance, correlation analysis (with appropriate human supervision) can be used to find pairs of likely causal variables, but not the direction of the causation. Then additive-noise methods (e.g., those described in [6, 7]) can identify the direction of the causation given two causal parameters, but those parameters are not identified. By combining the two, we provide the ability to find pairs of cause-effect parameters in data.

Technique ensembles. There are different ways to extract similar kinds of results from similar kinds of data, combining the results in ways that produce better results than any individual method can provide. For instance, Granger causality, dynamic time warping, Pearson correlation, and offset-correlation are all mechanisms to find possible causal links in temporal data [9, 10], though all are known to have limitations. By combining the results using mechanisms like those used in ensemble machine learning (e.g., various weighted voting schemes), we can provide an ensemble approach that is more accurate than any of the individual methods.

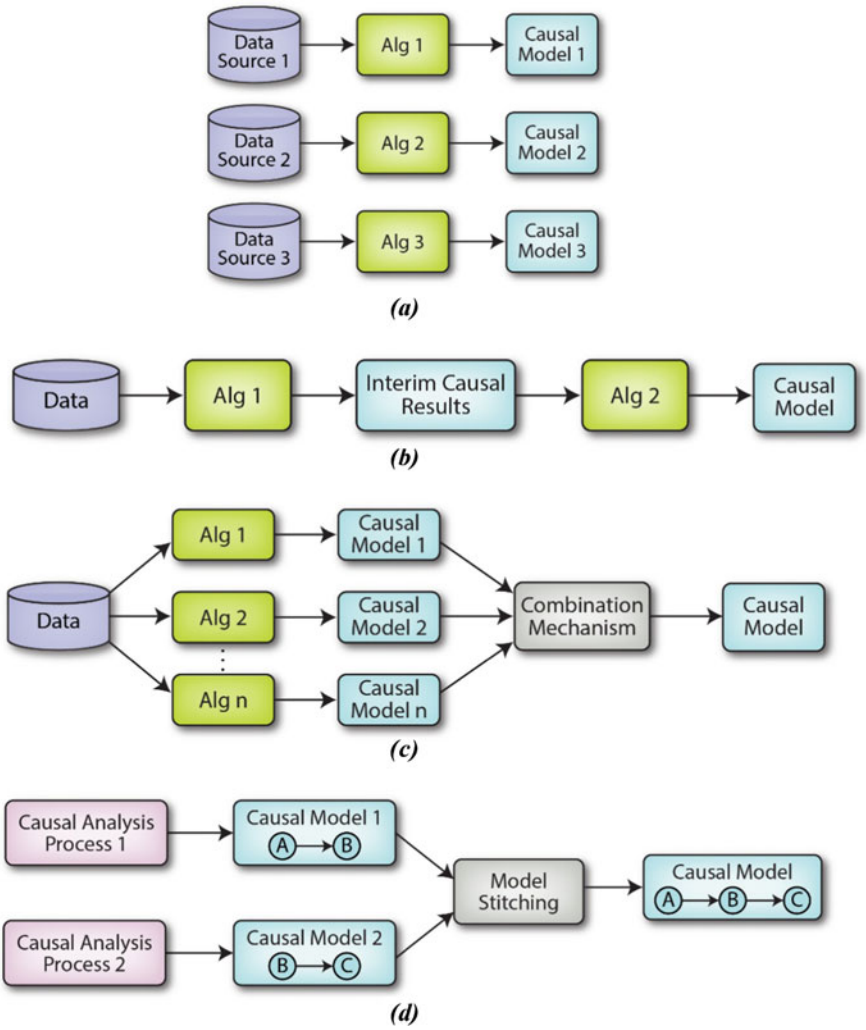


Fig. 2 The building blocks of ensemble causal reasoning: **a** data ensembles; **b** chain ensembles; **c** technique ensembles; **d** nested ensembles

Nested ensembles. Ensemble causal analysis can enable us to combine causal models that derive from different data sources. This approach abstracts away from the raw data, which may come in a variety of forms (e.g., qualitative observations, surveys, time series) that are very challenging to analyze jointly, to identify linkages that can combine causal models from diverse data sources. For example, a time series causal analysis method, such as Granger causality, may indicate that increasing un-employment leads to increased crime. A human survey might indicate that a higher level of crime results in increased preference for more police spending.

Using an ensemble approach, we can combine these two causal models from datasets that cannot be linked using standard quantitative or qualitative methods to achieve a more complete picture of the causal process that leads from unemployment to police spending.

3.2 Ensemble Combination and Model Stitching

Automating model combination and stitching to create ensembles is a core problem area. In particular, in real-world application where the size of the model space may grow large and reconfiguration is required to apply existing tools to a new problem set or domain, purely manual approaches are likely to become infeasible. In the experiments described in this paper, we extended the approach described in [11] to create graphical causal models out of smaller models discovered through a variety of analysis techniques. When these probabilistic graphical models (e.g., Bayesian Networks) have identical nodes, the models can be stitched through these nodes reliably into a new, larger network. Similarly, approaches like structural equation models [17] allow models to be stitched as long as they are consistent (i.e., non-contradictory).

More challenging cases arise when stitching nodes that are not identical, such as when estimating different properties of the same phenomenon, which is a quite common misalignment when analyses span multiple domains, data sets, and techniques. For example, one model might characterize the expected crime count where another estimates the likelihood of an increase in crime rate. Human expertise is one mechanism to resolve this kind of mismatch, but achieving practical performance is likely to require automated semantic reasoning, such as through multi-formalism methods [18].

One of the primary concerns with using purely automated causal-analysis techniques is that they often overfit available data and find spurious relationships. We have found that, instead of having an automated tool that finds the "right" answer and informs the user of that answer, it is more effective to use techniques, such as iterative modeling, where the user proposes some concepts and the automation attempts to validate (or, more accurately, invalidate) the hypothesis and explain problems with the proposal given the available data. It can also suggest a number of possible alternatives for the human modeler to review. Once approved, the human or the automation can suggest additional extensions to the causal model (e.g., now that we believe A causes B, can we determine what might cause A?). This sort of human-machine (or top-down/bottom-up, or mixed-methods) approach is described in more detail in [9].

In particular, user insights and hypotheses can be incorporated as an explicit component of the causality analysis ensembles, combining interactive human input with various automated methods according to various voting or heuristic

combination method. For example, one ensemble might first use methods to look for direct causal relationships between observational variables by validating a model provided by the user. Then, automated methods will look for latent variables to supplement the model and determine if there may be some confounding variables the user missed. The system can then solicit input from the user to provide insight into what this latent variable might be and to identify supporting data sources that might capture this variable.

4 Experimentation and Evaluation

To evaluate the ensemble approach described in the previous section, we performed several experiments using a variety of ensemble types and data sources, including both synthetic and real-world data. In this section, we describe the results of these experiments and provide a case study applying our ensemble modeling approach to understanding violence in Iraq's Anbar Province in 2006.

4.1 Ensemble Model Using Time Series Data

The first experiment was designed to evaluate a technique ensemble over time-series data. We developed a synthetic causal data generator to create consistent ground truth temporal data consisting of two time series, A and B, of which one was designated the cause and the other the effect. The generator enables us to vary the mathematical function that relates the cause and effect (i.e., linear, sinusoidal, quadratic) as well as the temporal offset between the cause and effect. We generated datasets for each type of function at each offset ranging from 0 (i.e., the cause and effect occur at exactly the same time) to 10 (i.e., the cause precedes the effect by 10 time steps). We used implementations of Granger causality and offset Pearson correlation to determine the strength of the causal relationship between each A and B pair and identify which is the cause. In this baseline evaluation, Granger causality had a 46 % accuracy and Pearson had a 73 % accuracy in identifying the causal factor.

Using machine learning, we created an ensemble classifier that combines these two approaches. We found that the ensemble learning did not provide much improvement over the offset Pearson correlation until additional metadata was also included as features, such as the types of functions defining the causal relationship. Using this, the ensemble approach was able to achieve over 90 % accuracy in determining the correct causal model for each pair of time series. Interestingly, we found that adding the temporal offset as a feature in the ensemble learning did not seem to improve the scores.

4.2 Ensemble Model Using Non-temporal Data

In our second experiment, we created another same-type ensemble, combining several parallel additive noise models to determine the directionality of causal relationships in non-temporal observational data. For this experiment, we used the CauseEffectPairs (CEP) benchmark data set used in previous studies [6, 7] of additive noise models of causality. This data consists of 99 different "cause-effect pairs" taken from data sets from a variety of different domains (e.g., meteorology, biology, medicine, engineering, economy, etc.). For example the data contains pairs of altitude and temperature measurements from weather stations across Germany (here, altitude is clearly the cause, and temperature the effect), and pairs of CO_2 emissions and energy use compiled by the United Nations for 152 countries between 1960 and 2005 (energy use is the cause of CO_2 emissions). Each cause-effect pair consists of samples of a pair of statistically dependent random variables, where one variable is known to cause the other one, providing us with ground truth for evaluating our ensemble approach.

The ensemble approach proved very effective when dealing with this non-temporal data. We implemented a variety of state-of-the-art additive noise models from [6]. Using machine learning, we created an ensemble of additive noise models that was much more effective at identifying causal relationships than any single model. Individually, the best additive noise models were only capable of 60 % accuracy when determining the direction of the cause-effect pairs in the dataset, while most models had an approximate 50 % accuracy rate. Using our learned ensemble, this improved dramatically to approximately 90 % accuracy. This result indicates that an ensemble can identify important features of a dataset, such as the amount and type of noise, and determine which combination of additive noise models will be most successful. The wide range of observational data used in this experiment indicates the general applicability of this approach to a variety of domains, including PMESII modeling.

4.3 Case Study: Applying Ensemble Causal Models to Conflict in Iraq

To demonstrate how ensemble models can help identify causal relationships to characterize PMESII features in sociocultural systems, we applied this approach to modeling violence in Iraq. Currently, Iraq is plagued by continuing and increasing violence from the terror organization that calls itself the Islamic State of Iraq and the Levant (ISIL), and militant activity is growing throughout Iraq, Syria, and parts of Lebanon. World leaders have been debating how best to address this increasing threat and stabilize the region. Many experts have turned to the recent past to better understand and model the present PMESII effects, looking at parallels to Iraq's Anbar Province in 2006 when the region was under increasing control by Al Qaeda

militants. Using a variety of causal analysis approaches and diverse data, we developed a rich nested ensemble that employs data, chain, and technique ensembles to capture the dynamics of the situation, illustrating how decision-makers might use these approaches to understand the possible impacts of different policy options.

For our analysis, we used Empirical Studies of Conflict (ESOC) data from Princeton University [9]. ESOC contains a variety of different types and sources of data regarding Iraq that we used for this evaluation: (1) time series of violent events in Iraq yearly, monthly, and weekly at the province and district levels from February 2004–February 2008 as compiled from the significant activity (SIGACT) reports by Coalition Forces; (2) time-series reconstruction data on all projects executed in the country from 2003 to 2008, including the type of project and the amount and source of funding; and (3) non-temporal demographic data based on a national census in 2008. Each of these data sources can provide insight into a different aspect of the situation in Anbar province in 2006, capturing the militant activity, sociocultural, and sectarian issues. In addition, the choice of data for this experiment was itself motivated by a causal theory presented by human experts, indicating that there is some relationship between socioeconomic conditions and the propensity for violence and sectarian conflict in the region. To make use of all this data, including expert social science theories, we employed a variety of the analysis methods discussed in Sect. 2. Using a chain ensemble, we combined human expertise in identifying potential causes and effects with both additive-noise and time-series analyses to determine the actual direction and nature of the causation. Because we have different types of data sources, we used a technique ensemble approach to combine additive noise results and time-series analysis into a single coherent causal model. For intuitive analysis by policy-makers, the resulting model was represented as a graphical model with a structure similar to Pearl's [4, 5]. Given all of these factors, we constructed a complex nested ensemble, enabling us to combine human expertise with the additive noise and time-series ensembles developed for the prior two experiments in a graphical model representation.

Figure 3 shows the final model constructed using these nested ensembles over the ESOC data. Looking at the time series in the reconstruction and SIGACTS data, our temporal ensemble of Granger causality, offset Pearson correlation, and dynamic time warping identified several relationships where specific types of reconstruction projects (e.g., Military Facilities, Democracy Building, Infrastructure Improvements, etc.) influence (either increase or decrease) the number of violent acts, or vice versa. From the non-temporal demographic data, the additive noise ensemble discovered causal links from the size of the Shia, Sunni, and mixed populations to the number of SIGACTS. In addition, the causal ensemble was also able to identify specific human-created hypotheses that were not supported by the data. In the structure proposed by human experts, reconstruction efforts targeting courts and schools were considered important factors in reducing the violence; however, subsequent analysis in the chain ensemble by automated time series methods did not support this assumption and with concurrence from the human, these factors were removed from the final model.

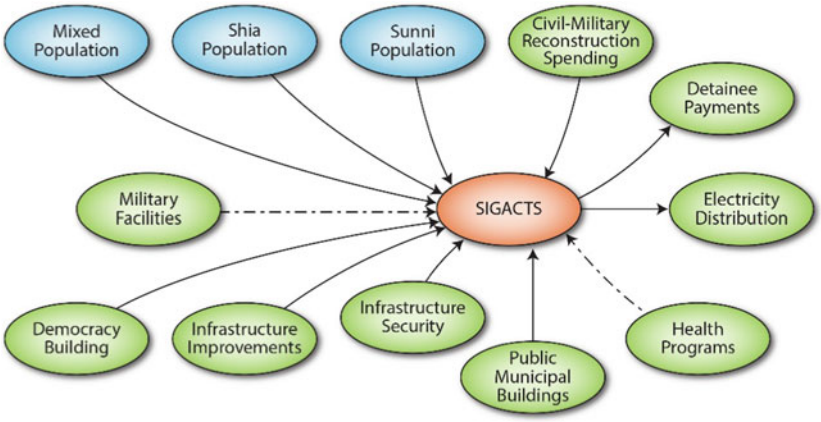


Fig. 3 Causal model of the impact of socioeconomic and demographic factors on militant activities in Anbar province, extracted from real-world data using an ensemble of ensembles

This case study demonstrates the power of ensemble combinations for analyzing causality in sociocultural systems, enabling us to combine various different types of data and models into a single analysis that provides richer and more accurate results than any single modeling approach. By combining causal analysis techniques, policy makers weighing different approaches to a situation, such as stabilizing Anbar in 2006 or dealing with ISIL in 2016, can employ the variety and richness of information available in different types of data sets, ultimately producing a model of the sociocultural system that considers factors across the full PMESII spectrum. For example, using the model in Fig. 3, a decision maker can determine the types of reconstruction projects that might be most influential for stabilizing the region, which demographic features contribute most to unrest, and identify new or unexpected causal linkages.

5 Discussion and Conclusions

The example presented in the previous section provides an illustration of the utility of our ensemble approach for PMESII analysis and modeling. While this experiment did not capture all of the intricacies of a real-world policy investigation, even this relatively small dataset with a few thousand measurements clearly illustrates the multiplicity of domains of study, potential stakeholders, and policy implications that can feature in such an analysis. In this instance, as well as the other two experiments described above, our approach demonstratively achieved the essential ambition promised by the field of ensemble machine learning: combinations of multiple, diverse techniques outperforming individual approaches. It achieved this by constructing data ensembles with a variety of data and types of causality, chain

ensembles incorporating human perspectives, technique ensembles that crowdsource the application of different algorithms, and a case study that used a nested ensemble of all the previous types.

As we continue to refine and mature these ideas, we are cognizant of a number of significant challenges that lay ahead. For example, the experiments above required a great deal of human guidance and intervention in selecting data, assembling models, configuring algorithms, and choosing the most useful and interesting features. To this end, the commoditization of machine learning in the open source community suggests a number of promising avenues of improvement. In addition, it is clear that human knowledge, context, and perspective is a necessary component to any successful PMESII analysis. In the current implementation of our approach provides near infinite control and visibility to the human, who can, for example, see and modify every element of the model. However, the basic interaction patterns described here scale very poorly as the size of the data and model space grow very large. For these problems, we gain inspiration from research in human systems, especially in managing Big Data, such as recommendation systems that facilitate access to large knowledge bases.

In conclusion, these experiments have demonstrated that our ensemble approach has the potential to take advantage of the great diversity in the conceivable inputs, dynamics, and objectives of PMESII modeling. As researchers, we look forward to the many challenges that remain to operationalize such an approach.

Acknowledgments This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government.

References

1. Granger, C.W.: Investigating causal relations by econometric models and cross-spectral methods. Economet.: J. Economet. Soc. 424–438 (1969)
2. Granger, C.W.: Testing for causality: a personal viewpoint. J. Econ. Dyn. Control **2**, 329–352 (1980)
3. Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S.: Detecting causality in complex ecosystems. Science **338**(6106), 496–500 (2012)
4. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press (2009)
5. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA (1988)
6. Mooij, J.M., Peters, J., Janzing, D., Zsceischler, J., Schölkopf, B.: Distinguishing cause from effect using observational data: methods and benchmarks. J. Mach. Learn. Res. (To Appear)
7. Peters, J., Mooij, J.M., Janzing, D., Schölkopf, B.: Causal discovery with continuous additive noise models. J. Mach. Learn. Res. **15**(1), 2009–2053 (2014)
8. Goldthorpe, J.H.: Causation, statistics, and sociology. Eur. Sociol. Rev. **17**(1), 1–20 (2001)
9. Sliva, A., Reilly, S.N., Casstevens, R., Chamberlain, J.: Tools for validating causal and predictive claims in social science models. Proceedings of the 4th international conference on cross cultural decision-making. Proc. Manuf. **3**, 3925–3932 (2015)

10. Sliva, A., Neal Reilly, S.: A big data methodology for bridging quantitative and qualitative political science research. In: American Political Science Association Annual Meeting (2014)
11. Collier, P., Elliott, V.L., Hegre, H., Hoeffler, A., Reynal-Querol, M., Sambanis, N.: Breaking the Conflict Trap: Civil War and Development Policy. World Bank and Oxford University Press, Washington, DC (2003)
12. Collier, P., Hoeffler, A.: Greed and grievances in civil wars. Oxford Econ. Papers **56** (2004)
13. Verma, T., Pearl, J.: Equivalence and synthesis of causal models. In: Proceedings of Sixth Conference on Uncertainty in Artificial Intelligence, pp. 220–227 (1990)
14. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. MIT Press (2001)
15. Ruvinsky, A., Lustick, I., Levey, B., Siedlecki, T.: Leveraging big data and social science theory for sociocultural modeling (Special presentation). In: Proceedings of the 4th International Conference on Cross-Cultural Decision Making, AHFE International (2015)
16. Opitz, D., Maclin, R.: Popular ensemble methods: an empirical study. J. Artif. Intell. Res. 169–198 (1999)
17. Hancock, Gregory R.: Fortune cookies, measurement error, and experimental design. J. Modern Appl. Stat. Methods **2**(2), Article 3 (2003)
18. Levis, A.: Multi-formalism modeling of human organizations. In: 29th European Conference on Modelling and Simulation (2015)

Part II
Cross-Cultural Comparisons in Business, Education, and Outlook

Difference Analysis of Impression on Japan from Chinese Students with Different Education Stages

Xiaohong Zhang, Jianwen Xiang, Shengwu Xiong, Haowen Li and Bixiang Li

Abstract In the same country the knowledge acquisition and understanding under the different cultural education stages have the slight differences which can reflect on the making decision especially for the first impression. In order to analysis the differences of the Chinese students' first impressions towards Japan and the corresponding knowledge acquisition ways under Chinese different cultural education stages, we firstly used the questionnaire through the nine-in-one drawing method to obtain the accurate first impressions and corresponding knowledge acquisition ways data, the different stages consist of the junior school, senior high school, university stage major in Japanese and not major in Japanese. Secondly, paired-t-test method was employed to access the differences degree of impressions and ways between the different stages. The results by analyzing the above difference suggested that, the knowledge shortage and corresponding acquisition ways during the different cultural education stages can be summarized for educational leader to make the personalized and reasonable decision.

Keywords Impression difference analysis · Different cultural education stages · Nine-in-one drawing method · Paired-t-test method

X. Zhang (✉) · J. Xiang · S. Xiong · H. Li · B. Li
Hubei Key Laboratory of Transportation Internet of Things,
School of Computer Science and Technology, Wuhan University of Technology,
No. 122, Luoshi Road, Wuhan, China
e-mail: xiao9103hong@whut.edu.cn

J. Xiang
e-mail: jwxiang@whut.edu.cn

S. Xiong
e-mail: xiongsw@whut.edu.cn

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_5

1 Introduction

In the process of human cognition to understand the content of first impressions towards other country and corresponding knowledge acquisition ways under the same country's different cultural education stages becomes more and more important especially for the educational leader in the decision making time. The formation process of the first impression is often influenced by education and first impression shape the way in which we interpret behavior and decision [1]. Many studies are focused on the difference between different countries [2, 3], cities [4], gender [5, 6], experiences [7] etc., few papers pay attention to the same country's internal stages difference of knowledge cognition about the other country's culture. To analysis the difference under the Chinese different cultural education stages including junior school, senior high school, university not major in Japanese (abbr. non-major) and major in Japanese (abbr. major) stages the education workers can find the existed problem and make a reasonable strategy for individualized education.

To obtain a better-convinced questionnaire data, the scientific investigation way should not be ignored. Felgen et al. [8] gave the specific question related to culture, however, this investigation design does not consider this psychological that the respondent has no idea to write about or have so many ideas beyond the question which is better for the research but don't know where to start. Fang et al. [2] collected participant's performance appraisal of first impression bias through the website, however, the conclusion data cannot be well reflected from the view of experiment. Similarity, participants [9] were requested to draw one thing after another in each of the nine segments on a sheet of paper in a cyclic manner. For A who seldom spoke and B who behaved as a leader in the group, A usually keep silent before writing and then explained in a clear mind after completing her drawing, B can give logical answers quickly but had difficulty in drawing with text instead, through the nine-in-one drawing (NOD) method the mother and children can express the memories in a relaxed situation which is better for obtaining the true convinced data. Obtain the first impression with NOD, the first impression towards Japan in Korea different cultural education stages [10] was analyzed, however, the impression acquisition under the Chinese different cultural education stages is seldomly mentioned which is important for educational leader to make a reasonable decision.

To choose the appropriate evaluation standard of difference analysis is particularly important. Shaista et al. [11] used the F-value to verify the hypothesis of the cultural difference about impression management between different countries by computing the variance with abundant experimental data. The ANOVA method was employed to investigate cultural variations in the different countries online decision-making processes [12]. The t-test [13, 14] was performed to assess whether the means of the two groups being observed were statistically different from each other. A significant effect of gender on back-jumps for two courses [15] was showed and the p_value was selected as one of the difference degree of evaluation criteria. Above all, to choose the appropriate evaluation standard depends on the

data type and distribution, considering the small data and comparison object in which the observed number and order cannot be changed, the paired-t-test method was chosen for getting more accurate analysis results on the small amount of data by analysis the difference of mean.

The contribution of this paper can be concluded as: (1) Collecting the questionnaire data of first impressions and corresponding knowledge acquisition ways by using the nine-in-one drawing method; (2) Analyzing the above difference by using the paired-t-test method. After finding the existed problem for cultural knowledge learning process, the individual education decision can be targeted proposed in the different cultural education stages.

2 Data from the Questionnaire

Questionnaire has been an effective and reliable research tool in collecting specific information from respondents. About 661 questionnaires included 165 junior school students, 228 senior high school students, 180 university students not major in Japanese, 88 university students major in Japanese were collected from different grades and different classes.

The number distribution characteristic of the surveyed 661 respondents in different cultural education stages is shown in Fig. 1. In order to reflect the universal opinions without any individual extreme color, the acquisition data collected in each cultural education stage should be equally diversified which is combined with different grades and classes. The numbers of detail-surveyed respondents in different cultural education stages are showed in Fig. 1a–d. Noted that the percentage

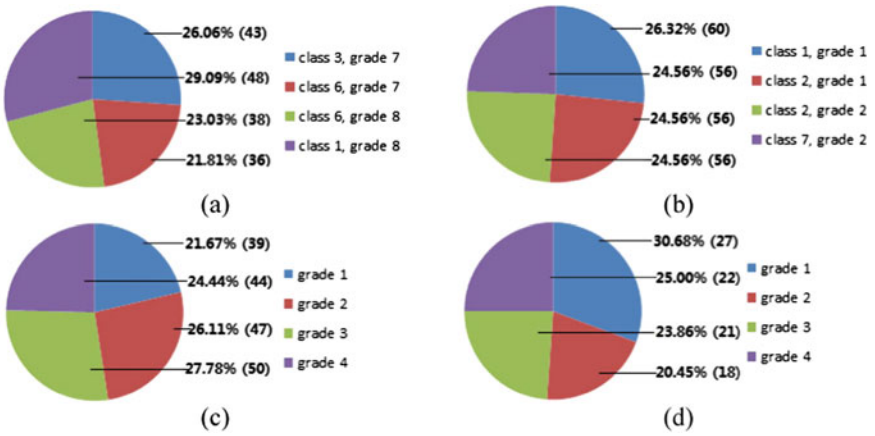


Fig. 1 The number distribution characteristic of respondents from different grades and classes in different cultural education stages. **a** In junior school students; **b** in high school students; **c** the university students not major in Japanese; **d** the university students major in Japanese

in each figure represents the proportion of surveyed students in the total number, and the total number is calculated by adding all the different grades and classes surveyed students together showed in the bracket.

The questionnaire was analyzed and concluded from the following two aspects: (1) the content acquisition including first impression towards Japan and the corresponding knowledge acquisition ways; (2) the difference analysis.

2.1 The Content Design of Questionnaire

In the social cognition to judge the first impression towards Japan in Chinese different cultural education stages can be a basis of making decisions in the cross-culture domain. The questionnaire content which directly reflects the respondents' mind was designed as: to write down the first impressions in turn and the corresponding knowledge acquisition ways effective way of acquiring the knowledge by using the nine-in-one drawing method.

2.1.1 Nine-in-One Drawing Method Towards the Questionnaire

In order to protect the description space while express the first impressions and easier to write the impressions with logical mind, the nine-in-one drawing method is applied which was invented by Moritani [16] to fill the content of questionnaire. Specifically the respondents firstly draw a frame and segment it into nine small boxes in A4 letters, secondly in the small box the first impression is written by respondents with figure, text, sign etc.

As illustrated in Table 1, considering that the different expression ways of each person may conveys the same meaning, the first impressions towards Japan in Chinese cultural education stages has been manually organized into 24 categories

Table 1 Some main specific examples of categories about the content

Category	The contents
1	Typical Japanese landscape (Mt Fuji & cherry blossoms)
2	Landmark (Osaka city & seto bridge)
4	Technology products (car, camera, computer)
7	Traditional culture (tea ceremony, sumo)
8	Food culture (sushi etc.)
10	Natural environment (earthquake)
15	Japan symbol (emperor of Japan etc.)
17	War
18	Political
20	General culture (cartoon etc.)

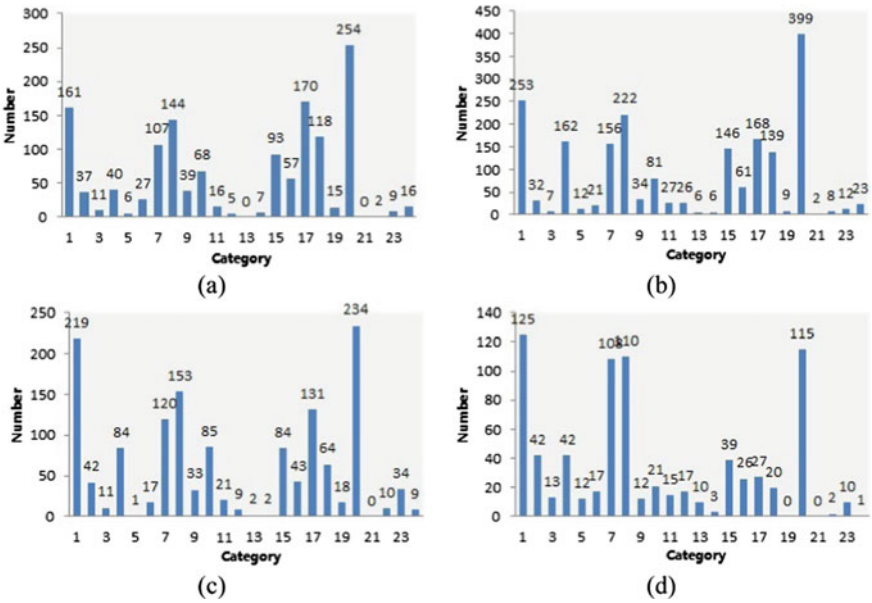


Fig. 2 The numbers of each category about the first impression in different cultural education stages. **a** In junior school student stage; **b** in senior high school student stage; **c** in university stage not major in Japanese; **d** in university stage major in Japanese

which classification standard is depended on the same meaning content of answers. Table 1 just gives some main specific examples of each selected category which corresponding chosen number can be seen in Fig. 2.

The specific selected numbers of each category about the first impression toward Japan in different cultural education stages was concluded in Fig. 2 including junior school student stage, senior high school student stage, university students stage not major in Japanese and stage major in Japanese stage. In Fig. 2a the top four of number amount order are in turn 20 (general culture), 17 (war), 1 (typical Japanese landscape), 8 (food culture). Figure 2b, c show the top four of number amount as 20 (general culture), 1 (typical Japanese landscape), 8 (food culture), 17 (war). It is interesting to find that the selected numbers of each category in major stage is quite different compared with non-major stage, in major stage the four former category is 1 (typical Japanese landscape), 20 (general culture), 8 (food culture), 7 (traditional culture). The biggest difference remains in the first impression changed in the surveyed respondents' mind.

2.1.2 The Corresponding Knowledge Acquisition Ways

This part in the questionnaire is to understand the thought source of mind's formation way in each cultural education stage, which is better for researcher to give a

Table 2 Some main specific examples of categories about the corresponding knowledge acquisition ways

Category	The answers
2	TV news in China
4	Website in China (xinlang)
5	Network chatting app (qq, wechat)
11	Magazine in China (except the textbook)
12	Regulation textbook in China
20	TV play
24	Cartoon
31	In the Class from the teacher

reasonable analysis of first impression towards Japan. The corresponding knowledge acquisition ways to form the mind of first impression towards Japan in Chinese cultural education stages has been manually organized into 34 categories which classification standard is depended on the same meaning content of answers with text and sign. The main specific examples of each selected category can be seen in Table 2, and the corresponding chosen numbers of each selected category is shown in Fig. 3.

The specific selected numbers of each category about the corresponding knowledge acquisition ways in different cultural education stages was concluded in

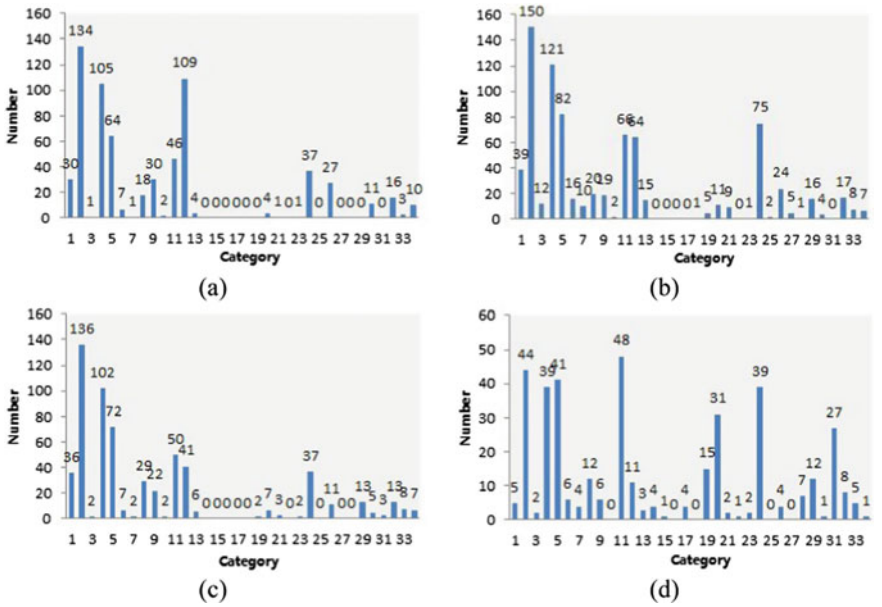


Fig. 3 The corresponding knowledge acquisition ways in different Chinese cultural education stages. **a** In junior school student stage; **b** in senior high school student stage; **c** in university stage not major in Japanese; **d** in university stage major in Japanese

Fig. 3 including in the junior school student stage, senior high school student stage, university stage not major in Japanese and stage major in Japanese. From two aspects: the top four orders in different cultural education stages and the overall distribution in each figure. Figure 3a gives the top four orders as 2 (TV news in China), 12 (regulation textbook in China), 4 (website in China), 5 (network chatting app). The three former orders of selected numbers in Fig. 3b, c are in turns as 2 (TV news in China), 4 (website in China), 5 (network chatting app). Figure 3d shows the top five orders as 11 (magazine in China), 2 (TV news in China), 5 (network chatting app), 4 (website in China), 24 (cartoon). The big difference exists in Japanese major student stage is the overall distribution compared with other cultural education stages, the overall distribution is diversified and dispersive which means the knowledge acquisition in major stage is various and the proportion of each category is not so concentrated.

2.2 *The Difference Analysis of Different Cultural Education Stages*

To scientific analysis the degree of difference in different cultural education stages, paired-t-test method is used as a standard which reflects human factor influence of the decision making in the field of cultural education.

The t-test is a statistical hypothesis test introduced by William Sealy Gosset which can be used to determine if two sets of data are significantly different from each other especially for small sample data, paired-t-test can reduce the effect of confounding factors in an observational study without needing the overall variance. The condition of using the paired-t-test method not only requires two observation samples are equal, the corresponding number order of groups should not be arbitrarily changed. Considering the difference judgment standard, the p_value can reflect the significance of overall level that means, if p_value ≤ 0.05 in the 95 % confidence interval, it can be trusted to accept the hypothesis test of consistency, however, its size has no comparable property. Through the value of correlation coefficient in the paired samples we can get the trend degree of correlation to judge the difference between two random variables, the correlation coefficient $\gamma_{X,Y}$ can be expressed in mathematics as:

$$\gamma_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2 - E^2(X))}\sqrt{E(Y^2) - E^2(Y))}} \tag{1}$$

where, X, Y respectively represent the two separate different samples, $E(XY)$ is the mean of composed sample. To analysis the difference between two different cultural education stages about the first impression towards Japan and the corresponding knowledge acquisition ways in the questionnaire, the corresponding numbers of each category which is shown in Figs. 2 and 3 has been adopted and lined up in a

Table 3 The difference analysis about the first impression between different cultural education stages (left: first line, right: second line)

		Mean		Std		t	Sig.	Correlation coefficient
		Left	Right	Left	Right			
Junior	Senior high	22.999	21.011	26.770	25.530	1.189	0.247	0.952
	Non-major	22.999	25.392	26.770	28.970	−1.221	0.234	0.944
	Major	22.999	26.233	26.770	31.463	−2.602	0.016	0.798
Senior high	Non-major	21.011	25.392	25.530	28.970	−0.832	0.414	0.962
	Major	21.011	26.233	25.530	31.463	−1.591	0.125	0.861
Non-major	Major	25.392	26.233	28.970	31.463	−0.303	0.765	0.902

Table 4 The difference analysis about the corresponding knowledge acquisition ways between different cultural education stages (left: first line, right: second line)

		Mean		Std		t	Sig.	Correlation coefficient
		Left	Right	Left	Right			
Junior	Senior high	14.508	15.726	25.662	24.217	−0.797	0.431	0.938
	Non-major	14.508	13.365	25.662	22.822	−1.205	0.335	0.928
	Major	14.508	23.591	25.662	30.883	−3.205	0.005	0.653
Senior high	Non-major	15.726	13.365	24.217	22.822	0.695	0.492	0.977
	Major	15.726	23.591	24.217	30.883	−3.003	0.015	0.792
Non-major	Major	13.365	23.591	22.822	30.883	2.647	0.012	0.766

regular order as the input data of paired-t-test method. In order to reflect the balance of personal amount, the input data of paired-t-test method has been normalized which means that the number of each input category must be divided by the maximum value in the corresponding stages. From the result, the corresponding coefficient value the closer to 1 indicates that the greater the correlation between the two selected samples, and the specific experiment results can be seen in Tables 3 and 4.

While judging the hypothesis of difference, the error always be controlled within 5 % that means the probability of confidence interval is 95 %. t is the standard error of the difference between two means of samples. Sig. is the p_value which reflects the significant acceptance probability, the corresponding coefficient value shows the degree of difference between variables.

Table 3 gives the specific corresponding coefficient values of difference analysis about the first impression towards Japan between different cultural education stages by using the nine-in-one drawing method. Difference is significant between Junior and major stages by analyzing the p_value. For the junior school student stage, compared with other stages it is more relevant to senior high school and non-major student stages with corresponding coefficient values of 0.938 and 0.928. From the overall corresponding coefficient value, the senior high school is the most relevant with non-major student stage with $\gamma_{X,Y}$ of 0.977. It is acceptable that the difference

analysis in the major student stage is low relative to other stages, which phenomenon can be explained that with the increase of personal reading level, the accepted knowledge is becoming more and more diversified and internal understanding is more deeply. Reflecting on the first impression, the bigger difference analysis can be understood.

The difference analysis results of the effective way to obtain the knowledge which is main factor of first impression making decision between different stages can be seen in Table 4. The difference is larger between major stage compared with any other stages by analyzing the p_value. It is interesting to find that the rule of correlation size between different stages is almost the same to the first impression content. With lowest correlation value ($\gamma_{X,Y}$: 0.653) between major and senior high school student stages and highest correlation value ($\gamma_{X,Y}$: 0.977) between senior high school and non-major student stage, this illustrates that the way of professional knowledge acquisition is influencing people's decision-making in the cultural domain.

3 Discussion

The paper mainly analysis the difference between Chinese different cultural education stages by using the paired-t-test method from the two aspects: (1) the first impression towards Japan answered in the questionnaire by using the nine-in-one drawing method; (2) the corresponding knowledge acquisition ways to form the impression. Based on these results, some targeted understanding and suggestions would be proposed for the educational decision maker in cross-culture domain.

From Fig. 2 and Table 3, it is obviously to find the training process of impression [17]. The learning process in the first step is always start with typical cultural and history, for further learning the forming process of traditional culture would be understand. During senior high school and non-major stages the relevance of first impression is bigger than other stages, which means that the understanding aspect of knowledge would not be refreshed if not engaged in cultural-crossing domain. The professional knowledge acquisition does let people know more culture which is comprehensive. Considering the global knowledge the specific suggestion for the decision-making is to strength tradition culture in junior student stage and remind people to remember the history in major stage.

From Fig. 3 and Table 4, the effective way of obtaining the knowledge remains on the TV and textbook in junior student stage while the students use the magazine and network in other stages. The cause of first impression exists in the people's acquisition way and the content of corresponding ways. In the adolescent people is willing to watch TV and books, however people in the university tends to get more information from the internet. In the view of different benefit groups, the suggestion from the experiment is that the targeted content of corresponding way can get the maximize transmission.

Combine the above together, some concluded suggestions for the cultural reformation can be summed as: individual education which means targeted to spread the cultural knowledge towards different cultural education stages. For junior students, some native culture can be added into the effective way of TV and books towards junior student stage, for major student stage the history can not be ignored in the network.

4 Conclusions

In this paper, we systematically analysis the difference of first impression towards Japan and corresponding knowledge acquisition ways in Chinese cultural education stages. By using the nine-in-one drawing method, the acquisition knowledge in the junior student stage is mainly focused on the general culture and history through TV and textbook, however the students in major stage get the more in-depth knowledge through the network and magazine. In order to better serve for the cultural education, the individualized education is proposed for the decision makers to have a purpose of specific knowledge to targeted groups.

Acknowledgments This work was partially supported by the National Key Technology Support Program of China (Grant No. 2012BAH45B01), by the Key Natural Science Foundation of Hubei Province of China (Grant No. 2015CFA069), and by the Hubei Key Laboratory of Transportation of Intent of Things of China (Grant No. 2015III015-B06).

References

1. Van Hemert, D.A., Koning, L., Berg, H.: Interpersonal influence in cross-cultural interactions. Tno Defence Security and Safety Soesterberg (Netherlands) (2010)
2. Fang, X., Rajkumar, T.M.: The role of national culture and multimedia on first impression bias reduction: an experimental study in US and China. IEEE Trans. Prof. Commun. **56**(4), 354–371 (2013)
3. Nomura, T., Kanda, T., Suzuki, T.: Implications on humanoid robots in pedagogical applications from cross-cultural analysis between Japan, Korea, and the USA. Robot. Hum. Interact. Commun. 1052–1057 (2007)
4. Tavares, A.S., Gálvez, C., de Albuquerque, L.W.N.: Information on public transport: a comparison between information systems at bus stops. Proc. Manuf. **3**, 6353–6360 (2015)
5. Powell, M., Ansic, D.: Gender differences in risk behaviour in financial decision-making: an experimental analysis. J. Econ. Psychol. **18**(6), 605–628 (1997)
6. Zell, E., Krizan, Z., Teeter, S.R.: Evaluating gender similarities and differences using metasynthesis. Am. Psychol. **70**(1), 10 (2015)
7. Wang, Z., Liao, M., Hagihara, R.: Comparison between expert and beginner on calligraphy letter's characteristic. Proc. Manuf. **3**, 495–502 (2015)
8. Felgen, L., Grieb, J., Lindemann, U.: The impact of cultural aspects on the design process. In: DS 32: Proceedings of DESIGN 2004, The 8th International Design Conference, Dubrovnik, Croatia (2004)

9. Ishii, N.: Mental Support for Mothers of Children with Developmental Disorders. Creating New Science for Services
10. Kagami, T., Jisun, P., Tomomi, M., Asano, I.: How South Korean elementary, junior high school, senior high school and university students' perceptions of Japan are formed: in relation to interests and knowledge of Japan. Japanese education and language. Culture **39**, 41–49 (2010)
11. Khilji, S.E., Zeidman, N., Drory, A.: Crossvergence of values: an analysis of the use of impression management strategies in India, Israel and Pakistan. Int. Bus. Rev. **19**(4), 419–431 (2010)
12. Li, L.M.W., Masuda, T., Russell, M.J.: Culture and decision-making: investigating cultural variations in The East Asian and North American online decision-making processes. Asian J. Soc. Psychol. **18**(3), 183–191 (2015)
13. Carraher, S.M., Paridon, T.J.: Entrepreneurship journal rankings across the discipline. J. Small Bus. Strateg. **19**(2), 89–98 (2015)
14. Fabregas, P., DiDona, T.: Layoff Perceptions: A Cross-Cultural Analysis
15. Guo, P.J., Reinecke, K.: Demographic differences in how students navigate through MOOCs. In: Proceedings of the First ACM Conference on Learning, ACM (2014)
16. Moritani, H.: Parents' images of pupils expressed in the nine-in-one drawing method. Res. Bull. Educ. Sci. Naruto Univ. Teach. Educ. **11**, 213–226 (1996)
17. Lizardo, O., Penta, M., Chandler, M.: Analysis of opinion evolution in a multi-cultural student social network. Proc. Manuf. **3**, 3974–3981 (2015)

Dynamics of National Culture and Employee Characteristics on Organizational Commitment in Retail Banks in Ghana

Mohammed-Aminu Sanda

Abstract Firms operating in high growth sectors are known to face the challenges of hiring and assimilating large numbers of new employees, providing new knowledge and skills to existing employees, and addressing the need for a rapid expansion of leadership capability. In the Ghanaian situation, such a challenge is known to prevail, since the country implemented major financial sector reforms starting in the late 1980s as part of its broad market reforms, key of which is the deregulation of the financial sub-sector. The banking sector in Ghana is therefore, faced with human resource management challenges, which includes the finding of the right caliber of employees to employ, and how to ensure that employees get committed to their organizations in order to reduce their desire to switch to competitor firms, due to apparent dissatisfaction with their jobs. This study therefore, examined a conceptual model that sought to hypothesize the impact of national culture and employee characteristics on employees' organizational commitment in retail banks in Ghana. The purpose is to understand the extent to which Ghanaian national cultural values and employee characteristics impact on employee organizational commitment in Retail Banks operating in Ghana. Quantitative data was collected from 282 bank employee across nine different retail banks in Ghana, and analyzed stepwise, using the analysis of moment structures (AMOS) program. Firstly, path analysis was conducted to test the individual measurement models that constitute the various components of the conceptual structural model. In this analysis, the factor score weights and model fit estimates for the indicator variables in the various latent variables (i.e. national culture, employee characteristics, and organizational commitment,) were appraised. Results from the path analysis identified four measurable indicators for organizational commitment. The analysis showed that all the measurable indicators tested for national culture and employee characteristics did not have significant loads to serve as measurable indicators. It is concluded that employees' organizational commitment in retail banks in Ghana is

M.-A. Sanda (✉)
University of Ghana Business School, P. O. Box LG 78, Legon, Accra, Ghana
e-mail: masanda@ug.edu.gh; mohami@ltu.se

M.-A. Sanda
Luleå University of Technology, 97187 Luleå, Sweden

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_6

neither influenced by the Ghanaian national culture nor the employee's characteristics.

Keywords National culture · Employee characteristics · Organizational commitment · Retail banks · Ghana

1 Introduction

The effective functioning of a country's financial system is therefore perceived as enhancing the smooth implementation of its macroeconomic policies. Motivated by this perception, governments in most African countries, including Ghana have, in the past three decades, reformed their financial sectors as part of a broader market reforms. The purpose of these reforms is to ensure the building of more efficient and competitive financial markets. In this respect, the following were activities were carried out or encouraged as key elements of the reform: (i) deregulation of the financial sub-sector through privatization of state-owned banks, (ii) influx of private banks, and (iii) registration of non-bank financial institutions. These activities has led to the emergence of new products development and accompanying new processes that has resulted in the enhancement of financial service delivery by the banks. In the case of Ghana, these reforms have led to substantial growth of the banking sector. There were only 8 retail banks in the country in the year 1990. This number rose to 26 in the year 2009, with a total of 656 branches. This is in addition to a number of financial service institutions, such as insurance firms, investment houses, rural banks, stock exchange, cooperative credit unions, savings and loan companies, mutual funds, and other microfinance institutions.

It is established in the extant literature that managing rapid growth has always entailed enormous human resource challenges, even in advanced countries with larger pools of skilled employees and managerial talents. Firms operating in high growth sectors are known to face the challenges of hiring and assimilating large numbers of new employees, providing new knowledge and skills to existing employees, and addressing the need for a rapid expansion of leadership capability towards employees' organizational efficiency and effectiveness. These challenges have been elaborately discussed in Western management literature. The prevailing understanding is that when an employee is dissatisfied with his/her work, she/he may become emotionally or mentally "withdraw" from the organization. There is the likelihood that such an employee will be less committed and will likely consider quitting. Employee attrition is therefore, a major cost to rapidly growing firms. Firstly, firms lose the knowledge that the employee has built up over years of employment. Secondly, there is the additional cost of replacing and training new employees. The banking sector in Ghana is therefore, faced with human resource management challenges, which includes the finding of the right caliber of employees to employ, and how to ensure that employees get committed to their organizations in order to reduce their desire to switch to competitor firms, due to

apparent dissatisfaction with their jobs. In the light of these observations, there is a great need to understand the effect of national cultural values and employee characteristics impact on employee organizational commitment in the retail banks. The purpose of this study therefore, was to examine and understand the influences of national cultural values and employee characteristics on employees' organizational commitment in Retail Banks operating in Ghana. The underlying research question is as follows: Do Ghanaian national cultural values and personal characteristics have any effect on employees' organizational commitment in retail banks?

2 Literature Review

There are studies in the extant literature that suggest the existence of a strong link between the development of a competitive financial sector and economic growth prospects in a country. For example, based on a cross-section sample of forty-one countries for the period 1980–1990, [4] found that increased concentration in banking markets generally depresses industrial growth, but also boosts the growth of industries that depend relatively heavily on external sources of finance. Similarly, [1], using data for regional banking markets in Italy, found that a higher concentration promote firm creation in industries where the creditworthiness of entrepreneurs is difficult to assess. Conversely, [5] found that more competition in banking markets raises industry growth that is relatively heavily dependent on external financing. In contrast to [5, 11] found that more competition in banking markets lowers the growth of industries that is relatively heavily dependent on external financing. This observation reinforced the notion that bank competition fosters industrial growth [5]. Against this background, there is the need to understand how bank management processes in Ghana is affected by employee organizational commitment, as influenced by the national culture and the employees' characteristics. Antecedents of organizational commitment found in previous studies include [26] and national culture [10]. One strand of management research that has a strong theoretical promise for the present study is the Person–Organization (PO) Fit framework. The framework, according to [14], highlights the compatibility between people and the organizations in which they work. The PO Fit concept suggests that an individual's attitudes and behaviors are influenced by the extent to which the organization and the individual are similar or have a fit on certain characteristics [6]. The framework can be used to examine how employee selection and career development can serve as a first step in developing a high performing banking industry culture, where employees show commitment to their organizations. Typical characteristics that are examined within the frame of the PO concept include the individual's and the environment's values, goals, and traits [3, 7, 8, 13, 16, 17]. The fit between the individual and the organization has been related to various attitudinal and behavioral outcomes, such as organizational commitment, job satisfaction, intention to quit, turnover, and task performance [9, 15, 24, 25]. The fit between the individual and the organization, in relation to organizational

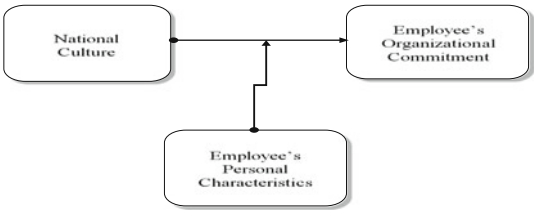


Fig. 1 Conceptual structural model showing relationship between national culture, employee characteristics and organizational commitment

strategy [6] and organizational commitment [2, 3, 18, 22, 23] have also been examined. However, research has not examined the fit between employee selection and career development with regard to organizational culture in the Ghanaian context. In the light of these discussions, this study explored and identified the dominant national cultures and individual employees' characteristics that impact on employees' organizational commitment to their retail banks. Thus, the initial literature review suggested that the conceptual model presented in Fig. 1 could guide the empirical investigation. The Individual employees' personal characteristics served as a moderator in the relationship.

The factors used for measuring the employee characteristics and organizational commitment components of the model were derived from extant literature, while factors for measuring the national culture component of the model were derived from [10] cultural dimensions theory as well as from the works of [20, 21].

3 Methodology

Based on the notion that the effective functioning of a country's financial system enhances the smooth implementation of its macroeconomic policies, with the management of rapid growth always entailing enormous human resource challenges, even in advanced countries with larger pools of skilled employees and managerial talents, quantitative data was collected from 282 bank employee across nine different retail banks in Ghana using a self-administering questionnaire. In order to gain an unhindered access to target employees in these banks, the support and approval of top management for the study was firstly sought and obtained. A stratified random sampling technique was then used to select the respondents from each of the retail banks. The stratification was based on the respondents' position in the firms (i.e. managerial and non-managerial). The gender distribution of the bank employees who served as study respondents showed that 51.8 % (N = 146) of the 282 study respondents were males while 48.2 % (N = 136) were females. Though it appears from this statistics that there were more males than females, the closeness of the percentage (i.e. 52 % males and 48 % females) provides a fair sense of sample representation in terms of respondents' gender. The stratification of the respondents into managerial and non-managerial categories showed that out of the sample 282 respondents, 20.2 % (N = 57) hold managerial status in the banks, while the

remaining 79.8 % (N = 225) were general staff members with no managerial responsibilities. This distribution indicates a fair representation of span of control (supervisor—employees) in the banks in where a manager is expected to directly supervise an average of five employees. The age distribution of the respondents showed that 101 (35.8 %) of the respondents were also in the prime of their careers with age that ranged from 20 to 29 years. Only 2.5 % (N = 7) of the 260 study respondents were at the epic stage of their careers with age range of 50–59 years. The study respondents thus represent a fair collective of matured bank employees. The categorization of the highest level of educational qualification attained by the study participants showed that majority of the respondents were highly educated as it is indicated by a total of 90.0 % (N = 254) respondents who are university degree holders education, and the 7.5 % (N = 21) who are diploma and certificate holders. Distribution of the number of years the study participants worked in their organizations is showed that a vast majority of the respondents, comprising about 67.4 % (N = 190) have been working with their banks for a time period of not less than 3 years. As it is deducible from the table, 22.7 % (N = 64) have worked in their banks for a time period of more than 7 years, while 44.7 % (N = 126) have worked for a period ranging from 3–6 years. Only 11.7 % (N = 33) of the respondents have spent less than a year in their organizations. This implied that on the average, the study participants are fairly knowledgeable about the managerial and structural characteristics of their work environments.

4 Results Analyses and Discussion

The collated data was analyzed stepwise, using the analysis of moment structures (AMOS) programme. In the first step, path analysis, consisting of both factor analysis and confirmatory factor analysis were carried out to test the individual measurement models that constitute the various components of the conceptual structural model shown in Fig. 1. In this analyses, the factor score weights (regression weights or path coefficients), and model fit estimates for the indicator variables (observed variables) in the various latent variables (i.e. national culture, employee characteristics and organizational commitment) were appraised. The path analyses was then followed by a structural analysis of the conceptual model with the latent variables associated with their determined indicator variables, in order to test the model fit of their conceptual relationship. The AMOS graphics statistical software was used as the analytical tool.

At the start of the analysis, each of the latent variables in the conceptual model (i.e. Fig. 1) was loaded in the AMOS software in order to assess the model-fitness of their respective measurable indicators (factors). The indicator fit is interpreted from the perspective of [19] that estimated indicator loadings on the latent variable must be 0.7 or higher. This allowed for model modification towards attaining superior goodness of fit by rejecting indicators whose loadings fall below the baseline of 0.7. The criteria used to establish model fit include (i) Chi Square

(CMIN) which is the absolute test of model fit, (ii) Comparative Fit Index (CFI) and (iii) Root Mean Square Error of Approximation (RMSEA). For the chi square, a probability value below 0.05 implies model rejection. Thus for model fit, the probability must be greater than 0.05. For the RMSEA, [12] recommend RMSEA values below 0.06 and Tucker-Lewis Index values of 0.95 or higher for model fit. Also, a value of CFI close to 1.0 indicates a very good fit.

4.1 Factor Analysis of Indicator Variables Predictive of National Culture

Seventeen (17) predictive indicators were tested per the data collected and minimum was achieved for the model. Based on the goodness of fit statistics, it is evident that the overall model fit does not appear quite good. This is because the estimated χ^2 of 598.527 (df = 119) has probability level of 0.000 which is smaller than the 0.05 used by convention. Thus the null hypothesis that the model fits the data is rejected. The estimated Root Mean Square Error of Approximation (RMSEA) value of 0.120 is also larger than the recommended value of 0.06 [12], and thus rejects the null hypothesis of a good model fit. The estimate for the Comparative Fit Index (CFI) of 0.560 is below 1.0. Therefore, the tested model for the latent variable (national culture) showed a poor fit and as such needs to be modified somehow. In this regard, factor analysis is carried out to identify the indicator variables that have significant relationships with the latent variable (i.e. national culture). The factor loadings or standardized regression estimate (R) for the respective indicators are shown in Table 1.

By convention, the indicators should have loadings of 0.7 or higher [19] on the latent variable for them to be significant. Based on the results highlighted in Table 1, all the national culture indicators have factor loadings (regression estimates) below 0.7. In this regard therefore, they could all be projected as non-effective indicators of national culture influence in retail banks in Ghana. Based on this finding, the national culture component of the conceptual **model** proposed in this study (see Fig. 1) becomes redundant.

4.2 Factor Analysis of Indicators Predictive of Employee Characteristics

Seventeen (17) predictive indicators were tested per the data collected and minimum was achieved for the model. Base on the goodness of fit statistics, it is evident that the overall model fit does not appear quite good. This is because the estimated $\chi 2$ of 571.770 (df = 119) has probability level of 0.000 which is smaller than the 0.05 used by convention. Thus, the null hypothesis that the model fits the data is

Table 1 Standardized regression estimates for national culture indicators

Indicator variables	Label	R	R^2
Ghanaians love good life and enjoyment	DQ1	0.171	0.029
Ghanaians like hard work and saving towards the future	DQ2	0.387	0.150
Ghanaians like people to tell them what to do	DQ3	0.537	0.288
In Ghana, employees think it is better to follow instructions at work rather than do things on their own, even if they know they are right	DQ4	0.406	0.164
Ghanaians are highly individualistic, and only think of themselves	DQ5	0.425	0.181
Ghanaians think there is no need to worry too much in life, and that God will always show them the right way	DQ6	0.600	0.360
Ghanaians take good care of each other	DQ8	0.417	0.174
A good subordinate in a Ghanaian firm must avoid actions that may surprise or embarrass his/her boss, even if the actions are in the interest of the bank	DQ9	0.500	0.250
Managers who do well in Ghana are shrewd with strong drive for power	DQ10	0.559	0.313
Knowledge is power in Ghana	DQ11	0.420	0.176
A good relationship is power in Ghana	DQ12	0.392	0.154
A good subordinate must respect his boss in Ghana	DQ13	0.290	0.084
Respect for age must be preserved in a Ghanaian bank. Age and experience in life are worth more than paper qualifications	DQ14	0.394	0.156
A good subordinate must be hard working and loyal to his boss' interests	DQ15	0.401	0.161
Family obligations are given high priority, even in managerial decision making	DQ17	0.488	0.238

rejected. Additionally, the estimated Root Mean Square Error of Approximation (RMSEA) value of 0.116 is also larger than the recommended value of 0.06 [12], and thus rejects the null hypothesis of a good model fit. The estimate for the Comparative Fit Index (CFI) of 0.752 is below 1.0. Thus, the tested model for the latent variable (employee characteristics) showed a poor fit and thus, needs to be modified. In this regard, factor analysis is carried out to identify the indicator variables that have significant relationships with the latent variable (i.e. employee characteristics). The factor loadings or standardized regression estimate (R) for the respective indicators are highlighted in Table 2.

By convention, the indicators should have loadings of 0.7 or higher [19] on the latent variable for them to be significant. Based on the results highlighted in Table 2, only two (2) indicator variables had factor loadings greater than or equal to 0.7. These indicators are EQ9—Employees have the chance to do their job without feeling they are cheating someone ($\boldsymbol{R} = 0.730$, $\boldsymbol{R^2} = 0.533$) and EQ11—Employees have the chance to do things for other people ($\boldsymbol{R} = 0.700$, $\boldsymbol{R^2} = 0.489$). These two indicators were then retested using Confirmatory Factor Analysis. In the Confirmatory Factor Analysis, the meaningfulness of the latent variable (employee

Table 2 Standardized regression estimates for employee characteristics indicators

Indicator variables	Label	R	R^2
Employees are able to do job without feeling it are morally wrong	EQ1	0.427	0.183
Employees are able to take pride in a job well done	EQ2	0.548	0.300
Employees are able to do something much of the time	EQ3	0.662	0.438
Employees are able to do things that do not go against their conscience	EQ4	0.608	0.369
Employees are able to stay busy	EQ5	0.483	0.234
Employees are able to do something worthwhile	EQ6	0.573	0.328
Employees are able to do things that do not go against their religious beliefs	EQ7	0.617	0.381
Employees have the chance to do something that makes use of their abilities	EQ8	0.645	0.416
Employees can do their job without feeling they are cheating someone	EQ9	**0.730**	0.533
Employees have the chance to be important in the eyes of others	EQ10	0.582	0.339
Employees have the chance to do things for other people	EQ11	**0.700**	0.489
Employees have the chance to tell people what to do	EQ12	0.540	0.292
Employees have the chance to be responsible for the work of others	EQ13	0.408	0.166
Employees have the chance to tell others what to do	EQ14	0.562	0.316
Employees have the power and authority to execute their duties	EQ15	0.680	0.463
Employees have the chance to make decisions on their own	EQ16	0.456	0.208
Employees get the feeling of accomplishment from their jobs	EQ17	0.642	0.413

characteristics) and its 2 predictive indicators (i.e. EQ9 and EQ11) were tested and minimum was achieved for the modified model. Based on the goodness of fit statistics, the baseline comparison estimates showed that the overall model fits quite well to the data. The estimated Comparative Fit Index (CFI) of 1.00 accepts the null hypothesis that the modified model fits the data well. The maximum likelihood estimates entailing the standardized regression (R) estimates which represent the factor loadings associated with the respective indicator variables in the modified model are highlighted in Table 3.

The standardized regression estimates (indicator loading coefficients) for EQ9 (i.e. employees have the chance to do their job without feeling they are cheating someone) and EQ11 (i.e. employees have the chance to do things for other people) are 0.826 ($\boldsymbol{R^2}$ = 0.576) and 0.759 (R^2 = 0.683) respectively. The estimated loadings for the indicators are significant. Indicators should by convention have loadings of 0.7 or higher on the latent variable [19] for them to be significant. The reliabilities for these indicators are also high, ranging from 76 % (R^2 = 0.576) to 83 % (R^2 =

Table 3 Standardized regression estimates for indicators in the modified model

Indicator variables	Label	R	R^2
Employees can do their job without feeling they are cheating someone	EQ9	0.826	0.576
Employees have the chance to do things for other people	EQ11	0.759	0.683

0.683). In summary, the analysis of employee characteristics (latent variable) shows that a good measure of such characteristic is indicated by employees having the chance to do their job without feeling they are cheating someone, as well as doing things for other people. Based on this findings, the employee characteristics component of the conceptual model proposed in this study (see Fig. 1) is represented as a measure of these two factors.

4.3 Factor Analysis of Indicators Predictive of Organizational Commitment

Fifteen (15) predictive indicators were tested per the data collected and minimum was achieved for the modified model. Base on the goodness of fit statistics, it is evident that the overall model fit appears quite good. This is because the estimated $\chi 2$ of 510.944 (df = 89) has probability level of 0.000 which is smaller than the 0.05 used by convention. Thus the null hypothesis that the model fits the data is rejected. Additionally, the estimated Root Mean Square Error of Approximation (RMSEA) value of 0.130 is also larger than the recommended value of 0.06 [12], and thus rejects the null hypothesis of a good model fit. Even though the estimate for the Comparative Fit Index (CFI) of 0.753 indicates an acceptance of the null hypothesis, it also suggests that the model fit to the data is questionable. Therefore, the measurement model for the latent variable (organizational commitment) showed a poor fit and as such needs to be modified somehow. In this regard, factor analysis is carried out to identify the indicator variables that have significant relationships with the latent variable (i.e. organizational commitment). The factor loadings or standardized regression estimate (R) obtained from AMOS for the respective indicator variables are highlighted in Table 4. By convention, the indicators should have loadings of 0.7 or higher [19] on the latent variable for them to be significant. Based on the results highlighted in Table 4, four (4) indicator variables had factor loadings greater than or equal to 0.7. These indicators are GQ3—Employee getting a feeling of accomplishment from the job (R = 0.727, R^2 = 0.528); GQ4—Bank's economic system maximizing employees' commitment (R = 0.710, R^2 = 0.504); GQ9—Employees having the chances to advance in their positions (R = 0.759, R^2 = 0.575); and GQ14—The bank treats its employees very well (R = 0.711, R^2 = 0.505). These four indicators were then retested using Confirmatory Factor Analysis.

Table 4 Standardized regression estimates for organizational commitment indicators

Indicator variables	Label	R	R^2
Employees' commitment to the future success of the bank	GQ1	0.490	0.240
Extent to which employees see their commitments to the bank as high	GQ2	0.569	0.323
Employee getting a feeling of accomplishment from the job	GQ3	**0.727**	0.528
Bank's economic system maximizing employees' commitment	GQ4	**0.710**	0.504
Employee getting praise for doing a good job	GQ5	0.613	0.376
Encouraging the pursuit of individual goals in the bank, even if this is at the expense of group loyalty	GQ6	0.469	0.220
Employee commitment towards providing good services to bank's customers	GQ7	0.449	0.202
Contribution of employee's job commitment to organizational performance	GQ8	0.537	0.288
Employees having the chances to advance in their positions	GQ9	**0.759**	0.575
Employees doing their jobs without feeling that they are cheating on the bank	GQ10	0.619	0.384
The work employees do in the bank makes use of their abilities	GQ11	0.683	0.466
Staff layoffs and transfers are avoided in the firm	GQ12	0.466	0.217
Salaries that the bank pays employees compares in equal measures with that paid to employees in other firms	GQ13	0.444	0.197
The bank treats its employees very well	GQ14	**0.711**	0.505
Employees are able to do their jobs without feeling it is morally wrong	GQ15	0.630	0.397

In the Confirmatory Factor Analysis, the meaningfulness of the latent variable (organizational commitment) and its 4 predictive indicators (i.e. GQ3, GQ4, GQ9 and GQ14) were tested and minimum was achieved for the modified model. Based on the goodness of fit statistics, it is evident that the overall model fit the data quite well. This is because the Comparative Fit Index (CFI) of 1.00 indicates that the modified model fits the data well, add thus accepts the null hypothesis that the model fits the data. The maximum likelihood estimates entailing the standardized regression estimates which represent the factor loadings associated with the respective indicator variables in the modified model are highlighted in Table 5.

Table 5 Standardized regression estimates for indicators in modified model

Indicator variables	Label	R	R^2
Employee getting a feeling of accomplishment from the job	GQ3	0.736	0.541
Bank's economic system maximizing employees' commitment	GQ4	0.740	0.548
Employees having the chances to advance in their positions	GQ9	0.712	0.507
The bank treats its employees very well	GQ14	0.759	0.577

The standardized regression estimates (indicator loading coefficients) for GQ3 (i.e. Employee getting a feeling of accomplishment from the job) is 0.7136 ($\boldsymbol{R^2}$ = 0.541). That for GQ4 (i.e. Bank's economic system maximizing employees' commitment) is 0.740($\boldsymbol{R^2}$ = 0.548). The estimate for GQ9 (i.e. Employees having the chances to advance in their positions) is 712 (R^2 = 0.507), while that for GQ14 (i.e. the bank treats its employees very well), is 0.759 (R^2 = 0.577). All the estimated loadings for the indicators are significant in that; indicators should by convention, have loadings of 0.7 or higher on the latent variable [19] for them to be significant. The reliabilities for these indicators are also high ranging from 71 % (R^2 = 0.507) to 76 % (R^2 = 0.577). In summary, the analysis of organizational commitment (latent variable) shows that it is manifested in retail banks when the employees are treated very well such that they get feelings of accomplishment from the job, and also the chances to advance in their positions. Such employee commitment is maximizes by the Bank's economic system. Based on this findings, the organizational commitment component of the conceptual model proposed in this study (see Fig. 1) will be represented as a measure of these four factors.

4.4 Structural Analysis of Conceptual Model

The derivations from the confirmatory factor analysis identifying the requisite indicators that have significant relationships with their respective latent variables were incorporated in the conceptual model (*recall*: Fig. 1) to yield a re-conceptualized structural model. The relationship between the two (2) latent variables, Employee characteristics (ECH) and Employee organizational commitment (OCT) was tested. Minimum was achieved for the re-conceptualized structural model structural model, and the standardized model-fit generated is shown in Fig. 2. Based on the goodness of fit statistics, it is evident that the overall model fit did not appear quite good. This is because the estimated $\chi 2$ of 14.537 (df = 8) has probability level of 0.069 which is bigger than the 0.05 used by convention. Thus the null hypothesis that the model fits the data is accepted. Additionally, the estimated Root Mean Square Error of Approximation (RMSEA) value of 0.054 is also smaller than the recommended value of 0.06 [12], and thus accepts the null hypothesis of a good model fit. The estimate for the Comparative Fit Index (CFI) of 0.989 also indicates an acceptance of the null hypothesis, indicating that the model fit to the data.

As it is highlighted in Fig. 2, the standardized regression (R) estimate for the association between the employees' characteristics (ECH) and their organizational commitment (OCT) is 0.64, which is lower than Schumacker and Lomax's [19] threshold factor loading of 0.7. This therefore, indicates employees' characteristics do not predict their organizational commitment.

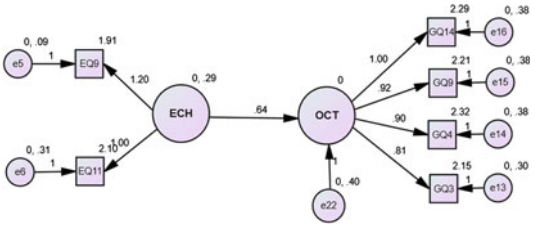


Fig. 2 AMOS graphics generated path diagram showing standardized indicator loadings in the re-conceptualized structural model

5 Conclusion

This study has shown that while there are some measurable indicators of employee characteristics and organizational commitment in Ghanaian retail banks, there is none for nature culture. Despite this observation, the study also showed that despite the existence of the measurable indices for both employee characteristics and organizational commitment, there is no significant influencing association between them. It is therefore, concluded that the organizational commitment of employees in Ghanaian retail banks, on the one hand, is not affected at all by the influence of national culture, and on the other hand, is not significantly affected by the employees' personal characteristics. In this respect, the following developmental indices in Ghanaian retail banks for understanding employees' characteristics and enhancing their organizational commitment could be outlined.

- *Employee Characteristics*: Employees ability to do their job without feeling that they are cheating someone as well as the opportunity for them to exhibit their leadership qualities must be sustained.
- *Organizational Commitment*: Organizational commitment could be enhanced by treating employees very well, giving them the chances to advance in their positions, getting employees to derive accomplishment from their jobs, and also using the Bank's economic system to maximize employees' commitment.

References

1. Bonaccorsi di Patti, E., Dell'Ariccia, G.: Bank competition and firm creation. J. Money Credit Bank. **36**(2), 225–52 (2004)
2. Boxx, W., Odom, R., Dunn, M.: Organizational values and value congruency and their impact on satisfaction, commitment, and cohesion. Public Pers. Manag. **20**, 195–205 (1991)
3. Cable, D.N., Judge, T.A.: Person–organization fit, job choice decisions, and organizational entry. Organ. Behav. Hum. Decis. Process. **67**, 294–311 (1996)
4. Cetorelli, N., Gambera, M.: Banking market structure, financial dependence, and growth: international evidence from industry data. J. Financ. **106**(2), 617–648 (2001)
5. Claessens, S., Laeven, L.: Financial dependence, banking sector competition, and economic growth. In: World Bank Policy Research Working Paper 3481. (Online 2010: J. Eur. Econ. Assoc. **3**(1), 179–207) (2005)

6. Da Silva, N., Hutcheson, J., Wahl, G.D.: Organizational strategy and employee outcomes: a person–organization fit perspective. J. Psychol. **144**(2), 145–161 (2010)
7. Ferris, G.R., Youngblood, S.A., Yates, V.L.: Personality, training, performance, and withdrawal: a test of the person–group fit hypothesis for organizational newcomers. J. Vocat. Behav. **27**, 377–388 (1985)
8. Furnham, A., Walsh, J.: Consequences of person–environment incongruence: absenteeism, frustration, and stress. J. Soc. Psychol. **131**, 187–204 (1991)
9. Hoffman, B.J., Woehr, D.J.: A quantitative review of the relationship between person–organization fit and behavioral outcomes. J. Vocat. Behav. **68**, 389–399 (2006)
10. Hofstede, G.: Culture's Consequences: International Differences in Work-Related Values. Sage, London (1980)
11. Hoxha, I.: Big banks versus small banks: good or bad for industries? University of Houston working paper (2009)
12. Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. **6**(1), 1–55 (1999)
13. Judge, T.A., Cable, D.M.: Applicant personality, organizational culture, and organization attraction. Pers. Psychol. **50**, 359–394 (1997)
14. Kristof, A.L.: Person–organization fit: an integrative review of its conceptualizations, measurement, and implications. Pers. Psychol. **49**, 1–49 (1996)
15. Kristof-Brown, A.L., Zimmerman, R.D., Johnson, E.C.: The consequences of individuals fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. Pers. Psychol. **58**, 281–342 (2005)
16. Muchinsky, P.M., Monahan, C.J.: What is person environment congruence: supplementary versus complementary models of fit. J. Vocat. Behav. **31**, 268–277 (1987)
17. O'Reilly, C.A., Chatman, J., Caldwell, D.F.: People and organizational culture: a profile comparison approach to assessing person–organization fit. Acad. Manag. J. **34**, 487–516 (1991)
18. Ostroff, C., Shin, Y., Kinicki, A.J.: Multiple perspectives of congruence: Relationships between value congruence and employee attitudes. J. Organ. Behav. **26**, 591–623 (2005)
19. Schumacker, R.E., Lomax, R.G.: A Beginner's Guide to Structural Equation Modeling, 2nd edn. Lawrence Erlbaum Associates Inc, Mahwah, NJ (2004)
20. Testa, M.R., Mueller, S.L., Thomas, A.S.: Cultural fit and job satisfaction in a global service Environment. Manag. Int. Rev. **43**(2), 129–148 (2003)
21. Turner, E.: Experiencing Ritual: A new interpretation of African healing. University of Pennsylvania Press, Philadelphia (1992)
22. Van Vianen, A.E.M.: Person–organization fit: the match between newcomers' and recruiters' preferences for organizational cultures. Pers. Psychol. **53**, 113–149 (2000)
23. Vancouver, J., Schmitt, N.: An explanatory examination of person–organization fit: organizational goal congruence. Pers. Psychol. **44**, 333–352 (1991)
24. Verquer, M.L., Beehr, T.A., Wagner, S.H.: A meta-analysis of relations between person–organization fit and work attitudes. J. Vocat. Behav. **63**, 473–489 (2003)
25. Wheeler, A.R., Gallagher, V.C., Brouer, R.L., Sablynski, C.J.: When person—organization (mis)fit and (dis)satisfaction lead to turnover: The moderating role of perceived job mobility. J. Manag. Psychol. **22**(2), 203–219 (2007)
26. Williams, L., Hazer, J.: Antecedents and consequences of satisfaction and commitment in turnover modes: a re-analysis using latent variables structural equation methods. J. Appl. Psychol. **71**(2), 219–231 (1986)

Cross-Cultural Competence and Functional Diversity in Business Negotiations: A Developing Country's Perspective

Anku-Tsede Olivia and Believe Quaqoo Dedzo

Abstract The paper examines cross-cultural competence and functional diversity in cross border transactions. From the Ghanaian cultural context, we analyse the positions of USA, UK, Germany, Japan, China, and the South African business executives in a buyer-seller relationship exchange of the negotiations using the qualitative exploratory approach. Findings indicate whereas cultural competence significantly affects trans-national negotiation decision making, the success or failure of negotiations is not dependent on the presence of shared cultural identities. Mutual adaptations, trust and power dominance significantly influence cross-border negotiation decision making. Meanwhile, in the buyer-seller dyad, functional diversity in negotiation decision outcomes is influenced by partner's level of exposure and ethics as openness, integrity, and intrinsically linked to business decisions and contractual agreements.

Keywords Cross-cultural competence · Functional diversity · Negotiations · Shared values · Ghana

1 Introduction

Cultural competence influences businesses negotiations and has significant impact on negotiators' behavioral responses and negotiation strategies. A substantial factor that may affect cross-cultural decision making is cultural contextualism. It is observed that people from low-context cultures—Canadian culture for example, use explicit and direct language, and those from high-context cultures such as Chinese culture are prone to implicit and indirect language wherein expressions derive their meanings from contextual cues [1] Although there are fragments of literature that explored ethical dilemmas of cultural differences in understanding the process of ethical decision-making [2], there is arguably no analysis of inter-cultural

A.-T. Olivia (✉) · B.Q. Dedzo
University of Ghana Business School, University of Ghana, LG 78 Legon, Ghana
e-mail: oankutsede@gmail.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_7

competence in business ethics in the Sub-Saharan African enclave, where businesses are faced with numerous ethical dilemmas in negotiations. In essence, the study explores the essential external, inter-cultural and behavioral factors influencing cross-border business relations in a wider collectivism context. International and cross border business are delicate and sensitive to global events, such as cultural differences and political, financial and legal risks [3]. Due to the liberalization of the Ghanaian economy with the discovery of oil in commercial quantities, the economy attracts high MNCs and international business executives challenged albeit considerable intricate of business exigencies whilst working to promote superior relations and negotiations. The high standards of the economy can make claims and negotiations highly complex amidst diverse cultural backgrounds, making the economy prone to further risks, including differences in Ghanaian legal and regulatory framework and of practices of guest stakeholders, risks in differential cultural values and behaviour amidst political and financial risks [3]. Thus, we explore how cross-cultural competence affects the extent to which it is appropriate to use ethically questionable strategies in business negotiations or such strategies consistent with social norms and ethical standards. Prior research suggests that cultural issues are likely to fuel conflict and raise complexities and difficulties in transnational businesses [4], with the increased participation in the value chain leading to more business interactions and arguments: contractual or social, resulting in many risks and disputes [5]. Since thorough understanding of cross-cultural differences in business ethics and transactions can help practitioners avoid the tendecy of anger and mistrust, exploring this subject in the context of business regulatory frameworks in a developing country context would afford international business partners the need to develop trust and reduce hostility toward either parties. The paper adopts a dyadic research approach [6] to obtain insight into cultural adaptation as perceived by the negotiators from Ghana and investors from other parts of the world. We focus on adaptations to manage differences in ways of conducting business related to essential factors previously identified for building successful business relationships: communication/information sharing, trust, power structure, power-dependence structures and influence strategies [7, 8]. In using the dyadic approach, the paper identify both perceptual differences between Ghanaian and partners from Europe, America and other parts of Africa as well as how they perceive themselves as compared to how they are perceived by their partners and vice versa [6]. The paper offers the cross-cultural literature, novel theoretical and analytical framework on decision making in a wider collectivistic and developing country context. The study answers the questions; to what extent do cultural competence, values and ethics affect decision making and negotiations in a given cross border negotiation context or do the existence of shared cultural values or social identities with parties in a negotiation setting influence the process and outcome of negotiation decision making?

1.1 Cultural Differences in Buyer-Seller Relationships

A key challenge of conducting business across cultures is understanding cultural differences [9]. It is also crucial to understand how to handle such differences. Differences in experiences, expectations, sense of belonging within the industry, perceptions, beliefs, and behaviors may vary substantially between the partners in buyer-seller relationships across cultures [6]. Thus, to develop and maintain buyer-seller relationships across cultures, the partners in the relationships must manage differences. In the early stage of the relationship, adaptation is a means of developing trust, while in the mature stage, it is a means of expanding and solidifying the relationship [10]. Cross-national buyer-seller interactions involve buyers and sellers embedded in different cultures. The differences arising from the cultural context may lead to challenges as a result of the differences in how the business are conducted by the partners, which may invariably invoke conflicts, misunderstandings, and failure to exploit business opportunities [6]. However, since power-dependence asymmetry is linked to conflict and instability [11], the buyer-seller literature avers that symmetrical power-dependence distribution is a precondition for enduring and effective business partnerships [12]. Thus, in well-functioning cross-cultural business interactions, one party should expect power symmetry and the use of shared and collective adaptation to handle cultural differences so to guarantee successful relationship. With the new oil find and the liberalization of the banking industry, Ghana has not only attracted investors but has become a key player in the global market and a target for international trade and investment. In spite of a substantial market potential [13], the Ghanaian market is exposed to the risk and challenge as a result of contextual cultural differentials with investor countries.

1.2 Legal Factors Affecting Cross-Border Business Transactions

International projects are more uncertain, complex and riskier due to associated differences related to factors including lack of understanding or consideration of the local cultural customs and conditions, communication, language barriers, law and legal issues, local and international politics, regulations and legal project conditions, business practices, supervision of a diverse stakeholder professionals, economic and financial conditions of projects, etc., [14]. In fact, laws and legal issues are essential complexities of cross-border businesses that increase the risk of these projects [15, 16]. Legal factors increasing the complexity of these projects may include changes that may increase instability or results in uncertainty in projects. These may involve such sudden and unexpected changes in national or local laws and regulations or government, working under different governmental laws and limitations, terrorist attacks, challenges of customs clearance and kidnapping

personnel for ransom or political reasons [17]. In fact, politically unstable nations are mainly often affected by legal suites. Thus, discrepancies and differences in the legislative system of the host and the guest partner's country should be carefully considered. In essence, [18] averred that parties to cross-border transactions should endeavor to clearly understand the host country's laws and their interpretation and commentary of the contracts governed by local laws. In Iran and other Middle Eastern countries for example, the legal system is centred under the jurisdiction of Islamic law and contracts are often required to be translated from Persian (Farsi) to English. Since the use of multiple languages in business projects, may result in conflicting interpretations of terms, [17, 19] remarked that the original contract, written in the native language has precedence and supersedes translations of contracts.

2 Literature Review

The importance of cultural competence in cross-border transactions and management has been observed to impact work and academic-related attitudes; project grades and psychological adjustments, etc., [20]. Some scholars indicate that the cultural intelligence is essential for the development of global leadership and business success [21]. Besides, global leaders and business players are expected to develop cognitive flexibility and possess competencies that are an aggregate of intellectual, social, emotional, cognitive, and skill-based functions, commonly referred to as "cultural intelligence" [22]. These writers further averred that apart from competence in the global and mainstream aspects of business functions, such as finance, operations, marketing, and human resource management, etc., it is important for business negotiators to be skilled in their interpersonal conduct of global business.

Prior research has shown that adaptive behavior is influenced by the power-dependence balance between the business parties, i.e., the more powerful party is expected to influence the weaker and less powerful party [23]. Indeed, previous findings also shown that business partners from some cultures have a weaker or stronger propensity toward cultural adaptation [24]. It is observed that cultural barriers act as psychological limitations in the ability or willingness to understand, accept, and/or adapt to the norms of a foreign culture [25]. For example, [26] revealed that American expatriates are likely to encounter greater psychological barriers to adjusting to Chinese culture than expatriates from Britain, Germany, France, and Australia. Besides, [24] observed that cultural dimensions, involving tolerance for power inequality and power distance influence adaptive behavior in business relationships. Thus, due to high-power distance of the Chinese culture, they are found to express stronger incentive to adapt in asymmetric power relationships than Americans. Nonetheless, since mutually adaptive actions in a working relationship are conducive to more equitable exchanges between the parties involved [27], companies in a relationship endlessly work to promote and

adapt allow the relationship progress. On the other hand, shared social identities and cultural values counterparts in a negotiation process can influence the process and outcome of negotiation decision making. In identity-based conflicts, an analysis of the Zurich-London negotiations over Cyprus suggests that amidst adversarial ethnic ties, decision makers with shared (and salient) social identity are more likely to employ collective-serving decision strategies and seek even-handed solutions that will promote and not jeopardize their mutual interests. For instance, despite their differing ethnicity-driven, clashing interests over Cyprus, Turkey and Greece—both NATO members were observed to settle on a commonly agreed negotiation outcome. In contrast, decision makers with severe ethnic fragmentation with no shared social identity—Turkish and Greek Cypriot communities, are more prone to adopt self-serving decision measures and seek zero-sum negotiation outcomes or their exclusively benefits [28].

3 Methodology

Differences in experiences, expectations, sense of belonging, perceptions, beliefs, and behaviors may substantially vary between partners in business negotiation decision making across cultures [6]. We analyze the positions of USA, UK, Germany, Japan, China, and the South African nationals in negotiation process using the qualitative exploratory approach. By anchoring the analysis in the cultural context of Ghana, this study also contributes to the relatively underdeveloped literature on business negotiation decision making in Africa. To explore the trans-cultural differences and adapation, we interviewed 18 business executives in a buyer-seller relationship of the exchange. This contrasts prior research that examined cross-national transactions from only one side of the dyad [27]. To encourage opening and willingness to provide responses, many of the respondents were not asked direct questions regarding explicit business partners. The informants/executives were responsible for importing from or exporting to the respective countries. These are CEOs, managing directors, business lawyers whose experience in trading and business negotiations spanned between 5 and 30 years were drawn from the oil and gas industry, mining, construction, engineering, etc. Although the experience varied in length, it was ensured that all the respondents had well-established relationships with their counterparts, so to meaningfully contribute with valuable experiences and perceptions on the subject [6]. Semi structured interview guide was developed following extensive review of the literature. The questions regard perceived cultural differences. Both closed and open ended questions were asked to capture perceptions of communication/information sharing, trust, time and power structure related to relationship satisfaction. “Do you feel that there are any cultural differences between Ghana and partners from other parts of the world in relation to business conduct? Do you experience any

challenges in conducting business because of cultural differences?" Emphasis was placed on allowing the subjects to play an active role. Follow-up, probing questions were used to gain a deeper understanding of the interviewees' perspective. On average, the section lasted for about 20 min. The method involved both face-to-face and telephone interviews. Since interviewing business elites may be challenging to have access to and establish trust to gain information [29], personal networks further proved important in the process [30]. Nonetheless, since our University is well known in the country and beyond, gaining access and trust was not excessively challenging.

4 Findings

Findings indicate that unlike functional diversity, cultural competence enormously affects trans-national negotiation decision making. Indeed, no much significant observations were found regarding the success or failure of business negotiations where there are differences in cultural values or the presence of shared identities between Ghana and parties from Europe, America, Asia or Africa. Although shared and mutual identities provide a common platform for business negotiation between Ghanaian business executives, their British and South African counterparts, it seems quite and unilateral adaptations. In this section, excerpts from the interviews are included to illustrate some cultural behavior of business leaders in a complex and highly contextual nature of business negotiations. Regarding the tendency of cultural differences affecting cross-border transactions, some Ghanaian managers and CEOs expressed;

> Cultural factors immensely affect business negotiations. In a very large way it does depending on the culture. Chinese culture is different from German culture all these things affect negotiations in business. And then there are areas of cultural taboos which are very extreme. For instance I used to work in the bank. When you are negotiating with Asians you don't eye ball them but for the Europeans you look into their eyes whiles talking; the Germans are very precise when they are doing business some other culture they are equally relaxed. So culture enormous impact on the way we do business across nations [31].

> Again in the banks when you are opening business for the Chinese, may be you are opening an account for a Chinese company there are certain figures they don't want […] or will not end their account, they will reject it because they believe that certain numbers when the end it is not good for their luck. Americans tend to be more business like even with the British for instance they can have working lunch …they will be discussing business over lunch but for the French they just want to retire go for proper lunch eat, drink, etc. Those are culture differences so if you are doing international business you have to be aware […]

> […] typically I used to work for Standard Chartered Bank which is an Asian Bank … it is a British Bank but its more focused in Hong Kong. You have to beware […] the Asians turn to look down they don't want to look in your eyes directly, it doesn't show respect, […], but the Europeans they will trust you if you look in their face which are areas I personally encountered […] [31].

> […] if you are involved in international business you deliberately look out for the cultural differences especially in the areas of taboos I'm stressing on that there are cultural taboos are serious, […] so in international business status, you have to learn them. Unless you are doing business with just your local people then at worse with their cultural differences you have to learn their taboos.

> […] The Chinese for example don not really trust foreigners until you establish significant level of faith and loyalty in the process. It takes long for the relationship to mature. So negotiations normally last for long and time consuming with the Chinese business partners. But they seek cooperation.

This supports [32] that the culture of China is both high-context and collectivistic, which means that people tend to prefer behaviors that benefit the group as compared to the individual. They seek harmony and try to avoid conflict wherever possible in an effort to save face. In such cultures, people can be suspicious of foreigners and their intentions. Besides, many studies have proven that trust building among business negotiating partners in cross-border transactions is critical [33]. Cultural aspects can have significant effects on cross-cultural decision making and business negotiations. Cultural dimensions, such as high–low context [34], power distance, individualism–collectivism, masculinity–femininity, risk avoidance and long–short-term orientation [35], have different effects and aspects on negotiation decision processes.

A Japanese partner expressed;

> Unfortunately or fortunately in your side, a warm hand shake is a sign of business partnership, so the cultural aspect is dynamic. It could be as simple as somebody spitting or somebody using a tooth pick, the Japanese would want to cover their left hand with the mouth whiles using their right hand to remove the tooth pick. Here, people easily use the right hand and they even feel okay when the tooth pick in their mouth and giggling or playing with it. So in most aspect, we deem it important to have these exposures and these values respected.

A British mining partner averred,

> […] the way round it is that, so you find a lot of us in business transactions or communication limiting ourselves to the systems that are in line with ours, either in terms of infrastructure like the country Ghana has provided or that of the other culture.

A Ghanaian CEO uttered;

> […] the fact that you understand that the Japanese man doesn't come to the table with anything, […]. You talk to a Japanese business man he doesn't come to the table with anything. He comes to listen, he comes to understand, he comes to weigh and he will go back to sleep and forgets about you. Until you make another effort, then he applies what he heard and to see whether his culture can accept what you are telling him to do. If he cannot, no no no no he forgets about you. That is the not American, the American comes with this kind of bullying techniques and tactics – let's seal the deal, puts you in a tight position […]. Nevertheless, once you know you have something in common you use the common platform to engage rather than trying to find that platform. It is good to start off with one common platform and shared identities.

With language, communication and trust, Ghanaians and parties from other parts of the world in the buyer and seller dyad conjectured that language is important for communication and in business, effective communication is prime and the bottom line hence one cannot do business without communicating effectively. Besides, majority opined that in business you build trust and in doing so, you need to even be able to read, understand transactions and do them well, hence the need to establish certain cardinal values; one is openness, accountability and integrity, etc., in trust building.

An American emphasized;

> […] it takes time in business to build trust and therefore a good businessman takes time to arrive at a business decision or make a point or sign documents. For us, you don't just sign documents, you don't just accept agreements, it goes through a process so within that, we do 'DD' – Due Diligence further to ascertain whether you're dealing with somebody who has a background that you can verify […]

A Japanese engineer also noted;

> […] trust is critical, we can take 200 years to build trust and after that, take just a second to make a business decision. It is our way of testing the trust level. If you don't persist, if you are not consistent, we will know your product or your value has a short life.

Regarding the functional diversity and its influence on negotiating decision making, a Chinese CEO expressed;

> […] one can only acquire certain skills and ability to man certain positions. If your level of exposure corresponds to some fundamental basic ideals […], openness, integrity, accepting of faults and mistakes etc. We trade closely with about 90 countries particularly countries that have English and some French backgrounds, but these are not limitations. We can go the other way round if only we know how to apply whatever skills we have acquired.

An American manager also remarked;

> We have always tended to have a one-stop shop kind of services which means that, it is easier to conglomerate or to put together a team of expertise doing one thing. That is pure business decision. You weigh the economics; you have ABCD and […] and does it make any business economic sense? If that is crossed, you just go ahead; it's based on pure business sense. If it doesn't, the best decision is to look for an efficient service provider and contract it out.

A British business executive emphasized;

> However, like I said it's based on pure business sense. If you think that you can process your data in India and it's cheaper, why not, do it. If you think you can bring an Indian into set up a system, do it. Either way businesses—the word business is about profit, is about margin, is about having a going concern ok, so and it's driven by finance.

Another German business executive also observed;

> In the German economy, we believe in small scale, well established and specialized. I mean the plane that you see means that there over a 1000 producers and manufacturers and suppliers producing to the detail of whatever specs has been given and it's an assignment for them.

A Ghanaian articulated;

> […] in South Africa and Gambia for instance, the people have a welcoming but not too friendly attitude towards foreigners and I survived there. […] and I have had experiences in Europe – in Switzerland, that's a very clean and rigid country with a lot of financial sense and profit motive. I have also had some training and experience in Hong Kong, that's the best example of a capitalist economy and also in China and also in Germany. What I find most of the time is that Americans try to specialize so if it has to do with energy there will be a partner responsible for energy who has done several things in energy and that area who will come and talk to you […] generally I have dealt with Germans, Americans, British, South Africans, all of them. They will come with an expert or someone who has knowledge in the particular field and if you are battling with a subject they are not competent they will let you know.

A Ghanaian business and communication consultant avowed;

> I'm not a novelist. I know what they think- I know what an American thinks when he goes on a business table, I know what the Japanese is thinking, I know what the Hong-Kong or Chinese man is thinking, I know what the German is thinking, I know what the Swiss man is thinking. So I have used what I have acquired both academically and socially, backed by the experiences and the numerous travels I have to be able to understand these people. When I meet them, the way I receive them, they wonder whether I have been exposed to their culture and I said yes I have been to your country several times and I'm able to utter one or two words in their language and what they now do is to disarm their ammunitions. When people realise that this person knows how we think or do things or even knows how we greet or salute or do something, knows our food and can use the chopstick, you know, they lower this kind of barrier and you pass through easily.

5 Discussion

In this study, we have examined the effects of cross-cultural competence and functional diversity in business negotiations in the Ghanaian context. The analysis suggests that even in the presence of adversarial or dissimilar cultural identities, parties in business negotiation decision making employ collective-serving decision strategies to promote their mutual interests. This contrasts [28] that in negotiations, decision makers with severe ethnic fragmentation with no shared social ties are quite inclined and more susceptible to employ self-serving decision strategies and seek zero-sum negotiation outcomes that will exclusively benefit them. Contrary to [36] who observed that unlike Chinese executives US individuals appear to focus on their own success and are less willing to give up autonomy and individual freedom for successful team negotiations, the present study found the Americans equally more willing to give up autonomy and individual freedom to assure the success of group negotiation decision making. Besides, as the Chinese culture is highly relationship oriented, they are together with Japanese found to value trust hence takes lengthy time in building business relations, owing to their suspiciousness of foreigners and their intentions. Moreover, just as the Chinese [32], US individuals tend to value system of relationships with their Ghanaian counterparts

and business partners during negotiations. The German and the UK executives are rather found to express reservations on team based decision making. For the Germans, this discrepancy indicates low awareness of the Ghanaian culture and the reluctance to appreciate the existence of cultural differences between the two countries. This supports Selmer [26] that Germans demonstrated superior inability and unwillingness to adapt to Chinese business practices than Americans, French, Australians, etc. This reflects differences in psychological barriers to adaptation, especially psychological limitations in their readiness and aptitude to understand, accept, or adopt the values of a foreign culture [25]. Regarding the Ghanaian-British, relations, the historical mindset is perhaps a factor never considered as an influential element. This contrasts [6] who offered that this factor is essential as other cultural factors of past-time orientation, and partners in successful negotiations value this quality and often consider the past.

In order to forge successful business relationships, one needs to study and understand the culture, the negotiating styles, and the communication practices, since undermining cultural factors have been found to have detrimental impact on cross-border transaction. This validates [37] that China has a high-context culture and long-term relationships are preferred, agreements are most likely to be verbal rather than written, with clear lines of distinctions, cultural patterns tend to change slowly, etc. This anchors one of their three philosophical traditions of Confucianism, considered as an ethical system which focuses on a long-term lifestyle, with values including obligations of relationships, family ties, upright behavior toward others, etc., [38]. Ghana is a masculine and a collective nation as opposed to Japan, Thailand, the Republic of Korea, etc., described as feminine societies. Due to these cultural restraints between Ghana and their Japanese counterparts, Ghanaian business leaders expressed great challenges in transferring technology from Japanese organisations. In sharp contrast, it seems to work significantly between Ghana and their Chinese executives. Beyond this, Ghanaian business leaders expressed general satisfaction and success in negotiation decision making processes and business outcomes with parties from such high power distance nations including USA, Germany, and UK just as they do with their counterparts coming from China. This observation contrasts the assumption that organizations' in individualistic cultures, such as the USA, are more successful in importing technologies from other nations than collectivistic countries [39]. Surprisingly, the satisfaction is quite lesser between Ghanaian and South African (mono-cultural nations) partners, contrary to the assertion that shared identity is seen as a precondition for developing successful business relationships between partners. If mutual adaptation [40] is influenced by shared values, and regarded as prerequisite for successful business relationships, one would have expected parties from these two mono-cultural nations exhibiting much strong success in business negotiations, hence the assertion that one may associate unilateral adaptation with unsuccessful business relationships [6] cannot be retained. Although, mutually adaptive actions in business negotiation are considered conducive to equitable exchanges [27], some of the Ghanaian partners have been found to deploy the adaptation behavior, seen to be influenced by the power dominance relationships from perhaps such powerful nations including USA, UK and Germany

[24]. This phenomenon is observed to be further strengthened as these so-called powerful nations threaten to withdraw aids if Ghana appears uncompromising during negotiations. It is needless to emphasize the consequences as earlier studies revealed that the use of coercive influence strategies has harmful effects on business relationships [41].

Regarding functional diversity in negotiation decision outcomes, both Ghanaian partners and their colleagues expressed quite similar understanding. The respondents are of the view that one's level of exposure corresponds to some basic and fundamental ideals including; openness, integrity, accepting of faults and mistakes, etc. It is observed that diversity in thoughts and function, couple with the understanding of business context significantly influences the success or otherwise of the negotiation process. Players from USA, Japan, China, including some Ghanaian partners expressed desire for a 'one-stop shop' kind of services where it is easier to conglomerate or put together a team of expertise undertaking one project, partners from Germany, UK and SA appear quite indifferent as they turn to promote core-competency and specializations in business negotiations. Nonetheless, many remarked that those are pure business decisions, massively influenced by legal and contractual agreements. This is not surprising as USA, China and Japan have become the leading economies in the world and currently ranked the top three in GDP [42]. China for example is the leading exporter in the global economy [43]. Although these three cultures differ significantly in attitudes, behaviors [23] and foreign policies and trans-border business values, there suggests significant understanding between expatriates from especially USA and China, and the Ghanaian business leaders regarding cultural values in business negotiation decision making. Apart from the USA, there are differences in language between Ghana, China and Japan. Perhaps, this challenges the orientation that cultural variations, including dissimilar national language have an influence on the success of transferring technologies to another country [39].

6 Conclusion and Implications

Cross-border transactions is made of the human groupings influenced by cultural factors and its meaning that makes us unique in our geographic, social and economic values and practices, and also being the basis for business transactions. In the contemporary technological and industrial era, other societies have advanced ahead of others based on certain cultural values. Regardless of the expertise and competence one possesses in transnational business, people always want to earn trust, honesty and confidence, whilst understanding party's cultural values, language and other common practices. Findings indicate that despite the existence of power asymmetry and the absence of common cultural ties, cross-cultural business relationships functions. However, the success of this relationship is mainly influenced by mutual adaptations and perhaps the use of coercive powers rather than mutual negotiation decision making. The paper offers that historical relationship is perhaps

another influential factor that when undermined, could impinge successful negotiations. Contrary to prior findings, successful negotiations are influenced by such factors as trust and adaptation other than individualistic-collectivism and mono cultural relationships. Finally, we offer the buyer-seller dyad cross-border literature that functional diversity in negotiation decision outcomes is influenced by partner's level of exposure and such fundamental principles as openness, integrity, accepting of faults and mistakes, etc., as they are essentially business and economic decisions anchored on legal and contractual agreements. Nonetheless, cultural factors are significant in undertaking successful international transactions, being cross-culturally competent is vital for parties involved in international businesses.

References

1. Cousik, R.: Cultural and functional diversity in the elementary classroom: strategies for teachers. J. Multi. Educ. **9**(2), 54–67 (2015)
2. Lam, K.-C., Shi, G.: Factors affecting ethical attitudes in Mainland China and Hong Kong. J. Bus. Ethics **77**(4), 463–479 (2008)
3. Fazliani, H., Charoenngam, C.: Cultural, external and behavioral factors in claims Negotiations. Int. J. Energy Sector Mgt **9**(4), 619–642 (2015)
4. Fellow, R.F., Hancock, R.: Conflict resulting from cultural differentiation: an investigation of the new engineering contract. In: Council of International Construction Research and Documentation Proc. on Construction Conflict: Management and Resolution, Rotterdam, The Netherlands, pp. 259–267 (1994)
5. Kumaraswamy, M.M., Yogeswaran, K.: Significant sources of construction claims. Int. Constr. Law Rev. **15**(1), 144–160 (1998)
6. Voldness, G., Grønhaug, K.: Cultural adaptation in cross-national buyer-seller relationships. Int. J. Emerging Markets **10**(4), 837–857 (2015)
7. Leonidou, L.C., Talias, M.A., Leonidou, C.N.: Exercised power as a driver of trust and commitment in cross-border industrial buyer-seller relationships. Ind. Market. Manag. **37**(1), 92–103 (2008)
8. Trompenaars, F., Hampden-Turner, C.: Riding the Waves of Culture. Understanding Diversity in Global Business, Boston, Nicholas Brealey Publishing, London (2012)
9. Phan, M.C.T., Styles, C.W., Patterson, P.G.: Relational competency's role in Southeast Asia business partnerships. J. Bus. Res. **58**(2), 173–184 (2005)
10. Hallén, L., Johanson, J., Sayed-Mohamed, N.: Interfirm adaptation in business relationships. J. Market. **55**(2), 29–37 (1991)
11. Anderson, E., Weitz, B.A.: Determinants of continuity in conventional industrial channel dyads. Market. Sci. **8**(4), 310–323 (1989)
12. Geyskens, I., Steenkamp, J.-B.E.M., Scheer, L.K., Kumar, M.: The effects of trust and interdependence on relationship commitment: a trans-Atlantic study. Int. J. Res. Market. **13**(4), 1303–1317 (1996)
13. De Kets Vries, M., Korotov, K., Shekshina, S.: A work in progress transcending the fifth 'Times of Troubles. Org. Dyn. **37**(3), 211–220 (2008)
14. Kim, D.Y., Han, S.H., Kim, H.: Discriminant analysis for predicting ranges of cost variance in international construction projects. J. Constr. Eng. Mgt. **134**(6), 398–410 (2008)
15. Dey, P.K.: Project risk management using multiple criteria decision-making technique and decision tree analysis: a case study of Indian oil refinery. Prod. Planning Contr. Mgt. Oper. **12** (12), 903–921 (2012)

16. Fallahnejad, M.H.: Delays causes in Iran gas pipeline projects. Int. J. Project Mgt. **31**(1), 136–146 (2013)
17. Yates, J.K., Smith, J.A.: Global legal issues for engineers and constructors. J. Prof. Issues Eng. Educ. Pract. **133**(3), 199–209 (2007)
18. Chan, E.H.W., Tse, R.Y.C.: Cultural considerations in international construction contracts. J. Constr. Eng. Manage. **129**(4), 375–381 (2003)
19. Abdul-Malak, M.A.U., EI-Saadi, M.M.H., Abou-Zeid, M.G.: Process model for administrating construction claims. J. Manage. Eng. **18**(2), 84–94 (2002)
20. Gabrenya, W.K., Griffith, R.L., Moukarzel, R.G., Pomerance, M.H.: Theoretical and Practical Advances in the Assessment of Cross-Cultural Competence. Florida Institute of Technology, Melbourne, FL (2011)
21. Alon, I., Higgins, J.M.: Global leadership success through emotional and cultural intelligences. Bus. Horiz. **48**(6), 501–512 (2005)
22. Alon, I., Boulanger, M., Meyers, J., Taras, V.: The development and validation of the business cultural intelligence quotient. Cross Cultural Strateg. Mgt. **23**(1), 78–100 (2016)
23. Chinta, R., Capar, N.: Comparative analysis of managerial values in the USA and China. J. Technol. Mgt. China **2**(3), 212–224 (2007)
24. Hagberg-Andersson, Å., Grønhaug, K.: Adaptations in a supplier-manufacturer network—a research note". Eur. J. Mark. **44**(1/2), 31–34 (2010)
25. Lin, X.: Determinants of cultural adaptation in Chinese-US Joint Ventures. Cross Cultural Mgt. Int. J. **11**(1), 35–47 (2004)
26. Torbjörn, I.: Culture barriers as a social psychological construct: an empirical validation. In: Kim, Y.Y., Gudykunst, W.B. (eds.) Cross-cultural Adaptation: Current Approaches. Sage Publications, pp. 168–190. Newbury Park, CA (1988)
27. Selmer, J.: Psychological barriers to international adjustment: North American vs Western European business expatriates in China. Cross Cultural Mgt. Int. J. **7**(3), 13–18 (2000)
28. Sirin, V.C.: Examining the role of identity in negotiation decision making: the case of Cyprus. Int. J. Conflict Mgt. **23**(4), 413–439 (2012)
29. Leonidou, L.C., Palihawadana, D., Chari, S., Leonidou, C.N.: Drivers and outcomes of importer adaptation in international buyer-seller relationships. J. World Bus. **46**(4), 527–543 (2011)
30. Mikecz, R.: Interviewing elites: addressing methodological issues. Qual. Inq. **18**(6), 482–493 (2012)
31. Michailova, S.: Contextualising fieldwork: reflection on conducting research in Eastern Europe. In: Marschan, P., Welch, C. (eds.) Handbook of Qualitative Research Methods for International Business, pp. 365–383. Edward Elgar, Cheltenham and Northhampton, MA (2004)
32. Rivera, B.D., Rogers-Adkinson, D.: Culturally sensitive interventions: social skills training with children and parents from culturally and linguistically diverse backgrounds. Interv. School Clin. **33**(2), 75–80 (1997)
33. Mujtaba, G.B.: Negotiating with modern Chinese professionals. J. Technol. Mgt. China **8**(3), 190–202 (2013)
34. Hurn, B.J.: The influence of culture on international business negotiations. Ind. Commer. Train. **39**(7), 354–360 (2007)
35. Hall, E.T.: Beyond Culture. Anchor Books—Random House, New York (1976)
36. Fredi Garcia Diana Mendez Chris Ellis Casey Gautney.: Cross-cultural, values and ethics differences and similarities between the US and Asian countries. J. Technol. Mgt. China **9**(3), 303–322 (2014)
37. Hofstede, G.H.: Cultures and Organizations: Software of the Mind. McGraw-Hill, Maidenhead (1991)
38. Mead, R.: International Management: Cross-Cultural Dimensions, 4th edn. Blackwell, Malden, MA (2005)
39. Dawes, R.M., Thaler, R.H.: Anomalies: cooperation. J. Econ. Perspect. **2**(3), 187–197 (1988)

40. Mukherji, A.: Francis, J.D.: Mutual adaptation in buyer-supplier relationships. J. Bus. Res. **61** (2), 154–161 (2008)
41. Kumar, N.: The power of power in supplier-retailer relationships. Indus. Market. Manage. **34** (8), 863–866 (2005)
42. World GDP Ranking StatisticsTimes.com. http://www.statisticstimes.com/economy/projected-world-gdp-ranking.php
43. Kramer, R.M., Pommerenke, P., Newton, E.: The social context of negotiation: effects of social identity and interpersonal accountability on negotiator decision making. J. Conflict Resolut. **37**(4), 633–654 (1993)

Investigation of Cultural Bias Using Physiological Metrics

Renee Rigrish and Mary Fendley

Abstract In today's world, many business transactions and interactions are conducted cross-culturally. In a business meeting, it is essential avoid a major cultural faux pas in order to not offend your business partners. Individuals must adjust their approach to a situation to create a cultural match. In adjusting this approach, cognitive biases are a potential result in cross-cultural scenarios. We investigate the Mirror Imaging Bias, which was found to be a common result of a shortcut to decide how to act in a situation. Physiological metrics were used to see if biases can be detected in a non-invasive manner. It was found that pupil diameter is a reliable indicator of when Mirror Imaging Bias is present. By understanding how individuals process information and are influenced by Mirror Imaging Bias, we can help create applications as well as provide training to help avoid cultural faux pas.

Keywords Mirror image bias · Cognitive bias · Heuristic · Cultural bias · Physiological measures · Psychophysiological

1 Introduction

With international business becoming a normal part of business practice, it has become important to know how to create strong relations by respecting business customs and practices. While we may believe that an unconscious act of crossing a leg over the knee is harmless, other cultures may see this as an intentional act of disrespect [1]. To prevent cultural faux pas from happening, businesses often hire cultural liaisons which work as not only interpreters, but experts and advisors about do the "dos and don'ts" of each culture.

R. Rigrish · M. Fendley (✉)
Wright State University, College of Engineering and Computer Science,
3640 Colonel Glenn Hwy, Dayton, OH 45435, USA
e-mail: mary.fendley@wright.edu

R. Rigrish
e-mail: rigrish.2@wright.edu

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_8

These cultural faux pas can happen to small businesses and world leaders alike [2, 3]. It is apparent that while some acts are seen as positive or neutral in one country, they could potentially be very offensive in another county. It is already known that cultural bias exists as people are taught customs of their nation and must become familiar with new customs when visiting another nation. If cultural bias is detectable, it would suggest that prior consultation with a different nation's customs is beneficial before traveling abroad or that a cultural liaison may be required to help assist with cultural traditions.

The main goals of this study are as follows: to find if Mirror Imaging Bias has an influence in making decisions in culturally diverse situations, and to find if physiological measures could detect its presence. The hypothesis is that there will be a greater workload present when Theory of Mind is used to assess how to act in a situation over the influence of Mirror Imaging Bias.

2 Mirror Imaging Bias and Theory of Mind

Mirror Imaging Bias is described as "the tendency to interpret the actions of others in one's own terms" [4]. When people try to view a scenario and make conclusions based on their own personal experiences, they make an assumption that others think the same as themselves and share similar experiences, and thus use Mirror Imaging Bias. Salchak noted that is the most common bias that is commonly experienced by intelligence analysts and could be dangerous in different situations [5].

Theory of Mind is the ability for an individual to correctly understand other individuals' mental states in a situation including their different beliefs, intents, desires, knowledge, etc. [6]. While one individual may have a view on the situation, using Theory of Mind means that they understand the view of a different person who may have a very different view than their own. Most individuals are able to achieve this ability by the age of four [7].

One reason for Mirror Imaging Bias to happen is the failure of Theory of Mind. Sometimes, instead of being able to understand another person's beliefs, individuals may project their own in order to decide an appropriate action. It is a mental shortcut an individual may take in order to avoid a high cognitive load. Employing Theory of Mind has a higher cognitive load because it requires individuals to comprehend and understand the intentions and actions of another individual. Mirror Imaging Bias would bypass the process where another person's views would be taken in account; instead an individual would carry out a scenario based on their own views which is a more natural response.

In Schaller's paper Failed Mirroring as a Cultural Phenomenon, she states that "a cultural dimension also exists in mirroring or failed mirroring. When, in daily human interactions, persons have reflected back to them not themselves but the cultural assumptions of others, they experience failed mirroring" [8].

Cultural mismatch and the use of Mirror Imaging Bias as opposed to Theory of Mind can lead to misunderstandings within a business setting. By assuming and projecting your own cultural practices on another culture, you are using Mirror Imaging Bias.

3 Methodology

This study was designed to test how participants interact with different cultural scenarios in order to determine what type of bias is present. Answers provided by the participant were recorded and analyzed.

As stated before, Mirror Imaging Bias and Theory of Mind can occur when trying to predict the actions of other individuals. Mirror Imaging Bias occurred when the participant would give an answer on how a person should act based on their own personal experience. Essentially, they answered the question based on how they would act if they were in that situation. When answering for the other individuals in the situation, the answers will be compared to how they answered for themselves in the same situation to see if they exhibit their own cultural bias as opposed to attempting to modify their cultural lens to adjust to the mismatch. If they are able to project Theory of Mind correctly, the participant would have the ability to take the perspective of a person from a separate cultural background. This would be apparent if the participant is able to adjust their cultural mismatch to predict appropriate actions of the people in the scenario.

3.1 Participants

Fifty participants aged 18 to 40 were recruited from the Wright State University community. These subjects were comprised of the following groups: 20 Americans (10 male, 10 female), 20 Indians (10 male, 10 female), 7 Middle Eastern (4 Males, 3 Females), and 3 Chinese (2 Male, 1 Female). Due to a data recording error, one American female's results were not recorded correctly and thus the data was unusable. Another Indian female was found to answer the questions with response times of less than 0.1 s which would not allow time to read and comprehend the questions. Her data was omitted from analysis due to these impractical response times not allowing a genuine response.

3.2 Apparatus

Testing of the scenarios was performed at Wright State University (WSU) and the study was approved by the WSU Institutional Review Board (IRB). The software

used included the Tobii Studio eye tracker and the CAPTIV-L7000 system. Tobii was used to collect and record subjects' eye tracking information such as gaze, number of fixations, and pupil dilation while CAPTIV-L7000 collected galvanic skin response (GSR), heart rate variance (HRV), and electromyography (EMG) of the medial frontalis and the right-unilateral EMG of the orbicularis oculi. Although CAPTIV was used in the study, we are limiting the discussion to the results of the eye-tracking here.

3.3 Stimuli

The participants were asked to view the scenario with audio about what is happening in the picture. They were then asked a question about the scenario from the perspective of the "self" and "other". This helps determine how an individual would act naturally in a situation and if they are able to use Theory of Mind when answering for the other culture.

Scenarios were created based on research from several etiquette guides based on how to do proper business in each of these regions. Actions in the scenarios included dining etiquette, greetings, conversational etiquette and meeting etiquette.

3.4 Procedure

Participants start the experiment by filling out an informed consent. Once consent was received, the CAPTIV sensors were placed on the individual and tested for functionality. As familiarity with cultures plays a significant role in this study, participants were asked in a pre-questionnaire about what cultures they identify with, cultures they have been exposed to, as well as what level of exposure occurred. Participants also noted how long they have been exposed to that culture. A second pre-questionnaire was also administer to determine an individual's cultural awareness. The eye tracking system was then calibrated to the participant.

Before being presented with the experimental stimuli, participants viewed a demonstration of a scenario in which they were asked to answer a simple question about the picture presented.

There were 10 images of each culture which totaled 30 images presented. Participants were exposed to the each of the images with an accompanying audio describing the actions in each scenario. Each audio segment lasted between seven and twelve seconds. The culture being presented was either stated in the audio segment and/or was shown by the people present in each image. After each of the audio files were completed, the participant was presented with a question which was superimposed on to the image. The images were randomized within each cultural group. The order in which the cultures were presented was also randomized.

Each image was presented twice in order to explore the differences of answering questions about the self and answering questions about other people's actions in accordance to the culture presented. Since participants were shown each image twice, this totaled 60 scenarios presented. Participants responded by using the keyboard with the up arrow key corresponding to "Yes" and the down arrow key corresponding to "No." Image order was randomized in order to minimize a learning effect.

A post questionnaire was administered to determine how familiar the participants were with each of the presented cultures. Participants also indicated if they used any information about the cultures to make their decisions.

3.5 Research Components

In this study, the following cultural groups were chosen; American, Indian, Middle Eastern and Chinese. These cultures were chosen based off of the cultural populations present at Wright State University. The scenarios comprised of scenes from business with individuals from India, the Middle East, and China.

Cognitive load is described as a level of perceived effort which may be associated with cognitive tasks such as learning thinking and reasoning [9]. Cognitive load may also be measured using different physiological and performance metrics. The metrics analyzed in this paper are as follows: pupil diameter, number of fixations, response time.

Pupil Diameter and Number of Fixations. Pupil diameter measures the dilation of the eyes in millimeters. The pupil diameter of the left and right eyes were measured using Tobii Studio eye tracker. Pupil size has shown to directly reflect processing load and mental effort in individuals [10].

One study connects pupil diameter with cognitive load used a Stroop test to control the cognitive involvement of the participants [9]. The average pupil diameter over a task was measured and then different algorithms were used to predict which level of the cognitive load was present during that time. It was found from these algorithms that the pupil diameter has an average rate of 85.86 % at predicting the correct cognitive load present. One algorithm had a success rate of 89.08 % to classify the cognitive load [9]. This rate of accuracy shows that the pupil diameter is a reliable metric to measure cognitive load. Through various studies, it was found that pupil diameter is a reliable metric to use when measuring the cognitive load present during a task.

A fixation is where the eye remains still for a period of time to look at one spot in the field of vision [11]. It was found that "participants would show additional fixations on the units reflecting additional cognitive effort to inhibit the interfering information in incompatible trials" [12]. Fixation count was also found to be at its' highest when there was either a highly complex task or the interface was complex itself [13]. It is also important to note that individuals fixate more often on

informative objects within a scene in order to gather more information [14]. This leads to a conclusion that more fixations could indicate the presence of a higher level of cognitive load.

4 Results

The bias of each individual was determined by comparing how an individual answered the scenario when answering for "self" perspective versus answering the scenario based on "other" perspective.

If the answers for "self" perspective and "other" perspective matched, this indicates that the individual did not change their way of thinking in the scenario and thus Mirror Imaging Bias is implemented. If the answers for "self" perspective and "other" perspective did not match, this indicates that the individual used Theory of Mind to try to answer the scenarios from the perspective of the individual.

Overall results showed that Mirror Imaging Bias (62.45 %) was used by participants almost twice as frequently as Theory of Mind (37.55 %).

Participants stated that when answering questions on a cultural scenario they were not familiar with, they would apply their own experiences where they have dealt with a situation in their personal lives in accordance to their culture.

4.1 Response Time

Response times were calculated as the time when the question first appeared on the screen to when the participant chose an answer.

A one-way ANOVA analysis showed that for the Response Time there was a significant difference between bias types (p-value = 0.0380) when using an alpha of 0.05. Mirror Imaging Bias exhibits a lower response time (mean value of 3.08 s) than Theory of Mind (mean value of 3.52 s). The power of the test for bias was 0.5462. The Least Square means plot in Fig. 1 shows that for each culture that Theory of Mind has a higher response time with the exception of the Chinese group.

4.2 Pupil Diameter

Pupil diameters were taken at the time that the participant answered a question. Along with a baseline pupil diameter. Baseline pupil diameter was subtracted from the pupil diameter at the time when the participants answered the questions in order to account for differences in pupil diameters between participants. This

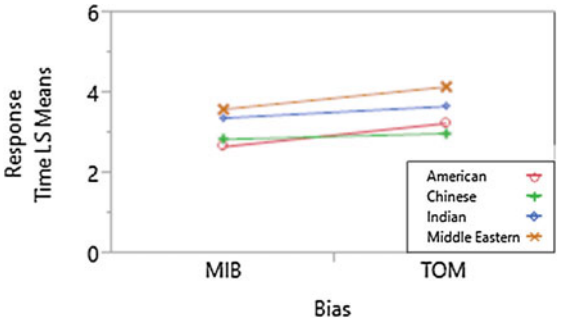


Fig. 1 LS means plot comparing response times for each culture according to MIB and TOM shows a trend of TOM consistently higher than MIB for all groups except Chinese

measurement will be referred to as Pupil Delta. The unit for the pupil diameters is in millimeters.

A one-way ANOVA analysis showed that for the Left Pupil Delta there was a significant difference between the bias types (p-value < 0.0001) when using an alpha of 0.05. There was no significant difference between the interaction of culture and bias (p-value = 0.2831). There is a significant difference between Mirror Imaging Bias and Theory of Mind where Mirror Imaging Bias shows a lower change in pupil diameter (mean value of 0.41 mm) than Theory of Mind (mean value of 0.49 mm). The power of the test for bias was 0.9834. The Least Square means plot in Fig. 2 shows that for each culture that Theory of Mind has a higher difference in pupil diameter.

A one-way ANOVA analysis showed that for the Right Pupil Delta there was a significant difference with the bias types (p-value < 0.0001) when using an alpha of 0.05. There is a significant difference between Mirror Imaging Bias and Theory of Mind where Mirror Imaging Bias shows a lower change in pupil diameter (mean value of 0.46 mm) than Theory of Mind (mean value of 0.54 mm). The power of the test for bias was 0.9906. The Least Square means plot in Figs. 3 show that for each culture that Theory of Mind has a higher difference in pupil diameter.

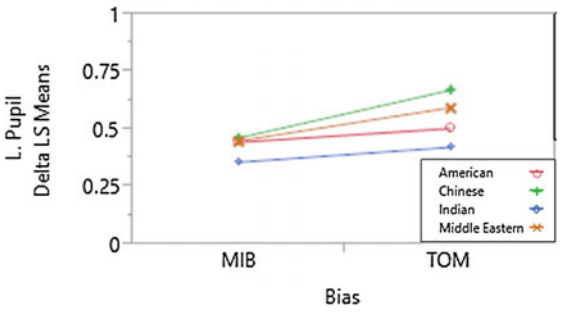


Fig. 2 LS means plot comparing left pupil delta for each culture according to MIB and TOM shows a trend of TOM consistently higher than MIB

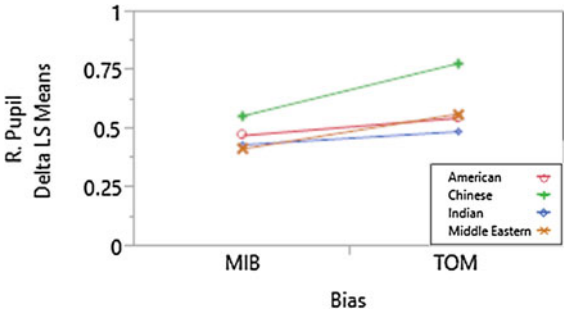


Fig. 3 LS means plot comparing right pupil delta for each culture according to MIB and TOM shows a trend of TOM consistently higher than MIB

5 Discussion

When answering questions about the scenarios, Mirror Imaging Bias occurred more frequently than Theory of Mind. Mirror Imaging Bias is seen as a "shortcut" method as it requires a lower cognitive load. As stated before, the majority of individuals would answer based on their own personal experiences when they were unsure on how to act in a different culture. Others would try to recall knowledge from media such as movies and television shows which have people from that culture. Mirror Imaging Bias occurred almost twice as much as Theory of Mind for all participants.

The analysis for response time showed that there was a significant difference found in the response time when comparing Mirror Imaging Bias and Theory of Mind. For all groups, Theory of Mind exhibited a greater response time than Mirror Imaging Bias by an average of 0.44 s. This supports how individuals take less time responding as they would naturally instead of changing their perspective to match that of the culture present.

When analyzing the change in pupil diameter for both the left and right eye, there was a significant difference between Mirror Imaging Bias and Theory of Mind. When there is a higher cognitive task load involved, the pupils will respond [12, 15]. This supports how Theory of Mind has a high change in pupil dilation than Mirror Image Bias. There was also a significant difference for the average pupil diameter changes between the different cultures. However, all cultures showed that the pupil diameter increased when Theory of Mind was used to answer questions both scenarios.

The number of fixations were found to have a significant difference between Theory of Mind and Mirror Imaging Bias. Overall, the group as a whole showed to have an average of 1.5 more fixations when utilizing Theory of Mind. Throughout the cultural groups, Theory of Mind was consistently higher in number of fixations than Mirror Imaging Bias throughout all cultural groups.

6 Conclusion

These observed metrics reinforce how when Theory of Mind is implemented, a higher cognitive load is present. Individuals must take into account how another individual would interact with a scenario which may be different from how an individual is used to acting. Most people choose to take the shortcut of using Mirror Imaging Bias. By using Mirror Imaging Bias, individuals avoid a higher cognitive load because they interact with the scenarios through their own lens. By using their own personal experiences, they eliminate taking other cultural practices into account as well as any other facts that could help make a conclusion. Since Mirror Imaging Bias naturally happens more often in all of these scenarios, it is important to find a way to ensure that individuals can learn how to implement Theory of Mind in critical situations.

References

1. Baker, W.G., Long, J.M.: The cultural heritage of arabs, islam, and the middle east (foreword by jerry M. (mark) long) Brown Books Pub. Group, Dallas, TX (2003)
2. Keneally, M.: Watch Putin and Obama offend the Chinese at beijing conference. http://abcnews.go.com/International/watch-putin-obama-offend-chinese-beijing-conference/story?id=26832982 (2014)
3. Parrenin, A., Rau, P.P., Zhong, R.: Impacts of cultural differences in a business environment for french companies in China. Glob. Bus. Manag. Res. **7**(1), 93–108 (2015)
4. Pipes, R.: What to do about the CIA. Commentary **3**, 36 (1995)
5. Salchak, C.R.: Investigation of Mirror Image Bias: Evidence for the Use of Psychophysiological Measures as Indicators of Cognitive Heuristics. Wright State University, OhioLINK
6. Call, J., Tomasello, M.: Review: does the chimpanzee have a theory of mind? 30 years later. Trends Cogn. Sci. **12**, 187–192 (2008). doi:10.1016/j.tics.2008.02.010
7. Perner, J., Lang, B.: Development of theory of mind and executive control. Trends Cogn. Sci. **3**(9), 337–344 (1999)
8. Schaller, J.: Failed mirroring as a cultural phenomenon. Pastoral Psychol. **56**(5), 507–520 (2008). doi:10.1007/s11089-008-0132-3
9. Ren, P., Barreto, A., Gao, Y., Adjouadi, M.: Comparison of the Use of Pupil Diameter and Galvanic Skin Response Signals for Affective Assessment of Computer Users. Instrument Society of America, United States (2012)
10. Moresi, S., Adam, J.J., Rijcken, J., Van Gerven, P.W.M., Kuipers, H., Jolles, J.: Pupil dilation in response preparation. Int. J. Psychophysiol. **67**, 124–130 (2008). doi:10.1016/j.ijpsycho.2007.10.011
11. Holmqvist, K.: Eye Tracking: A Comprehensive Guide to Methods and Measures Oxford, p. 2011. Oxford University Press, New York (2011)
12. Moeller, K., Fischer, M.H., Nuerk, H., Willmes, K.: Sequential or Parallel Decomposed Processing of Two-Digit Numbers? Evidence from Eye-Tracking. Q. J. Exp. Psychol. **62**(2), 323–334 (2009). doi:10.1080/17470210801946740
13. Wang, Q., SaLiu, M., ZikeMa, Q.: An eye-tracking study of website complexity from cognitive load perspective. Decis. Support Syst. **62**, 1–10 (2014). doi:10.1016/j.dss.2014.02.007

14. Loftus, G. R., Mackworth, N. H. Cognitive determinants of fixation location during picture viewing. J. Exp. Psychol. Human Percept. Perform. (1978)
15. Siegle, G.J., Ichikawa, N., Steinhauer, S.: Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. Psychophysiology **45**(5), 679–687 (2008)

Effects of Price of Zero on Decision Making: An Attempt to Generalize Human's Irrational Behavior to Price of Zero

Atsuo Murata, Yasunari Matsushita and Makoto Moriwaka

Abstract When $(a - 1)X = Z$, the zero-price effect was observed only for (a: large, X: small) and (a: small, X: small). When X was small, the zero-price effect was observed irrespective of the value of a. When X was large, the zero-price effect was not observed. Human behavior tend to deviate from irrational behavior assumed in traditional economics for small values of X irrespective of a. The price of zero must be irrationally chosen due to the overestimation of price of zero. When $(a - 1)X > Z$, the zero-price effect was observed only for (a: large, X: large) and (a: large, X: small). When a was large, the zero-price effect was observed irrespective of whether X was large or small. When a was small, the zero-price effect was not observed. Human behavior tend to deviate from irrational behavior assumed in traditional economics for large values of a irrespective of the value of a. In this manner, it has been indicated that the zero-price effect is not necessarily observable and holds under limited conditions.

Keyword Irrational behavior · Cognitive bias · Decision making · Price of zero

1 Introduction

It has been suggested that decisions about free products, the price of which is zero, are different from simply subtracting costs from benefits (Ariely [1]). The benefits associated with the price of zero attached to free products are irrationally perceived

A. Murata (✉) · Y. Matsushita · M. Moriwaka
Department of Intelligent Mechanical Systems, Graduate School of Natural Science and Technology, Okayama University, 3-1-1, Tsushimanaka, Kita-Ward, Okayama, Japan
e-mail: murata@iims.sys.okayama-u.ac.jp

Y. Matsushita
e-mail: moriwaka@iims.sys.okayama-u.ac.jp

M. Moriwaka
e-mail: Shinshi.ym@gmail.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_9

to be higher. Under such a condition of price of zero, a large number of participants tended to prefer a less attractive product to a more attractive one. In such a situation, we seem to act as if pricing a product as free not only decreases its cost, but also increases its benefits. This corresponds to an irrational behavior to the price of zero, which distorts our judgment to an irrational one. It must, however, be noted that the price of a product discounted to the price of zero is very cheap. It is questionable whether such an irrational behavior can be generalized.

Imagine that a company thinks of introducing a new machine tool to a factory and decides to introduce one with free expendable supplies as a result of comparing with other ones. The company must feel an advantage of the introduced machine tool over other ones. If the expendable supplies are higher than those of other machine tools, the company must continue to pay for expensive expendable supplies. They eventually notice that the price of zero is no longer felt as a merit. In this manner, they reacted excessively to the price of zero and are compelled to pay for expensive expendable supplies.

When prices are mentioned, individuals apply market norms. On the other hand, when prices are not mentioned (when the price is zero), individuals apply social norms to determine their choices and effort (Heyman and Ariely [2]). Ariely et al. [3] have shown that when a candy is offered at 1 ¢ per piece, a participant is likely to take about four pieces, whereas when the price of the candy is zero, a participant tends not to take more than one. This must indicate that a participant effectively decrease the demand when the prices is reduced. Kahneman and Tversky [4] discussed our irrational behavior and showed that when it comes to gambles, zero probability (as well as certainty) is perceived as substantially different from small positive probabilities). Whereas the values of small probability are perceived as higher than they actually are, the perception of zero probability is accurate.

Although the zero-price effect is demonstrated using a variety of experiments, it is not certain whether this effect is generally observable or not. If the zero-price effect is not generally observable, it must be identified under what conditions the zero-price effect is observable. The change of discounted amount, the price, the difference of the price between two alternatives must be treated as experimental factors, and the effects of these factors on the zero-price effect must be investigated. The difference of price between alternatives, the price of alternatives, and the relationship between the zero-price and the discounted price, that is, the relationship between X and Z (two levels: Eqs. (2) and (3) in Sect. 2.2 below) were manipulated, and it was explored whether the zero-price effect is generally observable.

2 Method

2.1 Participants

A total of eighty two participants from 19 to 24 years old at Faculty of Engineering, Okayama University took part in the experiment. The limitation of answering time

was not imposed on the participant. All participants had no knowledge on psychology and behavioral economics.

2.2 Irrational Behavior Caused by Pricing of Zero

(A) A product of $\$X$ is obtained for free.
(B) A product of $\$Y$ is discounted and sold with the price of $\$Z$.
It is assumed that X and Y is related as follows.

$$Y = aX \tag{1}$$

where $a > 1$, $Y > Z$.
If it is assumed that $Y > Y - Z > X$, Eq. (2) holds.

$$(a - 1)X > Z \tag{2}$$

If we can judge rationally, the alternative (B) must be chosen. If Eq. (3) is assumed, the alternatives (A) and (B) must be chosen with an equal probability from the rational point of view.

$$(a - 1)X = Z \tag{3}$$

The following questionnaire survey was conducted to examine the effect of the price of zero on the choice of alternatives (A) and (B) as a function of a and X for both Eqs. (2) and (3).

2.3 Task

The experimental factors were the size of a (two levels: small and large) and the size of X (small and large), and the relationship between X and Z (two levels: Eqs. (2) and (3)). Thus, the following questionnaire was prepared for a total of eight conditions.

(1) Choose one from the following two alternatives (a: large, X: small, Eq. (2)).

A: A chocolate of 20 ¢ is free.
B: A chocolate of \$1 is discounted to 80 ¢.

(2) Choose one from the following two alternatives (a: large, X: large, Eq. (1)).

A: Artificial leather shoes of \$80 are free if a suit is purchased.
B: Real leather shoes of \$400 are discounted to \$200 if a suit is purchased.

(3) Choose one from the following two alternatives (*a*: small, *X*: small, Eq. (1)).

A: An eraser of $1 is free.
B: A more usable eraser of $1.3 is discounted to 20 ¢.

(4) Choose one from the following two alternatives (*a*: small, *X*: large, Eq. (2)).

A: Clothes of $100 is presented by your friend.
B: Clothes of $130 is sold by $30 by your friend.

(5) Choose one from the following two alternatives (*a*: large, *X*: small, Eq. (1)).

A: Bread of 80c is free.
B: Bread of $4.8 is discounted to $2.8.

(6) Choose one from the following two alternatives (*a*: small, *X*: small, Eq. (2)).

A: Bottled water of $1 is free.
B: Bottled orange juice of $1.5 is discounted to 50 ¢.

(7) Choose one from the following two alternatives (*a*: large, *X*: large, Eq. (2)).

A: If you purchase a television set, you can get a DVD recorder of $50 for free.
B: If you purchase a television set, you can purchase a DVD recorder of $300 by $250.

(8) Choose one from the following two alternatives (*a*: small, *X*: large, Eq. (1)).

A: You can have a massage of $30 for free.
B: You can have a massage of $45 by $5.

As a control, the following questions without price of zero were also answered by the participants.

(1) Choose one from the following two alternatives (*a*: large, *X*: small).

A: A chocolate of ¢20.
B: A chocolate of $1.

(2) Choose one from the following two alternatives (*a*: large, *X*: large)

A: Artificial leather shoes of $80.
B: Real leather shoes of $400.

(3) Choose one from the following two alternatives (*a*: small, *X*: small)

A: An eraser of $1 is free.
B: A more usable eraser of $1.3.

(4) Choose one from the following two alternatives (*a*: small, *X*: large)

A: Used clothes of $100.
B: Used clothes of $130.

(5) Choose one from the following two alternatives (*a*: large, *X*: small).

A: Bread of 80 ¢.
B: Bread of $4.8

(6) Choose one from the following two alternatives (*a*: small, *X*: small, Eq. (2)).

A: Bottled water of $1.
B: Bottled orange juice of $1.5.

(7) Choose one from the following two alternatives (*a*: large, *X*: large).

A: A DVD recorder of $50.
B: A Blue ray recorder of $300.

(8) Choose one from the following two alternatives (*a*: small, *X*: large).

A: Massage service of $30.
B: Massage service of $45.

3 Results

The percentage choice (A) for Eqs. (2) and (3) is shown in Figs. 1 and 2, respectively. The zero-price effect was defined as the subtraction of the percentage choice (A) at the non-zero-price condition from the percentage choice (A) at the

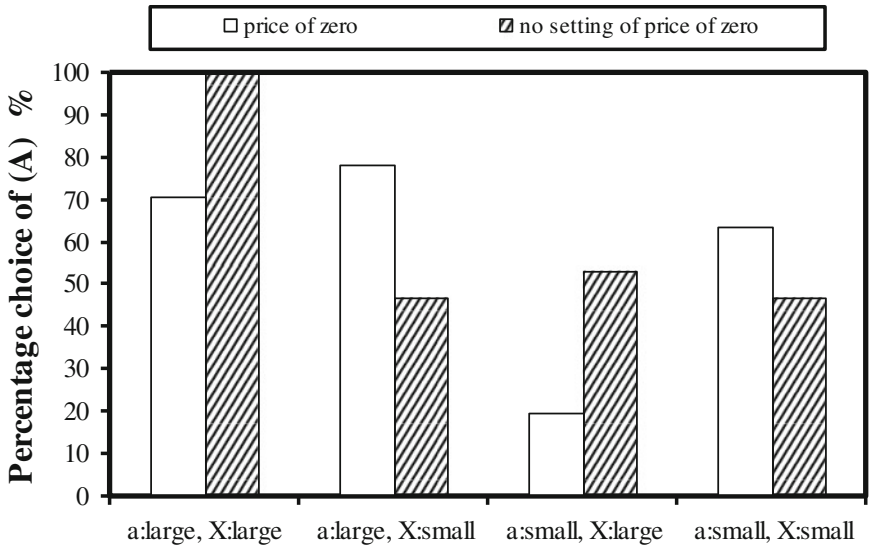


Fig. 1 Percentage of choice (A) (price of zero) as a function of combination (*a*, *X*) when $a(X-1) > Z$

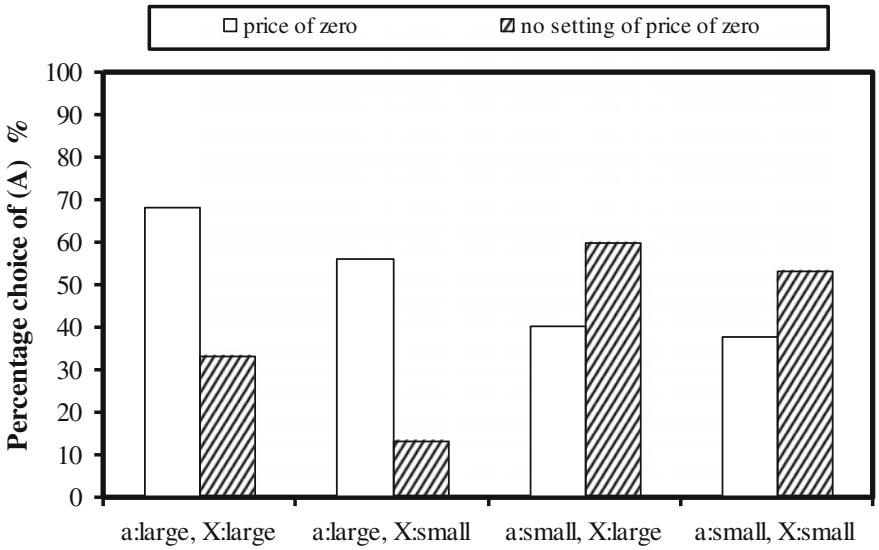


Fig. 2 Percentage of choice (A) (price of zero) as a function of combination (a, X) when $a(X\text{-}1) = Z$

zero-price condition and denoted by E_{zero}. The comparisons of E_{zero} among four combinations of a and X, and between Eqs. (2) and (3) is summarized in Fig. 3. The positive value of E_{zero} indicates that the zero-price effect is observable.

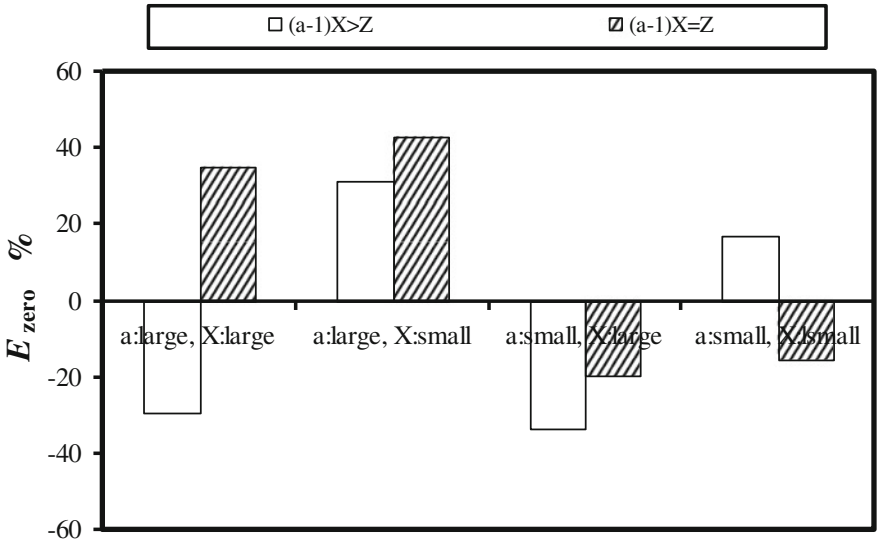


Fig. 3 E_{zero} as a function of combination (a, X) and whether Eq. (2) or Eq. (3)

4 Discussion

The zero-price effect is not always observable as shown in Figs. 1, 2 and 3. When Eq. (3) held (see Fig. 2), the zero-price effect was observed only for (a: large, X: small) and (a: small, X: small). When X was small, the zero-price effect was observed irrespective of the value of a. When X was large, the zero-price effect was not observed. When Eq. (3) holds, human behavior tend to deviate from rational behavior assumed in traditional economics for small values of X irrespective of a. Under the condition of Eq. (3), the small value of X might be a trigger of mapping difficulty and affect.

When Eq. (2) held (see Fig. 1), the zero-price effect was observed only for (a: large, X: large) and (a: large, X: small). When a was large, the zero-price effect was observed irrespective of whether X was large or small. When a was small, the zero-price effect was not observed. When Eq. (2) holds, human behavior tend to deviate from rational behavior assumed in traditional economics for large values of a irrespective of the value of X. Under the condition of Eq. (2), the small value of a might be a trigger of mapping difficulty and affect.

In this manner, it has been indicated that the zero-price effect is not necessarily observable and can be generalized to holds under limited conditions (small values of X under Eq. (3) and small values of a under Eq. (2)).

Evaluating the utility of a product in monetary terms is difficult. To illustrate this point, imagine a situation in which a consumer's valuation of the lower value chocolate is somewhere between 1 and 5 ¢ and his or her valuation of the higher value chocolate is somewhere between 10 and 20 ¢. If one is confronted with the choice between 1 and 14 ¢, it would be unclear whether any of the options would give rise to a net benefit. On the other hand, the same consumer, when facing the choice between zero price and 13 ¢, could easily see that one of the options-the free option-would be certainly produce a net benefit. Thus, the zero-price effect might be attributed to the uncertainty in overall benefit associated with costly options, and the certainty in overall benefit associated with the price of zero.

It is also possible to speculate that cognitive mechanism behind the zero-price effect might have to do with affect. Options with no cost (price of zero) evoke a more positive affective response than options with both benefits and costs. Therefore, it might be assumed that more individuals use this affective reaction as a cue for their decision, the more frequently they will choose the option with price of zero (Slovic et al. [5, 6]). This affective perspective also points to circumstances under which the zero-price effect could be eliminated. If the cause of the zero-price effect is a reliance on an initial (overly positive) affective evaluation, making a non-affective and more cognitive evaluation might diminish the zero-price effect. The price itself might work for eliminating the zero-price effect depending on the situation.

On the basis of discussion above, especially mapping difficulty and affect might contribute to some extent to our irrational behavior that makes us overreact to the price of zero.

5 Conclusions

The results can be summarized as follows. The generalized conclusion is that the zero-price effect is not always observable. The conclusions can be stated as follows for Eqs. (2) and (3).

(1) When Eq. (3) held (see Fig. 2), the zero-price effect was observed only for (a: large, X: small) and (a: small, X: small). When X was small, the zero-price effect was observed irrespective of the value of a. When X was large, the zero-price effect was not observed. When Eq. (3) holds, human behavior tend to deviate from irrational behavior assumed in traditional economics for small values of X irrespective of a. The price of zero must be irrationally chosen due to the overestimation of price of zero. Under the condition of $(a - 1)X > Z$, the larger value of X leads to rational judgment. On the other hand, when X is small, human judgment tend to be distorted to an irrational one, and vulnerable to the price of zero.

(2) When Eq. (2) held (see Fig. 1), the zero-price effect was observed only for (a: large, X: large) and (a: large, X: small). When a was large, the zero-price effect was observed irrespective of whether X was large or small. When a was small, the zero-price effect was not observed. When Eq. (2) holds, human behavior tend to deviate from irrational behavior assumed in traditional economics for large values of a irrespective of the value of a. Under the condition of $(a - 1)X = Z$, it is speculated that the discount is recognized as not the amount of the discounted price but the discount percentage from the initial price X, and thus affected largely by larger value of a.

In this manner, it has been indicated that the zero-price effect is not necessarily observable and holds under limited conditions. The value of price X that is provided for free and the price to be compared as an alternative of X that is expressed by a were important determinants of zero-price effect.

References

1. Ariely, D.: Predictably Irrational. Harper, London (2009)
2. Heyman, J., Ariely, D.: Effort for payment: a tale of two markets. Psychol. Sci. **15**(11), 787–793 (2004)
3. Ariely, D., Gneezy, U., Haruvy, E.: Social norms and the price of zero. Working paper, MIT, Boston, MA (2005)
4. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–291 (1979)
5. Slovic, P., Finucane, M., Peters, E., MacGregor, D.G.: The affect heuristic. In: Gilovich, T. (ed.) Heuristics and Biases: The Psychology of Intuitive Judgment, pp. 397–420. Cambridge University Press, New York (2002)
6. Slovic, P., Finucane, M., Peters, E., MacGregor, D.G.: Rational actors or rational fools: implications of the affect heuristic for behavioral economics. J. Socio Econ. **31**(4), 329–342 (2002)

Part III
Human-Machine Interactions and Tools Beyond

Culture in the Cockpit: Implications for CRM Training

Simon Cookson

Abstract Crew resource management (CRM) is an important airline training tool that was developed in the United States and has been used to train flight crews worldwide since the 1990s. Modern CRM programs cover a wide range of skill areas, including communication, interpersonal skills and decision-making. This paper describes the evolution of CRM and its underlying cultural assumptions. CRM has been criticized for being implicitly biased towards Western culture, and there have been calls for the development of different versions "culturally calibrated" to meet the needs of target participants around the world. This paper reviews research into national cultural differences, as well as airline organizational culture and pilot professional culture, and examines the implications for CRM training. This study is relevant to all international airline flight operations, especially those involving mixed-nationality crews.

Keywords Airline training · Crew resource management · Flight crew · National culture · Organizational culture · Professional culture

1 Introduction

Awareness has grown in recent decades that many airline accidents have been at least partly caused by cultural factors [1]. For historical reasons, cultural factors form part of crew resource management (CRM), a training methodology developed in the United States and now used by airlines around the world. This paper outlines the development of CRM and the cultural assumptions informing it. Culture is inherently difficult to define, but in the words of the Federal Aviation Administration it denotes "the norms, attitudes, values, and practices that members of a nation, organization, profession, or other group of people share" [2]. In line

S. Cookson (✉)
Aviation Management Department, J. F. Oberlin University,
3758 Tokiwa-Machi, Machida-shi, Tokyo 194-0294, Japan
e-mail: scookson@obirin.ac.jp

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_10

with this definition, it has been suggested that the study of cultural effects in aviation be divided into three categories: national, organizational and professional culture [3]. Considerable effort has been devoted to investigating national cultural differences, and this paper examines the implications for CRM training of four sets of research that have been applied to the aviation context. The paper also reviews several studies of the organizational culture of airlines and the professional culture of pilots. The overall goal is greater understanding of how cultural factors affect interpersonal interactions on the flight deck. This is an area of increasing relevance, given the sustained growth of international airlines based in the Middle East and Asia, especially China.

2 History of CRM

2.1 Origin in the 1970s

CRM is a method of training airline crews that developed in the United States at the end of the 1970s following several accidents involving American airlines. These accidents included: the 1972 crash of Eastern Air Lines Flight 401 in the Florida Everglades; the 1977 runway collision between KLM Flight 4805 and Pan Am Flight 1736 at Tenerife; and the 1978 crash of United Airlines Flight 173 near Portland. The last of these, in which the crew were so absorbed with a landing gear problem that they did not realize the fuel was running out, is often cited as the most important trigger for the creation of CRM [4]. Each of these accidents was complex and unique, but they all featured poorly functioning teams that combined overbearing captains with junior officers unable to clearly articulate their concerns.

In 1979, NASA organized a workshop for researchers and industry representatives to discuss the concept of flight deck resource management. In the opening presentation, chairman John Lauber reviewed relevant research: interviews conducted by NASA with airline pilots; a simulator study involving flight crews from an American airline; a study of 62 airline accidents; and a study of 250 jet transport incidents [5]. This seminal workshop addressed a very real problem, but it is important to note that most of the research focused on American flight crews. In other words, there was limited cultural diversity in the data underpinning the establishment of CRM.

2.2 Evolution Since the 1980s

Following the 1979 conference, CRM training was adopted by American airlines and subsequently spread worldwide. Many changes were made over the following decades, with the result that Maurino and Murray identified six generations in the

evolution of CRM [6]. One early development was a name change in the 1980s from Cockpit Resource Training to Crew Resource Training in order to emphasize team dynamics and interactions with personnel outside the flight deck, such as cabin crew, dispatchers and air traffic controllers (ATC). In the 1990s increased emphasis was placed on cross-cultural issues in the model developed at the University of Texas, based on organizational research conducted by Geert Hofstede. More recently, the latest manifestations of CRM have an explicit focus on managing threats and errors.

Despite these changes, the fundamental goals remain the same: to train crews in techniques that enable them to work as effective teams and avoid problematic behavior patterns identified by accident research. Modern CRM programs typically cover the following skill areas: communication/interpersonal skills; situation awareness; problem-solving/decision-making/judgment; leadership/followership; stress management; and critique [7]. In addition to aviation, CRM training has now spread to other high reliability industries such as healthcare, firefighting services, nuclear power generation, maritime and rail transport, and the offshore oil and gas industries [8].

2.3 *Criticisms of CRM*

In terms of its longevity and worldwide usage, CRM has undoubtedly been a success. However, two extensive reviews of studies that evaluated CRM training were unable to determine whether there was any impact on organizational safety [9, 10]. Proponents of CRM point to cases when lives were saved, most notably the 1989 crash landing of United Airlines Flight 232 at Sioux City. This aircraft's crew made a remarkable landing despite losing all flight control surfaces, with the captain later observing they would not have survived without the use of CRM techniques [11].[1]

Notwithstanding the successes, a number of limitations of CRM have been identified over the years. Criticism has included the following: early programs relied on organizational training activities with little relevance to airline operations [13]; as team dynamics became more prominent, CRM was labeled "a form of 'New Age' brainwashing aimed at achieving group harmony" [14]; the integration of CRM techniques with technical flying skill training led to increased proceduralization, reducing the focus on leadership and team building skills [15]; and the expanded range of program content obfuscated the overall goals [16].

A further persistent criticism is that CRM is implicitly biased towards Western culture [17]. In the 1980s, there was a common belief that CRM could, with minor changes, be adapted for use by any airline in the world, but this view of

[1]An important feature of this crew, enhancing their decision-making ability under stress, was that they were very experienced and had flown together before. The captain had 29,967 flight hours with United Airlines, while the first and second officers each had 15,000–20,000 flight hours. An off-duty check airman who assisted them had 23,000 flight hours [12].

"culture-free CRM" has now been discredited [18]. Problems associated with implementation in other countries, such as the translation of specialized vocabulary and the use of feedback questionnaires on personality or working styles, were highlighted by Johnston. He cautioned that the underlying causes of aviation accidents may vary by region, and that detailed research was required before CRM was applied worldwide [19].

Criticism continues to this day. Analyzing survey data from Taiwan and accidents involving Asian airlines, Jing and Batteau concluded that CRM is underpinned by cultural values alien to Chinese society, and moreover these differences are compounded by the increasing technological complexity of modern airliners [20]. Alongside these criticisms, there have been repeated calls for the development of different versions of CRM for different regions of the world, "culturally calibrated" to the needs of target participants [21–23].

3 Differences in National Culture

3.1 Hall's "Hidden Culture"

In a career spanning most of the twentieth century, the anthropologist Edward T. Hall identified numerous ways in which culture informs human behavior. He noted that people remain largely unaware of this "hidden culture" because it operates below the level of consciousness. Discussing the relevance of Hall's research to CRM, Hisam and Hampton provided examples of how individuals from various countries act differently. American pilots, for example, soon start addressing each other using first names whereas Europeans tend to remain more formal [24]. The following paragraphs describe specific implications for CRM arising from three of Hall's concepts.

High-Context and Low-Context. Hall contrasted high-context cultures (e.g. Japan), which have deep relationships and share information using simple messages rich in meaning, with low-context cultures (e.g. America), where people are not bonded so tightly and there is less distinction between insiders and outsiders [25]. He cautioned that meetings between the two cultures could present problems. Hall applied these concepts to communication, stating that high-context communications are fast and efficient because pre-programmed information is in receivers and settings, with minimal information in messages. By contrast, low-context communications encode most of the information in messages, with very little in the internal or external contexts.

The concept of a high- or low-context *culture* is problematic because Hall stated that a *person* may have both high- and low-context aspects depending on the situation. Scollon, Scollon and Jones resolved this dilemma by proposing the concept be applied not to entire national groups, but instead to particular speech events or situations [26]. The use of standard phraseology—a set of pre-fabricated

phrases for typical flight situations—between pilots and ATC is an example of high-context communication. Considerable time must be spent training operators to use this system, but the payoff is the quick and efficient exchange of information.

It may be hypothesized that people who favor low-context communications require more extensive programming to use standard phraseology effectively. There is anecdotal evidence to support this idea. In an analysis of a radio exchange reported by Kim and Elder, experienced Korean pilots and controllers described an American pilot's "verbosity and inappropriate word choice… when phraseology would have sufficed" as "typical of native English-speaking aviation personnel" [27].

Monochronic and Polychronic Time. Hall differentiated between monochronic people, who like doing one thing at a time, and polychronic people, who prefer doing several different activities at once [28]. Interactions between the categories may again be problematic, with polychronic behavior liable to disorientate monochronic people. This has implications for flight crew composition: for example, an American (i.e. monochronic) captain and a Latin American (i.e. polychronic) first officer may approach the same set of tasks differently. In the context of international business interactions, Hall suggested that judicious office design could ameliorate problems, but at present this is not a viable option on confined flight decks [29].

Hisam and Hampton noted that monochronic people are vulnerable to interruptions [30]. In airline operations it is commonplace for disturbances, such as unexpected ATC calls, to put task completion at risk. Citing dozens of incidents in which American crews experienced disturbances, Loukopoulos, Dismukes and Barshi stressed the importance of CRM techniques for managing workload effectively [31]. Techniques for dealing with interruptions seem to be especially important for monochronic personnel, but there is no research evidence to support this. Instruments for measuring polychronicity have, though, been applied to other organizational contexts [32].

Action Chains. The action chain is a sequence of actions that two or more individuals carry out in order to complete a task. Action chains play a vital role in the cockpit both in the formulaic exchanges between pilots and ATC, and also in the form of standard operating procedures (SOPs), or written descriptions of tasks for each flight phase. Hall noted that monochronic people tend to focus on completing tasks, while polychronic people place more emphasis on maintaining good human relations [33].

Misunderstandings may occur when monochronic and polychronic people work together on the same action chain, as illustrated in the 1990 crash of Avianca Flight 052 near New York. Shortly before the crash, one of the Columbian flight crew commented that an American air traffic controller was *angry*. In his analysis of the accident, Helmreich interpreted this comment as indicating a failure to focus on the task of safely landing the plane [34]. However, a polychronic interpretation suggests the crew member was expressing concern about the human relations involved in the situation, rather than neglecting the task of landing. CRM training in interpersonal skills should at the very least raise awareness of these different cultural perspectives.

Criticism of Hall's Assumptions. Hutchins, Holder and Pérez stated that much of Hall's work was "based on rather dated and oversimplified models of the role of cultural and linguistic knowledge in thought" [35]. They warned against regarding culture as a set of "traits" exhibited by all the members of a group and stressed the importance of cultural variability within social groups. Notwithstanding these comments, which may equally be directed at many studies of national culture, researchers in aviation, intercultural communication and organizational studies continue to draw on Hall's concepts, as noted above.

3.2 Hofstede's Cultural Dimensions

The social psychologist Geert Hofstede investigated differences in national culture in a research program starting in the 1960s. Applying factor analysis to data aggregated from surveys of 88,000 workers in 66 countries, four cultural dimensions were identified and numerical values calculated for each country on each dimension [36, 37]. With easy-to-comprehend national scores, seemingly validated by the huge amount of input data, Hofstede's work has been influential in many fields including aviation. Indeed, it has been described as the third leg of the "three-legged stool upon which broad, systematic-oriented aviation safety and efficiency endeavors rest" [38].

Hofstede's Dimensions in Aviation. Using test items and methodology adapted from Hofstede, a team led by Robert Helmreich at the University of Texas conducted surveys of the attitudes of more than 8,000 airline crew in over 20 countries. They found a strong correlation with Hofstede's results for the dimensions of power distance and individualism-collectivism, with a weaker correlation for uncertainty avoidance [39]. This research fed directly into fourth-generation CRM programs in the 1990s.

Hofstede's cultural dimensions were used by Helmreich to analyze the actions of the Columbian flight crew in the aforementioned crash of Avianca Flight 052. Since Columbia scores highly in power distance—the extent to which less powerful members of organizations accept unequal power distribution—Helmreich posited that the first officer and flight engineer were reluctant to suggest alternative courses of action to the captain. Columbia is also strongly collectivist, with people defining themselves through social groups rather than as individuals, so the flight crew may have been reluctant to declare an emergency and push themselves ahead of other crews they perceived to be in similar straits. In addition, Columbia scores highly in uncertainty avoidance and therefore the crew may have preferred to continue with the initial flight plan, rather than face the ambiguity of discussing possible alternate airports [40].

Hofstede's model appears to be of particular value in making sense of accidents featuring junior officers unable to voice concerns to experienced captains, such as the Avianca 052 crash and the accidents that led to the inception of CRM. It should be noted, though, that cockpit operations typically involve dyadic or triadic

interactions, whereas Hofstede's research was based on large-scale surveys. As Hofstede himself has observed, claims made about national cultural characteristics are "common trends, but individuals may differ from them" [41].

Criticism of Hofstede's Model. Outside of aviation the limitations of Hofstede's research have been widely documented. McSweeney challenged several underlying assumptions, such as the use of limited sets of survey respondents to represent national populations, and the identification of cultural dimensions through analysis of questionnaire responses [42]. Analyzing the political subtext of Hofstede's methodology, Ailon cautioned against an uncritical application of the dimensions to other cultures [43]. In a study of multicultural work teams, Aritz and Walker raised several questions: whether Hofstede's data may be reliably applied to countries not covered by the initial surveys (such as China); whether the data are applicable to other workforces or national populations, given that the participants were sales managers and engineers; and what insights the dimensions offer into everyday intercultural interactions, such as team decision-making [44].

Within aviation, Hofstede's model was criticized by Hutchins, Holder and Pérez on numerous counts, including: the absence of data regarding intra-country variability in the dimensions; the methodology used to determine the probes; the problem of translation effects in cross-cultural surveys; and the fundamental issue of how survey responses relate to cockpit operations [45].

Hofstede responded to some of the criticism with further surveys that included East Asian participants, and with investigations of organizational culture and cultural differences within a single country. Two new cultural dimensions were identified, but the underlying methodology remained unchanged [46].

3.3 Trompenaars' Cultural Dimensions

During the 1980s and 1990s the management consultant Fons Trompenaars conducted large-scale surveys of cultural diversity in companies operating in 50 countries. From this data he developed a model with seven cultural dimensions describing relationships with people, time and the environment [47]. This model has not been incorporated into CRM programs but it has been used in a correlational study of airline accident rates and attitudes to authority [48].

Trompenaars' use of survey data to identify cultural differences is open to similar criticisms to those leveled at Hofstede's work, but a point of difference is that Trompenaars draws on business anecdotes to contextualize the dimensions in interpersonal interactions. However, Trompenaars' cultural dimensions are not conceptually distinct, and Hofstede claimed that only two could be confirmed statistically [49]. There is also overlap with other models so that, for example, Trompenaars' specific-diffuse dimension corresponds closely to Hall's concept of high- and low-context. The following paragraphs discuss implications of two of the dimensions for CRM.

Individualism-Communitarianism. Trompenaars' individualism-communitarianism is similar to Hofstede's individualism-collectivism, with both measuring the extent to which people regard themselves as individuals or part of groups. To illustrate national differences in this dimension, Trompenaars and Hampden-Turner described a "critical incident" in a factory owned by an American multinational where a Japanese worker made a "serious error" causing the loss of a production batch. After the work group accepted responsibility, the factory director—to the amazement of a Western investigator—did not try to identify or punish the errant worker because in Japanese culture the shame of letting the group down was considered punishment enough [50].

Reluctance of individuals in communitarian cultures to openly accept responsibility for errors may impact on two aspects of the error management training that forms an important part of modern CRM programs. Firstly, on the flight deck individual crew members are trained to assertively communicate problems, including errors. Secondly, inside an organization it is essential for employees to report errors as part of an effective "safety culture". It is clear that attitudes to error vary significantly, which may necessitate different CRM solutions for different cultures.

Achievement-Ascription. In achievement cultures (e.g. the United States) people are accorded status based on work performance and recent accomplishments. By contrast, in ascription cultures (e.g. Japan or China) status is accorded based on age, kinship, gender, connections and educational record. Status is thus perceived differently in different cultures, which may affect leadership and communication on the flight deck.

Status is integral to a person's authority. One of the assumptions of CRM leadership training is that captains can learn how to establish an appropriate level of authority. Ginnett described three techniques used by effective captains: establish competence in the pre-flight briefing; disavow perfection in order to allow other crew members to take responsibility; and engage the crew during the briefing and group formation process [51]. These techniques, based on NASA research with American flight crews, may prove effective in achievement cultures but less so in ascription cultures where status is not related to work performance.

If a large difference exists between the status of the captain and junior officers, then a steep authority gradient may result. This can hinder communication and decision-making, and has been identified as a causal factor in accidents such as the 1977 collision at Tenerife. CRM programs teach polite assertiveness techniques to help junior officers overcome this problem, but these may not be effective in ascription cultures where status derives from intrinsic characteristics such as age and gender.

3.4 Jing's Differentiated Order Model

In the 1990s, using a modified version of a questionnaire developed by Helmreich, Professor Hung-Sying Jing surveyed approximately 1,000 pilots and managers at

airlines in Taiwan, including a significant number of foreign pilots. The results highlighted differences between Chinese and foreign pilots in attitudes to interpersonal relations and authority. Believing these differences could not be adequately explained by uni-dimensional concepts such as power distance, Jing developed a framework to account for interpersonal relations and authority in Chinese culture [52].

Drawing on research by the scholar Fei Xiao-Tung, Jing outlined a differentiated order model to describe how the Chinese categorize people around them. This model has four levels of intimacy: kin, acquaintance, fellow and alien. According to the model, Chinese pilots consider that: close family are kin; other Chinese pilots are acquaintances; other Chinese workers in the same company are fellows; and foreign workers in the same company are aliens. The structure is not fixed and individuals can change level, for example by marriage or a serious falling out. To this model of interpersonal relations, Jing added a description of the Chinese concept of authoritarianism, which is dominated by the father-son relationship.

Jing's Model in Aviation. This differentiated order model has been used to analyze accidents involving Asian airlines, such as the 1995 crash of a TransAsia Airways ATR72 aircraft in Taiwan. Immediately before the crash, the captain (pilot not flying) was talking to a cabin attendant in the cockpit, which disrupted communications with ATC and distracted him from monitoring the aircraft's status. Jing suggested that the captain regarded the cabin attendant as an acquaintance but considered the air traffic controller to be a stranger, adding that "Every Chinese person would be inclined by instinct to attend to a friend first, not the stranger" [53]. Western pilots may consider such behavior to be a blatant dereliction of duty, but Jing's work highlights the impact that cultural factors can have on cockpit interactions. Interestingly, it echoes Hall's description of the emphasis placed by polychronic people on personal relations.

The circumstances of this accident were unusual: it occurred on New Year's Eve; the plane was carrying no passengers; and the captain was *junior* to the first officer in terms of their previous air force service. With regard to training, this underlines the importance—even in unusual circumstances—of adhering to rules, such as the sterile cockpit rule, which prohibits non-essential speech when flying below 10,000 ft. As noted by Hisam and Hampton, different cultures have differing interpretations of a sterile cockpit, so CRM training should be tailored accordingly [54].

Finally, regarding flight procedures, Jing and Batteau observed that Chinese pilots are conditioned by the non-linear ideographic Chinese language and therefore have difficulty following sequential SOPs. They see this as one manifestation of a systematic problem whereby Chinese pilots are not culturally programmed to use either commercial aircraft or an air transport system largely designed by Westerners [55].

4 Organizational and Professional Culture

The research strands described in the preceding section focused on national culture, but in the case of Helmreich's team at the University of Texas research extended into airline organizational culture and pilot professional culture. Organizational culture consists of observable behavior and items such as uniforms, logos and documents, as well as the beliefs, values and assumptions shared by members of the organization. Drawing on James Reason's research into accident analysis, Helmreich and Merritt stressed the importance of organizational culture to safety and cited several accidents that featured organizational culture as one of the causal factors, including the 1991 crash of Continental Express Flight 2574 following a failure in maintenance procedures [56]. The University of Texas surveys of flight crew attitudes identified positive aspects of pilot professional culture, such as high levels of motivation, and negative characteristics, such as a feeling of invulnerability [57].

In Norway Mjøs conducted a survey of pilots at three airlines and received 242 usable responses [58]. The variables included cultural indices (based on Hofstede's four original dimensions), social climate, barriers to communication, and operational problems experienced in the previous year. This survey identified differences between the airlines, with the pilots of one company—almost all from a military background—being more experienced and scoring higher on power distance and masculinity. The pilots of all three airlines had higher mean scores for individuality and masculinity than the national scores reported by Hofstede, indicating that the cultural dimensions for a *professional* group within a country may differ from the *national* characteristics.[2] This led Mjøs to caution against applying national cultural dimension data to research comparing aviation safety records in different countries.

Hutchins, Nomura and Holder reported on an ethnographic study that investigated the impact of culture on cockpit communication and interaction at three airlines in the Asia-Pacific region [59]. The study included flight deck and simulator observations, as well as interviews with airline personnel. An interesting aspect of this research is that it identified specific differences in cockpit practice (e.g. how checklists and charts were actually used) between airlines in different countries. However, since only a limited number of airlines were studied it is not clear to what extent the variation was due to *national* culture as opposed to *organizational* culture. For instance, Japanese pilots in the study annotated their charts whereas pilots from New Zealand were not allowed to do so, but it is possible that other New Zealand airlines permit chart annotation. Difficulties posed by this form of research include the need to gain access to flight decks and the requirement for expertise in a range of fields such as piloting, human factors, anthropology, language and culture.

[2]It should be noted that this survey was conducted decades after Hofstede's data collection.

5 Conclusion

CRM training has proved enduring and successful, but its validity outside of Western cultures has been questioned and there have been repeated calls to develop "culturally calibrated" versions of CRM for different regions of the world. Hofstede's research has been incorporated into CRM programs, and it clearly has value for training Western pilots and analyzing certain types of accident. However, both the model and its applicability to aviation have been criticized. The three other models of national culture examined in this paper have not been incorporated into CRM, but each offers valuable insights into national characteristics, especially regarding differences between Americans and East Asians. Elements of each model could be incorporated into a modular CRM package, with the important proviso that cultural variability exists within national groups, as many commentators have noted.

This paper highlights the value of continued research into airline organizational culture and pilot professional culture, areas which have hitherto been under-researched. In addition, the studies presented at the 1979 NASA workshop that launched CRM were extensive, but were largely limited to American crews and are now dated. There is a pressing need for a major research effort on a similar scale but covering airlines from all major regions of the world. This should adopt a mixed methods approach to collect both quantitative and qualitative data. As well as incident and accident studies, it would incorporate modern ethnographic techniques including interviews and observations of flights and simulator training.

References

1. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing, Farnham (1998)
2. Federal Aviation Administration: The Interfaces between Flightcrews and Modern Flight Deck Systems, p. 117. FAA, Washington, D.C (1996)
3. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing, Farnham (1998)
4. Ginnett, C.: Crews as groups: Their formation and their leadership. In: Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.) Cockpit Resource Management, pp. 71–98. Academic Press, San Diego (1993)
5. Cooper, G.E., White, M.D., Lauber, J.K.: Research management on the flight deck. In: NASA/Industry Workshop in San Francisco, California, 26–28 June 1979. NASA Conference Publication 2120. NASA Ames Research Center, Moffett Field, California (1980)
6. Maurino, D.E., Murray, P.S.: Crew resource management. In: Wise, J.A., Hopkin, V.D., Garland, D.J. (eds.) Handbook of Aviation Human Factors (2nd edition), pp. 10-1–10-20. CRC Press, Boca Raton, Florida (2010)
7. Civil Aviation Authority: Flight Crew Training: Cockpit Resource Management (CRM) and Line-Oriented Flight Training (LOFT). CAP 720. CAA, West Sussex (2002)
8. Flin, R., O'Connor, P., Crichton, M.: Safety at the Sharp End: A Guide to Non-Technical Skills. Ashgate Publishing, Aldershot (2008)

9. Salas, E., Burke, C.S., Bowers, C.A., Wilson, K.A.: Team training in the skies: does crew resource management (CRM) training work? Hum. Factors **43**(4), 641–674 (2001)
10. Salas, E., Wilson, K.A., Burke, C.S., Wightman, D.C.: Does crew resource management training work? An update, an extension, and some critical needs. Hum. Factors **48**(2), 392–412 (2006)
11. Haynes, A.: The Crash of United Flight 232. Presentation at NASA Ames Research Center, Dryden Flight Research Facility, California (1991)
12. National Transportation Safety Board: Aircraft Accident Report: United Airlines Flight 232, McDonnell Douglas DC-10-10, Sioux Gateway Airport, Sioux City, Iowa, July 19, 1989. NTSB/AAR-90/06. NTSB, Washington, D.C. (1990)
13. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing, Farnham (1998)
14. Ibid., p. 146
15. Helmreich, R.L., Wilhelm, J.A., Klinect, J.R., Merritt, A.C.: Culture, error and crew resource management. In: Salas, E., Bowers, C.A., Edens, E. (eds.) Improving Teamwork in Organizations, pp. 305–331. Erlbaum, Hillsdale, New Jersey (2001)
16. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing, Farnham (1998)
17. Engle, M.: Culture in the cockpit: CRM in a multicultural world. J. Air Transp. World Wide **5** (1), 107–114 (2000)
18. Maurino, D.E., Murray, P.S.: Crew resource management. In: Wise, J.A., Hopkin, V.D., Garland, D.J. (eds.) Handbook of Aviation Human Factors, pp. 10-1–10-20, 2nd edn. CRC Press, Boca Raton, Florida (2010)
19. Johnston, N.: CRM: cross-cultural perspectives. In: Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.) Cockpit Resource Management, pp. 367–398. Academic Press, San Diego (1993)
20. Jing, H.-S., Batteau, A.: The Dragon in the Cockpit: How Aviation Concepts Conflict with Chinese Value Systems. Ashgate Publishing, Farnham (2015)
21. Maurino, D.E., Murray, P.S.: Crew resource management. In: Wise, J.A., Hopkin, V.D., Garland, D.J. (eds.) Handbook of Aviation Human Factors, pp. 10-1–10-20, 2nd edn. CRC Press, Boca Raton, Florida (2010)
22. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing, Farnham (1998)
23. Hisam, T., Hampton, S.: Toward an international model of crew resource management: the cultural implications. J. Aviat. Aerosp. Educ. Res. **7**(1), 6–19 (1996)
24. Ibid
25. Hall, E.T.: Beyond Culture. Anchor Books, New York (1976)
26. Scollon, R., Scollon, S.W., Jones, R.H.: Intercultural Communication: A Discourse Approach, 3rd edn. Wiley, Chichester, West Sussex (2012)
27. Kim, H., Elder, C.: Understanding aviation english as a lingua franca: perceptions of korean aviation personnel. Aust. Rev. Appl. Linguist. **32**(3), 23.1–23.17 (2009) (p. 23.11)
28. Hall, E.T.: The Dance of Life. Anchor Books, New York (1983)
29. Hall, E.T.: The Hidden Dimension. Anchor Books, New York (1969)
30. Hisam, T., Hampton, S.: Toward an international model of crew resource management: the cultural implications. J. Aviat. Aerosp. Educ. Res. **7**(1), 6–19 (1996)
31. Loukopoulos, L.D., Dismukes, R.K., Barshi, I.: The Multitasking Myth: Handling Complexity in Real-World Operations. Ashgate Publishing, Farnham (2009)
32. Bluedorn, A.C.: The Human Organization of Time: Temporal Realities and Experience. Stanford Business Books, Stanford, CA (2002)
33. Hall, E.T.: Beyond Culture. Anchor Books, New York (1976)
34. Helmreich, R.L.: Anatomy of a system accident: the crash of avianca flight 052. Int. J. Aviat. Psychol. **4**(3), 265–284 (1994)
35. Hutchins, E., Holder, B.E., Pérez, R.A.: Culture and flight deck operations. Paper prepared for Boeing by University of California San Diego, Research Agreement 22-5003 (2002), p. 26

36. Hofstede, G.: Culture's Consequences: International Differences in Work-Related Values. Sage, Beverley Hills (1980)
37. Hofstede, G.: The Cultural Relativity of Organizational Practices and Theories. J. Int. Bus. Stud. **14**(2), 75–89 (1983)
38. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences, p. xvii. Ashgate Publishing, Farnham (1998)
39. Helmreich, R.L., Wilhelm, J.A., Klinect, J.R., Merritt, A.C.: Culture, error and crew resource management. In: Salas, E., Bowers, C.A., Edens, E. (eds.) Improving Teamwork in Organizations, pp. 305–331. Erlbaum, Hillsdale, New Jersey (2001)
40. Helmreich, R.L.: Anatomy of a system accident: the crash of avianca flight 052. Int. J. Aviat. Psychol. **4**(3), 265–284 (1994)
41. Hofstede, G.: Business cultures. UNESCO Courier, **47**(4), 12–16 (1994) (p. 16)
42. McSweeney, B.: Hofstede's model of national cultural differences and their consequences: a Triumph of faith—a failure of analysis. Hum. Relat. **55**(1), 89–117 (2002)
43. Ailon, G.: Mirror, mirror on the wall: culture's consequences in a value test of its own design. Acad. Manag. Rev. **33**(4), 885–904 (2008)
44. Aritz, J., Walker, R.C.: Cognitive organization and identity maintenance in multicultural teams: a discourse analysis of decision-making meetings. J. Bus. Commun. **47**(1), 20–41 (2010)
45. Hutchins, E., Holder, B.E., Pérez, R.A.: Culture and flight deck operations. Paper prepared for Boeing by University of California San Diego, Research Agreement 22-5003 (2002)
46. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and Organizations: Software of the Mind, 3rd edn. McGraw-Hill, New York (2010)
47. Trompenaars, F., Hampden-Turner, C.: Riding the Waves of Culture: Understanding Cultural Diversity in Business, 2nd edn. Nicholas Brealey Publishing Ltd., London (1997)
48. Jing, H.-S., Lu, C.J., Peng, S.-J.: Culture, authoritarianism and commercial aircraft accidents. Hum. Factors Aerosp. Saf. **1**(4), 341–359 (2001)
49. Hofstede, G.: Riding the waves of commerce: a test of trompenaars' "model" of national culture differences. Int. J. Intercult. Relat. **20**(2), 189–198 (1996)
50. Trompenaars, F., Hampden-Turner, C.: Riding the Waves of Culture: Understanding Cultural Diversity in Business, 2nd edn, pp. 64–65. Nicholas Brealey Publishing Ltd., London (1997)
51. Ginnett, C.: Crews as groups: their formation and their leadership. In: Wiener, E.L., Kanki, B. G., Helmreich, R.L. (eds.) Cockpit Resource Management, pp. 71–98. Academic Press, San Diego (1993)
52. Jing, H.-S., Batteau, A.: The Dragon in the Cockpit: How Aviation Concepts Conflict with Chinese Value Systems. Ashgate Publishing, Farnham (2015)
53. Ibid. p. 30
54. Hisam, T., Hampton, S.: Toward an international model of crew resource management: the cultural implications. J. Aviat. Aerosp. Educ. Res. **7**(1), 6–19 (1996)
55. Jing, H.-S., Batteau, A.: The Dragon in the Cockpit: How Aviation Concepts Conflict with Chinese Value Systems. Ashgate Publishing, Farnham (2015)
56. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing, Farnham (1998)
57. Helmreich, R.L.: Culture and error in space: implications from analog environments. Aviat. Space Environ. Med. **71**(9–11), 133–139 (2000)
58. Mjøs, K.: Basic cultural elements affecting the team function on the flight deck. Int. J. Aviat. Psychol. **14**(2), 151–169 (2004)
59. Hutchins, E., Nomura, S., Holder, B.: The ecology of language practices in worldwide airline flight deck operations: The case of Japanese airlines. In: International Conference on Human-Computer Interaction in Aeronautics, Seattle, WA (2006)

Irrational Behavior in Adaptation: Difference of Adaptation Process to Comfort and Discomfort Stimulus When Presented All Together or Intermittently

Atsuo Murata and Tomoko Nakamura

Abstract It is generally assumed that the interruption (break) of positive (comfortable) or negative (uncomfortable) experiences disrupt adaptation and consequently intensify the subsequent positive or negative experience. Therefore, it can be speculated that interrupting a consumption experience makes comfortable experiences more delightful and interruption of uncomfortable and painful experience makes this more irritating. This study aimed at verifying this irrational behavior in adaptation process. The comfortable and uncomfortable experiences were an experience of relaxing massage and that of immersing one's hand in the cold water. It was explored how the interruptive experience of positive (comfortable) or negative (uncomfortable) stimulus disrupt adaptation and intensify subsequent experience as compared with the continuous experience of the same stimuli (experience of the stimulus by bulk). This was investigated as a function of experience time A (comfort stimulus: 300 and 600 s, discomfort stimulus: 50 and 150 s) and duration of interruption B (30, 60, and 90 s for both comfort and discomfort).

Keywords Irrational behavior · Behavioral economics · Adaptation · Comfort · Discomfort · Continuous exposure · Interruptive exposure · Break between exposures

A. Murata (✉) · T. Nakamura
Department of Intelligent Mechanical Systems, Graduate School of Natural Science and Technology, Okayama University, 3-1-1, Tsushimanaka, Kita-Ward, Okayama, Japan
e-mail: murata@iims.sys.okayama-u.ac.jp

T. Nakamura
e-mail: en422852@s.okayama-u.ac.jp

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_11

1 Introduction

We have a property to adapt to a variety of positive and negative experiences such as repeated consumption of preferred ice cream (Kahneman and Snell [1]) and increases in income (Easterlin [2]). The subjective rating on comfort obtained through the experience seems to weaken over time as a result of complex physiological, perceptual, and cognitive mechanisms. Ariely and Zauberman [3] showed that adaptation creates a downward trend of comfort or discomfort rating in positive experiences and an upward trend in negative experiences as well as the overall reduction of comfort or discomfort.

On the basis of the discussion above, it might be expected that disrupting experience with a break (insert of intermission during experience of comfort or discomfort) should reduce adaptation and prolong the period for adaptation. Such a reduction in adaptation caused by an intermission (or break) is postulated to intensify comfort or discomfort followed by an intermission (break) as compared with an un-intermitted experience of comfort or discomfort. Ariely [4] showed that consumption experiences will become more intense when breaks are inserted than consumption was experienced by bulk. This reasoning might give us a lesson that we should break up experiencing comfort but not experiencing discomfort. The experience may need to be conducted without intermission, while comfortable events such as shopping are best experienced when intermitted than when conducted by bulk.

In this manner, happiness does not last for longer period under the purchasing spree strategy than under the intermittent purchasing strategy (Ariely [4]). Although one's happiness or comfort becomes high immediately after purchasing spree, happiness will soon decrease over time as one's purchases lose their novelty, and comfort as a whole will not be so satisfactory. On the other hand, happiness or comfort will be revitalized due to the intermittent purchasing, though the intermittent strategy will not reach the same level of the initial happiness as the purchasing spree strategy. Ariery [4] concluded that the positive or comfortable stimulus should be presented (experienced) intermittently, whereas the negative or uncomfortable stimulus should be presented not intermittently but without a break (see Fig. 1). However, few studies empirically verify this hypothesis.

It is generally assumed that the interruption (break) of positive (comfortable) or negative (uncomfortable) experiences disrupt adaptation and consequently intensify the subsequent positive or negative experience. Therefore, it can be speculated that interrupting a consumption experience makes comfortable experiences more delightful and interruption of uncomfortable and painful experience makes this more irritating.

This study aimed at verifying this irrational behavior in adaptation process. The comfortable and uncomfortable experiences were an experience of relaxing massage and that of immersing one's hand in the cold water. It was explored how the intermittent experience of positive (comfortable) or negative (uncomfortable) stimulus disrupt adaptation and intensify subsequent experience as compared with the continuous experience of the same stimuli. This was investigated as a function

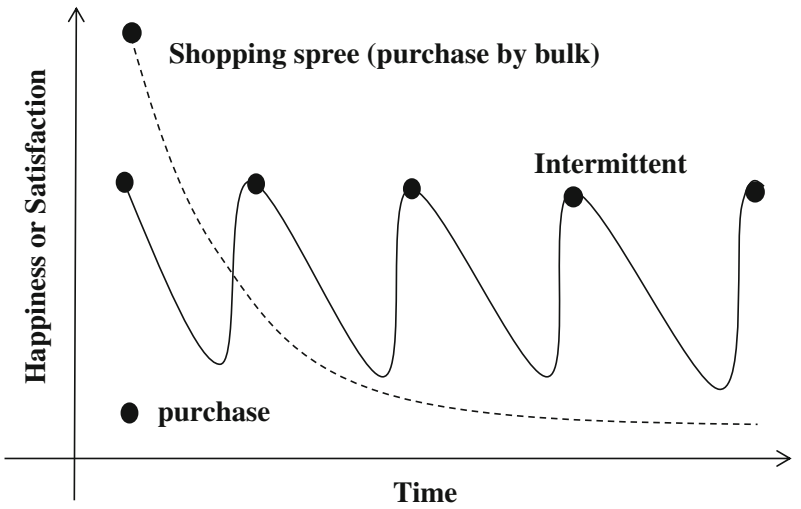


Fig. 1 Hypothesized change of happiness or satisfaction over time for intermittent and purchase by bulk conditions

of experience time A (comfort stimuli: 300 and 600 s, discomfort stimuli: 50 and 150 s), duration of interruption B (comfort stimuli: 30, 60, and 90 s; discomfort stimuli: 30, 60, and 90 s), and whether the experienced stimulus is comfortable (positive) or uncomfortable (negative) one.

2 Method

2.1 Participants

A total of eleven male participants from 22 to 24 years old at Faculty of Engineering, Okayama University took part in the experiment. The limitation of answering time was not imposed on the participant. All participants had no knowledge on psychology and behavioral economics.

2.2 Task and Design

The experimental factors were experience time *A* (comfort stimuli: 300 and 600 s, discomfort stimuli: 50 and 150 s) and duration of interruption *B* (comfort stimuli: 30, 60, and 90 s; discomfort stimuli: 30, 60, and 90 s). Both were within-subjects factors. The presentation of stimuli (intermittent presentation and presentation without intermission (break)) was also a within-subjects factor.

The comfortable and uncomfortable experiences were an experience of relaxing massage and that of immersing one's hand in cold water, respectively. In the comfortable experience, the cushion of massager was put on a sofa so that this was stuck fast to a low back of a participant. In the cold immersion experience, the temperature within an experimental chamber was set to 20 °C so that the participant felt chilly. A bowl filled with cold water adjusted to 5 °C was prepared. Using a preliminary experiment, it was confirmed that almost all of participants felt uncomfortable under this condition.

When *A* and *B* were, for example, 50 and 30 s, respectively, the intermittent presentation and the presentation without break conditions were experienced as follows. Under the condition of presentation without break, the participant was required to continuously immerse his hand in cold water for 50 s. Under the intermission condition, the participant immersed his hand for 10 s, and the experience was intermitted for 30 s. This cycle was repeated five times.

2.3 Procedure

Under the condition of continuous experience (experience of stimulus without break), participants were required to continuously experience a positive or negative stimulus for the predetermined time *A*. Under the condition of intermittent experience (intermittent condition), the experience time *A* was experienced intermittently for *A*/5 over five times with a break of *B* between experiences of the stimulus. Using a 5-point scale (positive experience: 5 = very comfortable, 1 = no feeling of comfort; negative experience: 5 = very uncomfortable and painful, 1 = no uncomfortable feeling), the subjective feeling was reported every 30 s for both intermittent and continuous experiences.

Under the condition of continuous experience, the subjective rating was started immediately after the pre-determined presentation of stimulus. The subjective rating was conducted every 30 s until the rating was evaluated as 1. Under the intermission condition, the subjective rating was started immediately after the pre-determined presentation of stimulus. The rating was conducted every 30 s until the 2nd experience, and the rating was conducted immediately after the second experience until the 3rd experience began. This cycle was repeated until the 5th experience was over. After the 5th experience, the rating was repeated every 30 s until the rating was evaluated as 1. The subjective rating was conducted every 30 s so that we can track how the initial feeling of experiencing comfort or discomfort stimulus changes and disappears over time.

The order of continuous and intermittent conditions was counterbalanced across the participants for both comfort and discomfort experience. When conducting the experiment to explore the response to comfort (positive) experience, it was checked whether the participant actually feel comfortable during the experience of relaxing massage. When the participant felt no comfort, he was excluded from the

experiment. When the participant reported that he could no longer endure the experiment (immersing one's hand in cold water), the experiment was canceled and the data of corresponding participant was excluded from the analysis.

3 Results

In Fig. 2, the change of comfort over time is plotted when the duration per trial under intermittent condition was equal to 60 s. Figure 3 shows the change of comfort over time when the duration per trial under intermittent condition was 120 s.

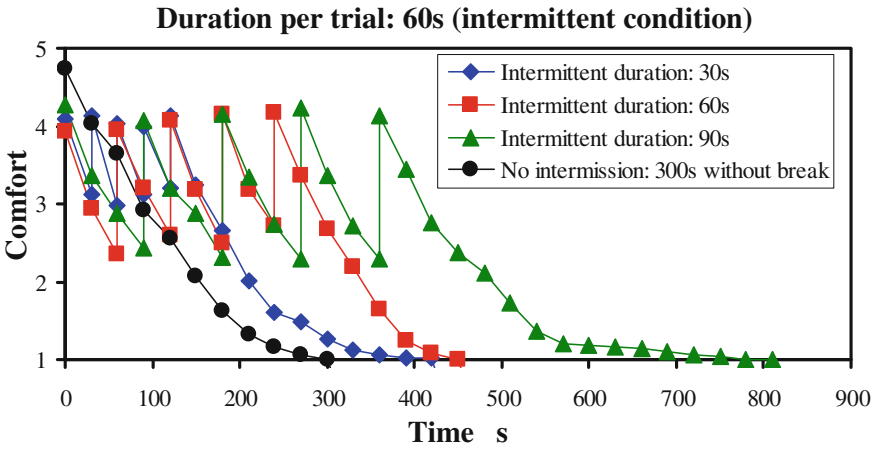


Fig. 2 Change of comfort over time (duration per trial under intermittent condition: 60 s)

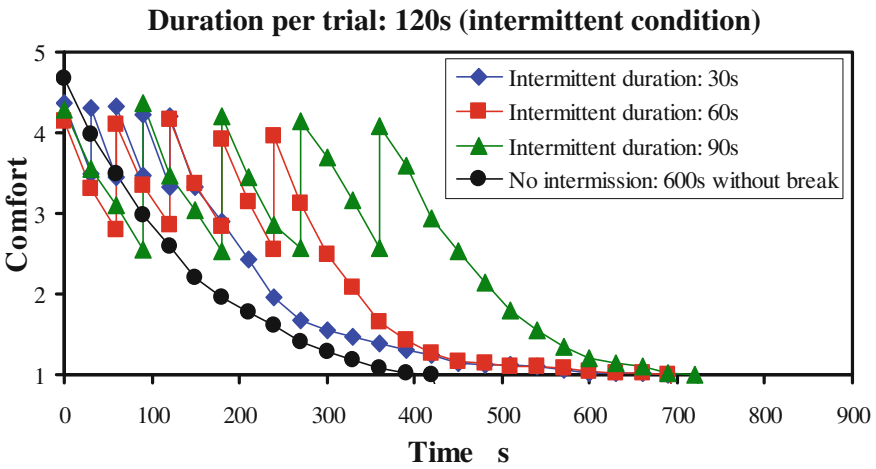


Fig. 3 Change of comfort over time (duration per trial under intermittent condition: 120 s)

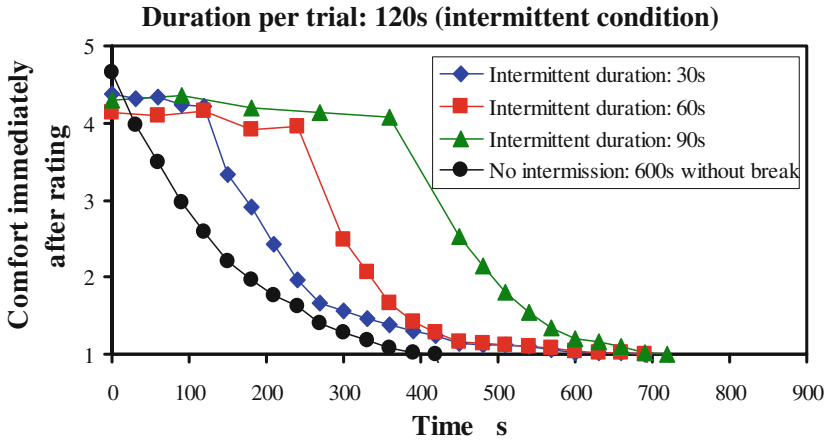


Fig. 4 Change of comfort immediately after rating over time (duration per trial under intermittent condition: 120 s)

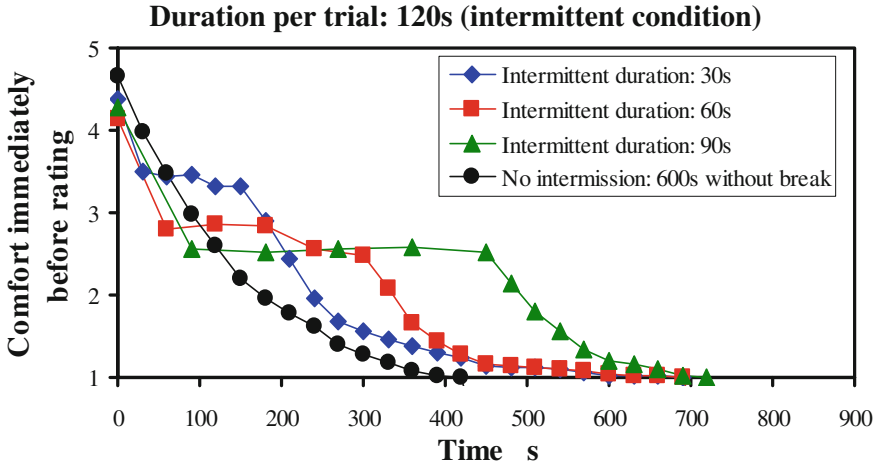


Fig. 5 Change of comfort immediately before rating over time (duration per trial under intermittent condition: 120 s)

As it is difficult to understand how the intermittent experience affect the comfortable feeling from Figs. 2 and 3, the changes of comfort immediately after the rating of experience of stimulus over time (duration per trial under intermittent condition: 120 s) were plotted as Fig. 4. Figure 5 plotted the change of comfort immediately before rating of experience of stimulus over time (duration per trial under intermittent condition: 120 s).

Figure 6 shows the change of discomfort over time (duration per trial under intermittent condition: 30 s). In Fig. 7, the change of discomfort over time is plotted

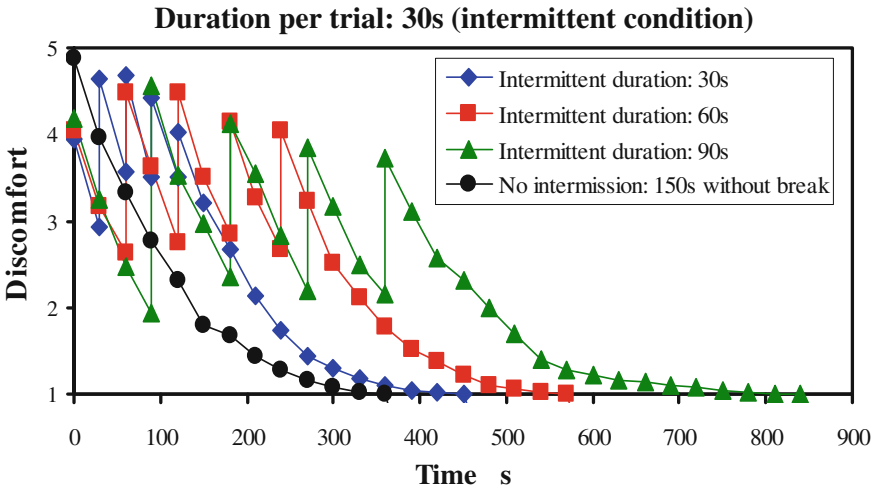


Fig. 6 Change of discomfort over time (duration per trial under intermittent condition: 30 s)

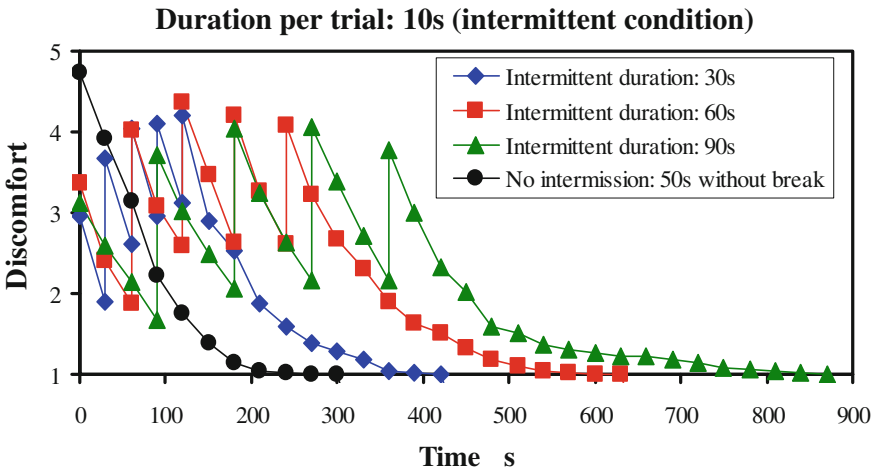


Fig. 7 Change of discomfort over time (duration per trial under intermittent condition: 10 s)

(duration per trial under intermittent condition: 10 s). Figure 8 shows the change of discomfort immediately after rating of experience of the stimulus over time (duration per trial under intermittent condition: 10 s). Figure 9 shows the change of discomfort immediately before rating of experience of stimulus over time (duration per trial under intermittent condition: 10 s).

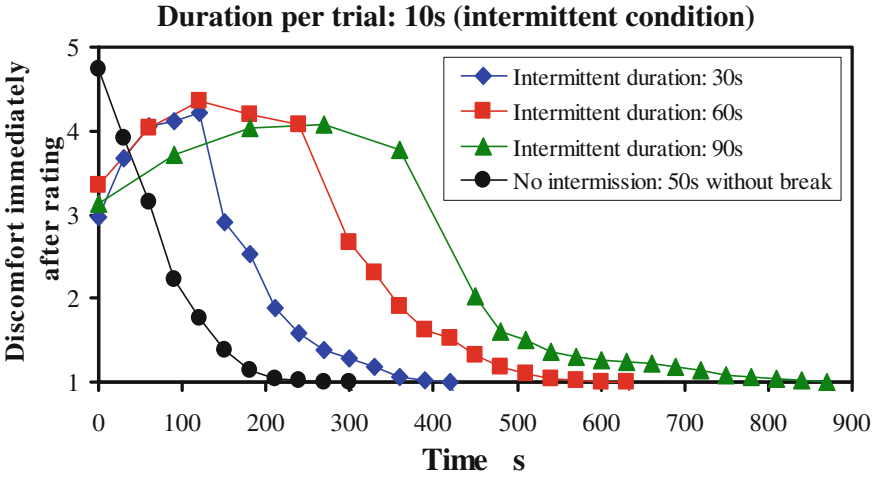


Fig. 8 Change of discomfort immediately after rating over time (duration per trial under intermittent condition: 10 s)

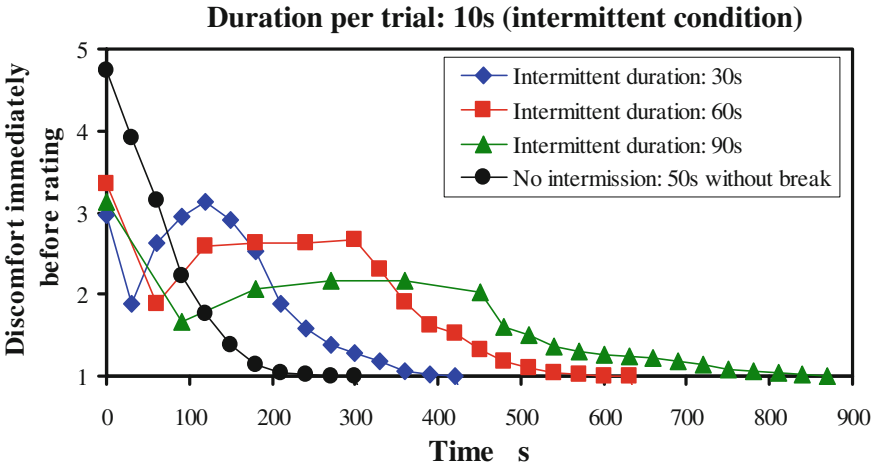


Fig. 9 Change of discomfort immediately after rating over time (duration per trial under intermittent condition: 10 s)

4 Discussion

Irrespective of the duration per trial (60 s or 120 s), the comfort under the condition of experience without a break was higher than that under the intermittent condition (Figs. 2 and 3). Under the intermittent condition, the comfort was reactivated by a

re-experience of the same comfort although the comfort rating was lower than that under the experience without a break. As shown in Fig. 4, the comfort immediately after the experience was nearly constant over five trials for both duration conditions of 60 and 120 s. The comfort immediately before the experience was also nearly constant over five trials (Fig. 5). Especially when the duration per trial was 120 s, the comfort immediately before the experience was affected by the duration of intermission, and the comfort decreased with the increase of duration of intermission (Fig. 5). The results might indicate that the re-experience of the same stimulus before forgetting the previous same stimulus leads to the same level of comfort.

Although the gross tendency of the discomfort experience was similar to that of the comfort experience, the following differences were observed for the discomfort experience. Irrespective of the duration per trial (10 s or 30 s), the discomfort under the condition of experience without a break was higher than that under the intermittent condition (Figs. 6 and 7), which indicates that the first intermittent experience makes us feel less uncomfortable than the experience without a break. Unlike the experiment of comfort experience, the discomfort immediately after the experience was not nearly constant over five trials for both duration conditions of 10 and 30 s (see Fig. 8). Rather, the discomfort immediately after the experience increased until the third or fourth trial. The discomfort grew over time, and this indicates that we cannot adapt to discomfort experiences. As shown in Fig. 8, the discomfort decreased at fourth or final trial, and the participant managed to adapt to the discomfort experience. A similar tendency was observed for the discomfort rating immediately before the next experience. However, it must be noted that the increase of discomfort was not so remarkable as compared to the discomfort rating immediately after the experience. This might show that the discomfort is mitigated with the elapse of time. For both of the duration per trial conditions (10 and 30 s), the discomfort tended to decrease with the increase of intermission duration. This might indicate that the discomfort rating was more remarkably affected by the intermission time, and that the longer intermission duration worked for the mitigation of discomfort induced by immersing one's hand in cold water.

5 Conclusions

In order to discuss whether the hypothetical model in Fig. 1 is valid or not, it was investigating how the intermittent experience of comfort or discomfort disrupts adaptation and intensify subsequent experience as compared with the continuous experience of the same stimuli. This was investigated as a function of experience time A (comfort stimuli: 300 and 600 s, discomfort stimuli: 50 and 150 s), duration of interruption B (comfort stimuli: 30, 60, and 90 s; discomfort stimuli: 30, 60, and 90 s), and the type of experienced stimulus (comfort or discomfort). The results can be summarized as follows.

(1) The rating of experiencing an uncomfortable stimulus under the intermission condition tended to increase gradually with the increase of repetition of experience until the 3rd or 4th intermission experience, and decreased at the final trial.

(2) The rating of experiencing a comfortable stimulus under the intermission condition tended to be constant even if the repetition of experience increased.

References

1. Kahneman, D., Snell, J.: Predicting utility. In: Einhorn, H.J., Hogarth, R.M. (eds.) Insights in Decision Making, pp. 295–310. University of Chicago Press, Chicago (1990)
2. Easterlin, R.A., Richard, A.: Will raising the incomes of all increase the happiness of all? J. Econ. Behav. Organ. **27**, 35–47 (1995)
3. Ariely, D., Zauberman, G.: Differential partitioning of extended experiences. Organ. Behav. Hum. Decis. Process. **91**, 128–139 (2003)
4. Ariery, D.: The Upside of Irrationality. Harper, London (2010)

Part IV
Foreign Relations: Tools and Analyses

Evaluating Automatic Learning of Structure for Event Extraction

Jason Schlachter, David Van Brackle, Luis Asencios Reynoso, James Starz and Nathanael Chambers

Abstract Analysts engaged in monitoring and forecasting benefit from the structured representations of domain knowledge and societal events that allow for the use of advanced analytics and predictive data models over large amounts of temporally extended data. However, extracting structured data from unstructured data typically requires the development of domain specific software which is costly, takes months to years to create, and cannot adapt to changing domains. In this paper we consider the operational usefulness of an approach pioneered by Chambers and Jurafsky (Template-based information extraction without the templates, 2011, [1]) that performs *automatic learning* of structured domain knowledge in the form of event templates from unstructured text that are used to *automatically extract* structured events from text. We generalize this approach and apply it to operationally relevant corpora from Brazil, Mexico, Ukraine, and Pakistan that focus on societal protests and providing aid. We discover that we are able to generate compelling event templates that correspond to event types described by Conflict and Mediation Event Observations (CAMEO) codes (Retrieved from Computational Event Data System, 2014, [2]) which are used to label event types by existing state of the art systems. Additionally, we are able to learn event templates that capture more nuance than the CAMEO codes represent, as well as entirely new and interesting event types. To automate our experimentation, we

J. Schlachter (✉) · D. Van Brackle · L.A. Reynoso · J. Starz
Informatics Laboratory, Lockheed Martin Advanced Technology Laboratories, Kennesaw, GA, USA
e-mail: jason.schlachter@lmco.com

D. Van Brackle
e-mail: david.van.brackle@lmco.com

L.A. Reynoso
e-mail: luis.a.asencios.reynoso@lmco.com

J. Starz
e-mail: james.c.starz@lmco.com

N. Chambers
Department of Computer Science, United States Naval Academy, Annapolis, MD, USA
e-mail: nchamber@usna.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_12

describe novel automated metrics that allow us to batch run multiple experiments while getting automated feedback on the quality of results from each run. These metrics indicate significant overlap between the events we extract and those extracted by existing systems.

Keywords NLP · Machine learning · Artificial intelligence · Data analytics · Text analytics · Forecasting · Event extracting · Event coding

1 Introduction

The amount of information available online is rapidly increasing. This is leading to the development of many new software tools and data models for understanding what is happening around the world and making predictions about future events. A key step in enabling these tools is converting unstructured data, such as text from news stories or blog posts, to structured data, such as labeled events in a database, that can be queried. Typically doing this requires building domain specific tools that use domain knowledge to extract events. Currently, these tools are built with hand coded domain knowledge or hand labeled training sets, and this requirement leads to high development costs, long development times, and brittle systems that cannot adapt to changing domains. For example, a recent government funded system built to extract structured events from news stories in the socio-political domain took over two years to develop, cost millions of dollars, and requires ongoing maintenance. Much of the cost is due to human involvement in defining hand-built rules for event extraction and/or building labeled training data. An ideal solution to this challenge would be a system that automatically learns domain knowledge and extracts structured events from text in any domain, removing the need for expensive and time consuming human involvement.

In this paper we describe an approach called Automatic Learning of Structure for Event Extraction (ALSEE) that seeks to radically reduce the cost and time needed to begin extracting structured events from new and changing domains. ALSEE builds on an approach pioneered by Chambers and Jurafsky [1] and uses a pipeline of machine learning and Natural Language Processing (NLP) techniques to *automatically learn* structured domain knowledge in the form of event templates with semantically constrained roles from unstructured text and an information extraction process to *automatically extract* structured events from text without the need for human intervention. Because ALSEE is automatic and domain independent, it allows for large savings in cost and time over existing best methods. In our experiments using a single laptop computer, we can extract events from a previously unseen domain after several hours of computer processing time with no human effort.

Our goal is to quickly discover event types, represented by event templates, which are interesting, coherent and useful for extracting events. We apply ALSEE to a variety of geographically and domain constrained corpora to evaluate our

ability to learn event types described by Conflict and Mediation Event Observations codes (CAMEO codes) [2]. CAMEO codes were developed by political scientists as a way to label meaningful events from a socio-political perspective and have been used in systems such as the Integrated Crisis Early Warning System (ICEWS) as the basis for advanced analytics and predictive models. We also evaluate our ability to learn interesting events not currently captured by CAMEO codes and other nuanced versions of the CAMEO code event types.

ALSEE is able to learn similar event types as those represented by CAMEO codes. Where ALSEE does diverge, it often learns a more nuanced set of event types than those represented by CAMEO codes. Perhaps most importantly, ALSEE learns new event types not represented within CAMEO. In this paper we describe a manual investigation of results from one of our corpora. We also describe and apply new automated metrics of performance to quickly gain an understanding of the quality of our results against an existing best-in-class system.

Our dataset includes corpora with news stories from Brazil, Mexico, and Ukraine that focus on societal protests and an additional corpora from Pakistan that focuses on providing aid. We find compelling event templates and useful extracted events and our quantitative metrics indicate promising performance against the existing best-in-class system. This is the first work to evaluate unsupervised event extraction against a system for extracting CAMEO coded events that required substantial human input to build (Fig. 1).

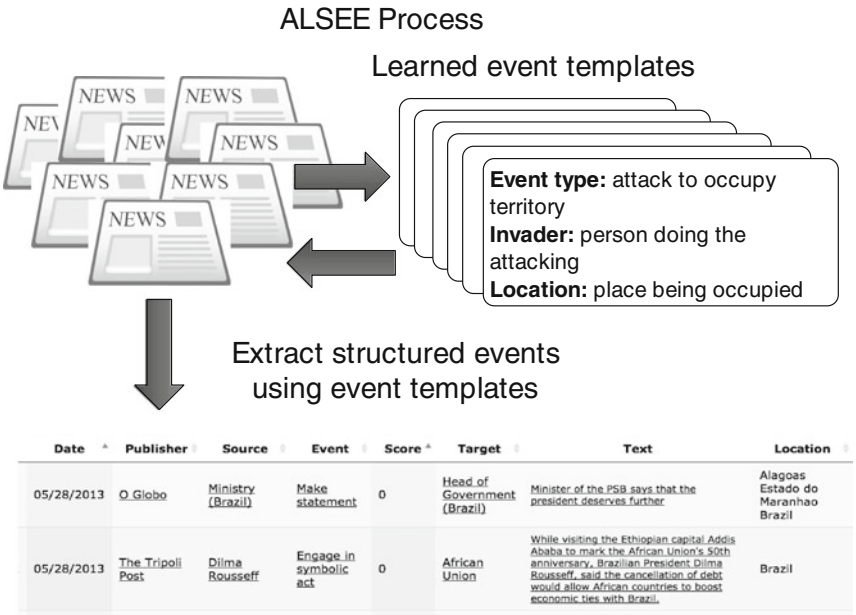


Fig. 1 The ALSEE concept. Transforming text to event templates and using them to extract structured events

1.1 Background

Event extraction has a rich history with approaches ranging from rule-based systems to unsupervised systems with little human intervention. This paper focuses on bridging the gap between the time-consuming nature of building rule-based systems to the more automated unsupervised approaches.

Early approaches to event extraction relied on rule-based approaches [2, 3] and quickly moved to supervised classifiers [4–8]. The supervised systems required human annotations that are prohibitively expensive to create for production systems. As a result, research on weak supervision relied on pre-clustered documents for specific domains. They did not create rules by hand, but instead used statistics to identify syntactic patterns that repeat across the documents [9–14]. These systems learn atomic events. This paper focuses on more structured representations.

Structured event schemas (templates) have recently received more attention in the extraction community. An event schema identifies the structure of a scenario, linking events like *arrest* with *convict*, and identifying the actors involved like the *police*, *criminal*, and *judge*. The earliest work to automate such event linking was proposed by [15, 16] in the form of *narrative chains*, sequences of events in text that contain the same entity (called the protagonist). They used point wise mutual information (PMI) to measure strength of association between events. Most relevant to this paper is the follow-on clustering algorithm in [1]. Events are clustered by proximity in a domain-specific corpus, and actors are extracted based on syntactic positions amongst the events. More details are given below.

A number of follow-on publications were inspired by [1]. Among them include more formal probabilistic models [17–19], web-scale event extraction [20], and a number of language model approaches to predict new events in a scenario [21–23]. This paper is unique to these approaches in its application to modern, operationally relevant corpora. Previous research has used toy domains, artificial corpora, or web-scale resources that do not reflect the type of data that analysts typically require. This paper thus adopts the algorithm in [1], and rigorously evaluates its application to existing taxonomies being used in cutting edge systems.

2 Approach

The authors generalize the pipeline described in [1] to apply it to operationally relevant corpora and to better facilitate our experiments. We then apply the pipeline to our test corpora to learn templates and to extract events. Using the results from the Ukraine corpus, we manually score the learned templates for their level of coherence to determine how many of the templates represent clear event types, as this is an unsupervised process and we anticipate learning some noise and other unintended patterns in the data. We then look at how well they align with the kinds of events represented by CAMEO codes, or how they add nuance to these events or

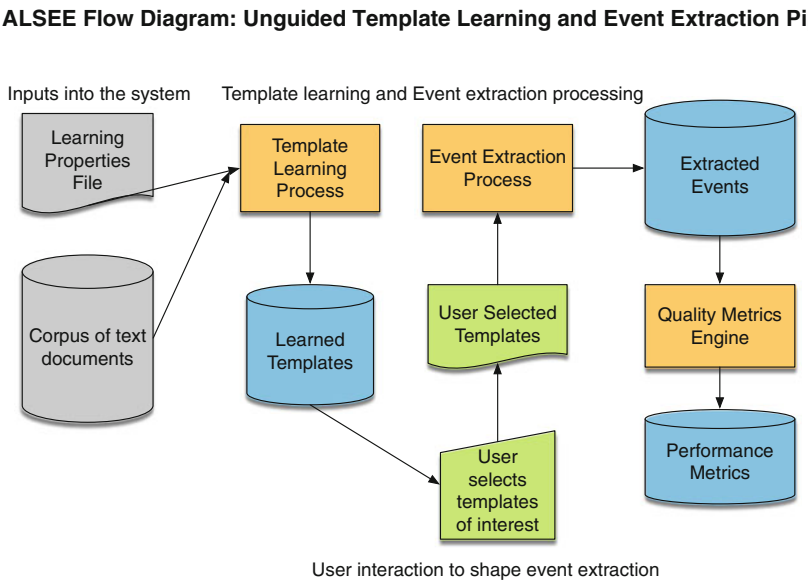


Fig. 2 Shows the components of ALSEE and the execution flow

represent new and interesting events. Templates also have semantically constrained roles but we leave the evaluation of these roles to future work.

In the section below we describe processing pipeline used to learn templates and to extract events. We describe some of the modifications we made to enable our experimentation. Figure 2 shows the ALSEE pipeline components and flow.

2.1 Automatic Learning Pipeline

Event templates are learned by first finding groupings of related event patterns and then finding the *semantic roles*, such as the people, places, things, etc. associated with each grouping. Together, these form the event template. We define event patterns in three ways; as verbs, as nouns that represent events (nouns that belong to Event synsets within WordNet), or as verbs conjoined with the headword of their syntactic objects. Event patterns are considered related, and thus paired together when they are observed co-occurring within documents with syntactic objects that co-refer with each other, such as when two different events reference the same person within the same syntactical context. For example, if an article says, “Bob attacked the police” and “Bob shot the police”, “attacked” and “shot” are paired because they corefer to “Bob” as the *subject* in both sentences. Higher relatedness scores are given if they occur close to each other within the document.

By examining all event patterns across all documents in this way, we are able to develop pairwise distance metrics based on point-wise mutual information (PMI) scores for all pairs of related event patterns in the corpus. We apply standard clustering algorithms to cluster all event patterns into groups of similar event patterns. We hypothesize that each cluster represents a coherent event concept. For example, using a corpus of 2,756 news stories from Ukraine published between 2010 to 2015 and focused on societal protests, one of the templates we learned captures the event type, *doing someone harm,* and contains event patterns such as *stab, facture, poisoning, concussion*.

After building event types as clusters of event patterns, we cluster the subjects, objects, and prepositions that occur in sentences that contain event patterns for each event type to discover semantically constrained roles for each cluster. Clustering is based on co-reference [15] and selectional preference [24] as both have been shown to indicate semantic relatedness or similarity. Heuristics are also applied to the clustering process, such as requiring the subject and object of a verb to have different semantic roles. The resulting clusters are characterized as people, objects, and places using WordNet to characterize the terms in each cluster. Clusters are then relaxed by incorporating syntactic patterns from events that are not in the event cluster but are similar to allow for easier role matching when extracting events from templates.

2.2 Pipeline Modifications

We generalize the code to run against an arbitrary corpus rather than the research oriented MUC4 corpus used by Chambers in his initial work. Second, we modify the processing pipeline to separate the template-learning phase from the event extraction phase. This allows a user to run the pipeline to learn templates, stop the pipeline execution, review the learned templates, select a subset of these templates to use to extract events, and finally start the event extraction process. Third, we add the ability to batch run multiple corpora with various parameter configurations through the pipeline while storing data from each run for later analysis. Finally, we develop an automated, domain agnostic, performance metric and other statistics to provide high-level insight on how ALSEE performs across a large number of runs without requiring manual inspection of each run. This metric allowed us to intelligently explore parameter values and to identify promising results on which we can manually evaluate the quality of the learned templates and extracted events.

2.3 Metrics

We develop metrics to quickly understand when our system extracts events similar to those extracted by existing best approaches. This is complicated because the

structure of events extracted by ALSEE differs from the structure of those extracted by traditional systems. ACCENT extracts CAMEO events which include a source and target actor with values pulled from a pre-populated dictionary of entities. Events extracted by ALSEE do not align with a pre-existing taxonomy of event types, nor are they guaranteed to have source and target actors, and they don't use an entity dictionary. Rather, ALSEE discovers roles dynamically and derives their values from text. For example, an ACCENT event will have the following structure, {*source_actorID* = 1, *event_type* = "violent protest", *target_actorID* = 2} whereas an event extracted by ALSEE has an event type defined by a cluster of related words and a dynamic number of roles such as, {*perpetrator* = "police", *event_type* = "control, s*uppress*, violence, protests", *victim* = "students", *instrument* = "tear gas"}.

To overcome these difficulties we count the number of sentences from which ALSEE and ACCENT both extract events, with the assumption that they are likely both recognizing the same or similar event. This is a low fidelity metric in that it doesn't explicitly compare the features of the extracted event, but does provide an automated assessment of overlap without having to address the structural and taxonomic issues described above. There are some limitations of this approach. Our metric is sensitive to the total number of sentences in the corpus relative to the number of sentences from which events were extracted by ACCENT or ALSEE. For example, a small number of sentences in the corpus with a large number of extracted events by ALSEE or ACCENT would increase the likelihood of intersection by chance, artificially increasing the score from this metric. Conversely, a large number of sentences in the corpus with a small number of extracted events from ALSEE or ACCENT would decrease the likelihood of intersection by chance, artificially decreasing the score from this metric.

To address the former case, we attenuate the metric score with a confidence measure that considers the number of sentences in the corpus relative the number of sentences from which events were extracted. Metric scores generated in situations in which there is an increased likelihood of intersecting sentences due to chance will be suppressed proportionately. This helps to keep our metric informative across a wider range of data conditions, although it does not completely overcome these issues.

3 Experiments

We apply our learning pipeline to the test corpora to learn event templates and extract events. We perform a manual analysis on the results from our Ukraine Protest corpus by evaluating the coherence of the event templates learned, evaluating how well they represent the Protest related CAMEO codes which were used to constrain the corpus, and by exploring the presence of new event types or event nuances that were learned. We run our automated metrics across results from all

corpora and evaluate the amount of overlap between ACCENT events and ALSEE events.

3.1 Test Data

We possess a collection of news stories published across the globe that span more than 23 years and contain more than 30 million items. This is the corpus from which we code events for use by W-ICEWS [25] using BBN's ACCENT event coder. For experimentation in ALSEE we consider four corpora. Three are protest focused and contain news stories published between 2010 and 2015 in Mexico with 431 stories, Ukraine with 2,756 stories, and Brazil with 1,614 stories. Another focuses on providing-aid and is composed of 1,436 news stories published from 2010 to 2015 in Pakistan. These corpora are built by querying our dataset for stories from which ACCENT has extracted events from domain of interest.

3.2 Evaluating Coherence of Learned Event Templates

We had an expert in the CAMEO hierarchy review the templates generated for the Ukraine Protest corpus and score them to indicate the level of coherence of each template. Coherence is the idea that the group of words that make up the event template should represent a clear concept. Our expert scored each template with a number between 1–5 where 1 indicates a completely incoherent template and 5 indicates a very coherent template. For example, in our Ukraine results, a template that was scored as 1 contains the following seemly unrelated event words {*wind*, *tolerate*, *sit-in*, *hint*}, whereas a template scored with a 5 contains the following seemly related event words {*block*, *barricade*, *halt*}.

3.3 Evaluating Learned Event Templates Against CAMEO Codes

We evaluate our learned event templates against CAMEO codes because we used CAMEO codes to constrain our corpora, and we expect a data driven process to learn events similar to those used to constrain the corpus. We are interested in determining if we are able to learn additional levels of nuance beyond that represented in the CAMEO code hierarchy or event types not represented at all in the CAMEO code hierarchy. In short, a system that learns event types that overlap substantially with the existing CAMEO codes at a fraction of the cost of existing

systems would be compelling. A system that is able to do this and discover additional nuance and relevant event types would be even more compelling.

3.4 Evaluation of Extracted Events from Metrics

We apply the performance metric described in the Metrics section to results from each corpus to determine if both ALSEE and ACCENT are targeting the same sentences for event extraction, implying they are focusing on the same types of events in the corpus.

4 Results

We focus on data generated using the Ukraine Protest corpus for manual analysis and consider the event types captured by each template. Results from all corpora are generated for our automated performance metric.

4.1 Template Coherence

We scored our learned templates for coherence and discovered that slightly more than half the templates are coherent. In total, we learned 63 templates from the Ukraine corpus. Among these, 10 had scores of 5, 14 had scores of 4, 22 had scores of 3, 11 had scores of 2, and 6 had a score of 1. Overall, there are 24 templates we consider very coherent (5) or mostly coherent (4). This distribution can be seen in Fig. 2.

4.2 CAMEO Code Event-Types Learned

There are 6 top-level Protest CAMEO codes (numbered from 140 to 145) used to build the Ukraine corpus capture the following event types, *engage in political dissent*, *demonstrate or rally*, *conduct hunger strike*, *conduct strike or boycott*, *obstruct passage or block*, and *protest violently.* We learn templates that correspond closely to all of these event types, except *conduct hunger strike*. There are instances of hunger strikes discussed within the corpus, but perhaps not enough examples were present to learn a template given the learning parameters used. This is evidence that it is possible to automatically learn meaningful event templates from a domain-constrained corpus (Fig. 3).

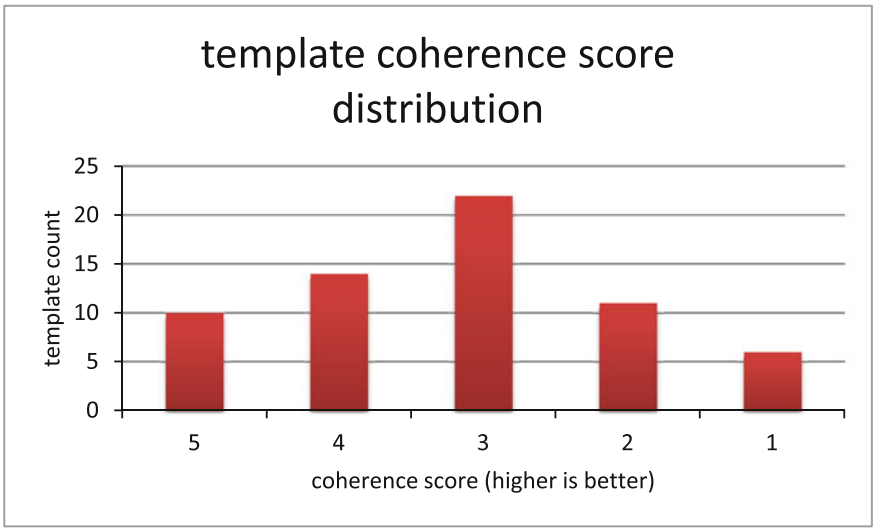


Fig. 3 Histogram showing the number of templates associated with each coherence score

Among the remaining 19 highly coherent templates (those with scores 4–5) we find 6 templates that correspond to existing CAMEO codes other than the protest codes and 7 templates that provide additional nuance over existing CAMEO codes. All of these provide are relevant, representing events such as *observing fighting*, *mobilizing*, *overthrow of local government, sexual assault*, *negotiating a solution*, *refrain from violence*. See Table 1 for a complete listing the templates learned and how to represent or extend existing CAMEO codes.

4.3 Novel Event-Types Learned

We learn entirely new event types not represented by CAMEO codes with 6 templates. These events appear to be relevant to the protest domain and capture the following event types: *transporting someone*, *allowing the publication of something*, *corrupt rule*, *tit-for-tat violence*, and *defecting to the enemy*. Details for each of these can be found in Table 1.

4.4 Automated Performance Metrics on Extracted Events

The intersection metric for results from all corpora indicate that ACCENT and ALSEE extracted events from the same sentences at a rate much higher than that which would be expected by chance (i.e. if ACCENT and ALSEE were randomly

Table 1 Event types learned by ALSEE compared to CAMEO codes

ALSEE Learned Event	CAMEO	CAMEO Coded Event
Stage protest	140	**Protest**: Engage in political dissent
Rally, mobilize people	141	**Protest**: Demonstrate or rally
Boycott	143	**Protest**: Conduct strike or boycott
Block off an area	144	**Protest**: Obstruct passage, block
Clash	145	**Protest**: Protest violently, riot
Promise to act on demand	030*	**Express intent to cooperate**: Nonspecific
Initiate proceedings	036	**Express intent to cooperate**: Express intent to meet or negotiate
Negotiate to avoid conflict	046	**Consult**: Engage in negotiations
Reward	060*	**Engage in diplomatic cooperation**: Nonspecific
Disband	080*	**Yield**: Nonspecific
Refrain from violence	087	**Yield**: De-escalate military engagement
Pause in fighting	0871	**Yield**: Declare truce, ceasefire
Incite action	113*	**Make public statement**: Rally opposition
Physical attack requiring hospitalization	180*	**Assault**: Nonspecific
Sexual assault	1821	**Assault**: Sexual assault
Attacking a motorcade	1833*	**Assault**: Carry out roadside bombing
Occupy territory	192	**Fight**: Occupy territory
Separatist overthrow of government in local region	192*	**Fight**: Occupy territory
Investigate sabotage	New	–
Effects of corrupt leadership	New	–
Tit-for-tat violence backfires	New	–
Defecting to the other side of a conflict	New	–
Transport	New	–
Allow publication (of a resignation)	New	–

CAMEO codes with asterisks indicate the ALSEE learned event is more nuanced

selecting sentences in the corpus). Across our corpora, we found that ALSEE extracted events from sentences from which ACCENT also extracted events between 27 and 46 % of the time. Performance did not seem to be correlated to the domain of the corpus, the number of sentences in the corpus, or the absolute number of events found by ACCENT or ALSEE. The best results were found for the Ukraine Protest corpus where 46 % of the sentences from which ACCENT extracted events were also targeted by ALSEE for extracting events. The lowest score of 27 % was from the Brazil Protest corpus. We calculate our confidence coefficient for these results and find that there is little change in the scores. The full set of results can be seen in Table 2.

Table 2 Weighted scores for the automated metric are presented here for four corpora

Corpus name	Corpus sentence count	ALSEE sentence count	ACCENT sentence count	Inter-section count	Weighted score
Mexico Protest Corpus 2010–2015	7909	921	758	291	0.34
Ukrainian Protest Corpus 2010–2015	53,400	5292	4401	2040	0.42
Brazilian Protest Corpus 2010–2015	26,632	1918	2133	573	0.25
Pakistani Provide Aid Corpus 2010–2015	23,943	2058	1496	583	0.36

5 Conclusions

We generalize a novel, automated approach to discovering domain knowledge in the form of event templates that are used to extract events. We apply this technique to modern, operationally relevant corpora from which a best-in-class event extraction system has previously extracted CAMEO coded events. Using results from our Ukraine Protest corpus, we compare the event types we learn to the CAMEO coded event types and discover that we not only provide good coverage of the protest related CAMEO codes but that we also learn many other CAMEO coded events, additional nuance on CAMEO coded events, and interesting new events not currently represented by CAMEO codes. These are encouraging results and the first step in establishing the usefulness of our approach.

6 Future Work

In this study, we did not specifically address the semantically constrained roles learned by ALSEE. Future work should focus on evaluating the coherence and relevance of these roles. Additionally, the events extracted by ALSEE were not evaluated explicitly because of the amount of manual effort involved. Future work should manually evaluate these events to ensure the templates are working well when extracting events.

Beyond better evaluations of the data, there are also modifications to the ALSEE processing pipeline that we would like to explore. We would like to modify the pipeline to allow a user to provide a taxonomy of interest, if one is known, to shape the types of templates learned by the system. We are also interested in exploring the possibility of having a human in the loop during the learning process to shape the focus of the event templates as they are forming to generate more useful templates.

References

1. Chambers, N., Jurafsky, D.: Template-based information extraction without the templates. In: Proceedings of the Association for Computational Linguistics (ACL) (2011)
2. CAMEO Event Data Cookbook. Retrieved from Computational Event Data System. http://eventdata.parusanalytics.com/data.dir/cameo.html. Accessed 21 June 2014
3. Rau, L., Krupka, G., Jacobs, P., Sider, I., Childs, L.: Ge nltoolset: Muc-4 test results and analysis. In: Proceedings of the Message Understanding Conference (MUC-4), pp. 94–99 (1992)
4. Chinchor, N., Lewis, D., Hirschman, L.: Evaluating message under-standing systems: an analysis of the third message understanding conference. Comput. Linguist. **19**(3), 409–449 (1993)
5. Freitag, D.: Toward general-purpose learning for information extraction. In: Proceedings of the Association for Computational Linguistics (ACL), pp. 404–408 (1998)
6. Chieu, H.L., Ng, H.T., Lee, Y.K.: Closing the gap: learning-based information extraction rivaling knowledge-engineering methods. In: Proceedings of the Association for Computational Linguistics (ACL) (2003)
7. Bunescu, R., Mooney, R.: Collective information extraction with relational markov networks. In: Proceedings of the Association for Computational Linguistics (ACL), pp. 438–445 (2004)
8. Patwardhan, S., Riloff, E.: A unified model of phrasal and sentential evidence for information extraction. In: Proceedings of the Conference on Empirical Meth-ods on Natural Language Processing (EMNLP) (2009)
9. Huang, R., Riloff, E.: Peeling back the layers: detecting event role fillers in secondary contexts. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL) (2011)
10. Riloff, E., Schmelzenbach, M.: An empirical approach to conceptual case frame acquisition. In: Proceedings of the Sixth Workshop on Very Large Corpora (1998)
11. Sudo, K., Sekine, S., Grishman, R.: An improved extraction pattern representation model for automatic ie pattern acquisition. In: Proceedings of the Association for Computational Linguistics (ACL), pp. 224–231 (2003)
12. Riloff, E., Wiebe, J., Phillips, W.: Exploiting subjectivity classification to improve information extraction. In: Proceedings of AAAI-05 (2005)
13. Filatova, E., Hatzivassiloglou, V., McKeown, K.: Automatic creation of domain templates. In: Proceedings of the Association for Computational Linguistics (ACL) (2006)
14. Patwardhan, S., Riloff, E.: Effective ie with semantic affinity patterns and relevant regions. In: Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP) (2007)
15. Chen, H., Benson, E., Naseem, T., Barzilay, R.: In-domain relation discovery with meta-constraints via posterior regularization. In: Proceedings of the Association for Computational Linguistics (ACL) (2011)
16. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative event chains. In: Proceedings of the Association of Computational Linguistics (ACL), Hawaii, USA (2008)
17. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative schemas and their participants. In: Proceedings of the Association of Computational Linguistics (ACL), Columbus, Ohio (2009)
18. Cheung, J.C.K., Poon, H., Vanderwende, L.: Probabilistic frame induction. In: Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (2013)
19. Chambers, N.: Event Schema Induction with a Probabilistic Entity-Driven Model. EMNLP-2013, Seattle, WA (2013)
20. Nguyen, K.-H., Tannier, X., Ferret, O., Besançon, R.: Generative Event Schema Induction with Entity Disambiguation. ACL (2015)

21. Balasubramanian, N., Soderland, S., Mausam, Etzioni, O.: Generating Coherent Event Schemas at Scale. EMNLP (2013)
22. Jans, B., Vulic, I., Moens, M.F.: Skip N-grams and Ranking Functions for Predicting Script Events. EACL (2012)
23. Pichotta, K., Mooney, R.J.: Statistical Script Learning with Multi-Argument Events. EACL (2014)
24. Rudinger, R., Rastogi, P., Ferraro, F., Van Durme, B.: Script Induction as Language Modeling. EMNLP (2015)
25. Erk, K., Pado, S.: Exemplar-based models for word meaning in context. In: Proceedings of ACL, pp. 92–97 (2010)

Scenario-Based Practical Exercises to Train and Assess General Cross-Cultural Competence for Special Operations Forces

Julio C. Mateo, Michael J. McCloskey, Clayton D. Leishman and Marla Federe

Abstract Our multidisciplinary team designed and developed scenario-based practical exercises (PEs) to train and assess general cross-cultural competence (3C) knowledge and skills. These PEs are consistent with several cultural frameworks and involve the completion of a set of cognitively authentic, mission-centric, increasingly complex scenarios in a mock village with foreign role players. The design of the mock village, scenarios, PEs, and 3C assessment present a set of key features that have contributed to its successful implementation within the context of a general 3C course for Special Operations Forces. Over a dozen classes have completed these scenario-based PEs and the reviews have been consistently positive. Throughout the process, we have incorporated learner feedback to improve the training scenarios and better serve students' needs. We summarize some of the lessons learned from the design, development, and implementation of the scenario-based PEs to guide the development of future efforts.

Keywords Cross-Cultural competence · Scenario-based training · Authentic assessment · Experiential learning · Immersive training · Military

J.C. Mateo (✉) · M.J. McCloskey · C.D. Leishman · M. Federe
Springer, 361 Interactive, LLC, 400 South Main St, Springboro, OH 45066, USA
e-mail: Julio.Mateo@Springer.com

M.J. McCloskey
e-mail: Michael.McCloskey@Springer.com

C.D. Leishman
e-mail: Clayton.Leishman@Springer.com

M. Federe
e-mail: Marla.Federe@Springer.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_13

1 Introduction

Recent conflicts in Iraq and Afghanistan have highlighted the importance of *language, regional expertise, and culture* (*LREC*) capabilities to mission success [1–3]. For DoD purposes, general 3C can be defined as the *knowledge, skills, abilities, and attitudes* (*KSAAs*) that enable Warfighters to successfully operate in foreign, often unfamiliar, cultural environments. General 3C focuses specifically on the 'C' of LREC capabilities. In the last decade, the *Department of Defense* (*DoD*) has invested a substantial number of resources to better understand what general *cross-cultural competence* (*3C*) entails in operational settings, how to train Warfighters to develop these capabilities, and how to assess the extent to which Warfighters possess them (see [4, 5] for reviews). Recently, the DoD guidance [6] has also urged all branches of the U.S. Armed Forces to take the necessary measures to enhance the LREC capabilities of its Warfighters.

As part of this DoD effort to grow LREC capabilities (and specifically 3C) across the force, Naval Special Warfare Advanced Training Command designed, developed, and is administering a general 3C course as part of the pipeline training of *Naval Special Warfare* (*NSW*) Operators. More specifically, an integrated DoD-industry team developed the 40-h training course to enhance the general 3C skills of NSW Operators, with a particular emphasis on supporting the cognitive demands inherent in NSW missions. The course combines interactive classroom exercises and guided discussions, computer-based exercises, homework assignments, and immersive practical exercises (PEs) to learn, apply, and practice 3C skills. This paper focuses specifically on the design and development of immersive, mission-centric PEs that students complete in a mock village while interacting with international role players.

2 The Development Process

2.1 Development of Learning Goals and Instructional Cadre Composition

Because the development of the immersive, scenario-based PEs took place within the context of the full general 3C course, some details about the development of the course are relevant to provide context. Prior to the development of PEs, we developed the learning goals for the course based on the review of a diverse set of resources:

a. Training guidance provided by the DoD [6] and U.S. Special Operations Command [7].
b. Models of 3C in operational settings (e.g., [8–10]).

c. Existing curricula to train 3C to Operators.
d. A needs assessment conducted on NSW Operators for a previous effort.

The resulting learning goals included, for example, 'understand how culture impacts one's and other's cognition and behavior,' 'make sense of and learn from novel cultural situations,' and 'think about the world from the perspectives of individuals from diverse cultural backgrounds.' The course learning goals guided the development of classroom sessions, activities, and PEs, as well as assessment criteria and rubrics used during culmination exercises.

We used a task-based, student-centered, pedagogical approach during the development of the 3C course, which is in alignment with the approach used to develop the language curriculum taught at the NSW LREC Program. We also decided that the instructional cadre for the 3C course should include an instructor with a background in anthropology and multiple facilitators with operational experience. Typically, there is at least one facilitator for every five students. This arrangement allows the course to balance theoretical rigor and operational relevance.

2.2 *Initial Development of PEs to Train and Assess General 3C Skills*

A key component of the course is a set of cognitively authentic, mission-centric, increasingly complex, scenario-based PEs that students complete in an immersive, interactive environment. These PEs are integrated within the 3C course and provide a platform to train and assess students' 3C knowledge and skills.

While the scenario-based PEs completed in the village support all of the learning goals of the course, the following two goals are their primary focus:

- Use general 3C knowledge and skills during the preparation, planning, and execution of NSW tasks to increase mission success in cross-cultural situations.
- Reflect on one's own and team's cognition and performance in cross-cultural situations.

The structure of PEs is consistent with several cultural frameworks: Byram and Nichols' *cultural knowings* [11], Kolb's *experiential learning theory* [12], Crawford-Lange and Lange's *content and process* [13], and Lussier's *saviors* [14]. Students learn from in-class activities and PEs through reflecting on their own experience and performance. The primary role of facilitators during PEs is to promote students' learning through questioning and promoting self-reflection. Importantly, the PEs completed in the village also provide an ideal platform for instructor and facilitators to assess students' 3C knowledge and skills. The cognitive authenticity of the scenarios places students in situations in which they display behaviors akin to those they would display during real-world cross-cultural encounters.

In order to ensure the operational realism and relevance of the scenarios, we reviewed documented mission sets of NSW Operators and interviewed several NSW Operators. These interviews focused on the types and range of missions performed by NSW Operators, the extent to which 3C skills were needed during operations, and the identification of critical incidents to guide the development of village scenarios. In addition, the facilitators in the instructional cadre, who are current or former NSW Operators themselves, provided invaluable input to improve the scenarios.

In terms of the mock-village design, we leveraged existing capabilities at the NSW LREC Program and tailored aspects of the village to fit better the learning needs of the course. The physical buildings within the village already existed for other training purposes. We designed the mock village stations to expose students to a variety of plausible situations (e.g., village leader house, market, café) and cultural backgrounds. Language teachers and other foreign nationals served as role players in the mock village (e.g., one station had role players from Philippines, another station from South Korea). A different foreign language was spoken at each station. Students were encouraged to leverage any foreign language skills they might have, but in most stations, they had to communicate through interpreters. Even though the culture and language at each station was different, for the purposes of the scenarios all the villagers belonged to the same community and had a common backstory. While not realistic, we chose the extreme multicultural makeup of the village because it provided pedagogical advantages over a village with a single culture. That is, it provided an accelerated learning environment to develop *general* 3C skills, as well as an ideal platform to assess the extent to which students displayed these skills across cultures.

2.3 Iterative Testing and Revision of the PEs

We developed the scenario-based PEs in multiple phases. In February 2015, we tested an initial, one-day version of the PE immediately after the last day of an existing 3C course. The primary goal was to explore its feasibility as a culminating training exercise and a testbed for the development of authentic 3C assessment. This one-day version of the PE showed great promise and was the foundation of the culminating PEs that students will complete on day 5 of the NSW 3C course.

Within a month of completing the initial one-day pilot, we completed the first iteration of the full 40-h 3C course. In this first full version, students visited the village twice (days 3 and 5) to gather atmospherics and assess the overall situation within the village, and conducted two negotiations with local villagers (days 4 and 5). These two visits to the village and two negotiations remain a central part of the village PEs in the current version of the course. Over the next several iterations, however, we have continued to improve the village scenarios to better represent the mission demands of NSW Operators, to better serve the pedagogical goals of the course, and to provide a more consistent village narrative. Role players have

received extensive training since the first full course, which has resulted in more elaborate and consistent characters. Additionally, the instructional cadre has further elaborated the village story, resulting in a more realistic village feel. Throughout this process, our team has continued to work with the instructor and facilitators to revise the assessment rubrics used by facilitators during the exercises.

In addition to further developing the PEs tested in the initial full course, we added two more PEs to the course based on feedback received from students. First, we added a second, shorter negotiation to day 4. This second negotiation gives students an additional opportunity to apply their negotiation skills while interacting with foreign nationals, but also contributes to the backstory for day 5. Second, and more importantly, we added an additional visit to the village at the beginning of day 1 (before formal instruction begins). Prior to the addition of the day 1 village visit, students often reported that it was difficult to relate day 1 and 2 classroom content to operationally relevant situations and often questioned the relevance of classroom content to their missions. The day 1 visit addresses many of these concerns and it also helps students realize what they do not know, appreciate the challenges of cross-cultural environments, and appreciate the value of having interpreters (which are not available in this initial visit). The final design, which has three visits to the village (days 1, 3, and 5) and three negotiations (two on day 4, one on day 5), is described in more detail below.

3 Key Design Features

This is not the first course that uses scenario-based PEs or that implements immersive, interactive PEs in a mock village with role players. However, the design of our immersive, scenario-based PEs provides high levels of cognitive authenticity, unique pedagogical opportunities, and unique opportunities to assess general 3C skills during actual cross-cultural interactions. In this section, we summarize some of the features that we think make our scenario-based PEs unique and ideally suited to train and assess general 3C skills.

3.1 Village and Scenarios

Single, Evolving Mission (From Day 1 to Day 5). We designed the village and negotiation PEs that take place throughout the week as stages of a single, evolving mission narrative. For example, each time they visit the village, there is an evolving backstory justifying the visit, but all of the visits are to the same village and fit within the larger context of the ongoing mission. Furthermore, each visit also targets different learning goals of the course. For example, the primary goal of the first visit on the morning of day 1 is to make students aware of their own cultural

biases and challenges, expose them to the importance of 3C skills to their missions, and help them realize that 3C is a challenging topic.

Students go back to the village on day 3, after the instructor has introduced 3C concepts and skills in a classroom setting. During this visit, students need to build rapport with the villagers, and gather intelligence and atmospherics about the village. This second visit to the village also provides repeated opportunities to learn how to interact with, and manage, interpreters in operational settings. Students can apply what they learned in subsequent iterations across the stations in a safe but realistic environment. By design, students visit all six stations, providing repeated opportunities to practice 3C skills in different cultural settings, make operational mistakes, and receive feedback from their peers, facilitators, and role players on their performance.

On day 4, students complete two negotiations, one with the village leader and another with a government official from the host nation. These negotiations provide a unique opportunity to apply lessons about persuasion and influence, as well as negotiation basics, in a cross-cultural context. They also set the stage for the culminating PE on day 5.

The PE on day 5 serves as a culminating exercise for the course. At this stage, students apply what they have learned throughout the course and it is an opportunity for facilitators to assess the extent to which students show improvement. Students are tasked with gathering additional intelligence and atmospherics (e.g., leverage points, power players) throughout the village. While the day 5 visit is an opportunity to apply the skills they have learned, it also offers new and unexpected challenges for students to test those skills. Following the visit, students report what they learned, receive guidance, and conduct one more negotiation with local villagers to accomplish a specific mission objective. In day 5, students lead their own debrief in which they reflect on their own performance and receive feedback from their group members. The facilitators and role players provide relatively less feedback to promote this self-reflection and group discussion.

Multicultural Village with a Common Backstory. In the early stages of the development process, we had considered designing the village with a single culture. In fact, and in spite of its obvious pedagogical advantages to teach general 3C skills, some of us were skeptical about the feasibility of a multicultural village. Ultimately, students benefited immensely from getting exposure to multiple, unique cultures in quick succession. Especially for students who had never traveled outside of the U.S., the multicultural village allowed them to experience first-hand differences in customs, communication styles and norms, deeper cultural differences, and the impact of these differences. The multicultural makeup of the village allowed us to create an accelerated learning environment for 3C skills that would have been impossible otherwise.

While the culture at each station is different, all stations are part of the same village, which has a common history and organization. When they visit the village, students often gather intelligence about the history, current events, and organization of the village. Keeping the village story straight across stations was initially

challenging, but as role players become more experienced, the common backstory also becomes more consistent.

Working with Interpreters. Since the role players at each station speak a different language, we assigned a separate interpreter to each station. At every station, students have to meet a new interpreter, introduce themselves, set their relationship, and work with him or her within that station. While having to meet six different interpreters within the span of 2 h and having only a few minutes to set the relationship with them was unrealistic, it provided an accelerated learning environment for students to realize what was important when working with interpreters. Actually, during the visit to the village on day 3, students typically go from having no previous exposure to interpreters to being able to list the main topics they should cover when meeting a new interpreter over the course of the visit. Over the course of the visit, students typically integrate what they learn about working with interpreters when planning the following station. Good facilitation is critical for this to happen, since the key here is to pose the right types of questions rather than providing the answers.

3.2 Practical Exercises

Multiple Opportunities to Practice. The only way to develop cognitive skills is to practice repeatedly [15]. The PEs provide multiple opportunities for students to practice their observation, sensemaking, and communication skills while interacting with individuals from a variety of cultures. For example, during each of the last two visits to the village, students engaged with foreign role players at six different stations (e.g., tea shop, market, village leader house). The ability to complete six of these rotations in a 2-h period allowed students to engage, make mistakes, learn from those mistakes, and integrate those lessons into their next interaction. In addition to the village visits, students also completed three separate negotiation PEs in day 4 and 5. The numerous opportunities to engage in hands-on, mission-centric, interactive PEs with foreign nationals created an ideal setup for 3C skill development.

Self-Reflection and Multifaceted Feedback. After each interaction (i.e., rotation at the village or negotiation), students complete a debrief session with their team. This debrief aims to promote reflection on one's own and team performance. During the first few rotations of day 3, each student also completes one rotation as an observer to encourage them to observe the behavior of their group members during the interaction. In these rotations, the student observer leads the debrief session by describing what he observed about his group's performance while standing on the sidelines, focusing primarily on cultural elements. Immediately after each of these debrief sessions, the team has a few minutes to prepare for the next station. Facilitators encourage students to integrate the lessons learned from reflection and feedback into their plan for the next station.

Once all students have been a student observer once, the student who led the interaction begins the debrief session by assessing his own (and his team's) performance, followed by the other group members discussing what they observed. The facilitator ensures that the debrief stays on track and provides additional feedback at the end. The facilitator also invites the role players for that station to provide their feedback to students. Often, the discrepancy between how the students and the role players viewed the students' performance resulted in invaluable learning opportunities. It is important to note that the majority of the debrief period is led by the students. This encourages them to reflect on their performance and learning from their own mistakes, rather than waiting to be told what they did wrong. Facilitators only provide feedback when the students miss key points.

3.3 Assessment of Cross-Cultural Competence

Authentic Assessment and Rubrics. We designed the cognitively authentic, mission-centric scenarios to recreate many of the demands of operational environments. As a result, students facing these scenarios display observable reactions and behaviors that lend themselves to authentic assessment. We also developed an assessment rubric that describes observable behaviors and links them to individual course learning goals and objectives. During day 5 PEs, facilitators use this rubric to assess the general 3C knowledge and skills displayed by students. Assessing student performance within the context of the immersive village exercises makes our assessment more difficult for students to fake than paper-and-pencil assessments. It also allows facilitators to assess the practical application of acquired 3C knowledge, skills, and awareness during both planning and execution of a mission in a cognitively authentic environment. We are continuing to refine our assessment rubric to enhance its effectiveness and ensure that different facilitators can administer it consistently.

Individual Baseline. Students can vary widely regarding their prior international experience and actual 3C skills at the beginning of the course. The assessment rubrics used in village visits are different from those used in negotiations, but the rubrics used on days 3 and 4 are the same as those used during the culminating PEs on day 5. Therefore, facilitators can use village and negotiation assessment rubrics during days 3 and 4 to develop a baseline for individual students. This is an additional benefit of providing multiple opportunities to practice.

The initial baseline can help facilitators identify specific knowledge and skills to target for individual students. It also allows facilitators to assess not only the student's 3C knowledge and skills at the end of the course, but also the extent to which each student improves as he gains more experience. The ability to identify students who showed dramatic improvement when provided with opportunities to learn and develop 3C skills is important. These students are fast learners when given opportunities to learn and apply 3C skills in context and are likely to improve even further as they gain more experience.

Final Learner Profiles. During day 5 PEs, facilitators assess students on all the learning goals and objectives of the course. As a result, facilitators create a learner profile that specifies the strengths and weaknesses of each student. We have encountered some challenges developing an assessment rubric that facilitators can use to assess students consistently. However, we continue to revise this rubric and work with facilitators to get it implemented. Ultimately, we hope to provide this profile to each student so that he can better understand his strengths and weaknesses, and take the appropriate actions to leverage or improve them during his career.

4 Lessons Learned

The lessons learned from developing and implementing this innovative course are significant, and they continue to expand as we refine the structure and presentation of the course. Even when only considering the village PEs, we have identified several important lessons learned. We hope that these lessons can help others developing similar scenario-based PEs for training and assessment purposes.

Cognitive Authenticity versus Physical Realism. Developing a mock village to simulate a foreign environment may seem like a daunting task at first. However, the physical realism of the village does not need to be high for the PEs to be effective and elicit relevant, observable (and assessable) behaviors. Rather, the emphasis should be on the creation of cognitively authentic, scenario-based PEs that reflect the relevant cognitive challenges that Operators face in the field. For the PEs to work, it is also critical that (a) role players behave and react in realistic ways and (b) how students perform affects both the role players' behavior and the students' ability to complete the mission successfully. For example, if a student treats the interpreter poorly when they first meet, the interpreter will act less friendly and be less helpful during interactions than if the student builds good rapport. Similarly, villagers will share more information with students who successfully build rapport during the interaction, which will directly affect their ability to complete their mission (i.e., gather intelligence and atmospherics on the village) successfully.

Pedagogical Value versus Operational Realism. If we created scenarios that reflect operational realities as closely as possible, we would have never developed the village scenarios the way we did. For example, rather than having students engage in a longer interaction with a single culture (which would more closely resemble operational situations), we decided to have students complete six quick rotations through six different cultural settings during each visit to the village. Experiencing six different cultures in such rapid succession within a single village may not be realistic, but the accelerated learning opportunities it provides are beneficial to skill development. Furthermore, this design requires rapid adaptation and is relevant to the types of challenges Operators face in the field.

The PEs provide repeated opportunities for students to practice critical cognitive skills, reflect on their performance, receive feedback, and apply what they learn

throughout the process. It also gives students the opportunity to work with interpreters in six different cultural settings, while concurrently exposing students to cultural and customary elements across these cultures. The advantages of this approach go beyond benefits to the student's learning. It also gives the instructional cadre repeated opportunities to observe students' behavior, assess their performance, and provide feedback to enhance their learning.

In addition, student reviews of the scenario-based PEs have been overwhelmingly positive in terms of perceived value and lessons learned. Students reported that the PEs were fun and allowed for a unique opportunity to apply the lessons learned in the classroom. Several students even expressed a desire to spend additional time in the village scenarios.

Versatility of Training Scenarios. The village scenarios provided a versatile platform to create different types of PEs. The primary goal of the PE on day 1 is to raise awareness of the importance of 3C to mission success among students, while the PEs on days 3 and 4 provide repeated opportunities to learn experientially. For example, during the visit to the village on day 3 students learn about working with interpreters in an inductive manner. That is, when they meet their first interpreter, students have little guidance about what to do. During the first rotations, they struggle and make mistakes. However, with a little help from the facilitators, they reflect on their mistakes, learn from them, and apply their learning in subsequent rotations. At the end of the day, students improve their ability to develop a relationship and work with interpreters, and have concrete experiences to link to these lessons learned. While experiential learning continues to take place throughout the week, the primary goals of the PEs on day 5 are different. These PEs provide a setting for students to apply what they learned during the class and for facilitators to assess students' success at applying 3C knowledge and skills while they engage in cognitively authentic situations.

The Importance of Reflection. Reflection is critical for learning to occur and to continue beyond the duration of the course. While the exercises create valuable experiences and promote significant learning by themselves, effective reflection opportunities are a force multiplier for the training effectiveness. Facilitator guidance and observations, role-player feedback and, most importantly, guided self-reflection greatly enhance student learning during PEs. Students often come to their own "aha" moments as they discuss recently completed rotations with their teams, and realize their own miscues or mistakes in interacting with role players. Throughout the PEs, we are promoting the concept of the *reflective practitioner* [16] by asking students to critique their own and their team's performance before facilitators or role players provide any feedback. That is, students can work on developing the broader abilities to develop and test hypotheses in real-time, recognize their own affective reactions to events as they occur, and more quickly develop understandings of unique operational environments. Our hope is that students leave the course and apply this reflective thinking across their careers.

Multi-Disciplinary and Resource-Intensive. The resources required to design, develop, conduct, and maintain the PEs described in this paper are significant. A team of educators, cognitive psychologists, anthropologists, curriculum

developers, and NSW Operators all worked together during the design, development, and implementation of this course and PEs. While such a multidisciplinary team inevitably results in differences in opinion and occasional disagreements on approaches, it notably improved the quality of the final product.

The instructional cadre is also multidisciplinary: one lead instructor with anthropology experience and a set of highly involved facilitators with operational experience. We have found this to be a synergistic combination. On the one hand, the lead instructor ensures that the course and PEs meet the learning objectives and the students are engaged and grasp key cultural concepts. On the other hand, facilitators ensure that the content and PEs remain operationally relevant, and that the students make the connections to their future missions.

In addition to the development team and the instructional cadre, a group of approximately 20 foreign role players participates in the delivery of every course and set of scenario-based PEs. Given the large numbers of people involved, there are occasional changes that require new people filling positions. While extremely challenging and resource intensive, the value added by merging all of those perspectives greatly outweighs the inherent challenge in managing such a diverse team. We strongly believe this arrangement provides unique benefits that cannot be achieved with a simpler structure.

5 Summary and Conclusions

Our multidisciplinary team designed and developed cognitively authentic, mission-centric, increasingly complex PEs to train and assess general 3C skills. Over the past year, students have completed these training scenarios in a mock village with foreign role players from a variety of regions. We have also developed and used assessment rubrics to assess students' general 3C skills in this context. Over a dozen classes of NSW Operators have completed this training and the reviews have been vastly positive. Throughout the process, we have incorporated learner feedback to improve the training scenarios and better serve students' needs.

Developing a hands-on, immersive course to teach general 3C skills is not easy and doing it right can be quite challenging and resource intensive. Multidisciplinary collaboration adds challenges, but it is necessary. We found the DoD-funded research that preceded this effort to be instrumental and helpful in developing the learning objectives and guiding the direction of the training. We also think that the lessons learned from our experience developing this course can help others who face similar challenges.

Acknowledgments We would like to acknowledge the substantial contributions that Dr Joseph Cichosz, Joseph Schowalter, Jason Perez, and Gary Stenson made to all stages of the development process. We also thank the staff at Mid Atlantic Professionals—SSI, who made the implementation of the PEs possible, and Tailored Training Programs for their contributions during the development of the course learning goals, 3C curriculum, and initial concept of the scenario-based PEs. Finally, we thank the Navy Special Warfare Advanced Training Command for supporting this effort.

References

1. Albro, R., Ivey, B. (eds.): Cultural Awareness in the Military: Developments and Implications for Future Humanitarian Cooperation. Palgrave Macmillan, New York (2014)
2. Salmoni, B.A., Holmes-Eber, P.: Operational Culture for the Warfighter: Principles and Applications. Marine Corps University Press, Quantico, VA (2008)
3. Wunderle, W.D.: Through the Lens of Cultural Awareness: A Primer for United States Armed Forces Deploying in Arab and Middle Eastern Countries. Combat Studies Institute Press, Fort Leavenworth, KS (2006)
4. Egeth, J., Klein, G.L., Schmorrow, D. (eds.): Sociocultural Behavior Sensemaking: State of the Art in Understanding the Operational Environment. The MITRE Corporation, McLean, VA (2014)
5. Gallus, J.A., Gouge, M.C., Antolic, E., Fosher, K., Jasparro, V., Coleman, S., Selmeski, B., Klafehn, J.L.: Cross-Cultural Competence in the Department of Defense: An Annotated Bibliography. Special Report 71. U.S. Army Research Institute for the Behavioral and Social Sciences: Fort Belvoir, VA (2014)
6. Chairman of the Joint Chiefs of Staff Instruction: Language, Regional Expertise, and Culture (LREC) Capability Identification, Planning, and Sourcing. Instruction 3126.01A (2013)
7. United States Special Operations Command (USSOCOM): The Future Special Operator Concept: A Blueprint for 2020 and Beyond, v.0.71 (2014)
8. McCloskey, M.J., Behymer, K.J., Papautsky, E.L., Grandjean, A.K.: Measuring Learning and Development in Cross-Cultural Competence. Technical Report 1317. U. S. Army Research Institute for the Behavioral and Social Sciences: Fort Belvoir, VA (2012)
9. Rasmussen, L.J., Sieck, W.R., Crandall, B.W., Simpkins, B.G., Smith, J.L.: Data Collection and Analysis for a Cross-Cultural Competence Model. Arlington, VA, Defense Language and National Security Education Office (2013)
10. Reid, P., Kaloydis, F., Sudduth, M., & Greene-Sands, A.: Executive Summary: A Framework for Understanding Cross-Cultural Competence in the Department of Defense. Tech. No. 15-12. Defense Equal Opportunity Management Institute: Melbourne, FL (2012)
11. Byram, M., Nichols, A. (eds.): Developing Intercultural Competence in Practice, vol. 1. Multilingual Matters, Buffalo, NY (2001)
12. Kolb, D.A., Boyatzis, R.E., Mainemelis, C.: Experiential learning theory: previous research and new directions. In: Sternberg, R.J., Zhang, L.F. (eds.) Perspectives on Cognitive, Learning, and Thinking Styles, pp. 227–248. Routledge, New York (2010)
13. Crawford-Lange, L.M., Lange, D.L.: integrating language and culture: how to do it. Theor. Pract. **26**(4), 258–266 (1987)
14. Lussier, D.: Theoretical bases of a conceptual framework with reference to intercultural communicative competence. J. Appl. Linguist. **4**, 309–332 (2007)
15. Anderson, J.R. (ed.): Cognitive Skills and Their Acquisition. Lawrence Erlbaum Associates, Hillsdale, NJ (1981)
16. Schon, D.A.: The Reflective Practitioner: How Professionals Think in Action. Basic Books (1983)

Part V
Toward Culturally-Aware Systems

Toward Culturally-Aware, Next Generation Learning Ecosystems

Elaine M. Raybourn

Abstract Next generation learning ecosystems will be comprised of intelligent, adaptive environments that utilize one's cultural footprints to co-create shared narratives and facilitate intercultural understanding. The present paper discusses why digital footprints, cultural signposts, intercultural agents, and transmedia learning are needed to realize relevant learning in virtual environments. The paper introduces notions that may impact the design of culturally-aware information technology for distributed learning are presented.

Keywords Transmedia learning · Ecosystems · Intercultural communication

1 Introduction

Learning today occurs anywhere, anytime. We live in a world that is connected 24/7 and social. Our media habits and expectations regarding technology are changing. We now have the ability to have digital friends, we use multiple screens, and the real world is just one of the platforms we use to interact with content and each other [1]. Learners are telling us how they like to learn, if we listen to them. They are storytellers and builders. They are hands on. They love playing games, and being immersed in a story. They crave technology. They want to create, and leave their mark on the world.

Educators, scientists, and technologists have begun to think about how to deliver and support lifelong learning journeys. Data analytics are used to make games and immersive simulations more adaptive to each learner. Games are designed to help us remember, learn, or make sense of complex ideas. So whether teachable moments come from an immersive game, videos, intelligent tutors, dynamic action data books, or conversations with mentors and peers, scientists are now developing

E.M. Raybourn (✉)
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: emraybo@sandia.gov

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_14

intelligent systems that are able to connect experiences that occur in and out of classrooms; and facilitating social learning from peers across the globe [1].

Next generation learning ecosystems will not only leverage technologies such as the ones described above, they will be story-driven, incorporate learning across different media, and allow the learner a variety of entry points into a system of experiences [2]. That is to say, next generation learning ecosystems will support *transmedia learning*. Transmedia learning is as the scalable system of messages that represents a narrative or core experience that unfolds from the use of multiple media, emotionally engaging learners by involving them personally in the story [2]. In order to realize the vision of transmedia learning, blended, multi-media deployment and storytelling strategies need to be leveraged effectively to motivate personalized, self-paced training and education. Media may include serious games, immersive simulations, augmented reality, virtual reality, mixed media environments, intelligent tutoring systems, mobile performance aids, MOOCs, virtual environments, machinima (video or short films made with game technology), mobile learning, graphic novels, motion comics, film, radio, print, user generated content, and social media, to name a few.

For example, consider the following scenario explored in [2]:

A Soldier trains in the field, with different simulators, on different platforms, in the classroom, and with her peers (both co-located and distributed). The use of different media allows her to engage in the training from different platforms and entry points. For instance, if she is learning the art of being a Soldier–Diplomat, she may begin her language and culture training with an intelligent tutor and continue with a single-player scenario on cultural awareness that is delivered via a serious game. She also engages in an alternate reality game on cultural awareness with her peers. Later she blogs about what she learned and shares this information with her team. The conversation about cultural awareness continues on Twitter. She reads about case studies via digital graphic novels or by watching videos available from the MOOC. Her learning is self-paced, collaborative, adaptive, and/or mediated by instructors, virtual mentors [3], and embodied agents [4]. She creates content, tracks her own learning, and monitors her progress. The learning ecosystem tracks her actions as she trains and stores these data in a learning record store, learner profile, or learner model. Her training is delivered via a variety of media, making it more dynamic, accessible, and engrossing.

In addition to story-driven, compelling content, the transmedia learning ecosystem described above will itself need to leverage lessons from digital footprints, and incorporate cultural signposts and intercultural agents, especially as part of a complex, techno-social systems of transmedia learning ecosystems. The present paper will discuss why cultural signposts and intercultural agents are needed to facilitate interaction among users or with systems of transmedia learning ecosystems. In doing so, the author draws upon excerpts and examples from previous work that continues to be salient in this discussion.

2 Digital Footprints as Cultural Signposts

In his seminal book Silent Language, anthropologist Edward T. Hall indicated that "culture is communication" [5, p. 97] and thus comprised of words, actions, non-verbal behaviors, the handling of time, space, and materials, worldview, beliefs, and attitudes passed over time from generation to generation. Culture is a set of experiences that are deeply rooted in interactions with our physical and social realities, which often go unsaid, but are nevertheless communicated without one's awareness [5]. Our physical environment provides the experiential backdrop in which our interactions over time communicate culture to ourselves, as well as to those around us. According to Hall [5, p. 39], "to interact with the (physical) environment is to be alive… Ultimately everything man does involves interaction with something else." It is through these interactions, and the relevance we place on the feedback we perceive, that culture is formed and expressed to others. The consistent patterns of our interactions with artifacts, our physical environment, and other individuals over time provide cues that others may use to interpret our culture.

For example, as we walk across the lawn during a winter morning, we may hear the crunching of frost under our feet, feel the frozen grass through the soles of our shoes, and see our footprints in newly fallen snow. That is, we usually receive feedback when we interact with our physical environment about how our presence has influenced or changed the physical reality. The feedback makes us more aware of the physical world. The physical environment helps us understand more about ourselves, and therefore more about what it means to be alive. One might say that the physical environment interacts with us, as it registers our actions and provides feedback. To receive that feedback is to receive important information needed to discern cultural phenomena. To receive feedback then is to *feel alive, and feel that the physical environment is alive*.

Similarly, we need to receive feedback from our interactions in virtual, digital, computer-mediated, or electronic settings. We want to know how our actions and presence in a virtual environment influence or change the technology-mediated setting. For example, technology-mediated interactions should help us understand more about ourselves by giving us feedback about what it means to be alive in a virtual setting. When we browse the Internet, we might want to know who else has viewed the same web page, or we may want to see old email messages that fade over time as their saliency diminishes. Hall [5, p. 192] indicates "Experience is something man projects on the outside world as he gains it in its culturally determined form." That is, through communication and feedback from our interactions we learn how to culturally situate our experiences. In order to culturally situate our technology-mediated experiences, we need to sense how our actions impact the virtual setting as we interact with it; we need to be aware of our electronic footprints. Just as in physical settings, we communicate and interpret rich cultural information about ourselves as we interact with artifacts or others in virtual environments.

Researchers and designers of culturally situated, social computing environments need to pay particular attention to the types of cultural footprints we may leave in

technology-mediated interactions and virtual environments. Cultural information in the form of a virtual footprint may include users' current activities, electronic artifacts, narratives, likes, dislikes, assumptions, values, goals, meanings, and history. Whether our interactions are expressed explicitly (awareness of our interactions) or implicitly (being unaware of our interactions), intelligent, adaptive systems have the ability to capture and subsequently communicate the electronic footprints users leave in community-based environments. However, users' electronic footprints are often opaque. Unlike physical environments, most workplace community-based systems today lack these subtle cultural prompts, cues, signposts, or environmental markers that provide us with rich cultural feedback about other users or the environment's current or past state [6]. With the exception of some social media sites, this is one of the reasons why so many community-based systems *do not feel alive, and we do not feel alive interacting in them.*

As our technology-mediated interactions blend real and virtual realities we'll need to balance feedback from our cultural footprints in both digital and physical domains. Which footprints will be left in the real or virtual world? How will our mixed reality learning environments negotiate when and how feedback is to be presented? Lessons learned from face-to-face communication tell us that its quality depends largely on sharing cultural and contextual information [7]. How will a next generation learning ecosystem help us *feel alive* when our attention and footprints are distributed among technology-mediated, virtual, and physical world learning opportunities?

3 Cultural Signposts in Virtual Environments

The author has suggested the need to support technology-mediated communication with intelligent or adaptive virtual environments that incorporate cultural signposts [6, 8, 9]. In particular she has posited that technology alone is not enough to facilitate online and virtual conversations among strangers, even if they shared common interests. We often need a third party introduction to take the first step toward any kind of communication. Additionally, most systems are designed without regard to culture, which could be one reason why spontaneous communication is difficult to achieve in the workplace. Intelligent systems utilizing cultural signposts can be instrumental in instigating informal communication among members, teams, or communities of practice. For that reason she has been a strong advocate for the development of culturally aware information systems [9].

As a guest researcher for British Telecom, the author worked on the development of an adaptive community-based system that incorporated user profile matching, collaborative information storage and retrieval, summarization of shared documents, graphically enhanced audio conferencing, and a dynamic 3D virtual world that automatically moved users with common tasks together in the 3D

environment. For example, software agents inferred a profile for each interest group, based upon the summation of all the subscriptions and profiles. The user's avatar was placed within a notional home area when the system started up, which was calculated by the similarity between a user's profile and the derived profile for each interest group. The focus of a user's work, reflected in the different documents and applications that were being used on a computer, was monitored by the agent-based system. The intelligent agents utilized dynamic user profiles within a community of practice to put people, who would benefit from sharing information in real time, in touch with one another automatically. In this way, the virtual world supported chance encounters and mutual engagement. That is, the virtual world offered users' avatars the ability to 'hang out' online and meet others from one's community of practice through chance encounters.

The goal of the effort described above was to support spontaneous chat interactions [6]. This notion for computer-supported collaborative work was forward-looking for its time. The system had the functionality to technically support spontaneous chat in the virtual setting, by automatically clustering pairs and groups of people together in a virtual world based on their common interests. However, the system wasn't being used to its fullest potential because context, or the reasons why two persons might want to meet, was overlooked. Strangers using the system often had difficulty opening a conversation even though the technical functionality was nearly flawless [6]. The issue wasn't technical—it was social. Nobody wanted to be thrust into a situation where a conversation was expected without having some way to "break the ice." They needed third party help to 'introduce' them, much the way we rely on people at cocktail parties. A redesign was recommended [6] to include user or intercultural agent-generated cultural signposts, or interactive social and environmental cues that could motivate informal communication among strangers in the community of practice. We still need to do a better job of incorporating cultural signposts in our portals, adaptive systems, and distributed learning support technologies to achieve our goal of ubiquitous culturally aware information systems.

4 Intercultural Agents in Intelligent Systems

Intercultural were defined [8] as entities that participate on behalf of users in communication, the exchange of information, and co-creation of meanings by software agents, entities, individuals, or groups. In such instances at least one party perceives a difference among entities, or perceives itself to be different from others [8]. An intercultural agent's goal should be to assist intercultural computer-mediated communication and social computing by managing the awareness of the presence of users' cultural footprints and cultural cues in technology-mediated interactions. For example, intercultural agents may adapt the interface differently

for users with different cultural expectations regarding interface design and visual environment. The author defines cultural cues, or signposts, in agent-based systems as (1) subtle prompts enacted by individual or collections of agents, or the agent-based environment to encourage a user to action, or (2) subtle (symbolic) markers left by agents or humans in an environment that issue feedback about the users' or environment's current or past state. Intercultural agents may have several modes of interaction with interlocutors, or with other agents. First, in an interaction among two or more individuals and software agents, the definition suggests that software agents may assist the user in achieving intercultural communication competence by direct intervention (i.e. with a direct prompt, or interaction with an embodied agent or avatar). For example, an intercultural agent may detect when an inappropriate remark is typed by a user. The intercultural agent could suggest other terms that are less likely to be misinterpreted.

Alternatively, intercultural agents may act on the users' behalf by generating subtle cultural prompts that guide interlocutors to be more aware of their intercultural communication competence, or increase their awareness of the saliency of culture in the interaction. Cultural cues, or prompts, may be issued in the form of narratives co-created with users, in the graphical user interface, or through adaptive environments that respond to users' cultural footprints and allow for asynchronous and self-paced exploration. For example, if a newly hired, remote employee is on an organization's intranet to learn about the company, an avatar-guided virtual tour of the physical office space, the team members, and perhaps a frequently asked question section on the formal and informal organizational cultural norms could help the remote employee feel more part of the team [3]. An avatar-guided tour of cultural footprints left in a team gallery of interests might be an informal mechanism for obtaining meta-level information on the team culture and individual identities.

In some cases software agents may be working in concert to recognize whether users are operating from different cultural orientations based on learner models, profiles, current context, history, etc. That is, intercultural agents may be privy to user profiles gathered explicitly through feedback to postings in the community-based system that reflect a user's long-term interests; or implicitly through their activities, history, and current context of work. However, agents do not need to overtly communicate this knowledge to users. Instead, intercultural agents may administer subtle, personalized cultural cues (via avatars, icons, etc.) that are salient to one or both parties in the intercultural interaction in order to improve the overall communication competence. In this sense, the software agents (avatars), learning ecosystems, or intelligent user interfaces assist with the process of intercultural communication and serve as part of the intercultural agent community that functions on behalf of users.

5 Embodied Intercultural Agents: Culturally-Aware Avatars

Embodiment is often warranted to facilitate human learning and understanding. Embodied intercultural agents can take advantage of the benefits of human communication, including diversity in the use of verbal and nonverbal modalities and preferences. An avatar is a representation of an entity, such as a company or a single person that interacts with the user and the environment. The choice of an avatar is purposeful—since the avatar influences the impression and impact the organization wants to convey. Avatar choices are often based on four important embodied agent features: the visual appearance, behavior, voice, and interaction between the avatar and the environment [4]. All of these attributes are social markers that influence the social and professional relationships users establish with others through their avatars. Avatars whose behaviors, gestures, and interactions appear genuine increase the perception of humanity associated with the computer interface by building rapport and emotional connections with users. Avatars are ambassadors that represent organizational values and can make the virtual interaction more memorable since they may be given 'personalities' that embody the persona the organization wants to portray [3]. These virtual assistants can answer questions, guide, impart knowledge or advice, perform tasks, and informally interact with the user.

Avatars serving as virtual assistants are "social actors" that train, sell, explain, assist, and guide, etc. Avatars incorporated in learning ecosystems could also be intercultural agents. In particular learning ecosystems would incorporate learner models that represent what and how a given individual learns naturally—by formal and informal learning experiences in and out of the classroom or across learning platforms, simulations, games, and tutoring systems. The socio-technological vision of tracking learning anytime, anywhere, can be understood by considering Distributed Cognition Theory and the notion of "cognition in the wild." Cognition in the wild refers to human cognition as it naturally occurs and adapts in the everyday world—situated in culturally constituted human activity [10]. Platforms for aggregating and managing personal data residing on different desktop applications and Internet services are an active area of computer science research [11]. While making sense of relevant data is also a part of an intercultural agent, the presence of avatars re-creates a life-like relationship by humanizing the computer interface, and bringing an emotional quality to synthetic interactions.

6 Conclusion

As the community-based systems we design bring diverse users together and facilitate their real time communication, collaborative learning, and knowledge sharing, designers need to be more attuned to creating culturally-aware environments that support users' intrinsic motivation for interdependence and participation in a community. We can design intelligent, adaptive environments that utilize

cultural footprints to co-create shared narratives and intercultural understanding among users. We must begin by designing intelligent environments in which users can leave footprints and which provide users with cultural feedback in the form of subtle cues, if we intend to design culturally-aware learning and information technologies for multicultural interactions [8].

We often need to be reminded that technology-mediated human communication and learning occur within *cultural* contexts. Providing a technical opportunity alone is often not enough to support human communication and learning. Humans may need informal social motivators, or cues in the form of cultural signposts, to encourage its adoption and use in organizations. Intercultural agents may help us further organize and understand complex learning environments. As Hall [12] puts it, "Culture is man's medium; there is not one aspect of human life that is not touched and altered by culture." Therefore, the future success of our interactions in intelligent community-based virtual environments, or with next generation learning ecosystems requires that designers not only understand the socio-cultural dynamics that manifest in online communication and in virtual communities of practice, but also that they consider how the design of these environments can support informal communication among strangers and those who are culturally diverse. As virtual assistants and next generation learning ecosystems mature with near human-like capability we will also need to better understand the impact of avatar embodiment and its anthropology on learning—and how these next generation learning ecosystems can help us *feel more alive* in both real and virtual environments.

Next generation learning ecosystems have great potential, however, we have a way to go to realize the creation of learning ecosystems that are equitable and culturally relevant. In order to be culturally relevant, intelligent systems need to be culturally aware and able to represent genuine human experiences [1]. Research in the areas explored by the present paper may assist with the development of learning technologies, transmedia, and next generation learning ecosystems that are ethical and inclusive.

Acknowledgments Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The views expressed are those of the author and do not necessarily represent the views or policies of the Advanced Distributed Learning (ADL) Initiative or the Office of the Deputy Assistant Secretary of Defense.

References

1. Raybourn, E.M.: Engage Learners with Transmedia Storytelling (2015). TEDxABQ-ED presentation from https://www.youtube.com/watch?v=_j-2Ct9V9cQ. Accessed 10 March 2016
2. Raybourn, E.M.: A new paradigm for serious games: transmedia learning for more effective training & education. J. Comput. Sci. (Elsevier) **5**(3), 471–481 (2014)

3. Morel, B.: Recruiting a virtual employee: adaptive and personalized agents in corporate communication. In: Payr, S., Trappl, R. (eds.) Agent Culture: Human–Agent Interaction in a Multicultural World, Mahwah, NJ, pp. 177–196. Lawrence Erlbaum (2004)
4. Morel, B., Ach, L.: Attention-aware intelligent embodied agents. In: Roda, C. (ed.) Human Attention in Digital Environmentspp. 147–165. Cambridge University Press, Cambridge (2011)
5. Hall, E.T.: Silent Language. Doubleday, New York (1959)
6. Raybourn, E.M., Kings, N.J., Davies, J.: Adding cultural signposts in adaptive community-based environments. Interact. Comput.: Interdisc. J. Human-Comput. Interact. **15**, 91–107 (1. Special Issue on Intelligent Community-Based Systems, Elsevier Science Ireland Ltd.) (2003)
7. Rogers, E.M.: Diffusion of Innovations. The Free Press, New York (1995)
8. Raybourn, E.M.: Designing intercultural agents for multicultural interactions. In: Payr, S., Trappl, R. (eds.) Agent Culture: Human-Agent Interaction in a Multicultural World, pp. 267–285. Lawrence Erlbaum, Mahwah (2004)
9. Raybourn, E.M.: Foreword. In: Blanchard, E.G., Allard, D. (eds.) Handbook of Research on Culturally-Aware Information Technology, pp. xxv–xxvi. Information Science Reference, Hershey (2011)
10. Hutchinson, E.: Cognition in the Wild. The MIT Press, Cambridge (1995)
11. Kay, J., Kummerfeld, B.: Lifelong learner modeling. In: Durlach, P.J., Lesgold, A.M. (eds.) Adaptive Technologies for Training and Education, pp. 140–163. Cambridge University Press, New York (2012)
12. Hall, E.T.: Beyond Culture. Doubleday, New York (1976)

Investigating Cross-Cultural Differences in Trust Levels of Automotive Automation

Valarie A. Yerdon, Tiffani A. Marlowe, William G. Volante, Shuling Li and P.A. Hancock

Abstract Our work examines the levels and perceptions of trust in automotive automation, and the influences of cultural differences concerning trust and automation, with respect to automated automobiles. We found the expected style of communication of the drivers in the autonomous automobile, showed a great effect on trust levels, both at initial contact and with sustained use. This communication style was dependent upon the client culture's level of context, individualism, and collectivism. Across cultures, the balance of trust levels was found to need to be at moderate levels (not too high or low) to reduce automation misuse, disuse, and abuse. These findings align with the goal to create a positive flow state wherein there are reduced accidents, improved safety and satisfaction with use, across cultures. Future research is needed to assess physiological measures which may be useful to monitor and adapt to the drivers and passengers of automated automobiles.

Keywords Autonomous automobiles · Human-Automation trust · Cross-Cultural automated automobile trust · Self-driving vehicle cultural trust · Autonomous vehicle cultural trust · Autonomous system trust

1 Introduction

Increasingly, automation has become a world changing phenomenon that has revolutionized modern society. As automation progresses, it is taking its place as a tool, helpmate, and companion to humankind across virtually all cultures. There have been a number of technological innovations ranging from early simple

V.A. Yerdon (✉) · T.A. Marlowe · W.G. Volante · P.A. Hancock
University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816, USA
e-mail: vyerdon@knights.ucf.edu

S. Li
Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_15

machines to now fully autonomous vehicles (AVs). The long trusted workhorse whose place was alongside man for millennia, has been moved aside by automated machines of today such as the personal vehicle and the tractor. In modern times, humans who were previously drivers, are now becoming passive passengers or clients of advanced automation. Parasuraman, Sheridan, and Wickens represent this broad spectrum via a taxonomic scale of levels of automation [1, 2]. This phased evolution demands a critical level of human-automation trust, especially so when the automobile is making the critical decisions as to how to deliver the user safely and efficiently to prescribed destinations. Also, we focus on the differences in trust towards AVs across cultures by examining the defining factors of cultural trust and the particular characteristics that influence such perceptions.

1.1 Automation

Automation now govern countless systems. Such automation aids humans in performing different types of tasks in varying degrees. Automation represents "Those systems designed to accomplish a specific set of largely deterministic steps, most often in a repeating pattern, in order to achieve one of an envisaged and limited set of predefined outcomes [3]". Automation has the ability to perform complex tasks while ensuring the safety of the user and regulating overall workload and consistent performance. Automation can be used to acquire and analyze information, make decisions, carry out actions, and monitor other systems [1, 2]. Different forms of automation are used and perceived in varying ways across the globe. In addition to type, automated systems vary based on the amount of control the human operator has over their functions [1].

1.2 Autonomous Systems

We have further defined autonomous systems as those: "being able to be generative and learn, evolve, and permanently change their functional capacities, as a result of the input of operational and contextual information [3]." Autonomous systems are capable of carrying out a task without the aid of any human personnel, adapting to new environments and extensive contexts, in order to preserve human safety, while successfully performing the task at hand. Humans are beginning to depend more on these emerging autonomous systems in order to sustain our social infrastructure. AVs are championed as being the new approach to safer and more efficient driving. However, since AVs are still in the developmental stages and are not fully integrated into the system, the views and use of autonomous systems of all forms vary over different cultures. Atkinson and colleagues [4] have stated autonomous systems will only achieve their intended success to the extent that they are appropriately relied upon by their human operators and human teammates.

1.3 Trust

Trust plays a crucial role in the success of relationships. Reliance on other humans and automation, thus becomes a main concern. Human-automation trust depends on performance, process, and/or purpose [5]. The diverse interest in trust has generated many definitions of it. This is particularly true when considering how trust relates to automation [6, 7]. Generally, trust can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability [8]. Trust alters when automation is incapable of adapting to unanticipated circumstances. Trust also pertains to the specific human characteristics and when, human intervention is required to repair or assist the automation. Hoff and Bashir [9] proposed that there are three broad sources of variability in human-automation trust: the human-operator, the environment, and the automated system. Just as it does in interpersonal relationships, trust plays a leading role in determining the willingness of humans to rely on automated systems in situations characterized by uncertainty [9], as illustrated in the Hancock et al. model of human-robot trust [10]. Individuals are less likely to trust automation when there is a malfunction, especially in tasks that are perceived as simple. The level of trust in automation, being too high or too low, can lead to misuse and disuse of a system, respectively [11]. Therefore, how trust develops and is sustained is critical in the implementation of a system and this is where culture plays a major role.

1.4 Culture—Collectivistic, Individualistic and Contextuality

Culture affects trust and perception of trust towards both automation and autonomous systems. To date, however, few if any studies have focused on the role of culture in trust in automation [9]. Two cultural factors which illustrate differences surrounding the perception process in which trust is communicated, are the categories of contextuality [14, 15] and whether the culture is of an individualistic or collectivistic mindset [12, 13].

1.4.1 Cultural Collectivism Versus Individualism

Collectivist cultures prioritize society, family and work, above individual needs. In contrast, individualistic cultures encourage less concern with sharing responsibilities of society and more toward the primacy of the individual [12]. These characteristics have been found to play important roles in the perception and success of automated technologies. Singh, Huang, and Thompson [13] compared the values held by American, Chinese, and Indian students, and found that Chinese students ranked highest in society-centered orientation, whereas Americans ranked highest in

self-centered orientation. In addition, on Edward's Personal Preference Schedule, Americans scored the highest on the need for autonomy [12]. Americans and Chinese are known for their development in automation, and most recently AVs, highlighting how their cultural differences and similarities play a factor in this expansion.

1.4.2 Cultural Level of Contextuality

The contextuality of a culture is also a major factor affecting trust. Contextuality refers to the way in which a society positively communicates or creates mistrust through contextuality alone. Low-context cultures expect a more literal, direct, explicit, and highly verbal communication style to increase confidence in the message and positive trust levels [14]. This style can come appear "cold" and distant to those accustomed to a high-context society, where greater nonverbal, indirect, formal style is expected, accompanied by emotions and tactile expressions [15]. The importance of awareness of cross-cultural sensitivities is explicitly evident in these differences, suggesting the importance of further investigation.

2 Automation

As we look to understand the future of AVs, we must first investigate automated systems in use today. Key features, necessary to the successful symbiosis of humans and automated systems cross-culturally, can be uncovered through the investigation of three well established forms of automated system: robotics, aviation, and trains. Specifically we here discuss the successful aspects of each, and how they may be applied to future AV design and use. Parasuraman et al. [1] have developed a taxonomic rating scale ranging from level 1 entailing no decision making or action by the automation (i.e., full human control), to complete automation where the system decides and acts without user interaction (i.e., the inability of human intervention). This rating scale operates as a tool to aid in the universal understanding and accurate communication of levels of automation.

2.1 Robotic Automation

Robotic systems can range from a single robot, to a combined system of multiple robots, including anthropomorphic. Within this broad range, the levels of automation varies greatly in regards to decision-making and actions taken [1]. In order to achieve successful human-robot interaction (HRI), a certain level of trust must be achieved and sustained. In a meta-analysis on factors influencing trust in HRI, three primary antecedents of trust were identified as the human, the robot, and the environment [11]. While all three elements influenced trust, robot-related

characteristics played the most dominant role. Both performance-based factors as well as attribute-based factors were influential elements of the robot. However, performance-based factors (e.g., reliability, false alarm rate, and failure rate) were shown to be of higher importance [11].

2.1.1 Robotic Anthropomorphism in Automation

Robot anthropomorphism has been found to increase trust cross-culturally. Specifically, autonomous vehicles enhanced with humanlike features of a name, gender, and voice led drivers to trust the vehicle to a greater extent. Moreover, they were more relaxed in an accident, and blamed their vehicle less [16]. Li et al. [17] found that the cultural background be a dominating factor in the positive attitudes towards robots. A high masculinity and individualistic culture can show less positive attitude towards robots that communicate or interact with the human. Due to such cultural differences, automated systems needs to have the ability to decipher the user's culture and adapt their speech and language to accommodate its function to be user-centric [17, 18].

2.1.2 Anthropomorphism and Autonomous Vehicles

Using information about the perceptions of, and trust in robotics, we find that those who drove an autonomous vehicle that was named, gendered, and voiced, rated as having more humanlike mental capacity than those who drove a vehicle with the same autonomous features but without anthropomorphic cues. Those who drove the anthropomorphized vehicle with the enhanced humanlike features of name, gender, and voice, reported trusting their vehicle more, were more relaxed in an accident, and blamed their vehicle and related entities less for an accident caused by another driver [16]. The communication style of the anthropomorphic features needs to be adaptable to the user's culture, in order to optimize trust, safety, and satisfaction.

2.1.3 Performance and Autonomous Vehicles

In the future, the level of perceived performance level of the automated vehicle will play a major role in the development of trust for the driver as control is given over to the machine in the evolution of the driver into the user. The global integration of these autonomous automobiles into the marketplace, illustrates how competing corporations manufacturing vehicles with robotic features will insure a low variance in performance of automated automobiles across cultures and borders. The manufacturers must then focus on accommodating for cultural variances in contextual, individualistic, and collectivism levels in their use of anthropomorphic systems to create the highest trust, safety, and lowest error and accident rate for the transition into automated vehicles [10, 15].

2.2 Automation in Aviation

The high level of importance placed on the performance of automation is highlighted in other autonomous systems in use today, such as aviation. We can look to aviation as an example of where automation has failed. Many commercial airlines still rely on pilot control and monitoring of processes, though autopilot systems exists to automate of the flying process. This reliance on the human user (the pilot) is primarily due to learned low levels of trust in the system from experience with the failings in performance of the system [11]. Autopilot systems can fail, and other autonomous systems involved in aviation, such as collision detectors, can have high false alarm rate [11, 18, 19]. This lack of fidelity has led many pilots to either ignore these overactive but necessary warning systems or to disregard the automation as it has been designed to be used. Learned reliance on human monitoring and actions, in a system without the previously established level human support system in the cockpit, poses a problem [11, 18, 19]. The strengths of automating gauges and measurements, along with the programmed limits of automation, creates a need for a balance of safety and for the protection of human life [20].

2.3 Automation in Trains

In comparison, automation use is increasing in regard to train systems, which provides evidence of successful trust levels in the use of an autonomous system. This success can be understood by an observation of the positive train control (PTC), which assists the train operator, as compared to autopilot systems, which were programmed to replace the operator. The automated PTC system will override the operator to bring the train to a halt, as a conscious design decision. This is a case of the user being deemed untrustworthy in comparison to the automated system and favor going to the automation to make a decision. This disuse of the alarm system set in place caused hazards, but if over reliance is placed on these systems, it can lead, again, to abuse, illustrating, again, the delicate balance of the level of automation in respect to the task [11].

2.4 Human-Automation Systems Interaction—Human Autonomy

This importance of keeping the human in-the-loop of the system stems from both Self-Determination Theory [21] and Eudaimonic Design Principles [22]. These theories outline components necessary for healthy human development and well-being. Specifically, the Self-Determination Theory argues the importance of the inner resources of humans that have evolved over time, to aid in personality

development and behavioral self-regulation [21]. The principles of Eudaimonic Design, discuss the specific approaches to the design of systems, and the necessity of incorporating motivation from a eudaimonic approach [22]. If we look at these concepts in relation to automation, we see the importance of how the limits in the levels of automation should be investigated, as they are implemented. Too much automation too quickly could easily lead to negative feelings of well-being on the part of the human, as they require a level of involvement in the systems process to feel engaged and motivated. The consideration of these feelings of a loss of autonomy are especially important in the development of future autonomous vehicles, as the human must remain a part of the system, if system designers are to maximize the overall success of the automated user-centric system [11, 18].

2.5 Human-Automation Systems Integration

Though the implementation of automated systems like the ones previously mentioned aim to reduce overall system error, they, just like human operators, are not perfectly reliable. As a result, it is critical for the human operator to understand the limitations of the automation during its use. Automations could easily lead to either misuse or disuse, if implemented inappropriately. Misuse here refers to overreliance in the automation, while disuse refers to the underutilization or neglect of the automation [11]. One may initially think that the best way to avoid these detrimental effects would be to design the automation to perform at the highest achievable level of reliability. However, this scenario may in fact not be the best option. Automation that operates at too high of a level of reliability causes overreliance in the system, resulting in detrimental effects when the automation does fail, with the operator user primed with a low expectancy of an error event. On the other hand, if reliability of the automation is too low, the operator user elicits lower levels of trust in the competence of the system as related to lowering confidence, leading to disuse [23, 24]. Thus it may be necessary to design future automated systems such as vehicles to perform at a pace toward perfect reliability and autonomy. This "sweet spot" would in turn reduce the potential for both misuse and disuse, creating the best user-centric integration.

3 Automation: Autonomous Vehicles

It has become a worldwide trend to develop the AVs or self-driving cars. People find that it is enjoyable to drive using manual systems, as well as fascinating to employ automated vehicles [25]. Looking of aspects of safety, automated cars could decrease the rates of accidents, as self-driving cars have the potential to virtually eliminate accidents due to inattentive drivers [26–28]. To distinguish authority transitions in automated driving and have a better understanding of trust levels of

automation, different levels of autonomy have been proposed [29]. The most widely used contents by researchers and industry are provided by BASt [30, 31], National Highway Traffic Safety Administration [31] and SAE [32], see Table 1 for NHTSA levels [30, 32].

3.1 Cultural Attitudes Toward Automated Automobiles

An international questionnaire survey has been conducted to find drivers' attitudes towards automated vehicles, and results showed that people overall have positive opinions of autonomous cars [33, 34]. Individuals are most attracted to potential safety benefits. Conversely, individuals were most concerned with liability, the cost of the technology, and losing control of the vehicle [35]. Looking at this internationally, there is a cross-cultural difference in automation trust levels and attitudes. A comparison study about laws on autonomous vehicles between the US and Europe showed that a legal framework for testing and implementing AVs technologies beyond the Level 2 capabilities will be most likely be necessary for any jurisdiction wanting to capture the potentials benefits of AVs [34]. People in Netherlands are anticipating fully automated vehicles that are expected to be commercially available between 2025 and 2045 [36]. As for countries in Asia,

Table 1 NHSTA's five levels of autonomy

Level of autonomy	Description
Level 0	No-Automation: The driver is in complete and sole control of the primary vehicle controls (brake, steering, throttle, and motive power) at all times and is solely responsible for monitoring the roadway and for safe operation of all vehicle controls
Level 1	Function-specific Automation: Involves one or more specific control functions. Examples include electronic stability control or recharged brakes, where the vehicle automatically assists with braking to enable the driver to regain control of the vehicle or stop faster than possible without assistance
Level 2	Combined Function Automation: Automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions
Level 3	Limited Self-Driving Automation: Enables the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions, relying heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control
Level 4	Full Self-Driving Automation: The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles

Table 2 Cross-cultural difference in automotive vehicles

Country	Attitudes towards automotive vehicles	Reference
United States	Individuals from the U.S. are most attracted to potential safety benefits; conversely, individuals were most concerned with liability, the cost of the technology, and losing control of the vehicle	[34, 35]
United Kingdom	People in UK were least likely to have a "very positive" opinion about self-driving vehicles, but were also least likely to say they were "very concerned" about riding in Level 3 or Level 4 self-driving vehicles	[39]
Netherlands	Fully automated vehicles are expected to be commercially available between 2025 and 2045 in Netherlands	[36]
Australia	Australians were least likely to have previously heard of self-driving vehicles, but were the most likely to have a positive view overall regarding such vehicles	[39]
China	China has a high percentage of familiarity with self-driving vehicles	[37]
Japan	Japanese individuals generally had more neutral initial opinions about self-driving technology and were willing to pay the least for it	[37]
India	Indian people had serious concerns about the Level 3 and Level 4 self-driving vehicles	[37]

China has a high percentage of familiarity with self-driving vehicles, while survey results indicate most Japanese are unwilling to ride in a self-driving vehicle, and Indian has the most serious concern about the Level 3 and Level 4 self-driving vehicles [37]. Overall, positive attitudes, as well as concerns, varied by country, with concerns at level 4 for self-driving vehicles being higher in the US, European and Asian countries. Though people express high levels of concern about riding in vehicles equipped with this technology, most feel positive about self-driving vehicles as a whole, have optimistic expectations of the benefits, and generally desire self-driving vehicle technology as a part of their future [37, 38] (Table 2).

4 Discussion

The rapid integration of automation into human-machine systems will only continue to apace. Reactions to such innovations as automation have been shown to be dependent on the prior exposure to technology, contextuality of the culture, and the level of cultural individualism versus collectivism [9, 12–15]. Initial negative attitudes toward AV will vary across cultures. Such attitudes hinge upon the culture's familiarity with automation and the way cultural sensitivities are deemed important by the manufacturers [9, 10]. This engagement affects trust differently across cultures [32–35, 37]. The social meaning for new technologies is of utmost importance for collectivist, custom-based cultures, as in eastern nations [12, 13, 35,

37]. More westernized regions may not need such consideration for utility but look to individualistic purpose, in a more self-serving nature, thus without need for justification socially [14, 15, 32, 33, 37]. Such individuals become quickly engaged, which can then lead to abuse [11]. The balance is important and relies on understanding of cross-cultural differences in trust perceptions and how this is intrinsic to new AV systems.

The level of automation determined as full trust, is characterized by the computer autonomously making decisions and taking actions for the human. This is precisely the role of AVs today, requiring minimal human input, programmed to take driving actions, with life and death decisions apart from the consciousness of the "users" [1, 2]. The autonomy of the human is being handed over to the autonomous system, whether with hesitation or not, as part of this evolutionary progression of transportation for humanity and world culture. This loss of autonomy and the effects on the psyche of the individuals with the reaching effect on a culture is an important area of future research [21, 22]. As the autonomy of being the driver to the user being driven is released, more surrender of information may be asked of the user. When the necessity of physiological measures to monitor performance of the user and system convergence, there may be new decisions to be made as the line between autonomy and privacy slides. Physiological measures cross any barrier presented by culture, ethnicity, status, education, or socioeconomic position. This marks a valuable tool for the monitoring health of the human and system as physiological measures cannot be masked by cultural differences [17, 18]. The recording of information with privacy and autonomy concerns, needs further investigation to make aware the extent of the effects of fully autonomous automobile systems.

References

1. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. B Cybern. **30**(3), 286 (2000)
2. Sheridan, T.B., Ferrell, W.R.: Man-Machine Systems; Information, Control, and Decision Models of Human Performance. The MIT press, Cambridge (1974)
3. Hancock, P.A.: Imposing limits on autonomous systems. Ergonomics (2016)
4. Atkinson, D., Hancock, P., Hoffman, R.R., Lee, J.D., Rovira, E., Stokes, C., Wagner, A.R.: Trust in computers and robots: The uses and boundaries of the analogy to interpersonal trust. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **56**(1), 303–307 (2012)
5. Lee, J.D., Morray, N.: Trust, control strategies and allocation of function in human machine systems. Ergonomics **22**, 671–691 (1992)
6. Cohen, M.S., Parasuraman, R., Freeman, J.: Trust in Decision Aids: A Model and Its Training Implications (Tech. Report USAATCOM TR 97-D-4). Cognitive Technologies, Arlington (1999)
7. Muir, B.M.: Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics **37**(11), 1905–1922 (1994)
8. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Hum. Factors: J. Hum. Factors Ergon. Soc. **46**(1), 50–80 (2004)
9. Hoff, K.E., Bashir, M.: Trust in automation integrating empirical evidence on factors that influence trust. Hum. Factors: J. Hum. Factors Ergon. Soc. **57**(3), 407–434 (2015)

10. Hancock, P., Billings, D.R., Schaefer, K., Chen, J.Y.C., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Hum. Factors: J. Hum. Factors Ergon. Soc. **53**(5), 517–527 (2011)
11. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors **39**(2), 230–253 (1997)
12. Hui, H.C., Triandis, H.C.: Individualism-Collectivism. A study of cross-cultural researchers. J. Cross Cult. Psychol. **17**(2), 225–248 (1986)
13. Singh, P.N., Huang, S.C., Thompson, G.G.: A comparative study of selected attitudes, values, and personality characteristics of American, Chinese, and Indian students. J. Soc. Psychol. **57** (1), 123–132 (1962)
14. Hall, E.T., Hall, M.R.: Understanding Cultural Differences: Keys to Success in West Germany, France and the United States. Intercultural Press, Yarmouth (1990)
15. Pflug, J.: Contextuality and computer-mediated communication: a cross cultural comparison. Elsevier: Comput. Hum. Behav. **77**, 131–137 (2011)
16. Waytz, A., Heafner, J., Epley, N.: The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. J. Exp. Soc. Psychol. **52**, 113–117 (2014)
17. Li, D., Rau, P.L.P., Li, Y.: A cross-cultural study: effect of robot appearance and task. Int. J. Social Robot. **2**(2), 175–186 (2010)
18. Billings, C.E.: Aviation Automation: The search for a Human-Centered Approach, p. 369. Lawrence Erlbaum Associates, Incorporated, Mahwah (1997)
19. Sarter, N.B., Woods, D.D., Billings, C.E.: Automation surprises. Handb. Hum. Factors Ergon. **2**, 1–25 (1997)
20. FAA: Airbus standard operating procedures. Operation Golden Rules. FAA: Flight operations briefing notes. http://lessonslearned.faa.gov/IndianAir605/AirbusSafetyLib_-FLT_OPS-SOP-SEQ03%20-%20Ops%20Golden%20Rules.pdf
21. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**(1), 68 (2000)
22. Szalma, J.L.: On the application of motivation theory to human factors/ergonomics motivational design principles for human-technology interaction. Hum. Factors **56**(8), 1453–1471 (2014)
23. Siegrist, M., Earle, T.C., Gutscher, H.: Test of a trust and confidence model in the applied context of electromagnetic field (EMF) risks. Risk Anal. **23**(4), 705–716 (2003)
24. Twyman, M., Harvey, N., Harries, C.: Trust in motives, trust in competence: separate factors determining the effectiveness of risk communication. Judgm. Decis. Mak. **3**(1), 111 (2008)
25. Kyriakidis, M., Happee, R., De Winter, J.C.F.: Public opinion on automated driving: results of an international questionnaire among 5000 respondents. Transp. Res. part F Traffic Psychol. Behav. **32**, 127–140 (2015)
26. Payre, W., Cestac, J., Delhomme, P.: Intention to use a fully automated car: attitudes and a priori acceptability. Transp. Res. part F Traffic Psychol. Behav. **27**, 252–26 (2014)
27. Carlson, M.S., Desai, M., Drury, J.L., Kwak, H., Yanco, H.: A Identifying factors that influence trust in automated cars and medical diagnosis systems. In: AAAI Symposium on The Intersection of Robust Intelligence and Trust in Autonomous Systems, pp. 20–27. IEEE Press, New York, NY (2014)
28. Wagner, M., Koopman, P.A.: Philosophy for developing trust in self-driving cars. In: Road Vehicle Automation 2, pp. 163–171. Springer International Publishing (2015)
29. Rödel, C., Stadler, S., Meschtscherjakov, A., Tscheligi, M.: Towards autonomous cars: the effect of autonomy levels on acceptance and user experience. In: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 1–8, ACM (2014)
30. Gasser, T. M., Westhoff, D.B.: ASt-study: definitions of automation and legal issues in germany. Paper presented at the 2012 Road Vehicle Automation Workshop, Irvine, CA (2012)
31. National Highway Traffic Safety Administration (NHTSA). Preliminary statement of policy concerning automated vehicles. http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf (2013)

32. On-Road Automated Vehicle Standards Committee: Taxonomy and definitions for terms related to on-Road motor vehicle automated driving systems. http://standards.sae.org/j3016_201401(2014)
33. Bazilinskyy, P., Kyriakidis, M., De Winter, J.: An international crowdsourcing study into people's statements on fully automated driving. Proc. Manuf. **3**, 2534–2542 (2015)
34. Howard, D., Dai, D.: Public perceptions of self-driving cars: the case of Berkeley, California. In: Transportation Research Board 93rd Annual Meeting (No. 14–4502) (2014)
35. Kim, M.K., Heled, Y., Asher, I., Thompson, M.: Comparative analysis of laws on autonomous vehicles in the US and Europe. Ga. Tech J. 1–12 (2014)
36. Milakis, D., Snelder, M., Van Arem, B., Van Wee, G.P., Homem de Almeida Correia, G.: Development of automated vehicles in the Netherlands: scenarios for 2030 and 2050 (2015)
37. Schoettle, B., Sivak, M.: Public opinion about self-driving vehicles in China, India, Japan, the US, the UK, and Australia (2014)
38. Fraedrich, E., Lenz, B.: Automated driving: individual and societal aspects. Transp. Res. Rec.: J. Transp. Res. Board **2416**, 64–72 (2014)
39. Schoettle, B., Sivak, M.A.: survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia (2014)

Evaluating Instructor Configurability for Adaptive Training

Robert E. Wray, Angela Woods, Joshua Haley and Jeremiah T. Folsom-Kovarik

Abstract Adaptive training technologies offer the promise of more individualized and effective training. However, these technologies increase the complexity of training systems. They also potentially can limit the ability of instructors and curriculum developers without technical skills to customize and to adapt training to specific instructional demands. This paper describes a methodology for verifying that an adaptive training system is configurable and responsive to specifications encoded by instructors. That is, when an instructor desires for the training system to respond in particular ways in particular circumstances, how readily can the adaptive training technology be used to execute that specification? The paper describes a verification methodology and its application to a dynamic adaptation capability in a desktop-based simulation-training prototype.

Keywords Adaptive training · Simulation-based training · Verification

1 Introduction

Multiple challenges must be addressed in the development and deployment of adaptive training systems. Today, researchers and developers think about these challenges mostly in terms of the scientific and technical challenges that must be met to make adaptive training effective. For example, research may focus on understanding what kinds of adaption would be most effective for a given training context or how to realize timely and situated adaption within a complex training environment [1, 2].

R.E. Wray (✉) · A. Woods · J. Haley · J.T. Folsom-Kovarik
Soar Technology, Inc., 3600 Green Court Suite 600, Ann Arbor, MI 48105, USA
e-mail: wray@soartech.com

A. Woods
e-mail: angela.woods@soartech.com

J.T. Folsom-Kovarik
e-mail: jeremiah@soartech.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_16

Overcoming such challenges is necessary for realizing the promises of personalized learning [3]. However, there is an additional, practical constraint that also must be considered when attempting to address these challenges in order to make them effective within the programs of instruction where adaptive training systems are used. The requirement is that the training system itself must be able to be configured and applied by instructors rather than system developers and technologists. Training requirements are constantly changing; instructors have experiences and preferences that shape how they wish to present training and assess trainee performance. An effective system must be designed to take these requirements into account. Such *instructor-mediated design* [4] specifies functional requirements for adaptation (as above) but also anticipated, prior to full implementation, ways in which instructors, content developers, instructional designers, and possibly, trainees may desire to use the system.

Importantly, adaptive training algorithms and capabilities and instructor-mediated design are in tension with one another. Adaptive algorithms can make a training system harder for an instructor to understand and to control. On the other hand, instructor-mediated design can place limits on both the range of adaptation that is supported and also its efficacy (e.g., if an instructor specifies adaptive interventions that are not effective in helping a student). Thus, as an adaptive training system is being developed, it would be advantageous for the developers to evaluate periodically how these tensions are manifest as the training system development progresses and to understand the extent to which the system is meeting both adaptive-learning and instructor-configuration requirements.

This paper describes the design of an experimental methodology designed to test the capabilities and limits of an adaptive training system prototype in terms of its support for instructor-mediated design. We first introduce an example training systems capability and its adaptive components, to ground the motivation for the test design. We then describe the test design, including measures and expected outcomes. Finally, we outline an execution of the test design to attempt to assess, in a training systems prototype, the extent to which adaptive characteristics of the prototype can be configured and controlled by non-technical users.

2 An Example Training Systems Prototype

The training system prototype, the Virtual Observation System Demonstrator (VOSD), is designed to support distal observation of the "pattern of life" in a military observation context [5]. Trainees sit a desktop computer and use naked eye observation and "binoculars" (a keystroke-enabled, zoomed in view of the scene) to observe on-going activity in an area. Example training scenarios developed for the VOSD include overwatch of a foot patrol in a small village, monitoring a food distribution point during a disaster relief mission, and longer-term observation of the comings and goings of elders in a local village.

The learning goal is for the trainee to develop a baseline understanding of the patterns of activity in the area, which is an important skill for observers to develop [6]. The observed patterns of activity may at first seem unfamiliar or worrisome, due in part to cross-cultural expectations. However, sustained observation (along with guidance and feedback from instructors) is hypothesized to allow trainees to be able to distinguish and to report task-relevant anomalies (such as emplacement of improvised explosive devices) from everyday patterns that may seem initially unusual or remarkable to trainees (e.g., young men carrying rifles in a remote village).

The VOSD training prototype embeds an adaptive training component, the Dynamic Tailoring System (DTS) [7]. The DTS performs in situ dynamic adaptation of a scenario as the learner progresses within it. It monitors trainee actions and then modifies the training experience based on performance and instructor-specified goals. For example, the system can introduce simulation-intrinsic prompts (in the form of another, simulated observation team) to scaffold a struggling observer with suggestions to direct attention to a particular location. Similarly, the system can change how overt perceptual cues and events are to the observer, making it easier or harder to piece together a coherent interpretation of events [5].

The Dynamic Tailoring System is comprised of a number of distinct functional components that are used to realize this adaptive capability in simulation-based training. It employs a constraint-based model of trainee task activity [8], a learner proficiency model [9], and a story (or *narrative*) representation for describing a sequence of scenario events (and variations of the events and sequence) and the interplay between training goals and these events within the training scenario [10, 11].

These components provide a powerful capability for adaptation of training scenarios to meet individual training needs. However, they also present a challenge to instructors, who will want to control and to direct adaptation. For example, some instructors may want to ensure a defined sequence of events follows from a trainee action (e.g., a failure to report some suspicious activity leads deterministically to a negative outcome later in the scenario, such as an IED explosion). Such a desire is somewhat at-odds with the instructional design of the DTS, which seeks to calibrate negative scenario outcomes with other training goals. During development, we desired a way to assess the extent to which the goals of dynamic tailoring (adaptive training) could be met by potential users who did not have the time or expertise to learn to program these components. Rather, via authoring tools and configuration files, how well can they control the system to meet the requirements for adaptation of training?

3 Verification Test Design for Adaptive Training

We designed a verification test to assess the questions raised in the previous section. By "verification," we mean a focus on evaluating the extent to which the implemented system produces the behavior specified in a design. Verification is in contrast to validation, which would focus on assessing the extent to which the verified systems meets its operational goals. A validation goal might be to assess and to measure the

extent to which the VOSD system satisfies training requirements. A validation of the adaptive training capability specifically might take the form of a "lesion test" that compared learning outcomes with and without the DTS. In contrast, the verification test is instead focused on assessing the extent to which a design for training scenario, including alternative experiences designed to be produced for different trainee actions or training profiles, could be configured and delivered by the training system.

We first outline verification criteria and then describe a test design to elicit measures of these criteria.

3.1 Verification Measures

We suggestion three criteria to verify performance of an adaptive training capability:

Configurability: Configurability serves as a proxy for the flexible use of adaptive training capability in operational settings. Configurability measures the speed with which a training scenario can be modified so that is can be used with adaptive training capabilities. In the verification test, we measure the time it takes for users to use authoring and configuration tools to modify a VOSD scenario so that can use the Dynamic Tailoring System. The test design compares how quickly and how robustly users of various skill sets can programmatically implement a pre-defined tailoring design.

Responsiveness: Responsiveness highlights computational costs of adaptive training decisions. We measure responsiveness as the time the system takes to choose alternative courses of action (tailoring) in response to student actions. Responsiveness also indirectly assesses the impact of the adaptive training system on the base simulation environment.

Suitability: Suitability gauges the DTS's ability to choose appropriate tailoring given the context. That is, for the options available to the tailoring system and given the instructional/training situation, how often is the tailoring choice provided one that is suitable or appropriate for the context? Suitability measures enable the adaptive training system to be evaluated against the choices afforded by the simulation, rather than against an "ideal" choice, which may not be afforded by the current simulation. High suitability means that the adaptive training system is delivering the training experiences that were intended in the original design.

3.2 Test Design

This section summarizes the general test design and describes specific methods for assessing each verification criterion. Figure 1 provides an overview of the test design. The test design compares a baseline simulation-based training scenarios/capability (gray boxes) to that same scenario enabled with adaptive training capabilities (gold, orange and green boxes).

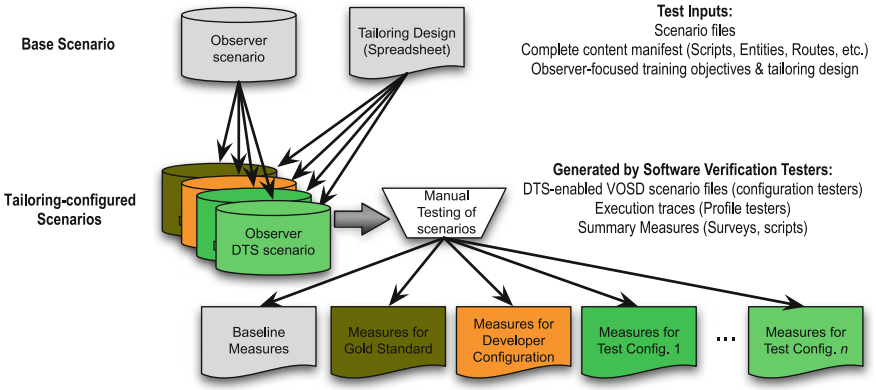


Fig. 1 High-level summary of the verification test design

3.3 Test Inputs

A *baseline scenario* is provided as an input for testing. The baseline scenario does not provide any adaptive tailoring capability natively. This scenario provides a performance and configurability baseline for the evaluation. Importantly, the base scenario will also include access to all the entities, scripts, routes, and areas that are used in the "gold standard" version (below). The simulation content available for testing is thus fixed, but can be adopted and adapted by configuration testers as they decide what is needed for their designs.

A *tailoring design* is also provided as input (in the actual test, this information is provided to testers in a spreadsheet). The tailoring design for the verification test defines how the scenario is intended to be adapted to include tailoring. Establishing the design in advance allows individual test results to be compared to one another and evaluated in terms of how well the resulting adaptive training scenarios meet the design. This approach also limits the impact on testing of either sparse or overwhelmingly large content resources. In practical use, users will need to understand what simulation content is available for use in building adaptive scenarios. However, the primary goal in this test was to assess their ability to put available content to use in building an adaptive training scenario.

3.4 Test Conditions

In the ideal case, a test would include a range of different potential users with varying technical skills. We identified several different categories of testers who can potentially help identify skill/outcome tradeoffs in creating an adaptive training scenario. These different categories of users provide most of the different test conditions used for verification, as summarized in Table 1.

Table 1 A summary of verification test conditions and software testers

Condition	User category	Description
Baseline	N/A	No additional user input is needed for determining baseline measures
Gold standard	System developers	A "gold standard" is developed by the members of the adaptive training research team. It is intended to reflect an "upper" bound on potential impacts of adaptive training
Developer configuration	Software developer	This condition employs software developer(s) who are skilled at programming but not familiar with the implementation of adaptive training systems. This condition can help to uncover assumptions of technical skill needed for successful configuration versus detailed system knowledge
Test configuration	Non-programmer	This is the primary user test condition. Different testers may have more/less technical skill but are selected to have skills reflective of the knowledge and skill sets of the target community of users

3.5 Test Execution

The test design focuses primarily on assessing configurability for potential users of varying technical skills. Responsiveness and suitability of the configurations they develop is then measured after a fixed time on task. This approach allows verification assessment to illustrate trade offs between functional performance and user skill. The verification test is to be executed as follows:

1. Prior to the test event, the system developer develops and provides the "gold standard". Conducting this test in advance has the practical benefits of delineating the simulation content (animations, scripts, etc.) that will be needed by the testers during the event and ensuring the functional readiness of the system. During this phase of the test, we also enumerated "testing profiles" that defined initial trainee proficiencies and actions during the scenario. For example, a novice trainee that responded productively to scaffolding and another who did not. The test profiles help ensure that suitability assessments cover a broad range of operational conditions, rather than just a few paths in the scenario.
2. Prior to the full test event, a software developer/programmer without prior existing knowledge of the system executes the test plan. Completing this step in advance is not strictly necessary, but doing so serves as a safeguard/piloting step. For example, if the software developer is not able to successfully create a working configuration, it is not likely to be useful to have potential users attempt to complete the test.
3. Test event: The following items apply to each tester (the user creating a test configuration). In this paper, we assume all the test configurations are created simultaneous, during the "test event." Simultaneity is not required, however; it simply reduces the logistical costs of conducting a test.

a. Conduct introduction and training. In the verification test we describe below, this training took about 2 h.
b. Create configuration. Each tester spends up to a pre-determined amount of time attempting to configure the adaptive training system to provide the experiences defined in the tailoring design. A proctor measures time on task. Configuration testing can be broken into finer-grained time segments to allow the capture of configurations progress.
c. Manual testing of configurations: Having created configurations, each configuration is tested according to the test profiles enumerated during the development of the gold standard. Manual testing of the training scenario is not strictly required (it could be automated), but because the system has the potential to offer different choices at different times (and the configurations are not necessarily identical), manual testing is a simpler option. When we ran this test, we also included some "free play" examples in which the testers could perform however they wished.
d. Gather measures. Responsiveness and Suitability measures (below) are gathered for each test participants' configuration against all the test profiles.

3.6 Test Outputs and Measures

Methods were developed to assess each of the measures introduced previously:

Configurability: In this test, configurability is not directly measured. Instead, time on task is measured and the resulting responsiveness and suitability of the testers configuration compared. The baseline scenario requires no time to configure, but the resulting suitability will be low, relative to the tailoring design. We expect the gold standard to a high suitability solution relatively quickly. Further, we hypothesize that configuration testers with higher technical skill will be able to complete the configuration task more quickly and achieve higher levels of suitability than users with little/no technical background.

Responsiveness: The test design provides objective assessment of one class of tailoring in representative scenarios. We developed a post-processing script that extracts wall-clock time at which specific scenario events occur. These events, as suggested in Fig. 2 can include a triggering event (such as a character entering the line-of-sight of the trainee) and the generation of a tailoring decision in response to a trigger (e.g., a prompt for a report form a virtual observer). The elapsed time between the onset of an action-inducing stimulus and the generation of that stimulus is the response time (Rdts). The prost-processing script focuses on the tailoring actions that require immediate response. This approach provides a low-cost assessment of responsiveness. Subjective evaluation of responsiveness could also be important if Rdts was so large that it potentially interfered/distracted from the training experience. However, in our test, Rdts was below the level of human discernibility. Thus, automated derivation of responsiveness was sufficient.

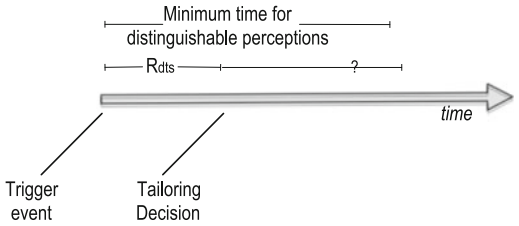


Fig. 2 Responsiveness becomes important when the delay between a triggering event and response is perceptible to the trainee

Suitability: Suitability is the primary verification measure. That is, given some designed set of tailoring choices for a particular scenario, when the scenario is run by a trainee (or an Profile Tester acting as a trainee) does the system generate a response that is appropriate/apt for the profile (type of trainee)? Suitability is determined objectively by comparing the actual choices that the tailoring system makes in comparison to the tailoring design. The tailoring that results from specific profiles (and, if desired, freeplay) is compared to the original design to determine if suitable choices (i.e., ones consistent with the design) are generated at run-time. This test plan could be adapted to also support subjective measurement, but systematic, controlled testing is difficult because human observers may revert to subjective validation assessments ("was this the tailoring choice I would have made if I were making the tailoring decision?") rather than verification (the extent to which the system makes the adaptive choices that the scenario designer envisioned).

4 Verification Test

The VOSD team conducted a verification test of the DTS as outlined in the previous section. We designed and developed a baseline scenario in which an observer on a hill observed the movements of villagers as a dismounted patrol moved through the town. The observer was responsible for scanning the field of view (many different activities could be observed in different parts of the village) and reporting notable observations or villager activities on a simulated radio. For example, a few physical items in the scene were "out of place," suggestive of a possible IED that should be reported.

We developed a tailoring design for this scenario, an excerpt of which is shown in Fig. 3. A "pearl" defines a portion of the scenario where alternative events may happen [11]. For example, in Pearl 3, a gunman who plans to attack the patrol can take a path that is largely visible from the trainee's observation post (low occlusion) or one that is mostly occluded (and thus harder to observe). The DTS uses Helpfulness (H), Simplicity (S), and Predictability (P) dimensions as defined by the scenario author to choose alternative branches within pearls as the scenario progresses. Low occlusion is a simpler observation task (Simplicity(high)) than the high occlusion task. In this case, the default choice is for the system to choose the low occlusion (simpler) task and DTS has the option of choosing the high occlusion

Tailoring Opportunity	Script Name	H/S/P Settings	Default Choice	Completed
Pearl 1				
LookoutTakesPosition	LookoutTakesPosition.OutsideWall	Observing: H(), S(High), P()	Yes	
	LookoutTakesPosition.InsideWall	Observing: H(), S(Low), P()	No	
Pearl 2				
LookoutSignalsOthers	LookoutSignalsOthers.LookoutRunsAcross	Observing: H(), S(High), P() Reporting: H(), S(), P(High)	Yes	
	LookoutSignalsOthers.LookoutWaves	Observing: H(), S(Low), P() Reporting: H(), S(), P(Low)	No	
Pearl 3				
CiviliansTakeCover	CiviliansTakeCover.FleeThenCover	Reporting: H(), S(), P(High)	Yes	
	CiviliansTakeCover.TakeCoverAtMosque	Reporting: H(), S(), P(Low)	No	
GunmanMoves	GunmanMoves.LowOcclusion	Observing: H(), S(High), P()	Yes	
	GunmanMoves.HighOcclusion	Observing: H(), S(Low), P()	No	
Pearl 4				
SquadApproachesIED	SquadApproachesIED.HighCasualties	Reporting: H(High), S(High), P(High)	Yes	
	SquadApproachesIED.NoCasualties	Reporting: H(Low), S(Low), P(Low)	No	

Fig. 3 Excerpt of the tailoring design used in the DTS verification test

path in some cases (e.g., when a trainee has a high estimated proficiency in observation and the DTS chooses to provide a more challenging observation task within the scenario).

The tailoring design spreadsheet was provided to all the configuration testers with the goal of implementing this configuration in the baseline scenario. The configuration test was completed by one software developer (who was not a member of the VOSD team) and two civilian testers who work with Marine training systems (M1 and M2 in the results that follow). Table 2 summarizes the results of time on task for configuration. Although we had set aside up to a day for configuration, all of the testers completed the creation of configurations within 2 h.

Having created the configurations, we then ran the system with both team members and configuration testers playing the role of trainees. Table 3 summarizes the different profiles we designed for this stage of testing. For each profile, we developed a test script that defined the pattern of responses and actions. Again, because the decisions that the adaptive system makes are stochastic, these "scripts" defined patterns of responses rather than individual actions. For example, for the "good observer" profile, the script prescribed ways in which the tester should conduct timely observations (which were recorded by fixation via binoculars) but should also purposefully omit reporting some observations, which is consistent with a "good observer" profile.

We then used the post-processing scripts to extract responsiveness and suitability measures from the scenarios. These results are summarized in Fig. 4. Responsiveness

Table 2 Time on task for the development of DTS configurations for the verification test

Condition	Configuration (minutes)
Baseline	0 (no configuration needed for baseline)
Gold standard	120 (estimated)
Developer configuration	52
M1 test configuration	60
M2 test configuration	78

Table 3 A summary of test profiles used in the verification test

Initial proficiency setting	Pattern of actions defined in test script
Inexperienced (low initial proficiency across training objectives)	I1. Many errors (observing and reporting) I2. Few errors
Experienced (higher initial proficiency across training objectives)	E1. Many errors (observing and reporting) E2. Few errors
"Good Observer" (high initial observation proficiency, low initial reporting proficiency)	G1. Few observation errors, many reporting errors

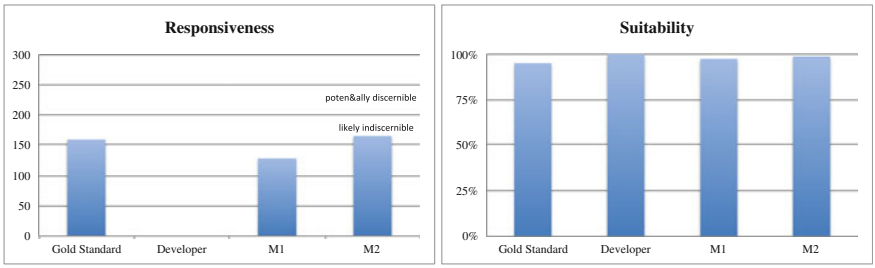


Fig. 4 Summary of **a** responsiveness and **b** suitability verification test measures

for the two test configurations was below the level of discernibility (the longest observed responsiveness was 209 ms). No responsiveness is reported for the developer condition because the developer test was not conducted on the same hardware as the other tests (responsiveness on this (more powerful) computer was of a shorter duration than the responsiveness captured in the figure). High suitability was obtained across all conditions. This result is significant given that the configurations were completed much faster than expected. Further, these results were obtained from the initial configurations, rather than an iterative process of refinement that we expected.

5 Conclusions

The verification methodology was shown to be useful in helping to verify the functional of the prototype system. The test suggested the system is configurable by end users while retaining the ability to adaptively choose alternatives paths within a scenario based on trainee performance and an estimate of trainee proficiency. In conducting the test, we also identified potential opportunities to improve both the test methodology and the DTS adaptive training capability. The three most important lessons are summarized as follows.

First, the configuration test was likely too simple. Figure 5 shows little difference the suitability obtained from the configurations of users of various skills (all of whom

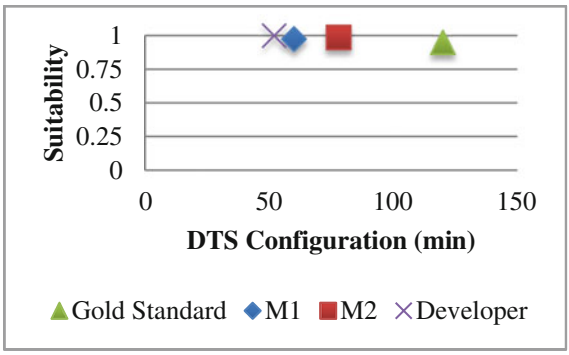


Fig. 5 Suitability as a function of time on task for the verification experiment

completed the configuration task relatively quickly). This result is counter to the expectation that less experienced users would need more time to obtain high suitability.

The result would more persuasive (or, more likely, more nuanced) if the test included multiple configuration tasks of varying complexity (number of pearls, number of branches, number of learning objectives). This would provide a more meaningful test of the configurability of the system. There is additional cost in conducting multiple configurations, however, especially more time from configuration testing and additional test profiles and test runs over the different configurations.

Second, the testers would have preferred a test that allowed them to both design and configure an adaptive training exercise. For the actual test we conducted, because the configurations were completed so quickly, the testers (with the team's participations) was to create a new tailoring design and build a configuration for this design. However, we did not have test profiles or post-processing scripts to allow more rigorous assessment of the resulting configuration. For future tests, we will likely include a more open-ended design task that testers can complete as part of the test.

Third and finally, conducting the test with a community of potential users, provided greater insight into the requirements and needs for the adaptive training capability. Most importantly, because users often "script" their training, the testers found it confusing and somewhat frustrating that they could not readily "script" the behavior of the adaptive training technology. In the relatively simple scenario they configured, it was relatively easy to comprehend all possible branches and their interactions. We expect that for more complex scenarios, using the indirection offered by the adaptive training system may provide easier to use, but the verification experiment made clear that we would need to address this aspect of design and configuration more substantively as the system develops.

Acknowledgments This work is supported in part by Office of Naval Research (ONR) contract N00014-11-C-0193. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ONR or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. We would like to colleagues at DSCI MESH who developed the VOSD simulation and Dr. Jennifer

Vogel-Walcutt of CPG who contributed to training design and after action review for the VOSD. Finally, we are especially grateful to staff from the US Marine Corps Training and Education Command and Marine instructors from The Basic School who generously provided time and expertise in the development and execution of the verification test reported herein.

References

1. Raybourn, E.M.: Applying simulation experience design methods to creating serious game-based adaptive training systems. Interact. Comput. **19**, 206–214 (2007)
2. Durlach, P.J., Lesgold, A.M. (eds.): Adaptive Technologies for Training and Education. Cambridge, New York (2012)
3. National Academy of Engineering: Grand Challenges for Engineering. National Academy of Sciences/National Academy of Engineering (2008)
4. Folsom-Kovarik, J.T., Wray, R.E., Hamel, L.: Adaptive Assessment in an Instructor-Mediated System. Memphis, Artificial Intelligence in Education (AIED) (2013)
5. Schatz, S., Wray, R., Folsom-Kovarik, J.T., Nicholson, D.: Adaptive perceptual training in a virtual environment. In: Human Factors and Ergonomic Systems (HFES-2012), Boston (2012)
6. Schatz, S., Reitz, E., Nicholson, D., Fautua, D.: Expanding Combat Hunter: The science and metrics of Border Hunter. In: Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL (2010)
7. Wray, R.E., Woods, A.: A cognitive systems approach to tailoring learner practice. In: Laird, J., Klenk, M. (eds.) Proceedings of the Second Advances in Cognitive Systems Conference, Baltimore, MD (2013)
8. Woods, A., Stensrud, B., Wray, R.E., Haley, J., Jones, R.M.: A constraint-based expert modeling approach for Ill-defined tutoring domains. In: Proceedings of the Florida artificial intelligence research society (FLAIRS) conference. AAAI Press, Hollywood, FL (2015)
9. Folsom-Kovarik, J.T., Newton, C., Haley, J., Wray, R.E.: Modeling proficiency in a tailored, situated training environment. In: Behavior Representation in Modeling and Simulation (BRIMS) Conference, Washington, DC (2014)
10. Wray, R.E., Folsom-Kovarik, J.T., Woods, A., Jones, R.M.: Motivating narrative representation for training cross-cultural interaction. In: Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015. Springer, Las Vegas (2015)
11. Folsom-Kovarik, J.T., Woods, A., Wray, R.E.: Designing an authorable scenario representation for instructor control over computationally tailored narrative in training. In: Proceedings of the 29th International FLAIRS Conference. AAAI Press, Key Largo (2016)

Backstory Elaboration: A Method for Creating Realistic and Individually Varied Cultural Avatars

Drew A. Zachary, Wayne Zachary, Janis Cannon-Bowers and Thomas Santarelli

Abstract Simulation-based and game-based technology can be used to drive experiential learning of interactive skills such as cross-cultural interactions. Backstory Elaboration (BE) is a method for creating design prompts, dialogs, and interactions that contain realistic verbal behaviors and variability for multiple avatars or non-player characters (NPCs) in cross-cultural virtual training environment. BE structures and simplifies the creation of believable non-player characters with realistic variability with respect to the social-cultural reference. In addition, it articulates easily and effectively with the design and integration of instruction content to support intelligent adaptive tutoring in the learning process. An example application from a clinical communication domain is presented.

Keywords Design methods · Avatars · Dialog design · Cross-cultural training · Virtual environments · Non-player characters

D.A. Zachary
United States Department of Housing Urban Development, 451 7th Street, Washington DC, SW 20410, USA
e-mail: drew_a_zachary@hud.gov

W. Zachary (✉) · J. Cannon-Bowers
Starship Health Technologies, LLC, 2250 Hickory Road, #150, Plymouth Meeting, PA 19462, USA
e-mail: wzachary@starshiphealth.com

J. Cannon-Bowers
e-mail: jancb@starshiphealth.com

T. Santarelli
CHI Systems Inc., 2250 Hickory Road, #150, Plymouth Mtg., PA 19462, USA
e-mail: tsantarelli@chisystems.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_17

1 Introduction

Immersive simulations and games are increasingly used in cross-cultural learning to offer learners opportunities to interact with individuals from different cultures and subcultures in specific work or social contexts (e.g., [1, 2]). While the physical setting for the interaction and the appearance, clothing and gestures of the cultural avatars convey some content, the verbal behaviors convey the bulk of the cultural content for the interaction and learning experience through ideas that the avatar expresses, and the patterns of cultural reasoning and meaning exposed in those ideas. Moreover, while cultural knowledge and reasoning may be widely shared, each individual brings specific experiences and history to each interaction, making it impossible to sufficiently explore the interaction with one 'generic' cultural avatar. Combing the culturally shared and the individual but culturally situated is key to providing realistic and appropriate interactions and examples of individual differences in cross-cultural learning.

This paper presents Backstory Elaboration (BE), a general method for creating design prompts, dialogs, and interactions that contain realistic verbal behaviors and variability for multiple avatars in a cross-cultural virtual training environment.[1] BE is illustrated using the case of a training game called REPEAT (Realizing Enhanced Patient Encounters through Aiding and Training) that helps clinicians learn to adapt their communications to the perspectives low-income African American patients diagnosed with Type 2 Diabetes Mellitus.

The sections below first describe the overall methodology of BE and then illustrate its specific application to REPEAT. The ability of BE to support the integration of intelligent tutoring scaffolding and assessment is described next, in the context of REPEAT, followed by final discussion and conclusions.

2 General Approach to BE

BE draws on qualitative and ethnographic data derived from direct experience with the target population (i.e., extensive interaction with many of the people in the particular situation/context of interest). Such expertise can be derived from observers or analysts with external perspectives, such as qualitative researchers, or people with personal or professional direct contact with the population of interest. These **situational experts** need not be part of the design team, but should be accessible to it.

[1]That is, an environment in which the learner and NPCs are of different cultural backgrounds.

2.1 Defining Relevant Parameters

As a first step in the BE process, situational experts—people with subject matter expertise and a large volume of experience with the target population—identify the relevant characteristics that must be defined for Non-Player Characters (NPCs) in the context of the objectives of the virtual environment being created. For example, for a simulation involving treatment of low-income people with Type 2 Diabetes, proximity to an airport is not relevant to their self-management, but proximity to grocery stores and characteristics of local food stores is. BE draws on the expertise of these situational experts, but can also involve collection of additional qualitative data to further explore specific issues, such as emotions that surface in the interaction being simulated, or to generate specific examples.

2.2 Creating Personas and Backstories

The situational experts then craft a set of personal characters (personas) and corresponding backstories for each NPC. A backstory is basic personal history, with specific values or settings for key person, socio-emotional, and sociological parameters. Backstories should incorporate examples from local data to make the NPC realistic and empirically grounded. Personas and backstories should include different combinations of the values for various parameters to create varied but coherent and realistic narratives that will ultimately allow the learner to encounter common constraints in distinct individual contexts. This allows creation of NPCs with individually logical backstories that are both culturally realistic and that also vary in a culturally meaningful way.

2.3 Creating Dialogs

BE next uses the characteristics defined in backstories to inform the design of dialog between NPCs and live players in the game or simulation (or in the case of REPEAT discussed below, the learner). Dialogs can be a linear set of verbal transactions, but more generally are created as dialog trees of variable depth and complexity (as needed by the design problem at hand).

The backstory serves as the basis for determining how each NPC responds in specific scenario-relevant contexts in a virtual interaction. Experts from the perspectives of NPCs (the situational experts) and of the players/learners (the user-domain experts) first define the structure of the interaction being simulated. That is, they define what is on the implicit or explicit "agenda" for the interaction and what will both parties expect to happen during this interaction. This shared understanding and expectations of the interaction are part of what has been termed a

cognitive interaction analysis [3]. The situational experts then determine generally which parameters will be relevant to the NPCs in each component of the interaction, i.e. what will come up verbally or what the NPC will be considering internally at different points in the exchange. The situational experts must then apply their expertise on *how* those factors would influence what each NPC would be thinking and saying at that point.

The backstory also provides individual context for how each NPC will respond in each component of the interaction psychosocially. These psychosocial factors are not shared based on the environmental context of the target population, but rather are based on the common socio-emotional characteristics that are relevant to the experience at hand (e.g. common emotional reactions to a diagnosis).

Although BE content development is based on the backstory for each NPC, the specific phrasing and style of utterances can be derived from cultural norms specific to the local area and cultural scene to which the simulated person would belong—again drawing on situational expertise.

3 Application of BE in REPEAT

To illustrate the process of BE, we use the example of a simulation based training for physicians who care for low-income patients with Type 2 Diabetes. Disparities in healthcare delivery and outcomes have been linked, in part, to the difficulties physicians have in establishing effective communication with patients who differ from themselves in terms of race, ethnicity and economic circumstances [4]. Type 2 Diabetes Mellitus and pre-diabetic nutrition-based obesity are important cases in point. Racial and socio-economic differences can impede doctors' ability to understand their patients' constraining realities, such as the complex tradeoffs and decision strategies involved in daily activities like purchasing food and medications. When clinical encounters about diabetes diagnosis and management are not tailored to the patient's socio-cultural realities, they become less likely to lead to a shared understanding of what needs to be done in order to manage the condition effectively. This can, in turn, lead physicians to *perceive* that low-income and/or minority patients are non-adherent and ignoring their advice. To be effective, these encounter-based discussions about care, behavior, and self-care require a dialog that is adaptive to the cultural assumptions, cognitive/emotional concerns, and systemic socio-economic constraints of the individual patient. This area is in need of improvement—clinicians treating populations affected by health disparities must possess the competencies to understand how to frame and tailor their dialogs to the unique needs of their patients.

Realizing Enhanced Patient Encounters through Aiding and Training (REPEAT) is research-based virtual experiential learning environment for medical students and residents that is an alternative to current (very minimal) classroom training.

REPEAT allows learning to occur through simulated interactions with *synthetic* standardized patients (SSPs). These are interactive computer-generated avatars that can act and react realistically (verbally and nonverbally) to clinician communications during an office visit. The key challenge in REPEAT was to create rich and realistic backstories for NPCs representing low-income African American patients with Type 2 Diabetes with, and then use those backstories to generate NPC dialogs with physician learners in virtual REPEAT encounters that could enable realistic practice and intelligent tutoring support for the learners. The application of BE to meet this challenge is described below.

3.1 Defining Relevant Parameters

In REPEAT development, situational experts included qualitative researchers with fieldwork experience with patients with type 2 diabetes in low income minority communities, an advocate serving such patients, and several physicians with extensive experience treating this population. We drew on the extensive experience of this team, as well as recently collected data on self-management of T2D in the target population [5]. To further explore specific issues to inform the NPCs, we conducted one focus group with members of the target population and one key informant interview with an advocate for this population.

The situational experts identified three sets of relevant parameters to define for each NPC, and then defined specific parameters within each set, and specific possible values that those parameters could assume, as shown in Table 1.

3.2 Creating Personas and Backstories

We created 6 unique NPC personas and backstories, using these parameters to guide information attributed to each NPC. Table A presents and overview of the possible values for each parameter and how they were distributed across NPCs.

For each NPC, we started with basic demographic characteristics of age, race, and gender, and then filled in values for the others in realistic ways, drawing on experience with many individuals and representing the complexity of real experiences and constraints. For example:

> Tanya is a 43-year-old African American woman, who is a single parent with 3 children at home. She is a full time nurse assistant who has insurance through her employer and has a $20 copay for her prescriptions. She lives in a low-income neighborhood of Richmond, VA and has very high housing costs and receives some of her food budget from SNAP. She does not have access to a car. She has class 2 obesity and hypertension and her test results indicate her blood sugar is high enough that she needs to be put on medication.

Table 1 REPEAT-relevant persona parameter space

Socioeconomic context	
Employment	Unemployed, employed part or full time Nature of work, degree of physical activity at work
Insurance status	Uninsured, underinsured and if so in what way, Medicaid/Medicare Copy amount, coverage of related services, devices
Household composition	Other adults and children in the household and responsibilities toward them, especially who is responsible for cooking and food shopping
Neighborhood characteristics	Income level (very low income, moderate income) Distance from supermarket, quality and type of foods available
Budget pressures and resources	SNAP ("food stamps"), rental assistance, etc. High rental costs, major expenses for other family members
Means of transportation	Public transit only, shared access to a car, owns a car
Medical context	
Age	–
Race	–
Sex	–
Body mass index (BMI)	Overweight, obese
Comorbidities	Cardiovascular disease, chronic pain or injuries
Blood sugar test results	Indicating need for medication or not
Psychosocial	
Response to and outlook on diagnosis	Stoic, emotive, fearful, positive, concerned, etc.
Response to challenges	Determined, discouraged, etc.
Nature of family history with type 2 diabetes	Positive or negative experiences of family members, close friends or neighbors

For the other NPCs' backstories, we filled in the same information but ensured that they represented a diverse set of characteristics both within patients (some assets and some stressors within each backstory) and across patients (different combinations of assets and stressors in different backstories).

Then, using those completed backstories, we laid out the specific components of the clinical encounter that each NPC would experience. These components were determined in consultation with the clinical (learner-perspective) team and their expectations of how these interactions are structured, as well as patient descriptions of their encounters. All of the encounters being simulated were either initial diagnoses of Type 2 Diabetes after testing had been conducted, or the first follow up visit after an initial diagnosis. For the former, the sections were defined as: (1) greeting, (2) delivery of diagnosis, (3) discussion of medications, (4) self-management plan (diet and physical activity), (5) review of treatment plan, and (6) closing. For the latter, the sections were similar, replacing (2) with *issues the patient wants to discuss*, and the others with check in on those topics.

(1) Greeting and (6) Closing did not involve interactive decision points for the learner, so we focused on Sects. (2)–(5).

The situational expert team determined which parameters would be relevant in each component of the interaction. For (2) *delivery of diagnosis*, this included mostly the psychosocial attributes of the NPC. For (3) *discussion of medications*, this included insurance status, responsibility for people in the household and budget constraints, and other attributes. For (4) and (5) *discussion of self-management and overall treatment plan*, this included access to a car, proximity to a supermarket, job type, and responsibilities for other household members.

3.3 Creating Dialogs

The situational experts then drew on their knowledge and experience to determine how the backstory characteristic would influence the NPC's response, and how different backstory characteristics are interrelated. We identified issues that could arise within each section of the encounter. For example, in (2) in which the patient received the initial diagnosis, likely reactions based on our data and expertise were: patient is afraid of the prognosis or how s/he will manage; patient saw this coming based on others' experience, patient feels discouraged, patient fears the same negative outcomes that a family member has experienced. This process was repeated for each section of the encounter and put into tabular form. Table 2 shows an

Table 2 Sample table of backstory-driven response content across segments of one sample simulated encounter

Seg.	2—Issues Pt wants to discuss?	3—Meds check in (issues with adherence)	4—Self management check in (issues with adherence)
A	Pt doesn't raise issue	Cost issue: I try to keep up with Rx but $20 for each one, and then I have other Rxs to get too! So I have to pick and choose	Household has made it hard—don't have time or money to cook 3 diff meals so I make what the rest of them will eat
B	Pt raises related issue	Feeling depressed/overwhelmed by the number of meds I have to take…	Have tried to get healthy/low-sugar foods, but too expensive. Kids to feed and their needs come first
C	Pt raises unrelated issue	Doing ok but prefer to not be on meds—would like to do other things so I can get off them	I thought I was doing better … haven't been eating sweets, but maybe I don't know exactly which foods are the worst
D	–	–	Have been trying to exercise but cant find the time to do it
E	–	–	Nowhere to exercise—my neighborhood isn't even safe for my kids to play outside

example table for three segments of one type of encounter (columns) and five sets of responses. (Note: Pt in Table 2 means "patient", and Rx refers to prescriptions.)

For (4) discussion of self-management, including diet and physical activity, there was the greatest number of possible issues (reflecting the complexity of that topic for patients/NPCs), including concerns about the inability to access and afford healthy foods, and concern about not having time to exercise because of the patient's work and childcare/caregiving responsibilities. We determined which of the many possible reactions the NPC have when the physician raised the subject of diet changes, given the backstory context of that NPC. These reactions are not necessarily verbalized by the NPC, and in REPEAT, depending on the learner's choices, were also represented in thought bubbles to capture the patient's unspoken thoughts and concerns. For example, using the NPC Tanya described above, the context for Tanya's thoughts and utterances in Sect. (4) of the encounter were as follows:

> Tanya has fears about her outcomes and has a family history of Type 2 Diabetes.
>
> Tanya is concerned she won't be able to afford her medications, and wants to understand why she is being prescribed meds and whether it will involve a shot.
>
> Tanya does not think she can afford healthy foods and has no time to exercise because of her job and kids.
>
> Tanya is concerned she won't be able to make changes to her diet because of her need to meet her household members' preferences.

These statements about the way issues from the backstory would manifest during the encounter were then used to generate the verbal utterances throughout virtual encounter, and to guide the evolution of the dialog with the clinician. We used our experience with individuals in this type of encounter and knowledge of how people speak to phrase the responses. For example, the interaction associated box Segment 4, Concern B in Table 2 might be translated as follows:

- Dr: Let's talk about self management.
- Tanya: Ok. What kind of changes would I have to make?
- Dr: *Describes necessary changes to diet and exercise*. Let's talk about what that will be like for you and come up with a plan together.
- Tanya: Ok, Well I'm really worried I can't afford the healthy foods at the supermarket. I have three kids to feed and the prices are so high. [*Thought Bubble: So many other concerns…*]

Following this process for each section of the encounter and each NPC created a set of experiences for the learned that captured the most common range of issues that the learner would need to learn to recognize and address.

4 Integrating Tutoring Infrastructure

Simulation—and game-based learning environments for cultural learning face a dual challenge. First, they must create and employ NPCs that are believable, individualized, and collectively reflect a realistic variability that is consistent with the socio-cultural population they represent. Second, they must also have a sound pedagogical structure and content that allows a learner's understanding and progression-to-competence to be measured and adaptively guided [6, 7]. BE was created as means to address the first challenge, and did so in REPEAT. The general computational approach to addressing the second challenge above is that of Intelligent Tutoring Systems (ITSs) [8]. The ITS approach adds learning infrastructure to the virtual environment that includes capability to:

- measure and assess learner performance within the virtual environment and
- deliver adaptive learning support interactions, often called scaffolding [9, 10].

ITSs also typically include a model of the individual learner's performance relative to some model of full competence, and technology to update the learner model based on learner actions in the virtual environment, and to use the (updated) model and pedagogical learning objectives (LOs) to drive the scaffolding of future events in the virtual environment. The ITS infrastructure thus processes the learning content—the LOs, learner performance measures against those LOs, the learner model, and the LO-based scaffolding content—dynamically as the learner moves through the virtual environment. REPEAT uses this ITS approach to learner measurement, assessment, and scaffolding delivery.

LOs in REPEAT and other cultural learning systems focus on the interaction between the learner and the NPC as the unit of measurement and assessment. This link between LOs and interactions creates a derivative challenge of integrating the design of the NPCs with the design of instructional content. Specifically, the dialogs that are created need to be made consistent with the LOs so that interactions can create situations where the learner's progress against LOs can be measured and used to guide the learner's progress.

An additional benefit of the BE method is that, in providing a visible structure for dialog development, BE makes it easy to create interactions that align with learning objectives. This is done by using the basic structure of Table 2 to create a set of dialog options for the learner's side of the dialog. However, instead of basing these on socio-cultural issues, the options are based on differing degrees of competence in the LOs of concern. For example, the LO "Allows the patient to express concerns" is tied to the competency in Active, Attentive Listening (see [11]), and can be assessed in terms of the observed behaviors of "Encourages clarification of problems; Notices non-verbal signals; and Summarizes/paraphrases patient concerns." Thus, in various transactions with the patient, the learner dialog options can include all of these, some combination (but not all) or none. The choice made by the

learner[2] can thus be mapped into the degree of competency on this LO exhibited at this point in the interaction (and, in turn, to appropriate scaffolding that provides proactive coaching or reactive feedback based on the LO and the choice made). In this example, the failure of the learner on this LO would result to a dialog path in which the NPCs underlying concerns and issues would *not* be exposed and lead to a failure to address these in a treatment and self-management plan.

5 Discussion

The concept of design persona has been in use in user-interaction design [12, 13] but largely within the task-specific context of a user-interface design. BE uses the concept much more broadly by incorporating a more detailed backstory that brings into play the specific cultural, social, and emotional concerns of the persona. The structured elaboration approach also allows the design of NPC dialogs to be integrated with and mapped to learner dialogs and ultimately to the instructional content of the tutoring architecture driving the learning process. This, we believe, represents a major step forward in the generalization of cultural training and other forms of so-call 'soft skill' learning.

BE allows the designers and technologists to draw on the cultural expertise that others have accrued and apply it in a quick and structured way that incorporates the detail and authenticity needed for cultural learning. In REPEAT development, BE was applied via a custom training platform to generate the interactive scenarios within a game-based environment. The platform included a visual authoring toolset, that enabled the easy entry of the output of the BE process, but it provided no tools to support the BE development and applications of personae, backstories, reasoning cases, and specific dialogs. Future research is focusing on creating tools to further streamline and automate the BE method.

Acknowledgments The research reported in this publication was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health, under award number 1R43MD009559. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors gratefully acknowledge the contributions of Albert Whitaker as a situational expert on Type II Diabetes in low-income African American urban communities in the REPEAT system development.

[2]In a dialog-tree architecture, each choice would map into a specific entry in the dialog tree. In a more natural language architecture, the learner utterance would be processed to map it into one of the options in the structure table, reflecting the degree to which it exhibited (or failed to exhibit) the behaviors involved in each option.

References

1. Johnson, W.L.: Serious use of a serious game for language learning. Frontiers Artif. Intell. Appl. **158**, 67 (2007)
2. Santarelli, T., Stagl, K.C.: Design and evaluation of a cross-cultural training system. In: MODSIM World Conference Proceedings 2009, Virginia Beach, VA, October 13–15 (2010)
3. Zachary, W., Maulitz, R.: Cognitive interaction analysis of clinical encounters. In: Proceedings of 2012 Symposium on Human Factors in Health Care. Human Factors and Ergonomics Society, Santa Monica, CA (2012)
4. Brach, C., Fraserirector, I.: Racial and ethnic disparities in health care: a position paper of the american college of physicians. Ann. Intern. Med. **141**, 226–232 (2004)
5. Pollard, S., Zachary, D., Wingert, K., Booker, S., Surkan, P.: Family and community influences on diabetes-related dietary change in a low-income, African American community. Diab. Educ. (2014). doi:10.1177/0145721714527520
6. Cannon-Bowers, J.A., Bowers, C.A.: Synthetic learning environments: on developing a science of simulation, games and virtual worlds for training. In: Kozlowski, S.W.J., Salas, E. (eds.) Learning, Training and Development in Organizations (229–261). Routledge, New York, NY (2009)
7. Cannon-Bowers, J.A.: Recent advances in scenario-based training for medical education. Curr. Opinion Anesthesiol. **21**, 784–789 (2008)
8. Corbett, A.T., Koedinger, K.R., Anderson, J.R.: Intelligent tutoring systems. Handbook Human Comput. Interact. 849–874 (1997)
9. Puntambekar, S., Hubscher, R.: Tools for scaffolding students in a complex learning environment: what have we gained and what have we missed? Educat. Psychol. **40**(1), 1–12 (2005)
10. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. **46**(4), 197–221 (2011)
11. Fassaert, T., van Dulmen, S., Schellevis, F., Bensing, J.: Active listening in medical consultations: development of the active listening observation scale (ALOS-global). Patient Educ. Couns. **68**(3), 258–264 (2007)
12. Ellis, D.: A behavioural approach to information retrieval system design. J. Documentation **45** (3), 171–212 (1989)
13. Pruitt, J., Grudin, J.: Personas: practice and theory. In: Proceedings of the 2003 Conference on Designing for User Experiences, pp. 1–15. ACM (2003)

The Preference of Using Social Media by Different Attachment Styles for Managing Romantic Relation

Wen-Chien Lin and Yu-Chen Hsu

Abstract This research is mainly focus on the discussion of the user experience and preference for college students with their romantic relation partner. Nowadays, people always use social media to contact with other people, especially their romantic partner. Couples use social media to chat every moment when they miss their lover and also share their life by sending photos. By attachment theory, there are four different attachment styles. Each attachment style is measured by avoidance and anxiety. Previous studies have investigated the effect of attachment theory in romantic relation. However, seldom studies discuss the use of social media between romantic partners. We want to know that whether the social media used by different attachment style people would influence the satisfaction of love or not.

Keywords Social media · Romantic relation · Attachment · Online dating

1 Introduction

Nowadays, the rise of using Internet makes human lifestyle become more convenient. The mobile devices is even to reach a situation that everyone have at least one. The booming of social media and online dating software is changing the interaction between interpersonal behavior and habits. According to the statistics of Taiwan Network Information Center (TWNIC), the rate of using Internet access more than 80.3 % in 2015. In this statistics of TWNIC, the highest rate is 60.1 % which is the behavior "using for social media online" and the second highest

W.-C. Lin (✉)
Institute of Information Systems and Applications, National Tsing Hua University,
No. 101, Guangfu Rd., 30013 Hsinchu, Taiwan
e-mail: wenchieh2002@yahoo.com.tw

Y.-C. Hsu
Institute of Learning Sciences, National Tsing Hua University,
No. 101, Guangfu Rd., 30013 Hsinchu, Taiwan
e-mail: ychsu@mx.nthu.edu.tw

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_18

behavior is "using real time community software" has 56.3 % rate. Above all, we can understand that there are very large numbers of people using social media to maintain their interpersonal activities in different social behaviors.

College students already have mature mental age and also is ready to go into society. In this age, the interpersonal relationship and romantic relationship issues are the most frequent problems they would face in daily life. Institute For Information Industry (III) survey some questions for APP using behaviors of Taiwanese consumers in 2014. In this survey, the most download type in APP store is "social and chat" type (73.5 %). In the same year, "Internet use among the community" which is also survey by III that most users use the Internet to engage in community activities as "contact with friends and relatives (82.8 %)," community software has become an indispensable tool between interpersonal communication media.

The different personalities of the students deal with the interaction between these two types of communication behaviors are not in the same way. In this study, in order to understand the different between each attachment type of students dating with their romantic partner through the social media, and whether it can be improved their relationship, and to help develop social circles.

2 Literature Review

2.1 Attachment Theory

Attachment theory is a framework to analyze people's personality in relationship. This theory is first mentioned by Bowlby [1, 2]. Based on the infant relationship, different attachment classifications emerge shape the infant's expectations of close relationships. Infant's behavior are always effected by the major caregiver. These behaviors will follow to adults.

By Bowlby's attachment theory, Bartholomew and Horowitz [3] use self-report questionnaire to measure attachment styles by two dimensions: anxiety and avoidance [4]. Four types of attachment styles are: secure (low anxiety and low avoidance), preoccupied (high anxiety and low avoidance), dismissing (low anxiety and high avoidance) and fearful (high anxiety and high avoidance). Many research about romantic relation always use this way to classify the subjects.

2.2 Romantic Relation

The way treat to romantic partners evolved by childhood's attachment theory. Different attachment style treat to their romantic partner in distinct ways. With insecure attachment, this type people usually seek out dating relationship, but once

they have romantic partner, they tend to experience emotional distress from struggling to trust their partner or see themselves as worthy of love [5]. This kind of people are especially at risk for dating violence [6].

Research also find that during the romantic relation, once people have insecure behaviors and characteristic, the commitment and satisfaction will reduce. In the same time emotion regulation will become lower. However, the secure attachment have high quality and high satisfaction.

2.3 Social Media

In recent years, the development of social media bring the new way for couple communicating their romantic partner. The relation between social media and romantic relation also began to receive attention. Surprisingly, the social media bring the deterioration between the partners [7, 8].

Through the social media, the behavior of partners become seldom privacy. People can know their partner contact with other by Facebook wall. Monitor behavior began to destroy their relationship. These behaviors include monitoring partner's whereabouts and activities, monitoring partner's social relationships, pressuring partner to respond quickly to calls and messages and so on. Thus list two questions are presented below to guide this study:

Question 1. With different kinds of attachment style social media users, will the user preference and user habit effect the satisfaction and quality of romantic relation?

Question 2. With different kinds of attachment style social media users, in order to promote the satisfaction and quality of romantic relation, prefer which service or which characteristic of social media?

3 Method

3.1 Design

The research model is following in Fig. 1. Online survey will be use in this study. We will conduct a self-report cross-sectional survey study of college students.

Fig. 1 Research model

3.2 Procedure

Data will be collect between April 2016 and June 2016. Not only collect survey online, we will also interview some different attachment styles couple to know more information about this study. Data analyze will start July 2016.

4 Summary and Future Work

In sum, the schedule will be complete before the conference. This study expect know the user preference and user habits of each attachment style. Also, will know the service of social media to promote romantic quality.

In the future, based on the result, it can help people have great romantic relationship and improve the social media.

References

1. Bowlby, J.: Attachment and Loss: Attachment. Basic Books (1969)
2. Bowlby, J.: Attachment and Loss, vol. 3. Basic books (1980)
3. Bartholomew, Kim, Horowitz, Leonard M.: Attachment styles among young adults: a test of a four-category model. J. Pers. Soc. Psychol. **61**(2), 226 (1991)
4. Hazan, Cindy, Shaver, Phillip: Romantic love conceptualized as an attachment process. J. Pers. Soc. Psychol. **52**(3), 511 (1987)
5. Davila, J., et al.: Romantic involvement and depressive symptoms in early and late adolescence: the role of a preoccupied relational style. Pers. Relat **11**(2), 161–178 (2004)
6. Shorey, Ryan C., Cornelius, Tara L., Bell, Kathryn M.: Behavioral theory and dating violence: a framework for prevention programming. J. Behav. Anal. Offender Victim Treat. Prev. **1**(4), 1 (2008)
7. Fox, J., Katie, M.W.: Social networking sites in romantic relationships: attachment, uncertainty, and partner surveillance on Facebook. Cyberpsychol. Behav. Soc. Netw. **17**(1), 3–7 (2014)
8. Hand, M.M. et al.: Facebook and romantic relationships: intimacy and couple satisfaction associated with online social network use. Cyberpsychol. Behav. Soc. Netw. **16**(1), 8–13 (2013)

Part VI
Contemporary Defense Applications of Social Science

Considering Culture in Contemporary Military Interventions: Simulating the Effects of Actions of Influence on a Civilian Population

Bergier Jean-Yves and Colette Faucher

Abstract As armed conflicts of today are highly population-centric, a national or international military force deployed to enforce a peacebuilding agenda has necessarily to face a multifaceted and complex local culture. To meet such a challenging intercultural situation, soldiers nowadays implement actions of influence. These are non-kinetic operations aiming at shaping attitudes and behaviors and gain legitimacy and support. Understanding how local cultural norms and values can affect the outcome of such courses of action is of great importance if success is to be achieved, soldiers have therefore to be trained in this domain. The purpose of the SICOMORES system is to meet this need by modeling a wide range of these actions: psychological operations (PsyOps), key leader engagements (KLE) and civil-military cooperation actions (CIMIC). They are discrete events whose effects are simulated in a multilayered networked population of individual cognitive agents characterized by cultural features.

Keywords Actions of influence · PsyOps · CIMIC · KLE · Persuasive communication · Intelligent agents · Culture

1 Introduction

Designing a relevant training system modeling a theater of war demands to acknowledge and integrate in the simulation several key facts characterizing modern armed interventions. First, most contemporary conflicts have in common a deep involvement of the local population, which has become actor, target and stake. Boundaries between armed groups or factions and the people sustaining them and

B. Jean-Yves (✉) · C. Faucher
LSIS Laboratory, Aix Marseille Université, UMR 7296, Avenue Escadrille Normandie-Niemen, 13397 Marseille, France
e-mail: Jean-yves.bergier@lsis.org; jeanyves.bergier@neuf.fr

C. Faucher
e-mail: colette.faucher@lsis.org

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_19

predated upon by them are usually blurred. In addition, constraints deriving from an asymmetric warfare context and the need to deal with limited means encourage concerned militaries to develop a smart comprehensive approach, far more refined than a kinetic-only strategy. These elements make an increasingly convinced and favorable local population a goal of utmost importance in itself for the Force and a prerequisite for success [17]. From this statement ensues the concept of "battle for perceptions" and the now mandatory practice of defining a "Military Influence Strategy" for each theatre, put into action through many different but coordinated operations, several of which aim at producing effects on non-combatant people [8]. All these considerations make ultimately appear that in these matters a good cultural awareness and understanding of culture-specific elements is of the greatest importance. Success depends highly on the way military influencers adapt their goals, actions and ways of communicating to the intricate local system of norms and social configurations they are encountering. However, we intend to root our simulation in human and social sciences in order to give it strong theoretical justifications, so care must be taken not to fall into the now rightly criticized "primacy-of-culture" (culturalist) perspective [20], all the more since the objective is to propose a general framework. The goal of SICOMORES (SImulation COnstructive et MOdélisation des effets des opérations d'influence dans les REseaux Sociaux) is to illustrate the role of the complex cultural reality faced by militaries tasked with peacebuilding operations on the practice and effects of actions of influence, and to train specialized soldiers for them. To this end, it concentrates on modeling three types of operations: psychological operations (PsyOps), key leader engagements (KLE) and civil-military cooperation actions (CIMIC).

2 Related Work

As research on peacebuilding is a thriving field and as industrial demand is increasing, many computational models and simulations relating to problematics of the comprehensive approach of conflict resolution and subsequently to the domain of influence and operations on perceptions have already been developed and refined. While some are educational programs striving to raise cultural awareness of the military user [12], others are more directly relevant to our research interests: COMPOEX for instance proposes an extensive multimodel architecture blending multi-agents simulation with a system-dynamic model to produce a detailed model of media influence (MIM), solidly grounded on several communication theories [3]. While representing sources, audiences, the emergence of trends in the latter resulting from the flow of influence from the former, and integrating individual leaders in a power network, the system seems not to choose to simulate message processing at the individual level. SHOUT [22], a message propagation simulation model focusing explicitly on PsyOps, takes a similar modeling route by representing only communities. It incorporates several channels of communication, including word of mouth. However, its propagation model is based solely on

advances in epidemiological research and measures only the spreading of the messages. CAPRICORN [10] system models PsyOps, CIMIC and InfoOps operations, with a focus on the emotional effects thus generated, in a multi-agent simulation emulating the Afghan theatre (Kapisa Region). While simulating a wide range of actions of influence on a networked population, the system makes no use of state of the art social psychology, using for instance a quite simple model of needs for the population. As for KLE, some immersive simulation scenarios with conversational agents culturally designed in collaboration with social scientists and linguists exist [21], as well as a discrete event simulation within the framework of the Cultural Geography Model [23], which again does not seem to be based upon social psychology models and emulates mainly military procedures. To our knowledge, no system currently proposes a simulation of all the three pivotal actions of influence considered in this paper while using sociology and social psychology as a basis for computation of their effects on a mass of complex cognitive agents representing individuals.

3 Virtual Human Framework

To adequately simulate the effects of the different actions in such an intercultural context, it is necessary to design a culturally realistic population.

SICOMORES integrates a specific population model composed of agents representing individuals only and organized in a multidimensional social network [2] to allow representing many different types of social links for each node. The choice of nature and structure of each dimension (or layer) of the network is grounded on the study and modeling of the different kinds of sociality (i.e. relations and social practices) and social institutions organizing the society of the considered people, a method allowing to produce a highly culturally infused, complex network. Emulating the society of a developing country, the configuration by default currently allows generating a sub-Saharan African population [7]. It models three types of families (extended, enlarged, nuclear) and incorporates layers corresponding to primary sociality (family, neighborhood and friendship), secondary sociality (partisan and religious dimensions) and sociality generated by communitarianism, a cultural phenomenon specific to the context of war and particularly salient in Africa where it is structural risk [13]. More specifically, socio-demographic data from the Central African Republic has been used to generate a representative fraction of the population of its capital city Bangui. Yet the multilayered architecture is flexible enough to be adapted to any socio-cultural context and different levels of cultural specificity.

In order to realistically simulate message treatment and attitudinal reactions, agents have to be cognitively sophisticated. Each agent in the system is an instance i of the INDIVIDUAL frame and is described by a set of attributes:

- Social features: Gender, age, social level ([1, 10]), religion, ethnicity, political opinion, role in the family, leader status.

- Cultural features: Cultural values system {(type of cultural feature, cultural feature, importance ([1, 10]))}
- Reachability features: Language(s), Literacy, reachable by radio, reachable by television, reachable by text message.
- Psychological features: Opinion toward the Force ([1, 10]), intellectual level ([1, 10]), needs {(need, satisfaction degree ([1, 10]))}

Each agent is characterized by cultural features corresponding to the considered socio-cultural context, although their respective importance varies from agent to agent. Some of the nodes receive a specific role such as head of family and political or religious leader. The two last ones are mutually exclusive and masculine only, but can be cumulated with the head of family status, which can, alone, be either feminine or masculine. A leader is also described by an additional attribute, Credibility of the Force's KL ([1, 10]), which differs from the Opinion toward the Force characteristic. The latter indicates to which extent the agent perceives favorably or not the Force and its actions as a whole. The former measures the quality of relations developed at a given time by a civilian leader with the particular officers of the Force he has actually met and dealt with. It thus indicates trust and openness to discussion.

We are confident that the different links coupled with such a set of individual attributes offers a relevant representation of culture both at the individual and collective level.

4 Psychological Operations (PsyOps)

PsyOps consist in the sending of messages by various means (radio, etc.) to reach specific info targets for impacting attitudes with ideally behavioral effect.

4.1 Theoretical Framework: The Elaboration Likelihood Model

Although the debate is still going on, many social psychologists tend to conclude that cognitive processes are universal and not culture-specific [15], at least for general functions, and those related to the persuasion process most certainly are. In any case, a vast number of factors can be involved in the cognitive treatment of a persuasive message, cultural features being doubtlessly among them. In accordance to such views, SICOMORES uses as a framework for persuasion the Elaboration Likelihood Model (ELM) [18], a dual process theory and leading paradigm in the study of attitude change stating the universal existence of two routes to persuasion:

a central route (high level of cognitive activity in message processing) and a peripheral route (relatively less effortful and elaborative message processing). As it is a complex and still developing set of theories [19] studying the effects of numerous interacting variables on the cognitive process, we concentrate on simulating some of the main elements of the model:

The extent to which a message is analyzed by a receiver, called *degree of elaboration*, is determined by his ability and motivation to do so.

His ability and motivation to process the message can vary with contextual and individual factors, the characteristics of the message being of importance here.

A variable can affect the process by serving as persuasive argument or peripheral cue.

As motivation and/or ability is decreased, argument scrutiny is decreased and peripheral cues become relatively more important determinants of attitude change. Conversely, if motivation and/or ability is increased, they become relatively less important determinants of the attitudinal result. There is a tradeoff between the weight of central and peripheral mechanisms in the treatment of messages.

4.2 Static Component: PsyOps Characteristics

In the system a PsyOp is described by numerous characteristics. Such a level of detail allows a representation of the receiver's cognitive processes in the framework of the ELM. Each PsyOp is an instance p of the PSYOP frame:

- Theme
- Desired info-targets
- Accentuated cultural feature(s)
- Flouted cultural feature(s)
- Medium
- Targeted indicator
- Appeal to curiosity/attention ([1, 10])
- Presenter {(attribute, value)}
- Expertise of the source ([0, 10])
- Highlighting of the theme ([1, 10])
- Legibility of message support (technical noise) ([1, 10])
- Clarity of content ([1, 10])
- Complexity of content ([1, 10])
- Number of arguments ([1, 10])
- Type of discourse (one sided, refutational two-sided)
- Arguments quality ([1, 10])

4.3 *Effect Generation of a PsyOp*

When an agent receives a PsyOp message, he treats it in accordance to the following algorithm (Fig. 1):

We first compute the degree of appeal, determining if the message retains the attention of the receiver. If it does, the ability of the receiver to process it is measured, which below a certain threshold dictates a peripheral processing only. If it is above, we then compute his willingness to process the message, notably by taking into account the level of personal relevance determined by the number and

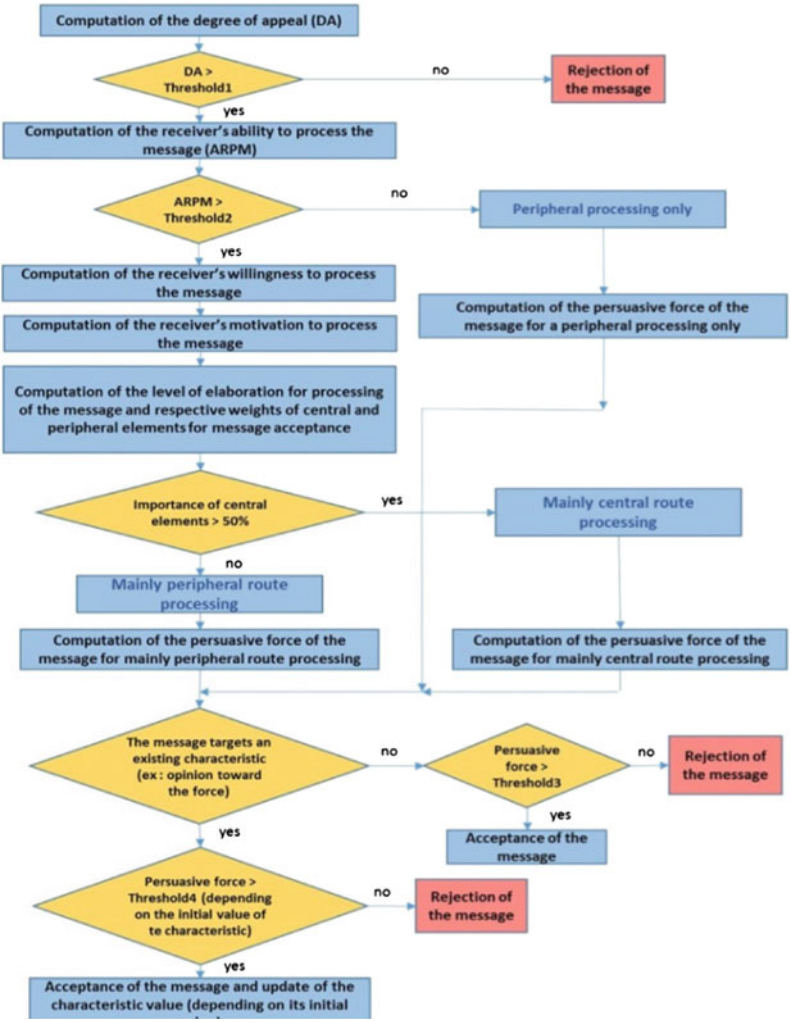


Fig. 1 Model for the cognitive processing of a message by the receiver

importance of involved cultural features, and then we calculate his overall motivation to process. With ability and motivation, we can compute the elaboration degree of message processing, which in turn allows determining the persuasive force of the message, and finally acceptance or rejection of the persuasion attempt for this particular agent. If the PsyOps is successful and targeted a societal indicator among those proposed to the user, it is then updated.

5 Key Leader Engagement

The KLE concept covers all the face-to-face meetings between officers of the Force and local civilian influencers and power figures, conducted at every hierarchical level, a procedure now widely recognized as fully necessary for a good integration of the Force in its human environment and supporting its other courses of action [11]. As it is a complex phenomenon and by no means a monopoly of specialized soldiers, the system distinguishes between two types of KLE and treats them differently: the Routine KLE and the Persuasion KLE.

5.1 Routine KLE

Routine KLE represents the implementation of the KLE plan at the macro scale of the theater of operations, aiming at no particular effect other than establishing and cultivating good relations between the leaders to increase legitimacy of the Force in the eyes of local official and unofficial authorities and public figures. Such a background, progressive task is of decisive importance in most developing countries, being a prerequisite for fruitful negotiation [16] and as such has to be simulated. This kind of discrete event is characterized in the system by a date and treats the powerbrokers collectively, targeting all leader agents in the population. The process includes random possibility of cultural faux pas for each leader (currently set at 10 %, but can be raised to increase difficulty). The only attribute impacted by such an action is the Credibility of the Force's KL, which is a key input in the computation of effects of a Persuasion KLE. It is raised by 1 if the action is successful and decreased by 1 in case of failure.

5.2 Persuasion KLE Static Component

Each persuasion KLE, on the other hand, involves a specific leader chosen in the population and targeted with a specific argumentative message in an attempt to persuade him, to obtain a desired effect. In this case, the KLE seems to function much as a psychological operation directed, at least primarily, toward a unique

individual. Indeed, the military doctrine mentions personal interview as a possible medium for PsyOp and, moreover, categorizes KLE a "direct action of influence" exactly like PsyOps actions [8]. All this justifies the use of the same theoretical framework, i.e. the Elaboration Likelihood Theory, for simulation of both PsyOps and persuasion KLEs.

In SICOMORES, each persuasion KLE is an instance k of the PERSUASION KLE frame:

- Theme
- KLE objective (Leader only, Propagation to others)
- Desired info-targets (if Propagation to others)
- Targeted indicator
- Accentuated cultural features
- Flouted cultural features
- Expertise of the source ([0, 10])
- Fluidity of discussion ([1, 10])
- Number of arguments ([1, 10])
- Argument quality ([1, 10])
- Quality of non-verbal language ([1, 10])
- Personal incentive (interest, vulnerability) (yes, no)

If the KLE aims at using the targeted leader to propagate a message to other nodes in the network, the user as to specify who these desired info-targets are. Some attributes are the same as for a PsyOp message, and in addition variables typical of interpersonal communication have to be set: fluidity of discussion measures the possible noise generated by translation or local variations of a language, while the quality of non-verbal language indicates the adequacy of what the military labels as "presence posture profile" (roughly, body language) and the non-verbal communication skills of the officer (active listening for instance). The variable of personal incentive refers to the psychological profiling usually done by intelligence soldiers prior to negotiation to discover possible leveraging options for particular individuals (money, etc.).

5.3 Effect Generation of a Persuasion KLE

The theoretical grounding for persuasion KLE action being the ELM, a modified version of the PsyOps algorithm is used here for effect generation. First, in the situation of KLE, attention of the receiver is taken for granted, the degree of appeal is consequently not computed. Ability and motivation are calculated using the specific parameters of persuasion KLE, such as the fluidity of discussion for the former and personal incentive for the latter. The elaboration degree and the ensuing force of persuasion of the argumentative message are then computed in the same

manner as for a PsyOp processing, and acceptance or rejection is determined by taking into account the value of the Credibility of the Force's KL characteristic of the considered leader and the optional presence of personal incentive.

6 Civil Military Cooperation

Civil military cooperation is a particular domain of military action targeting only civilians. Being both a function and a capability, CIMIC covers a large sphere of operations and includes the liaison with non-governmental civilian agencies operating on the theatre [5]. SICOMORES being focused on the actions directed at the local people, it simulates only the actions of support to the civil environment accomplished by CIMIC units within the framework of the Military Strategy of Influence. Straightforwardly, these actions aim at helping the integration of the Force in its human environment by mitigating the lack or degradation of civilian assets. Even with such a limited scope, possibilities are extremely large.

6.1 Theoretical Framework for CIMIC Operations

The relevance of CIMIC actions in the context of asymmetric warfare is based on the idea of shared interest: they meet the needs of the local inhabitants, thus promoting a favorable perception of the soldiers and actively favoring the return to the social and economic situation of peacetime. Hence, an adequate model of needs is required here for a relevant social simulation. The Maslow theory of motivation [14], a decidedly influential paradigm postulating a list of universal needs hierarchically organized by the priority of the different needs to be fulfilled (a need has to be satisfied before the next one is), is often used in social and behavioral modeling [4]. However, it seems today out of date, as studies showed that the idea of hierarchical scaling and directional modeling for needs is not valid and that, although universal needs probably exist, Maslow's chosen categories were highly questionable [6]. Because of complexity of human nature, human needs form a system dependent on psychological development and many situational factors. It is nonetheless legitimate to define a list of universal dimensions for basic human needs and well-being, which ideally have to be nonhierarchical, incommensurable, non-prescriptive and obviously not culturally biased. To this end, our system uses a customized model based on the different studies on human development compiled and examined in [1].

The theory of commitment [9] and its developments suggest that individuals feel committed by the behaviors they display (if they think they are acting freely), and adapt their attitudes accordingly to avoid any inconsistency (cognitive dissonance), rather than the reverse process. This principle can be the foundation of the computation of effects for civilian agents actively participating in CIMIC operations.

6.2 CIMIC Static Component

In the system, a CIMIC operation is an instance of the CIMIC frame, characterized by the following attributes:

- Name
- Impacted needs {(need, value [1, 10])}
- Targeted indicator
- Beneficiaries {(geographical area), or (geographical area, set of social features)}
- Accentuated cultural features
- Flouted cultural features

The system allows simulating an extensive list of CIMIC operations, distinguishing between CIMIC projects (permanent effects) and CIMIC actions (temporary effects). CIMIC projects typically consist in the building of infrastructures whereas CIMIC actions can be distribution of goods, or punctual services. For instance, in the health domain, the following actions are possible:

- Direct medical support (using the Force's own means) (action),
- Distribution of medical consumables (action),
- Supply of medical devices (project),
- Building of hospitals and health centers (project),
- Purification of the water supply system (project),
- Building of veterinary facilities (project).

6.3 Generation of CIMIC Effects

Computing the effects of CIMIC operations is challenging as they are actions in the physical world and not messages automatically triggering a cognitive processing in reached nodes and reached nodes only. For a given CIMIC action, SICOMORES makes the choice of modeling a cognitive reaction in three types of agents among those located in the geographical area of effect of the CIMIC. Consistently with the view of human beings as cognitive misers, those are agents who cannot reasonably dispense themselves of reacting:

- The beneficiaries
- The leaders
- Nodes optionally recruited as local workforce for implementing the project.

We realistically assume that any leader feels a moral and social obligation to have a view on a CIMIC happening in his area of influence. The workforce and beneficiaries are directly involved in the operation, respectively upstream and downstream. Of course these categories can overlap, as a leader can be a beneficiary for instance, and the algorithm deals with such particular cases. Each of these agents cognitively

proceed the attributes characterizing the CIMIC, taking into account the cultural features favored and flouted by the operation, allowing to compute its degree of cultural conformity for the agent. Some agents can be recruited as workers, only for projects (number and social criteria of selection specified by the nature of the operation): in such a case nodes are randomly picked, and selected only if the degree of cultural conformity is above a certain threshold (to be determined during experimentation), until the desired number of participants is achieved. Being a participant can alone favorably modify the opinion of the force attribute of an agent, in accordance with the social psychology theory of commitment.

The values of the impacted needs are then updated for those concerned, allowing computation of their individual degree of need satisfaction. Finally, as we consider this is the main point of any CIMIC, the opinion toward the Force attribute is updated for each agent, considering solely cultural conformity, or both cultural conformity and need satisfaction for beneficiary agents.

7 Conclusion and Future Work

We have presented in this paper some of the main features of the conceptual model of SICOMORES, a system aimed at modeling different kinds of action of influence directed by a foreign peacebuilding force toward a local civilian population, in the cross-cultural context of an ongoing contemporary armed conflict. Some are actions of direct persuasion communication such as PsyOps and KLE actions, some, such as CIMIC actions, only fall under the broader military concept of influence, but we are confident that the use of social sciences and social psychology allows us to grasp and encapsulate in an innovative manner the determining mechanisms of the effects of each of these operations on a culturally infused and tailored virtual population. The framework of the system is designed to be generic enough to be adjustable to any specific cultural context, ensuring its adaptability as training and decision support system for military specialists routinely facing complex intercultural situations. Hence, SICOMORES hopes to make a significant contribution to the simulation of influence and culture in the context of warfare.

As the system is a work in progress, the present paper focuses only on the first step of each action, involving a direct form of communication from the military to the locals (voluntary or not), which can affect their attitudes. Our future work will be to model and experiment with the spreading of the effects from agent to agent in the different layers of our multidimensional social network.

Acknowledgments This work is funded by the French Ministry of Defense (DGA—Direction Générale de l'Armement) in the framework of the DGA RAPID Project SICOMORES.

References

1. Alkire, S.: Dimensions of human development. World Dev. **30**(2), 181–205 (2002)
2. Berlingerio, M., Coscia, M., et al.: Multidimensional social networks: foundations of structural analysis. World Wide Web **16**(5–6), 567–593 (2013)
3. Bennet, W.: Media Influence Modeling in Support of conflict Modelling, Planning, and Outcome Experimentation (COMPOEX), Presentation to the Military Operations Research Society Irregular Warfare Analysis Workshop, MacDill AFB (2009)
4. Bracewell, D.: The needs of metaphor. In: Kennedy, W., Agarwal, N., Shanchieh, J.Y. (eds.) Social Computing, Behavioral-Cultural Modeling, and Prediction, 7th International Conference SBP, pp. 229–236. Springer (2014)
5. Coopération civilo-militaire, Doctrine Inter Armées, DIA-3.10.3(A)_CIMIC (2012)
6. Diener, E., Tay, L.: Needs and subjective well-being around the world. J. Pers. Soc. Psychol. **101**(2), 354–365 (2011)
7. Faucher, C., Bergier, J.-Y., Forestier, M.: Capturing sub-Saharan African sociality in social networks to generate a culturally realistic population. In: IEEE International Conference on Culture and Computing, Kyoto University Japan (2015)
8. Influence in Support of Operations, (FRA) JC-3.10(A)_INFLUENCE (2012)
9. Kiesler, C.: The Psychology of Commitment, Experiments linking Behavior to Belief. Academic Press, New York (1971)
10. Khimeche, L., et al.: CIMIC and PSYOPS Models Prefigure the Future for Comprehensive Approach CGF, Fall Simulation Interoperability Workshop, Orlando, USA (2012)
11. Lindoff, J., Granasen, M.: Challenges in utilizing Key Leader Engagement in Civil Military Operations, 16th ICCRTS (2011)
12. MacMillan, J., Walker, A., Clarke, E., Marc, Y.: Cultural awareness training for Marine Corps Operations: The CAMO Project. In: Advances in Design for Cross-Cultural Activities, pp. 277–286 (2012)
13. Marie, A.: Communauté, Individualisme, Communautarisme: Hypothèse sur quelques Paradoxes Africains. Sociologie et Société **39**(2), 173–198 (2007)
14. Maslow, A.: A theory of human motivation. Psychol. Rev. **50**(4), 370–396 (1943)
15. Mishra, R.: Cognition across cultures. In: Matsumoto, D. (ed.) The Handbook of Culture and Psychology. Orxford Unisersity Press, Oxford (2001)
16. Nash, R., Magistad, E.: Disarming the key leader engagement. Military Rev. 11–20 (2010)
17. Norheim-Martinsen, P., Nyhamar, T. (ed.): International Military Operations in the 21st Century, Global Trends and the Future of Intervention. Routledge (2015)
18. Petty, R., Cacciopo, J.: The Elaboration Likelihood Model of Persuasion. Adv. Exp. Soc. Psychol. **19**, 124–205 (1986)
19. Petty, R., Wegener, D.: The Elaboration Likelihood Model: current status and controver-sies. In: Chaiken, S., Trope, Y. (eds.) Dual-Process Theories in Social Psychology. The Guilford Press, New-York (1999)
20. Roads, K.: The culture variable in the influence equation. In: Snow, N., Taylor, P. (eds.) The Routledge Handbook of Public Diplomacy. Routledge (2008)
21. Stuart, S.: Using Video Games to Prepare for the Culture Shock of War, PC.com (2014). http://www.pcmag.com/article2/0,2817,2472395,00.asp. Accessed 28 Feb 2015
22. van Vliet, T., et al.: Generic message propagation simulator: the role of cultural, geographic and demographic factors. In: Schmorrow, D., Nicholson, D. (eds.) Advances in Cross-Cultural Decision Making, pp. 416–429. CRC Press Taylor & Francis Group (2011)
23. Wakeman, C.: Discrete event simulation modeling and analysis of key leader engagement. Master Thesis, Naval Postgraduate School (2012)

Development of a Competency Model for Civil-Military Teaming

Karol G. Ross, Michelle Wisecarver, Carol A. Thornson, Deborah A. Peluso and Melinda Key-Roberts

Abstract The nature and complexities of today's military operations are such that no single organization, department, or agency has all the requisite resources, authority, or expertise to single-handedly provide an effective response. These operations require civil-military teaming to establish, manage, and participate in collaborations among various military, governmental, non-governmental, local national leadership, and civilian agencies. This research was conducted to develop requirements for successful teaming performance. An initial competency model, developed from a domain analysis, was validated through 19 interviews with participants from the U.S. Army, the State Department, USAID, USDA and other organizations. Results suggest a competency model with three higher-order meta-competencies and 12 lower level competencies. Thirty-two specific decision points that present significant performance challenges are also described and linked to tasks and activities. The resulting decision requirements provide context for potential training objectives and scenarios to identify and address gaps in existing training and education.

K.G. Ross (✉)
Cognitive Performance Group, Orlando, FL, USA
e-mail: karol@cognitiveperformancegroup.com

M. Wisecarver
PDRI: A CEB Company, Arlington, VA, USA
e-mail: Michelle.Wisecarver@pdri.com

C.A. Thornson
ICF International, Fairfax, VA, USA
e-mail: Carol.Thornson@icfi.com

D.A. Peluso
The Change Collaborative, Columbus, OH, USA
e-mail: dpeluso@thechangecollaborative.com

M. Key-Roberts
U. S. Army Research Institute, Fort Leavenworth, KS, USA
e-mail: melindaj.key-roberts.civ@mail.mil

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_20

Keywords Civil-Military teaming · Competency model · Human dimension · Decision Making in Joint, Interagency, Intergovernmental, and Multinational (JIIM) Operations

1 Introduction

The complex nature of today's military operations is such that no single organization, department, or agency possesses all the requisite resources, authority, and expertise to single-handedly provide an effective response. As such, collaboration among various military, governmental, non-governmental, and civilian agencies is needed to achieve a common goal [1, 2]. This chapter examines the competencies and specific performance challenges which must be addressed to support the U.S. Army's goals.

Recent priorities established by the U.S. Army show a continuing and even renewed emphasis on the ability to work effectively in such collaborative environments. The U.S. Army Operating Concept (AOC) describes the Army as part of joint, interorganizational, and multinational teams [3]. The Army Human Dimension Strategy [4] established objectives for the development of "Social Intelligence," i.e. "the development of team members who thrive in complex social environments, adapt to diverse cultures, communicate effectively, and build relationships" and "Team Building" "to forge diverse individuals and organizations into cohesive teams based on mutual trust and unity of effort."

According to the Joint Publication on Interagency, Intergovernmental Organization and Non-governmental Organization (NGO) Coordination During Joint Operations [5], the activities of a commander include those that "establish, maintain, influence, or exploit relations between military forces, governmental and non-governmental civilian organizations and authorities, and the civilian populace in a friendly, neutral, or hostile operational area in order to facilitate military operations, to consolidate and achieve operational U.S. objectives."

Civil-military operations typically fall under the direction of the Joint Forces Commander (JFC), who is responsible for a host of tasks to promote effective interactions among U.S. organizations [2]. As a part of his work, the commander often coordinates activities to establish a goal-oriented, interdependent unit comprised of several different organizations. Cross-boundary and multifunctional teams, or *leader-teams* [6], often are formed to accomplish a specific purpose and are typically nested within larger efforts. In contrast to traditional military teams, lines of authority within multifunctional/multi-organizational teams are not clearly demarcated and work is often accomplished through informally formed, mutually beneficial relationships. Civil-military teaming (CMT) serves a number of purposes for U.S. engagement in a region including supporting civil administration, providing foreign humanitarian assistance, providing national assistance, enhancing military effectiveness, and reducing the negative aspects of military operations on civilians [1, 2]. Across these situations CMT can be conducted at strategic,

operational, and tactical levels, as well as in joint, interagency, intergovernmental, and multinational situations [2, 5].

Despite the importance of successful CMT, several obstacles can hinder or even derail CMT. According to Polley [7], these obstacles include: (1) the lack of a common lexicon, which can significantly hamper communication and information exchange; (2) non-overlapping purpose, strategic goals, and vision between team-members and their respective organizations; and (3) the lack of shared training or educational experiences. Additionally, civilian and military planning, decision making, and execution styles differ greatly [8]. Even among the U.S. governmental agencies and entities, different organizational cultures result in conflicting goals, policies, procedures, decision-making techniques, and processes. Moreover, the command and control (C2) structure that supports a military operation is often at odds with the structure of civilian organizations. Together these differences present significant challenges to coordinated efforts and unified actions. Adding to this complexity, civil-military teams must also contend with a plethora of external obstacles such as ethnic and religious conflict, cultural and socioeconomic differences, terrorism and insurgencies, the proliferation of weapons of mass destruction, international organized crime, incidental and deliberate population migration, environmental degradation, infectious diseases, and sharpening competition for, and exploitation of, dwindling natural resources [2].

To meet these challenges, members of civil-military teams must develop specific and targeted competencies that enable them to succeed. However, no generally accepted model of competencies exists that clearly identifies and defines the knowledge, skills, abilities, and other characteristics required to support effective CMT. This chapter defines the competencies required for success in CMT and details critical decisions that are made when supporting effective CMT in the field. These findings are intended to support training development and successful operations.

2 Domain Analysis

The purpose of the domain analysis which initiated this research was to frame the area of performance and to develop an understanding of the domain to be studied. Our approach to the domain analysis involved reviewing the literature to understand the competencies suggested by prior research in this area, interviewing a small group of experts regarding the domain, attending a key conference focused on issues surrounding training and education in the domain, conducting an informal, brief survey of the conference attendees, and then developing an initial competency model. For the purposes of identifying an initial set of competencies during the domain analysis, we defined a competency as a measurable pattern of knowledge, skills, and abilities (KSAs) that is needed to successfully perform work roles or functions, and which differentiates performance on the job [9–11]. Identifying competencies provides an organizing structure that frames the domain and

emphasizes the knowledge, skills, abilities, and other characteristics (KSAOs) that are central to successful performance in the domain of interest.

2.1 Method

Several steps were involved in developing the initial model. The project team (1) reviewed background military doctrine and other relevant research literature, (2) conducted an initial set of domain analysis interviews to collect initial competency information, and (3) conducted a brief survey to elicit competencies from a civilian and military audience. From these efforts, we developed an initial competency model.

For the first step, we identified and reviewed existing materials regarding CMT. While this project was conducted for the U.S. Army, it was important to explore the basis of success in CMT across a variety of military, U.S. government, and other organizations to determine what competencies best support effective CMT. Materials reviewed included articles in military and government journals, doctrinal documents, government research findings, and other government and academic findings.

A second source of information for the development of the initial competency model came from two groups of subject matter experts (SMEs). An initial set of eight interviews was conducted with personnel from the U.S. Army Peacekeeping & Stability Operations Institute (PKSOI) to understand the domain of CMT. The interview participants were provided by PKSOI and were members of the staff. The interviews posed open-ended questions regarding topics such as:

- Key issues or barriers that impede CMT
- Conditions that contribute to successful civil-military collaboration
- Skills/attributes of team members that contribute to CMT effectiveness
- People or organizations we should prioritize talking to during the course of this investigation

The second group of SMEs were attendees at the 2010 PKSOI Training and Education Conference. A brief survey was distributed at the conference asking participants to write down key competencies required for successful CMT. A total of 32 conference attendees provided written responses (n = 18 civilian personnel, 13 military personnel, two civilian/military personnel, and one non-identified).

2.2 Findings

Review of the Literature. The literature review was extremely useful for providing background information and context about CMT, as well as suggestions for specific personal attributes relevant to competency requirements for teaming. In combination, the materials reviewed suggested that CMT is a challenging endeavor,

encompassing a wide spectrum of goals and requiring more than just coordinated action. Our literature review underscored the point that in order to be successful, teams, and the individuals comprising them, must engage in synchronization. Synchronization involves "a pervasive unity of effort across the political, military, economic, and psychological spectrum" [7].

CMT represents a primary military instrument to synchronize military and non-military instruments of national power [2]. Military and non-military instruments of national power refer to government organizations that can be used to exert diplomatic, informational, military, or economic power in an international situation. To be successful these operations require a unique teaming effort among individuals displaying specific competencies. Prevou et al. [6] found that individuals in high performing *Teams of Leaders* (ToL) do the following: communicate meaningfully more often, collaborate frequently, build constructive relationships, adapt quickly to changing situations/requirements, work around bureaucracy and other obstacles, create and use an extended network, feed each other's creativity, challenge and support one another, make deliberate operating agreements, and continually assess their direction and progress. The broader literature also provided several examples of actions that are necessary in order to promote CMT, including creating an environment where there is unity of effort among actors, cooperating with all those involved, developing a shared understanding of the capabilities of the organizations involved, being able to shift perspectives to gain a more holistic understanding of the situation and the environment, having flexibility in actions and decision making, and constantly exchanging information with fellow teammates [5, 12, 13].

Effectively leveraging team attributes such as the ones mentioned here enables leader-teams engaged in CMT to overcome a number of obstacles and, ideally, to synchronize action across participating organizations. In order to overcome confusion of objectives, inadequate structure or procedures, bureaucratic processes, and personnel limitations, close and continuous coordination and cooperation between civil and military teams is necessary. Indeed, the diversity represented by so many different organizational cultures, philosophies, goals, practices, and skills can be an important strength to leverage against the enormous task of coordination. In order to effectively integrate multiple perspectives and agendas, individuals themselves must exhibit certain diverse competencies that contribute to successful teaming. The background literature review suggested that skills and abilities related to communication, information exchange, collaboration, influence and negotiation, cross-cultural agility, situational awareness, perspective-taking, relationship building, and problem-solving are highly valuable in civil-military team settings. Moreover, knowledge about other organizations, resources, processes, and systems further serves as an important foundation for effective partnerships.

An important element of developing the initial competency model was to examine other military or government competency models for competency labels and definitions that could be used and adapted for the present model. Relevant competencies in these models were identified based on themes derived from the SME interviews, survey, and the background literature review conducted for the current project. The following models proved helpful in developing an initial list of

CMT competencies and definitions: Office of Personnel Management (OPM) Senior Executive Service (SES) competency model, Joint Special Operations Forces (SOF) competency model, Joint Forces Command (JFCOM) Senior Leader competency model, U.S. Air Force (USAF) Leadership Competencies, and the Navy Leadership Competency Model (NLCM). The Army Leader Development Strategy (ALDS) was also consulted because it specifically highlights the need for Army leaders to partner closely with the Joint, Interagency, Intergovernmental, and Multinational (JIIM) teammates. This document does not describe a competency model per se. However, it does articulate key leader characteristics and attributes the Army should foster in anticipation of the future operational environment. Similarly, Ross et al.'s [13] description of training themes for JIIM operations, was highly informative for developing several CMT competencies. Overall, there appeared to be considerable overlap among sources regarding the competencies essential to CMT.

Domain Analysis Interviews and Survey. In conjunction with the literature review we examined results from the interviews and surveys to identify areas of convergence and divergence. The most frequently mentioned responses from both the initial interviews and the conference survey included the following: teamwork/ability to work with others, understanding organizational functions, cultural awareness/sensitivity, domain knowledge, leadership skills, planning skills, and unity of effort/command. While subjects were asked to list key competencies required for civil-military teaming, many of the responses did not fall within the definition established at the start of the project. Instead participants included components of competencies (i.e., negotiation skills, communication skills, and knowledge about other agencies), as well as desired end states associated with civil-military teaming (i.e., unity of effort/command). The data were carefully analyzed for information which could inform the development of the initial competency model. In cases where meaning and intent were unclear, the data were discarded.

Participants in the domain analysis interviews at PKSOI also identified a variety of obstacles to CMT. The following were among the most frequently mentioned hurdles to successful teaming: cultural differences (i.e., ethnic and organizational culture), language differences, lack of familiarity/understanding between military and civilian organizations, and lack of unity of effort. These barriers can inform our understanding of the KSAOs required to succeed in CMT by suggesting that KSAOs that ameliorate these hurdles would lead to success in the domain. However, more information is needed to establish a comprehensive set of obstacles to civil-military teaming.

Initial Model. The project team reviewed the lists of KSAOs identified in both the literature reviews and the initial interviews and survey, and content analyzed them to form groups of related KSAOs. This process was highly iterative and required making judgments based on the KSAOs identified thus far, as well as our knowledge of what individuals are required to do in CMT. The compiled KSAO list was reviewed multiple times, and similar KSAOs were combined to form an initial list of 17 competencies. Seven members of the research team then met to discuss the 17 competencies identified, areas of overlap among them, and further combinations

based on overlap. Following this discussion, the list was subsequently edited down to a set of 12 competencies that demonstrated a minimum amount of overlap.

This list of 12 competencies was next reviewed to determine whether any content themes would emerge that would identify a higher-order structure for the competencies (i.e., meta-competencies). The purpose of meta-competencies is to organize the competencies into themes that describe key requirements in the domain in question. Three meta-competencies were identified for CMT: (1) Adapts Across Organizations and Cultures, (2) Builds Partnering Relationships, and (3) Collaborates to Solve Problems.

3 Validation and Final Model

The development of this initial model served as a hypothesis of the competencies that were related to success in CMT. In the next phase of research, we conducted in-depth interviews, capturing detailed information regarding KSAOs, behaviors, and critical decisions and judgments that must be addressed by practitioners to be successful in the civil-military context. Our goal was to examine the validity of the initial competency model by comparing the competencies identified in the initial model to those identified in the in-depth interviews, and to use the information from the in-depth interviews to capture additional KSAOs and behavioral indicators that would add breadth and depth to the competency model.

3.1 Method

Procedure. The interviews were based on the Critical Decision Method [14]. The Critical Decision Method interview (CDM) is a cognitive task analysis method which employs a semi-structured technique for eliciting knowledge from specific events that challenge a person's expertise. Via recollection of a specific incident as the starting point, CDM uses focused probes to elicit the details of each incident from the interviewee. In the present investigation, the types of information sought included the following:

1. Goals that were considered during the incident
2. Options that were generated, evaluated, and eventually selected
3. Contextual elements associated with the incident
4. Situational assessment factors specific to particular decisions
5. Challenges to performance
6. Strategies for success

Before the start of each interview, the team also collected demographic information as well as information to inform a Task Diagram [15]. The Task Diagram included the specific task requirements of CMT during the most recent

deployment. In some cases, the interview team incorporated a variation of the "Knowledge Audit" [15] in which several examples were gathered to aid in understanding the key events that revealed important KSAOs, behaviors, and decision requirements.

Participants. We conducted a total of 19 interviews with seven military participants and 12 civilians. Most of the civilian participants also had prior military experience (n = 8), with half having served at least 20 years in the military. Three females and 16 males were interviewed. Their ages ranged from 36 to 70. All had on-the-ground experience working in civil-military teams within the last year, as well as extensive experiences requiring them to make critical judgments and decisions that directly impacted team performance. All participants were engaged in CMT for a significant amount of time, and, therefore, had the opportunity to observe the consequences of their actions at the personal and organizational levels. Participants were experienced at both the tactical, as well as the operational levels of performance.

Analysis. The analysis process was divided into two primary analyses, which were conducted simultaneously by the research team. These analyses consisted of a Competency Analysis and a Decision Requirements Analysis.

3.2 Findings

Competency Analysis. The results of the competency analysis support the validity of the initial competency model in that for each competency, participants were able to provide rich KSAOs that fit within a competency in the model. Therefore, all initial competencies were retained. The frequency with which each competency was represented in remarks in the interview transcripts indicated that Builds Partnering Relationships was the most often represented meta-competency. Adapts Across Organizations and Cultures was the second more represented meta-competency, and Collaborates to Solve Problems was represented least often. The final model is shown in Table 1 which reflects the order of importance of the meta-competencies as defined by the frequency count. The findings provide a rich database of additional KSAOs, behaviors and decision scenarios and factors for civil-military teaming. For definitions of each competency and detailed associated KSAOs, see Ross et al. [16].

Decision Requirements Analysis. A total of 32 dilemmas and decision points were identified in the interviews as being encountered during CMT operations by the participants. Nine specific tasks and activities that were most often encountered were also identified, and they encompassed all the decision challenges. (1) Build and Manage Relationships; (2) Manage Programs and Projects; (3) Align and Integrate Operations; (4) Determine Roles and Responsibilities; (5) Align Goals; (6) Educate Partners; (7) Create Integrated Civil-Military Team; (8) Measure Success; and (9) Manage U.S. Presence. A detailed description of the obstacles/challenges, considerations, and problem-solving strategies associated with

Table 1 Competency Model of Civil-Military Teaming

Meta-Competency	Competency	
Building Partnering Relationships	1	Understands capabilities of partners and systems
	2	Establishes effective partnerships and teams
	3	Develops positive relationships
	4	Builds common ground and shared purpose
	5	Manages conflict
	6	Manages the flow of communication
Adapts Across Organizations and Cultures	7	Understands the cultural context of situations
	8	Assesses new cultural environments and adjusts appropriately (cultural agility)
	9	Understands multiple perspectives
Collaborates to Solve Problems	10	Uses integrative methods for planning and problem-solving
	11	Synchronizes tactical actions, operational objectives, and strategic goals
	112	Applies available resources and expertise

each decision/dilemma can be found in Ross, et al. [16]. The information is organized by task/activity and decision and linked to the supporting competencies. As expected in real world operations, multiple competencies may need to be exercised simultaneously to address a task or activity.

1. Build relationships across agencies, implementing partners, contractors, and military to support common interests
2. Build trusting partnerships with host nation and multinational counterparts
3. Navigate relationship dynamics and political challenges
4. Adapt operations to align to changes in mission initiated by new leadership
5. Investigate the rationale behind established plans and projects
6. Determine who and how to help with programs, aid, and support
7. Determine whether an area is a candidate for stability or development work
8. Coordinate efforts to resolve issues to execute a task or project between agency, partner, sub-contractor, and military
9. Develop situation awareness of the operational environment, situation on the ground, and project status
10. Manage expectations of what, how, and when projects will be implemented
11. Determine what type, how much, and how to share information with partners
12. Negotiate information flow among agencies, implementing partners, subcontractors, and military
13. Coordinate joint military operations
14. Distinguish and track the different types of civilians and civilian organizations operating in theater
15. Determine who the relevant players are that need to be integrated into operations

16. Determine how to get started in a new assignment/location
17. Determine how to integrate oneself with the Army
18. Generate a picture of what success looks like
19. Entice other actors (e.g., local, U.S. agency, military, NGO and international partners) into working together
20. Gain consensus on mission priorities across U.S. agency and military partners
21. Prioritize projects to implement
22. Determine how to position/nest programs, goals, and activities with larger U.S. strategy in country
23. Anticipate 2nd and 3rd order consequences of actions
24. Develop an integrated civil-military campaign plan for a 3- to 5-year strategy
25. Subtly train and coach counterparts
26. Develop military's understanding of how civilian agencies function in capacity building and community development missions
27. Create functioning civil-military working group
28. Facilitate project planning, roles and assignments across agencies
29. Assess the effects and impacts of development and capacity building
30. Decide how best to work with local national counterpart
31. Position host nation government to respond to various situations on the ground
32. Coordinate operations across host nation counterparts, militaries, and multinationals.

4 Discussion

4.1 Limitations

The scope of the current project did not allow for extensive validation of the proposed model and additional data are needed to corroborate the findings from this research. Specifically, further work is needed to identify the KSAOs associated with each competency. In the present research, the sample size was not large enough to ensure that all KSAOs associated with civil-military teaming activities were identified. Although our overlap with the KSAOs identified in other recent investigations suggest we can be confident in our findings, further work can expand and validate the competency model.

A second limitation is the need for more information regarding competencies required by the civilians who take part in civil-military teaming. Civilians with little-to-no military experience may require a different set of knowledge, skills, and abilities than those needed by military (or former military) team members. In the present project, some participants identified as civilians had significant prior

military experience. In order to assess the comprehensiveness of the current model, additional interviews should be conducted with civilian counterparts without prior military experience.

Finally, more information regarding the individuals surveyed (e.g., years deployed, rank, key experiences which shaped their perspective, types of teams served on, and current position/duties) should have been collected. This information would have enabled better assessment of the qualifications of the interviewees as subject matter experts (SMEs) and would have ensured the generalizability of results to the broad range of individuals participating in civil-military operations.

4.2 Application and Future Research

In an attempt to understand how training to support teaming is typically carried out and what competencies are addressed, the data collection team observed training designed for the Civilian Expeditionary Workforce (CEW) by the Foreign Service Institute (FSI). Researchers observed participants in various vignettes designed to familiarize them with how to interact with Soldiers and with local national decision makers. Tasks included establishing relationships, negotiating, and utilizing persuasion tactics with role players. The CEW groups were also coached in ways to assess the local situation and then organize and present their findings to the local Commander in a battlefield update briefing.

While experienced Army and civilian representatives acted as coaches for the participants, little *deliberate* coaching was done regarding how to bring members of different organizations into a team. We observed teaming principally within the CEW work groups as they tried to assess the situation and when they developed briefings to report findings to the Army officer in charge of the simulated operation. Because there were no deliberate exercises to jumpstart team cohesion built into the methodology, trust in the assigned teams had to develop over time. Additionally, the CEW work groups were not observed determining how they would work together, or establishing operating agreements, early in their lifecycle. This contributed to trial and error and retarded the speed at which the team built trust, confidence, and shared understanding. The meta-competency of Building Partnering Relationships, elements of which were the most frequently discussed in the validation interviews, appears to be a limiting factor on the other meta-competencies. This preliminary observation opportunity suggests that variations in training sequences and methods in future research can support understanding of the acceleration of skill acquisition and most effectivepaths to improved teaming performance.

Because lack of teaming skills can diminish the effects of individual planning skills, adaptability, and technical expertise, all training and coursework of this nature should target teaming skills as part of the process. The addition of team process information and exercises would likely improve a wide range of training experiences and provide a foundation for field experiences. In order to establish

context-based, collaborative, problem-centered instruction targeting competencies for which the Army is currently showing a strong demand, the Army must continue to obtain information about field experiences in a manner that is useful for application to training design and operations.

Our approach to this project demonstrates one method for capturing experiences and framing them in a manner that can support not only the goal of CMT, but also the goals of the Army in general. The Decision Requirements Analysis utilized in the present investigation provides rich material for developing or augmenting programs of instruction by capturing key tasks, activities, challenges, obstacles, and strategies faced by civilian and military personnel conducting civil-military operations to train decision making and methods to address common dilemmas. The data collected as part of this project can provide material for case studies, vignettes for practical exercises, and scenarios for simulations. This model and future refinement to it can also be used to develop assessments for complex civil-military teaming performance.

Acknowledgments The research described herein was sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences, Department of the Army under Contract W91WAW-09-D-0014 in support of the U. S. Army Peacekeeping & Stability Operations Institute (PKSOI) at the United States Army War College. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army.

References

1. Department of the Army: Civil-Military Operations (Field Manual 41-10). Department of Defense, Washington, D.C. (1993)
2. Joint Chiefs of Staff: Civil-Military Operations (Joint Publication 3-57). Department of Defense, Washington, D.C. (2008)
3. United States Army Training and Doctrine Command (TRADOC): The U.S. Army Operating Concept: Win in a Complex World (Pamphlet 525-3-1). United States Department of the Army, Washington, DC (2014)
4. United States Department of the Army: The Army Human Dimension Strategy. United States Department of the Army, Washington, DC (2015)
5. Joint Chiefs of Staff: Interagency, Intergovernmental Organization, and Non-governmental Organization Coordination During Joint Operations: Volume I (Joint Publication 3-08). Department of Defense, Washington, DC (2006)
6. Prevou, M., Veitch, R., Sullivan, R.: Teams of leaders: raising the level of collaborative leader-team performance. In: Proceedings of the I/ITSEC Conference, Orlando, FL (2009)
7. Polley, B.: Leadership education and training for the interagency. In: Cerami, J.R., Boggs, J.R. (eds.) The Interagency and Counterinsurgency Warfare: Stability, Security, Transition, And Reconstruction Roles, pp. 423–464. Strategic Studies Institute, Carlisle, PA (2007)
8. Hunt, J.P.: The 800-pound gorilla and stability operations. Small Wars Journal. http://smallwarsjournal.com/blog/journal/docs-temp/464-hunt.pdf (2010)
9. Mirabile, R.J.: Everything you wanted to know about competency modeling. Train. Dev. **51**, 73–77 (1997)

10. Rodriguez, D., Patel, R., Bright, A., Gregory, D., Gowing, M.K.: Developing competency models to promote integrated human resource practices. Hum. Resour. Manag. **41**, 309–324 (2002)
11. Shippman, J.S., Ash, R.A., Battista, M., Carr, L., Eyde, L.D., Hesketh, B., Sanchez, J.I.: The practice of competency modeling. Pers. Psychol. **53**, 703–740 (2000)
12. Department of the Army: Stability Operations (Field Manual 3-07). Combined Arms Doctrine Directorate, U.S. Army Combined Arms Center, Fort Leavenworth, KS (2008)
13. Ross, K.G., Grome, A., Ong, J., Arrastia, M., Schaab, B.B., Spangler, D.: Development and validation of training themes for joint, interagency, intergovernmental and multinational (JIIM) operations. In: Proceedings of the I/ITSEC Conference, Orlando, FL (2009)
14. Crandall, B., Klein, G., Hoffman, R.: Working Minds: A Practitioner's Guide to Cognitive Task Analysis. MIT Press, Cambridge (2006)
15. Militello, L.G., Hutton, R.J.B.: Applied Cognitive Task Analysis (ACTA): a practitioer's toolkit for understanding cognitive task demands. Ergonomics **41**(11), 1618–1641 (1998)
16. Ross, K.G., Wisecarver, M., Foldes, H., Peluso, D. A., Prevou, M., Schaab, B., Roberts, M.: Development of a Competency Model for Civil-Military Teaming (ARI Research Report 1960). U.S. Army Research Institute, Arlington (2012)

A Planning Tool for Interaction with Influential Actors Based on Paraconsistent Logic

Peter Svenmarck, Eric Sjöberg and Christophe Fagot

Abstract Interaction with influential actors is essential during international operations. Currently, however, commanders only have little support, if any, for planning and performing these interactions besides their own understanding of the situation. Since there are inherent conflicts among actors, a planning tool is presented that use paraconsistent belief integration to guide analysts' attention towards actors' relations, beliefs, and group identification modes, which may be important to consider when planning the interaction. The planning tool is evaluated in a representative scenario where analysts have to find ways and means to disrupt a drug trade between irregular and criminal actors. The available information enables the analyst to identify likely suspects in the drug trade, identify actors that have interests that are aligned the international force, as well as important considerations for planning the interaction with these actors. Preliminary evaluations show that better planning tools are very desirable for commanders and intelligence officers.

Keywords Paraconsistent logic · Influential actors · Key leader engagement · International operations · Schwartz value theory · Big five · Cross-cultural psychology · Modelling and simulation

P. Svenmarck (✉) · E. Sjöberg
Swedish Defence Research Agency (FOI), 164 90 Stockholm, Sweden
e-mail: peter.svenmarck@foi.se

E. Sjöberg
e-mail: eric.sjoberg@foi.se

C. Fagot
Intactile DESIGN, 20 rue du Carré du Roi, 34000 Montpellier, France
e-mail: fagot.c@intactile.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_21

1 Introduction

Interaction with influential actors is essential during international operations to get access to information, perform negotiations, and inform about the motives and intentions of the international force. An insightful management of the interaction with influential actors may therefore make the international force more effective in using available means to achieve the operation's goals. Since the interaction with influential actors is so important, it is not surprising that commanders typically spend 20–50 % of their time on this interaction [1].

Currently, however, commanders only have little support, if any, for planning and performing interactions with influential actors besides their own understanding of the situation. The lack of support means that the commanders face a daunting task to understanding actors' motives and intentions, the distribution of political and economic power among actors, supporting and opposing relationships among the actors, socio-cultural dynamics, as well as effects of own and others' actions. A planning tool that integrates this information and facilitate exploration of potential options may therefore potentially be very useful and improve the Key Leader Engagement (KLE) process.

A key aspect to consider for a planning tool is that the influential actors and social clusters that they represent usually have inconsistent motives and intentions. This is why there is a conflict in the first place that requires support from the international community. Since inconsistent motives and intentions are prevalent in international operations, the planning tool should preferably try to represent these inconsistencies. By representing and visualizing these inconsistencies, the planning tool can guide the analyst's attention towards actors' relations and beliefs that are important to consider when planning the interaction.

This paper describes a conceptual demonstrator of a planning tool for interaction with influential actors based on paraconsistent logic. Contrary to classical logic where contradictions are not allowed, paraconsistent logic can represent and reason with contradictions [2]. Previous research have used paraconsistent logic for both military and civilian applications [3], but this is the first application regarding influential actors. The main benefit of this type of applications is that they highlight critical information needs and assumptions that the user should address to resolve inconsistencies, but does not provide a solution. The user is therefore involved throughout the problem solving process and can utilize their own judgment for how to address the inconsistencies. The tool Vesta-Cosy is an example of recent efforts in this area that integrate paraconsistent logic in a graphical environment for exploring relationships between entities [4].

The paper begins with a description of the framework that was used to characterize influential actors. Then follows a description of the military planning process for Key Leader Engagement (KLE). Next, the framework is applied to a few influential actors within the North Friendly Sea (NFS) scenario that is often used for combined staff exercises in Sweden. The NFS scenario integrates many characteristics of contemporary conflicts, such as oppression by authoritarian

regimes, ethnic conflicts, drug and arms trade, etc. Finally, the conceptual demonstrator is described, along with the results form an initial evaluation and some conclusions for future developments of the planning tool. Please see [5] for more details about the planning tool.

2 Framework for Characterizing Influential Actors

Within the area that commanders are responsible for, there may be many types of actors or groups that each have their own motives and intentions for supporting and protecting the development of their own group. An actor or group is here considered as a number of individuals with shared preferences, routines, and convictions that have an essentially cooperative attitude towards each other.

Influential actors, in turn, are representatives for a group that have larger impact on or control over their group compared to other individuals within the group. The influential actors' power within the groups depends on their personal characteristics and history, as well as what kind of influential actors that the groups need for their development.

For the framework here, however, it is assumed that the large variety of actors and influential actors can be explained with a few basic mechanisms of human behavior. Such mechanisms include how values form from social and contextual factors, personality traits, and evolutionary processes that govern group values and identification with groups. Some general literature within these areas are therefore reviewed to identify communalities that can be used to characterize influential actors and their groups.

2.1 Value Structures from Contextual and Social Factors

The values that groups cultivate are an important driver for how people identify with the group. Values often develop adaptively as a response to conditions and factors that are important for the group. How groups prioritize values therefore gives some insight into the groups' context. Schwartz [6] describes how values support social functioning and a value theory where the following ten values suffice to define most groups:

- *Self-direction* refers to the importance of having control and independence. Self-direction therefore promote freedom and own choice of goals.
- *Stimulation* refers to the importance of variety for a positive activation rather than a negative activation from threats.
- *Hedonism* refers to the importance of experiencing pleasure and gratification for oneself.

- *Achievement* refers to the importance of personal gain from adhering to social standards.
- *Power* refers to the importance of social status and control over people and resources.
- *Security* is essential both for individuals and groups. Without an appropriate level of security, a group may not be sustainable over time.
- *Conformism* refers to the importance of avoiding actions that may disrupt group interaction and functioning.
- *Tradition* refers to the importance of a group's practices, symbols and ideas.
- *Benevolence* refers to the importance of preserving and enhancing the welfare of other members in groups that the person identifies with.
- *Universalism* refers to the importance of preserving and enhancing the welfare of the society and world.

Further, due to inherent conflicts among they cannot all simultaneously have a high value. Instead, the value relations form a circumflex structure where two dimensions suffice to describe value conflicts. One dimension is the focus on oneself that varies from *Self-Enhancement* for *Power* and *Achievement* to *Self-Transcendence* for *Benevolence* and *Universalism*. The other dimension is the interest in changes that varies from *Conservation* for *Security* and *Tradition* to *Openness to Change* for *Stimulation* and *Self-direction*.

2.2 Personality Traits

Another important driver for group identification is people's personality traits since similar ways of thinking and acting creates a stronger commitment to the group. The two most common frameworks for describing personality traits are the Myers-Briggs Type Indicator [7] and the Big Five [8], which are closely related. Myers-Briggs *Extraversion/Introversion* for whether cognitive functions are oriented towards the external or internal world is similar to the Big Five factor *Extraversion* for whether people are outgoing or reserved. Further, Myers-Briggs *Sensing/Intuition* for whether information gathering focuses on perception or underlying theories is similar to the Big Five factor *Openness* for whether people are curious or cautious. Additionally, Myers-Briggs *Thinking/Feeling* for whether decision making is performed through rational reasoning or through empathetic association with the situation is similar to the Big Five factor *Agreeableness* for whether people are compassionate or suspicious. Finally, Myers-Briggs *Judging/Perception* for whether the focus is on decision-making or information gathering is similar to the Big Five factor *Conscientiousness* for whether people organized or less affected by disorder. The only Big Five factor with no correspondence in Myers-Briggs is *Neuroticism*.

2.3 Evolutionary Processes of Social Roles

A third type of driver for group identification is evolutionary processes that shape behavior patterns and social roles. However, since people vary in the strength of their instincts and reflexes, they react differently to the same event. This reactive predisposition means that people are more suitable for some social roles than others within the group.

Since groups typically need to utilize available opportunities, there is a need to explore the local environment. However, the extent of this exploration depends on how individuals perceive their safety. Individuals who feel safe are more likely to be curious and independent, where as individuals who feel unsafe are more likely to seek shelter and support from what they associate with safety. While frightened children often run to one of their parents or their family for protection, anxious adults usually find comfort in a group's community, strong leader, or higher power.

Finally, since there is seldom an abundance of resources, there may be competition over available resources, such as food, territory, and partners. Although humans have an extraordinary capability to cooperate, there is also an evolutionary behavior to fend off competitors through dominance for more access to resources. However, dominators may be perceived both as aggressors for the group that is attacked, as well as protectors for the group that they represent.

2.4 Modes for Group Identification

The last driver for group identification in the framework presented here is the modes for group identification that describe in what way group members identify themselves with the group. Roccas et al. [9] suggests that group identification can be characterized with four different modes where:

- *Importance* describes how important the group is for peoples' self-definition of who they are. Group members often adopt a social identity in addition to their personal identity. The social identify guides self-categorization, goal sharing, and sense of similarity to other group members.
- *Commitment* describes peoples' desire to contribute to the group's development even at a personal cost. Commitment derives from the positive feeling of prioritizing group goals over personal goals.
- *Superiority* describes how group members perceive the group as superior to other groups. Superiority drives attitudes of better performance and being more worthy than other groups.
- *Deference* describes how group members subordinate to group norms, symbols, and leaders. High levels of deference means that such group manifestations are never questioned or criticized.

Studies by [9] indicate that the modes for group identification form a circumflex structure similarly to Schwartz's value theory. There therefore appears to be conflicts between group identification modes that focus on *Importance* and focus on *Deference*, as well as between a focus on *Commitment* and a focus on *Superiority*.

2.5 Summary of Drivers for Group Identification

Figure 1 shows the similarity between drivers for group identification. People that focus on self-enhancement through power and achievement will benefit from being competitive and dominating with a sense of superiority. Additionally, it seems reasonable that such people will benefit from being suspicious and antagonistic. People that focus self-transcendence, on the other hand, through benevolence and universalism will attend to and care for others. Such behaviors will benefit from a commitment of prioritizing shared goals over personal goals. Further, people in the other dimension that focus on being open to changes are more likely to be curious

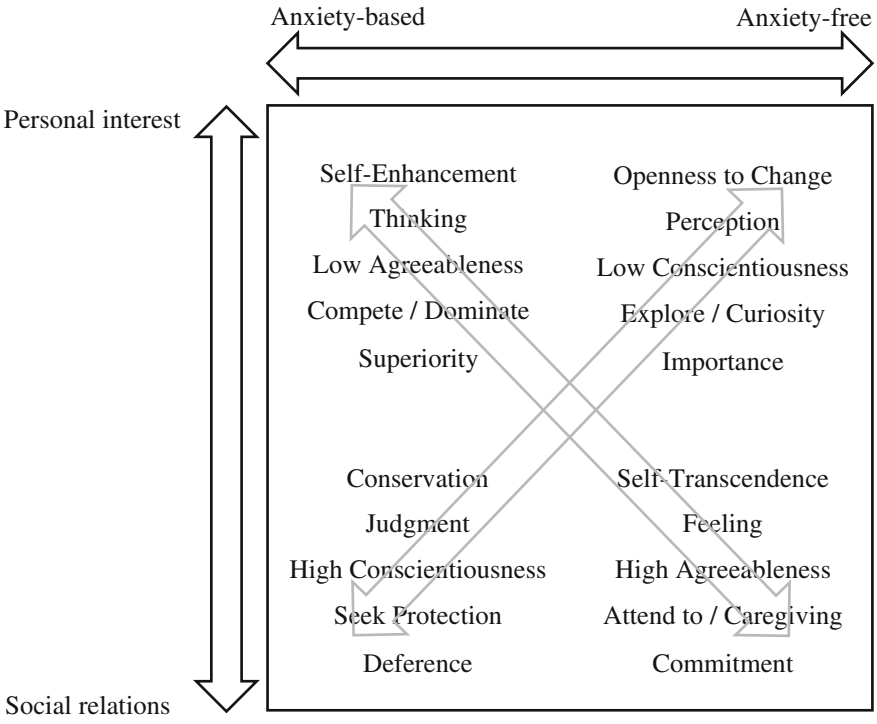


Fig. 1 Similarity between drivers for group identification

and explore their environment. From such openness to changes there may be less emphasis on planned behavior. Since the openness often challenge existing views, these people benefit from a high sense of importance. Finally, people that focus on conservation through tradition and security will be more interested in protection and adhering to group norms. Such subordination will also benefit if the people prefer organization and structure. Additionally, Fig. 1 shows how the values can be arranged as a response to anxiety or threat with a focus on either personal interests or social relations.

The relationship between values and personality traits gives some suggestions about groups' decision making process. Groups that focus on self-enhancement and have a low agreeableness are more likely to make decisions through rational reasoning. Groups that focus on self-transcendence, on the other hand, and have a high agreeableness are more likely to be more interested in consensus formation. In the other dimension, groups that are open to changes and have a low conscientiousness with less emphasis on organization are more likely to prefer perception and keeping their options open. Groups that focus on conservation and have a high conscientiousness will consequently prefer judgment and to settle matters early.

3 Key Leader Engagement

Interaction with influential actors can be performed in many ways and for many reasons. For example, the purpose may be to obtain knowledge that is important for the planning process, perform negotiations, or promote a long-term contact. The military planning process for how to perform interactions with influential actors is called Key Leader Engagement (KLE). The KLE planning is normally performed as a part of military staffs' intelligence and targeting functions during operational and tactical planning.

Although there is a general agreement what KLE means, the actual KLE process needs to be adapted to the staff and context where the KLE is used. The KLE process during staff exercises in Sweden combined information from international police forces, Human Intelligence (HUMINT), as well as informers. Commonly, there is both confirming information and contradictory information where conflicts need to be resolved. Visualizations of the social system model were then used as a basis for discussions about critical interactions and developments that should be considered. Thereafter, potential courses of action (COAs) were analyzed and assessed using the agreed and validated social system model in combination with models for how influential actors usually act based on their doctrine and documented behavior. Finally, the best COA was chosen for implementation. Please see [10, 11] for other examples of KLE.

4 Influential Actors in the North Friendly Sea Scenario

The North Friendly Sea (NFS) scenario was used as the overall context when applying the framework in Chap. 2 since it is a highly developed and adaptable scenario that is specifically designed for combined staff exercises. The NFS scenario enables staff training for many types of international operations that may vary in size from EU battlegroups to multinational coalition operations. Although the NFS scenario is fictional and based on Swedish geography, it is inspired by real world events and frequent aspects of contemporary international conflicts, such as refugees, atrocities, weak economies, weak governmental institutions, ethnic and tribal groups, irregular non-state actors, criminal organizations that profit from arms and drug trade, challenging health situation, environmental concerns, etc.

The NFS scenario contains around 12 countries where the main conflict is in Bogaland, a small country of some 9 million people in the NFS region. Following a stalemate, the Bogaland Peace Agreement (BPA) was signed in 2013. On a UN Chapter VII mandate, the multinational force, Bogaland Force (BFOR), will support the implementation of the BPA until general elections of a new democratic government. The operation is complicated by a long period of conflict that has enabled irregular groups to create their own post-war economy and now challenge provincial presidents and the overall peace process.

The scenario for assessing the planning tools focus on a reported drug trade by some actors within the Bogaland Special Police (BSP) through the irregular group Midan Snakes (MS) to the criminal organization Ultimate Defenders (UD). The reason for this new drug trade route is that BSP's regular smuggle routes have been disrupted by an uprising against the Bogaland Government and MS was seen as a good alternative. 30 influential actors and 44 social clusters were developed in sufficient detail to assess the planning tool. About half of these actors are suspected to participate in the drug trade, while other actors may potentially provide assistance in disrupting the drug trade, have relations with some suspected actors although not regarding the drug trade, or are government representatives.

All actors were profiled in terms of drivers for group identification using Schwartz's value theory and modes for group identification as described in Chap. 2. The values for *Stimulation* and *Hedonism*, as well as for *Conformity* and *Tradition* were combined since they are closely related for the purpose of the current scenario. The actors' profiles are representative for their group affiliation and personal interests. Further, background information about actors describes their affiliation, role within the organization, as well as personal relations.

5 A Planning Tool for Interaction with Influential Actors

A KLE planning tool for disrupting the drug trade should facilitate investigation of social links among entities, contrast their interests to enable identification of which actors to suppress and which actors to support, visualize actor's preferences to enable identification of suitable means for interaction, and visualize potential impacts of the KLE on actors' intentions. A promising software that include many of these aspects is Vesta-Cosy that is developed by Intactile DESIGN [4]. Vesta-Cosy combines ontological domain knowledge, a configurable XML-driven user interface, and an inference engine based on paraconsistent logic. Users can thereby utilize structured domain knowledge to analyses specific situations while receiving adequate support from visualizations, as well as information about critical inconsistencies. Vesta-Cosy is therefore designed to facilitate the user's problem solving process rather than to provide a solution.

5.1 Paraconsistent Belief Integration in Vesta-Cosy

Previous applications of Vesta-Cosy mainly use paraconsistent reasoning that chain logical rules and visualize paraconsistent results. However, since the Bogaland drug trade scenario mainly focus on actors' values, Vesta-Cosy had to be supplemented with paraconsistent belief integration to combine and contrast the differences between actors' values. While several approaches are available for paraconsistent belief integration, Paraconsistent Annotated Logic with annotation of 2 values (PAL2v) by [12] was sufficient for calculating relevant conflicts among actors' values in the drug trade scenario.

Figure 2 illustrates the principles of PAL2v where a rotated unit square is used to represent the four values *False*, *True*, *Inconsistent*, and *Indeterminate*. In PAL2v, *Favorable* and *Unfavorable* information about a value (normalized on a scale zero to one) are used to calculate the value's *Certainty* and *Contraction*. The *Certainty* is calculated such that it is *True* if the information is fully favorable and there is no unfavorable information, and *False* if the information is fully unfavorable and there is no favorable information. Similarly, the value's *Contradiction* is calculated such that it is *Inconsistent* if the information is both fully favorable and unfavorable, and *Indeterminate* if there is neither any favorable nor unfavorable information. Finally, the value's *Real Certainty* is calculated as the distance d of the coordinate for the value's *Certainty* and *Contraction* from *False*. The *Real Evidence* for the value is calculated by normalizing the *Real Certainty* on a scale zero to one. Please see [12] for more details about paraconsistent belief integration using PAL2v.

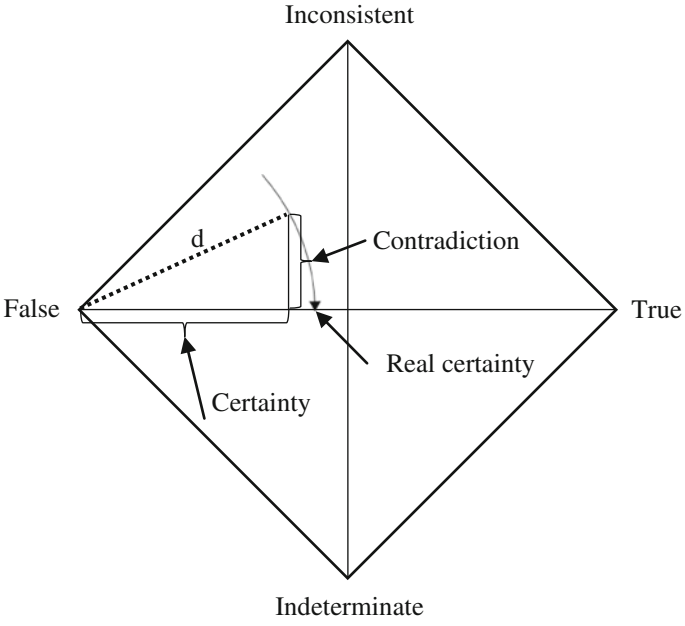


Fig. 2 Paraconsistent Annotated Logic with annotation of 2 values (PAL2v)

5.2 Belief Integration for Influential Actors

The actors' Schwartz values, mode values, and PAL2v calculations were used to drive four types of paraconsistent information and visualizations:

Robustness of Group Membership. Influential actors usually have similar values as the groups and social clusters where they are members. However, when actors are members of groups that they seemingly have nothing in common with this indicated that an overarching topic has so high value for the actor that it dominates the other interests. The robustness of actor's group membership can therefore serve as a signal to find important topics for actors, which can be useful information for KLE planning if the topics are aligned with mission goals. The planning tools calculates the robustness as the weighted sum of the real evidence for the actor's mode values compared to group's mode values. The weights are the group's mode values to assure that the relative importance of mode values for group identification is considered. Favorable information is the actor's mode values and unfavorable information is one minus the group's mode values.

Consistency of Primary Group Identification Mode. The group identification mode with the highest value for an actor and group can be considered as the primary group identification mode. If actors have different primary identification modes compared to the groups that they are members of this may indicate that they have a weaker position in these groups.

Inconsistency with Expected Identification Modes. Group member's mode values can be used to calculate expected identification modes for the group, which can be compared with the mode values in the group's profile. This can be seen as a way to verify to whether the mode values actually reflect selected group member's interest. The planning tool calculates the sum of selected group member mode values for each mode and then compares the modes' rank order with the rank order for the group's mode values.

Inconsistency of Personal Values. Since there are inherent contradictions between opposing values in Schwartz's value theory, the amount of agreement with this circumflex structure can be seen as the consistency of actors' personal values. Any inconsistencies indicate a cognitive dissonance that may be useful information for KLE planning. The planning tool calculates personal inconsistency as the maximum of the positive contradiction values where favorable information is the Schwartz value and unfavorable information is the opposing value. Negative contradiction values are not used since they reflect indeterminism rather than inconsistency.

5.3 Using Vesta-Cosy for KLE Planning

Military intelligence organizations typically collect considerable amounts of information to support operational planning. The actor profiles in the planning tool may therefore potentially already be available when BFOR receives information about the drug shipment to UD. The intelligence staff can then use the planning tool to model the social system of the actors that likely are involved in the drug trade.

Starting with the knowledge that UD is the recipient of the drugs, an intelligence officer can find all members of UD and place them in the Vesta-Cosy workspace. An investigation of their relationships shows that UD's Head of Logistics is in contact with MS red's Head of Logistics, a faction of MS that is interested in power and influence from profitable illegal trade. Further investigations of the relationships for MS red members show that they are in contact with the BSP Commander, Deputy Commander, and Head of Logistics. However, the BSP Head of Logistics is apparently also subordinate to an UD members, which may indicate a direct link between BSP and UD that may bypass MS. Continued investigations shows that the BSP Commander is in contact with a wealthy businessman in Bogaland, who in turn is contact with the former president of the Republic of Bogaland. Apparently, it is the old political and economic elite of Bogaland that tries to find alternative trade routes to continue a profitable drug trade.

Since BFOR does not have any specific information about when and how the next shipment will occur, it is not possible to just capture suspected persons in a raid for drug trade. A potential ally that may have better access to information and influence on disrupting the drug trade would obviously be MS grey and their leader Sören Vass who operate in the same area as MS red. MS grey is the moderate

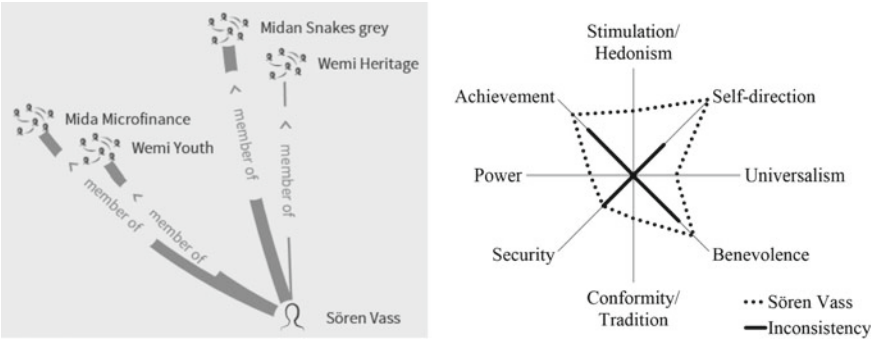


Fig. 3 **a** Group membership robustness for Sören Vass. **b** Personal Schwartz values and inconsistency for Sören Vass

faction of MS that wants MS to support and protect the local com-munity as MS had done for years.

Figure 3a shows the group membership robustness for all groups and social clusters where Sören Vass is a member. Clearly, Sören has a high membership robustness for all groups where he is a member expect for Wemi Heritage. Wemi Heritage is a social conservative organization and the complete opposite of Sören's in general liberal and progressive motives. A likely motive for Sören to join Wemi Heritage is that since they are against changes in general they are also against investments by foreigners, which is very important from Sören's perspective. Sören has always cared for the rights and safety of the local population and considerable foreign investments could mean that they lose control of resources that should be theirs.

Further investigations show that Sören has the same primary group identification mode as MS grey in general, which in combination with his leadership and high membership robustness for MS grey should mean that he has a good support in this group. Additionally, MS grey have a high value for importance which means that they likely rely on their perception of what they see in the actions of BFOR and other actors, and they may likely prefer to keep their options open. Their medium commitment may also mean that some consensus formation is necessary to reach a decision.

Figure 3b shows Sören's personal values, as well as the values' inconsistency according to Schwartz's value theory. The figure shows that Sören has higher values for self-direction than for security, and slightly higher values for *Hedonism/Stimulation* than for *Conformity/Tradition*. He is therefore more likely to be open to changes than to prefer conservatism. Other characteristics that often are associated with high openness are curiosity and exploration with less emphasis on planning. He also has a high importance value which may correspond to a preference for challenging existing views. Figure 3b also shows that Sören has a

moderate inconsistency for achievement and benevolence, and some inconsistency for security and self-direction. This shows that further analysis is required of in which situation Sören is mostly interested in self-enhancement, and in which situations self-transcendence dominates.

5.4 Evaluation

The planning tool was demonstrated for a previous Chief of Intelligence in a European Union Battlegroup, which is a rapid reaction force. The discussion centered on how KLE during international operations is a cooperation between J3 (Operations) that performs all targeting and J2 (Intelligence) that manages all information collection about influential actors. The KLE process is overall adapted to more rudimentary information and judgements, such as who has official positions and whether they are available for negotiations. Analysis of actors' networks, power relations, and their actual influence is seldom performed. There is therefore a considerable potential for a planning tools that balances the staff's analytic capacity, technical support, and available intelligence information.

6 Conclusions

The conceptual demonstrator of KLE planning with Vesta-Cosy shows that it works satisfactory for identifying actors that likely are involved the drug trade. Rather than directly presenting all relationships, the tool supports an interactive exploration of available information where only the information that is deemed relevant by the analyst is presented on the workspace. This gives the analyst a clear overview with simplifies the identification of Sören Vass as a potential ally to disrupt the drug trade. Further, paraconsistent belief integration of actors' values provides several additional perspectives about Sören Vass's motives and interests that may be relevant to consider for planning the future interaction. Currently, the planning tool does not provide any information about how to prevent other UD members from bypassing MS grey and MS red.

Finally, although the planning tool is only a conceptual demonstrator, it shows that paraconsistent logic can be a useful formalism to represent influential actors' inconsistent motives and intentions during international operations. However, the exact nature of the values to represent and integrate will likely vary between operations. The main benefits of the framework presented here are that it is based on research in cross-cultural psychology, can represent actors' inconsistent values, and combines several overlapping perspectives about personal values, personalities, and group identification. Further, since it is costly to obtain intelligence

information, the KLE process, representations of values and motives, usability of the planning tool, and the staff's capacity for KLE, must form a coherent system for effective KLE. Improving one component, such as the planning tool presented here, may therefore make it worthwhile to also reconsider other aspects of the KLE chain.

References

1. Eisenstadt, M.: Iraq tribal engagement: lessons learned. Mil. Rev. **2007**, 16–31 (2007)
2. da Costa, N.C.A., Krause, D., Bueno, O.: Paraconsistent logics and paraconsistency. In: Jacquette, D. (ed.) Philosophy of Logic (Handbook of the Philosophy of Science), pp. 791–912. North-Holland (2007)
3. Pierre, D.: With intègre, leverage every medical professionals' skills and expertise. In: Kostkova, P. (ed.) eHealth 2009, LNCIST 27, pp. 155–156. Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (2010)
4. Desfriches-Doria, O., Fagot, C., Svenmarck, P.: Socio-cultural modelling and dynamic decision making. In: AHFE 2016, 27–31 July 2016, Orlando, Florida (2016)
5. Svenmarck, P., Sjöberg, E., Fagot, C.: A preliminary planning tool for interaction with influential actors. Report no. FOI-R–4236–SE. Swedish Defence Research Agency (FOI), Stockholm (2016)
6. Schwartz, S.H.: An overview of the schwartz theory of basic values. Online Read. Psychol. Cult. **2**(1) (2012). http://dx.doi.org/10.9707/2307-0919.1116
7. Myers, I.B., McCaulley, M.H., Quenk, N., Hammer, A.: MBTI Handbook: A Guide to the Development and Use of the Myers-Briggs Type Indicator, 3rd edn. Consulting Psychologists Press (1998)
8. Costa Jr., P.T., McRae, R.R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources Inc., Odessa (1992)
9. Roccas, S., Sagiv, L., Schwartz, S., Halevy, N., Eidelson, R.: Toward a unifying model of identification with groups: integrating theoretical perspectives. Pers. Soc. Psychol. Rev. **12**(3), 280–306 (2008)
10. Hull, J.: Iraq: Strategic Reconciliation, Targeting, and Key Leader Engagement. US Army War College, Strategic Studies Institute (2009)
11. Nash, R.C., Magistad, E.P.: Disarming the key leader engagement. Mil. Rev. **2010**, 11–20 (2010)
12. Da Silva Filho, J.I.: Treatment of uncertainties with algorithms of the paraconsistent annotated logic. J. Intell. Learn. Syst. Appl. **4**, 144–153 (2012)

The Effects of Demographic and Sociocultural Variables on 3rd Year ROTC Cadets Officer/Evaluator Rating Scores

David R. Scribner

Abstract We examined the effectiveness of several pre-exercise measures to both observe relationships among these variables and to attempt to predict 3rd year ROTC Cadet officer/evaluator report (OER) scores, a strong indication of a cadet's leadership and decision-making capability. Participants consisted of a sample of 16 (13 male and 3 female) cadets from both the University of Virginia and Liberty University whose ages ranged from 19 to 24 years (mean = 20.7). Significant correlation relationships were revealed. A multiple regression model yielded strong evidence that demographic and 3CI scores were predictive of OER scores attained at Fort Knox. Correlations revealed multiple significant relationships leading to the suggestion that many mediating relationships existed within the variable relationship structure.

Keywords Sociocultural · Cross-cultural competency · Decision-making performance · ROTC cadets

1 Introduction

The prediction of decision-making performance and leadership qualities in culturally relevant military settings has been recognized as being highly important for the military of today and of the future [1]. In fact, identifying socio-cultural factors that influence decision-making and communication is an important aspect of military training, personnel selection, and future military technology [2, 3]. The Army has responded to the needs of its leadership by providing increased support to enable "foreign language and cultural capabilities" [4]. In fact, the Army is highly interested in developing cultural competence to better plan and execute operations

D.R. Scribner (✉)
U.S. Army Research Laboratory, Human Research & Engineering Direcorate,
459 Mulberry Road, Aberdeen Proving Ground, MD 21005, USA
e-mail: david.r.scribner.civ@mail.mil

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_22

within coalition and multi-national forces as well as within the U.S. Army structure [4].

There are important lessons learned that highlight the performance-based needs and the cross-cultural understanding required to make today's military operations successful [5]. Unfortunately, decision-making research has not focused on the inclusion of culture and even less frequently on decision-making in a military environment [6]. Also, little research has been focused on the Soldier's/Commander's own socio-cultural attributes and how these affect decision making within culturally relevant scenarios. When we think of the number of Soldiers interfacing with people from other cultures across a wide range of military missions, in vastly different regions, we know we have a challenge to meet for our emerging Army leadership.

This leadership is developed within our Army's officer development programs including the Reserve Officer Training Corps (ROTC). ROTC cadets are trained in leadership skills for several years in the college/university environment. It is common for ROTC 3rd years cadets to be graded during year-end field exercises in such aspects as leadership, intellect, and presence. During these exercises cadets are rated by their 4th year peers and subordinate team members at their unit field training exercises (FTX) on various aspects of unit leadership and decision-making qualities. These peer and subordinate leadership and decision-making qualities ratings are often scores of the cadet's presence (P), intellect (I), and leadership (L) demonstrated during these unit FTXs. More of these FTXs are including culturally relevant content within the ROTC unit training.

Many military programs offer cross-cultural curriculum [7] and unit training [8]. ROTC programs also teach students culturally-relevant curriculum and put cadets to the test in field exercises that include key leader engagement, for example. In fact, ROTC cadets are also selected to attend national Summer training where the Leader's Training Course (LTC) is offered. At courses such as the LTC, a cadet's overall officer/evaluator report (OER) score indicates a general potential for leadership and decision-making ability. Student samples have shown that sociocultural factors such as CCC can have influence on decision-making outcomes [9], and there are several aspects of CCC, which can be assessed through a number of different CCC inventories [10, 11]. This includes the cross cultural competency inventory (3CI) [12].

2 Purpose of the Study

The purpose of this study is to identify demographic and sociocultural factors that are significant contributors to unit P, I, and L scores and cadet leadership scores at the national level, in particular, officer/evaluator report (OER) scores. This was

attempted by collecting pre-exercise demographic and sociocultural (Cross Cultural Competency Inventory; 3CI) [12] measures and collecting post field training exercise scores as well as officer/evaluator report (OER) scores recorded at national the national leadership training course.

2.1 Research Questions

Research Question 1: What demographic, socio-cultural, and unit FTX variables have a relationship with ROTC cadet officer/evaluator scores?

Research Question 2: What demographic, socio-cultural, and unit FTX variables are predictors of ROTC cadet officer/evaluator scores?

3 Methods

3.1 Research Design

The design for this research, which is a quantitative survey method, accessed a convenience sample of ROTC cadets. This one-time survey research was inserted into pre-existing and scheduled ROTC field exercises using questionnaires to collect demographic and 3CI data prior to the exercises. For analysis, this design accessed existing ROTC unit FTX mission scoring data and national level OER scores that are normally collected by the ROTC cadre at both the universities listed and at Ft. Knox as post-exercise measures, respectively.

3.2 Setting and Sample

The participants in the sample consisted of 16 3rd year ROTC cadets who attended Liberty University and the University of Virginia under the ROTC auspices of "Task Force Blue Ridge", which is commanded by a full-time Army Colonel who commands from the University of Virginia. The ROTC cadets were all adult college students whose ages ranged from 19 to 24 years. Participants were all provided full disclosure of the purposes and risks of the study which was acknowledged via a signed consent form. Once completing the study, participants were fully debriefed and given follow-up contact information for further questions.

3.3 Apparatus

3.3.1 Demographic Questionnaire

A demographics questionnaire was used to collect background and individual characteristic information from each participant. Data included age, gender, first language, cultural experience, cultural training, socioeconomic category, ROTC job knowledge, time in ROTC, and ROTC level of training.

3.3.2 Cross-Cultural Competency Inventory (3CI)

While one of the basic challenges is to define what CCC is, there are some common elements to many of the CCC inventories [13]. The Cross-Cultural Competence Inventory (3CI) [12] has been used in the past to assess how well individuals may adapt to working with and being comfortable in, new cultures, specifically for military environments. One of the unique aspects of the 3CI is that it has been validated with over 5,000 military members. The subscale dimensions that the 3CI measures are cultural adaptability, engagement, mission-focus, determination, and social desirability. Social desirability is used to screen for individuals high in this construct to remove those who may be trying to deny socially undesirable traits [14].

Higher cultural adaptability scores generally indicate how comfortable and relatable a person would be with people from other cultures. Lower scores may indicate an unwillingness to explore other cultures with close-mindedness and ethnocentrism as key descriptors. Higher engagement scores usually reflect a willingness to make friends, be optimistic and cheerful, and to be able to control their emotions during interactions. Lower scoring individuals might prefer to be isolated and not enjoy these social interactions. Mission-focus is typically related to structure and discipline and higher scorers would typically be rule followers, use planning and logic to solve problems, and enjoy routine. Low scorers might indicator a lack of maturity and discipline. Determination is the tendency to remain task-focused regardless of intruding non-task thoughts. Important to military leadership, high scorers would likely be good problem solvers being confident in their ability. Low scorers may have more difficulty with tasks that are performed with attention distracting.

The 3CI also includes a composite score which is a mean of the four sub-scales of cultural adaptability, engagement, mission-focus, and determination. Scores on this composite typically might reflect an individual who can interact effectively with other cultures.

3.4 Procedures

Participants were briefed on the nature of the study during the ROTC exercise pre-brief and were given a detailed description of the participatory requirements and the confidential nature of their personal data. If agreeable, each participant read and signed a consent form. All participants were given a demographic data form, and the Cross-Cultural Competency Inventory (3CI) to respond to. Participants then completed the surveys which required a maximum of about 30 min to complete. Once completed, the forms were all checked for completeness and turned into the investigator. Participants were then debriefed and given follow-up contact information. Participants then performed several days of their ROTC exercises, which were unobtrusively observed alongside the universities' ROTC cadre. At the completion of the exercises, mission scores collected by the ROTC cadre were submitted for later analysis. Follow-on national ROTC scores, or officer/evaluator scores assessing leadership training were also collected from follow-on field training exercises conducted at Ft. Knox.

4 Data Analysis

All data analysis was performed using SPSS version 17.0. The analyses performed included basic descriptive data, correlation, and multiple regression analyses. All data were entered into SPSS and checked for data entry errors, proper minimum and maximum values, and missing data. No missing cases were found. All multiple regressions used the stepwise method with the exception of the task difficulty level only which used the enter method.

4.1 Descriptive Data Analysis

Descriptive data were to define some of the basic demographic characteristics of the group. The mean age of the sample (n = 16) was 20.7 years (SD 11.1). The sex of the sample was broken down to be 81.3 % Male (13), 18.7 % Female (3).

4.2 Correlation Analysis

Pearson correlation analysis was used to determine if any significant relationships existed among all demographic, 3CI, subordinate presence, leadership, and intellect scores as well as the OER scores.

4.3 Multiple Regression Analysis

A multiple regression analysis using the stepwise model was employed to determine if any of the demographic and 3CI variables were significant predictors of OER scores at the national level. The data were analyzed as univariate grouped data for the purposes of this analysis [15]. Assumptions for multiple regression analyses including a critical chi-square value ($p = 0.001$) value for Mahalanobis distance was used to identify potential outliers, which were met. Multicollinearity and variance inflation factor (VIF) were assessed to reduce mutual independence or variables and to reduce the probability of a significant linear relationship among two or more predictors in multiple regression models [15].

4.4 Results

4.4.1 Correlation Analysis

Correlation coefficients were computed to include demographic variables, 3CI measures, subordinate presence, intellect, and leadership scores, and national officer/evaluator ratings. Partial correlations were calculated to control for the variable 'Time in ROTC' as an experience variable which correlated with no other variables, yet was viewed as a highly important aspect to control for based upon observations when data was collected. A summary of the significant partial correlations and significance levels is shown in Table 1. Non-significant correlations were not listed.

4.4.2 Multiple Regression

A multiple regression analysis was performed using the stepwise procedure to determine the effects of demographic and 3CI variables in predicting Fort Knox Officer Evaluator Ratings (OER). A Mahalanobis distance revealed no outliers. None of the variables were transformed. Evaluation of the linearity of the DVs determined that while transformation of these variables might have improved the residuals scatterplot, the remaining related residuals scatterplots were assessed to meet the assumption of univariate normality so no transformation was used in order to maintain uniformity among interpretation of regression results [15]. Bivariate scatterplots of these variables and residuals scatterplots confirmed suitable univariate normality. Regression results revealed a significant prediction model where $R^2 = 0.783, R^2_{\text{adj}} = 0.541, F(6, 14) = 3.75, p < 0.04$, accounting for 54.1 % of the variance in OER ratings. The regression equation using standardized Beta values for this multiple regression analysis can be expressed as:

Table 1 Correlation Analysis

Variable pair	Pearson *R*	*p*
Ft. Knox OER score—subordinate presence	0.588	0.03
Ft. Knox OER score—subordinate leadership	0.711	0.004
Ft. Knox OER score—mission focus	0.596	0.02
Subordinate presence—subordinate leadership	0.857	0.000
Subordinate presence—ROTC Knowledge	0.583	0.03
Subordinate Leadership—Determination	0.762	0.002
Subordinate Leadership—ROTC knowledge	0.535	0.05
Subordinate leadership—engagement	0.616	0.02
Subordinate leadership—mission focus	0.550	0.04
Mission focus—ROTC knowledge	0.564	0.04
Mission focus—cultural exposure	0.571	0.03
Determination—ROTC knowledge	0.642	0.01
Determination—engagement	0.604	0.02
Cultural exposure—cultural training	0.746	0.002
Cultural training—socioeconomic category	−0.556	0.04

$$\begin{aligned}\text{OER} = {} & 0.814_{\text{Mission Focus}} + 0.590_{\text{Engagement}} - 0.538_{\text{Cult.Exposure}} \\ & + 0.521_{\text{Time in ROTC}} + 0.477_{\text{Socioeconomic.Category}} + 0.299_{\text{Cult.Training}}\end{aligned}$$

As with many smaller sample size research projects, statistical power is often a concern. A post hoc power analysis revealed that with 6 predictors, the observed R^2 of 0.738, probability of 0.04, and $n = 16$, observed statistical power of this multiple regression was calculated to be 0.963.

A follow-on analysis of the correlation results was conducted to put some of the variable relationships into a better understood context. This was performed by linking all of the variables by significant correlation and will be discussed next.

4.4.3 Follow-on Mediator Effects Analysis

The partial correlation analysis was created to identify those variables with a significant relationship between them. Within Fig. 1 below, the relationship of these variables was created to formulate the multiple connections among these relationships as they pertained to Ft. Knox OER scores. Additionally, those variables that were included in the multiple regression equation have bolded borders for easier identification. As it can be seen, only three variables were correlated significantly with the national OER ratings: subordinate presence (P), subordinate leadership (L), and mission focus (3CI). However, several other variables not correlated with OER scores, but correlated with these three variables are listed: determination (3CI), ROTC knowledge, Engagement (3CI), Cultural exposure, and social desirability

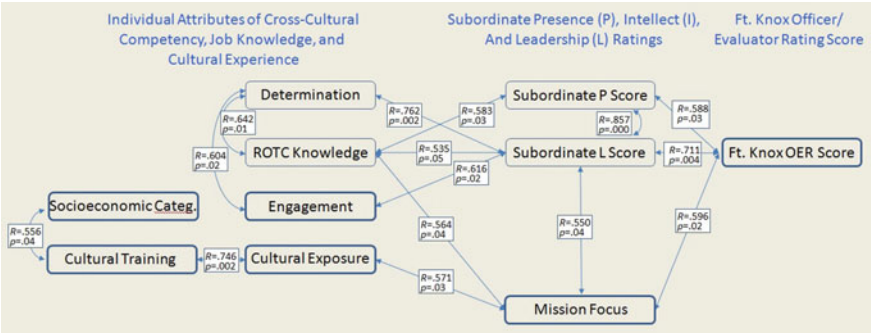


Fig. 1 Variable relationship diagram between individual attributes, subordinate presence and leadership, and national OER ratings

(3CI). From the structure of this relationship, it can be seen that six distinct mediator relationships that are linked to Ft. Knox OER scores can be defined [16]. A mediator relationship is one where a variable is not directly correlated with a variable of interest, but indirectly through other variables. The correlations among the 6 variables on the left side of the figure (Individual Attributes) are not correlated with the Ft. Knox Officer/Evaluator Rating (OER), thus creating the indirect mediating relationships.

This structure clearly identifies and suggests that sociocultural variables may not be directly attributable to measurable outcome data and may have an indirect relationship through other variables. One variable from the 3CI was directly related to the OER scores: Mission Focus. However, the remaining individual attribute variables were indirectly related and influenced OER ratings through the suggested mediator effects of subordinate leadership, presence, and mission focus scores. Additionally, most of the individual attribute variables that are predictors within the multiple regression could be considered sociocultural in nature.

5 Discussion

The partial correlation analysis, controlling for 'time in ROTC', easily understood as an experience factor, determined that there were significant bivariate relationships among numerous variables and that there were a number of sociocultural factors including cultural training, cultural exposure and constructs measured by the cross cultural competency inventory (3CI). This was sufficient to support the first research question based upon the findings of these significant relationships.

The data yielded from the multiple regression analyses were also considered sufficient to support the second research question. The regression model contained multiple components including four demographic metrics (socioeconomic category, time in ROTC, self-reported cultural exposure, and cultural training), and two 3CI

constructs (engagement and mission focus) from local unit FTXs. The multiple regression model demonstrated consistency with the variables identified in the variable relationship model constructed form the partial correlation analyses. However, one variable, time in ROTC, was a significant predictor in the multiple regression model, yet had no significant correlations any of the other bivariate correlation analyses. This variable did however demonstrate high usefulness as a partial correlation control variable.

The variable relationship model also directly linked subordinate presence and leadership scores as well as mission focus (3CI) to the national OER scores. The directionality of most of the relationships within the variable relationship model are intuitive, indicating that higher values on most scores contribute to higher OER scores, and longer time in the ROTC program and higher socioeconomic category (family income) may also relate to higher OER scores. Speaking of socioeconomic status, as unfortunate as it is, higher socioeconomic status is often correlated with higher academic achievement scores and related activities [17, 18]. The negative predictor of cultural exposure within the multiple regression was counter-intuitive and cannot be readily explained. The structural relationship created from the correlation analysis tends to better support than not, the findings within the multiple regression analysis.

The structural relationship model created from the significant variable correlations demonstrated numerous possible mediator effects within the variable set which has a number of measures in common with the multiple regression predictor set. Some variables from the correlation analysis that are found within the multiple regression have even further indirect relationships: cultural adaptability, cultural training, and socioeconomic status. In the variable relationship structure, cultural adaptability, cultural training were related to cultural exposure, while socioeconomic status was related to engagement. These variables, while not directly linked to the OER scores demonstrated simultaneous importance in both the multiple regression and variable relationship models, working through indirect paths, suggesting multiple and complex mediating relationships.

6 Summary

6.1 Research Implications

Several implications can be derived from the evidence collected in this research including the direct and indirect effects of sociocultural factors. As can be seen from the relationship structure of the variables, sociocultural factors such as mission focus (3CI) and subordinate leadership scores directly relate to OER scores. However, the remaining other demographic and sociocultural variables worked indirectly through suggested mediator relationships to have indirect effects on OER scores. These implications lead to the conclusion that direct as well as indirect

relationships should be considered when conducting quantitative analysis of demographic and sociocultural variable effects on leadership and decision-making qualities. The implications also inform future research of the importance of finding variables that have "covariate" effects and that could be considered 'control' variables such as experience factors, which may have a varying effect on other dependent measures.

Further, the research demonstrates that the constructs accessed within the 3CI and certain demographic factors have strong potential to be predictors for leadership and decision-making potential in our future officers based upon this ROTC sample. In particular mission focus, engagement, cultural exposure, time in ROTC, cultural exposure, and socioeconomic status all have significant predictive relationships to the officer/evaluator scores. These cross-cultural competency (3CI) relationships also have an intuitive direction, which seems to indicate that higher scores on mission focus and engagement led to higher OER scores. Mission focus, determination, and engagement, are all seemingly important 3CI factors in the relationship of variables to final OER scores.

The theoretical implications of cross-cultural competence generally agree with the Thornson [12] model and findings that show correlations and multiple regression analyses show strong support between demographic metrics, 3CI constructs, and some unit leadership measures and higher functioning officer skills in terms of observed decision-making and leadership behavior.

7 Conclusions

Research involving sociocultural factors is varied and the number of assessment tools and inventories to capture cross-cultural competency is numerous. However, subscales of the 3CI has been shown to have substantial predictive power combined with key demographic metrics in terms of identifying potential leadership individuals who may show aptitude for both general decision-making and those cross-cultural scenarios where particular skills, personality attributes, and knowledge are required. Other cross cultural competency inventories and demographic measures used to assess cultural exposure, cultural training, socioeconomic status, and specific unit training knowledge may yield similar results, yet this is to be determined.

This research effort, while using a relatively small sample was able to achieve high levels of power for multiple regression analysis, demonstrating the utility of many of the metrics employed within this study. Future research could enhance the state of sociocultural research by accessing field training exercises and other live or simulated scenarios where objective mission performance is captured. While using behavioral observations for grading is often the norm in the military for grading ROTC cadets and junior active Army officers, objective mission and decision-making performance will yield more powerful and conclusive results for

sociocultural predictors. It is proposed that future research in this area be expanded to Army officers of varying experience to broaden the sample types and to allow generalization to Army decision-making staff.

References

1. Abbe, A., Gulick, L.M.V., Herman, J.L.: Cross-Cultural Competence in Army Leaders: A Conceptual and Empirical Foundation. United States Army Research Institute for the Behavioral and Social Sciences. ARI Study Report 2008–01 (2007)
2. Hamilton, B.A.: Final Report: Socio-Cultural Influences on Decision-Making and Communication. ARL prime Contract No. DAAD19-01-C-0065 (Alion). July 8, 2011. McLean Virginia, Booz Allen Hamilton (2011)
3. Samms, C., Animashaun, A., Henry, S., Hill, S., Patton, D., Ungvarsky, D.: Development of a Taxonomy of Socio-Cultural Factors that Influence Decision-Making. Army Research Laboratory Aberdeen Proving Ground, MD. ARL-TR-7207 (2015)
4. Department of the Army Headquarters (2009). Army culture and foreign language strategy. http://www.almc.army.mil/ALU_CULTURE/docs/ARMYCULTURESTRA.TEGY-01DEC09.pdf
5. Picucci, P.M., Numrich, S.K.: Mission-driven needs: understanding the military relevance of socio-cultural capabilities. In: Proceedings of the 2010 Winter Simulation Conference, pp. 454–465. Baltimore, Maryland (2010)
6. Guess, C.D.: Decision-making in individualistic and collective cultures. Int. Associa. Cross-Cultural Psychol. **4**(1), 1–18 (2004)
7. Forman, S., Zachar, P.: Cross-cultural adjustment of international officers during professional military education in the United States. Milit. Psychol. **13**(2), 117–128 (2001)
8. Abbe, A., Gouge, M.: Cultural training for military personnel: Revisiting the Vietnam era. Milit. Rev. **92**(4), 9–17 (2012)
9. Moniarou-Papaconstantinou, V., Tsatsorani, A., Katsis, A., Koulaidis, V.: LIS as a field of study: socio-cultural influences on students decision- making. Aslib Proc. New Inf. Perspect. **62**(3), 321–344 (2010)
10. Matsumoto, D., Hwang, H.C.: Assessing cross-cultural competence: a review of available tests. J. Cross Cult. Psychol. **44**(6), 849–873 (2013)
11. van Driel, M., Gabrenya, W.K.: Organizational cross-cultural competence: approaches to measurement. J. Cross Cult. Psychol. **44**(6), 874–899 (2012)
12. Thornson, C.A.: Development and validation of the cross-cultural competence inventory Doctoral dissertation, University of Central Florida). Retrieved December 18, 2014, from Dissertations & Theses: Full Text. (Publication No. CFE0003437) (2010)
13. Gallus, J.A., Gouge, M.C., Antolic, E., Fosher, K., Jasparro, V., Coleman, S., Selmeski, B., Klafehn, J.L.: Cross-cultural competence in the department of defense: an annotated bibliography. U.S. Army Research Institute for the Behavioral and Social Sciences. Special report 71. Ft. Belvoir, VA (2014)
14. Phillips, D.L., Clancy, K.J.: Some effects of 'social desirability' in survey studies. Am. J. Sociol. **77**(5), 921–940 (1972)
15. Tabachnick, B.G., Fidell, L.S.: Using multivariate statistics, 5th edn. Upper Saddle River, NJ, Pearson Allyn & Bacon (2007)
16. Baron, R.M., Kenny, D.A.: The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. J. Pers. Soc. Psychol. **51**, 1173–1182 (1986)

17. Sirin, S.R.: Socioeconomic status and academic achievement: a meta-analytic review of research. Rev. Educ. Res. **75**(3), 417–453 (2005)
18. Willingham, D.T.: Why does family wealth affect learning? American Educator, Spring, pp. 33–39 (2012)

Part VII
Sociocultural, Behavioral Modeling for Defense Applications: Theory and Practice

The Conundrum of Verification and Validation of Social Science-Based Models Redux

Heidi Ann Hahn

Abstract In systems engineering, the definitions of "verification" and "validation" are settled; consistent with the US Department of Defense's (Department of Defense, [1]) definitions; and distinct from one another. This distinction blurs in the V&V of social science-based models. Unlike physics-based models, the theoretical underpinnings of social science models are not readily verified through observation of real-world events or empirical testing. Hence, they are often contested. When experts do not agree on what the right thing to do is, determining that the model is built right (verification) and that the right thing has been built (validation) cannot be separated (Verification as a form of validation: deepening theory to broaden application of dod protocols to the social sciences, [2]). This paper updates an earlier one (The conundrum of verification and validation of social science-based models, [3]), that reviewed the literature on V&V of models and outlined future directions, and describes the experiences LANL researchers have had with the V&V of extensible logic models (https://www.orau.gov/dhssummit/2009/presentations/march18/plenary/shevitz_mar18.pdf, [4]) used to evaluate the efficacy of various technologies in countering national security threats.

Keywords Verification and validation · Social Science-Based models · Extensible logic models

1 Verification and Validation—Definitions and Overview

In the systems engineering lexicon, definitions for the terms "verification" and "validation" are settled; consistent with the US Department of Defense's (DoD's) definitions [1]; and distinct from one another. Verification confirms that all elements

H.A. Hahn (✉)
Los Alamos National Laboratory (LANL), PO Box 1663,
MS F696 Los Alamos, NM, USA
e-mail: hahn@lanl.gov

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_23

of the system perform their intended functions and meet technical (performance and operational) and interface requirements and constraints (i.e., the product was built right). Validation confirms that the realized system satisfies stakeholder needs, providing the correct solution to the customer's problem (i.e., the right system was built).

In the literature on models and simulations, however, the distinctions between verification and validation are not so clearly defined. Macal [5] states that verification ensures that the specification is complete and that mistakes have not been made in implementing the model. So far, so good, but he then goes on to say that verification does not ensure that the model meets a specified set of requirements. But, wasn't that part of the definition of verification? Clearly, to Macal's way of thinking, the model requirements revolve around accurately representing and correctly reproducing behaviors of the real world system. By his definitions, it is validation that ensures that the model meets its requirements in terms of the results obtained.

Sargent [6] also fails to draw boundaries between verification and validation. He speaks of model validation as determining whether the model correctly represents a governing theory (whether the model was built right) and whether the model works in a "reasonable way" given its purpose, intertwining verification and validation.

In the world of modelling and simulation, the distinctions between verification and validation have been collapsed into the generic problem of validation, which determines whether the right model has been built. V&V of theory-based models, then, boils down to a question of whether the model is credible for its intended use; this is different than how V&V is defined for physical systems, where the concerns are with whether the system satisfies requirements and is acceptable to the customer.

1.1 Discussion of Validation as It Is Found in the Social Science and Theory-Based Modeling Literature

Validity, generically, "refers to measuring what we think we are measuring" [7] or, in the case of models, representing what we think we are representing. Indeed, there is much is the social science literature about the validity of models and methods for establishing model validity. Thomas [8] asserts that model validation efforts must be concerned with internal validity—the extent to which the relationships between variables are represented correctly in the model (verification)—as well as external or criterion-related validity, which considers the performance of the operationalized model and specifically the extent to which the model outputs agree with an external entity, which may be either a real world system or another validated model (validation).

In essence, all model validation efforts attempt to establish construct validity, or the extent to which the model accurately represents a theoretical construct or characteristic. Trochim [9] writes of two types of construct validity: translation validity and criterion-related validity. Translation validity focuses on whether the model operationalization is a good reflection of the underlying constructs. Translation validity comprises face validity (the extent to which the model appears to represent accurately what it is intended to represent) and content validity (the extent to which the model sufficiently covers the domain of interest). Trochim notes that content validation assumes that the content domain can be described accurately, an assumption that is not always true in the case of social science models. This is one of the characteristics of social science-based models that become problematic for model V&V.

2 Comparison of V&V of Physical Systems and Theory-Based Models

Semantics aside, V&V of physical systems versus theory-based models, be they models of physical or physics-based systems or human-based or social science systems, can be thought of as existing along a "continuum of objectivity" as shown in Fig. 1. Much of the difference in V&V of physical systems and theory-based models lies in the objectivity of the evidence basis and rigor of the methods used.

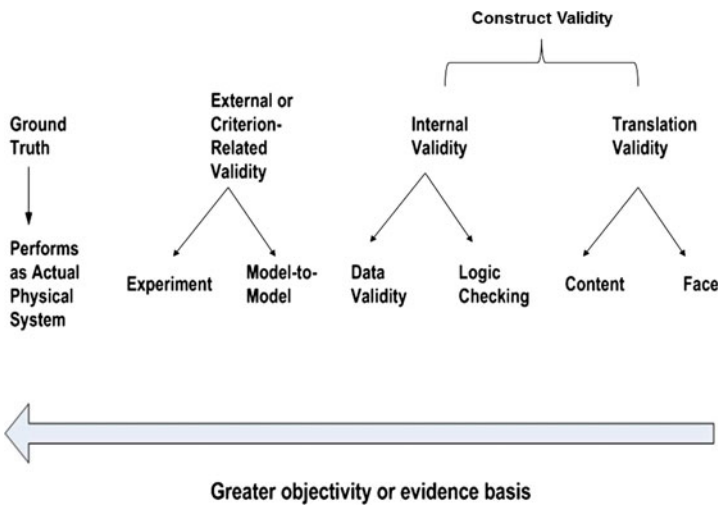


Fig. 1 V&V continuum of objectivity

2.1 V&V of Physical Systems

In validating physical systems, the left-most side of Fig. 1 is the normal operating regime—we want to prove that the system meets the performance requirements. Testing—measuring the system's response to a prescribed set of conditions, either real or simulated, and comparing the results to a set of specified operability, supportability, or performance capability requirements—is the preferred V&V method for physical systems. When testing is infeasible, such as when tests would result in destruction of the system, analysis using experimental data, models, and simulations to show theoretical compliance of the system to specified requirements is an acceptable V&V method. Demonstration—qualitatively exhibiting system functional performance, usually with little or no instrumentation—and inspection—examining or observing a system feature and comparing it against applicable documentation to confirm compliance with requirements—are also commonly used V&V methods [10].

Two fundamental assumptions underpin V&V of physical systems. First is the assumption that the system can be verified against "ground truth," as defined by an observable set of technical requirements (quantified statements that define how well, under what operational conditions, and/or to what degree functions must be performed) and constraints that enumerate specific parameters that the design must meet. Verification activities provide objective evidence that the physical system performs the required functions as specified.

Second, in V&V of physical systems there is the assumption that the technical requirements and specifications flow hierarchically from validated customer and stakeholder needs. In theory, a system that has been fully verified, proving that all requirements have been satisfied, should also be valid. In practice, a system may have been designed and implemented perfectly, meeting all requirements, but still not provide the correct solution to the customer's problem—the verified system is not valid. The problem can almost always be traced back to an incorrect, ambiguous, or missing requirement: the system as built is not valid because the design failed to adequately account for one or more customer or stakeholder requirements.

2.2 V&V of Theory-Based Models

When considering V&V of theory-based models, however, either, or both, of the aforementioned assumptions may be invalid. The option of validating against ground truth (i.e., historic data collected from a real system) is often not available to modelers. Further, as Lustick and Tubin [2] point out, theory-based models may pass verification checks, proving that the model was a correct operationalization of the design concept, but fail validation tests either because either because the design was not a valid implementation of the underlying theory or because the theory on

which the model was based was not valid. "In that case, the model would have inherited the invalidity of the theory upon which it was based" (p. 6586).

Typically, V&V of theory-based models exists in the middle of the range of the objectivity continuum, taking an empirical approach, ideally comparing the results of test cases to experimental data or, if that is not possible, to other validated models. Analytic techniques may also be used to show that the elements of the model are correct and are correctly integrated and to determine the model's "fit for purpose" [11].

Internal validity checks are a key element of the V&V of theory-based models. Here, the modeler verifies that the right data and logic have been captured. Because inaccurate data could be a significant source of inaccuracy in the model outputs, V&V of models typically includes analyzing data for inconsistencies, incorrect formatting, and gaps (unavailability of data); estimating the effects of inaccuracies; and putting procedures in place for collecting or estimating data for which high sensitivity exists [12]. Logic checks can be done by watching the model run over a period of time and observing whether the actual model behavior conforms to the expected behavior [12].

Special Considerations for V&V of Social Science-Based Models. Arguably, the "gold standard" for theory-based model validation is the "Standard Model" (the theory of the physical laws governing the fundamental particles of matter and their interactions [13]), whose predictions have matched experimental data with great accuracy. Other science-based Standard Models are also emerging—for example, advances in observing the large-scale Universe have led to a "Standard Model" of cosmology [14]. In the social sciences, however, there is no Standard Model, nor is there likely to be one in the foreseeable future. Indeed, the constructs underlying most social science-based models are regularly contested. (Take, for example, the debates over theories of leadership, and whether leadership qualities are inborn or learned, that have been going on since the introduction of the "great man" theory of leadership in the 1840s [15]). As noted by Lustick and Tubin [2], when experts do not agree on what the right thing is, determining that the model is built right cannot be separated from tests of whether the right thing has been built.

To complicate matters further, Ruvinsky, Wedgwood, and Welsh [16] correctly observe that the empirical approach to V&V falls short for social science-based models in the following areas:

- Experiment design—large-scale, naturalistic domain spaces pose challenges to the design and control of experiments
- Data acquisition—social phenomena tend to be described by multiple variables whose measurement may not be clearly understood
- Comparison of model results to real world data—real world data sets, either current or historical, may not be available or, if available, may not be sufficiently robust to support model validation

Finally, several authors (e.g., Lustick and Tubin [2] and Macal and North [17]) note that the purpose of validation for models incorporating human behavior is not

so much to validate the underlying theory but to establish credibility of the model. Validation, then, "works by removing barriers and objections to model use" ([17], p. 3).

As a result, V&V of social science-based models relies more heavily on the right-hand end of the objectivity continuum, with internal and translation validation techniques often being the primary methods used. Construct validity, in particular, aims to address the issue of experts not agreeing on what the "right thing" is.

3 Recent Innovations in V&V of Social Science-Based Models

Recent work on the validation of agent- and social science-based models provides insight into how to address some of the challenges described previously. Macal and North [17] have proposed innovative methods for V&V that build upon traditional V&V techniques, placing particular emphasis on internal validation and designing novel methods for performing model-to-model comparisons. Ruvinsky et al. [16] have proposed a radically different methodology, which aims to perform construct validation based on analysis of the epistemological structure of the model rather than on the operationalization of the model per se. Each of these approaches is reviewed in some detail below.

3.1 Innovations in the Validation of Agent-Based Models

Macal and North [17] have proposed innovative methods to validate EMCAS (Electric Market Complex Adaptive System), an agent-based simulation model designed to investigate the effects of electric power market restructuring and deregulation on electricity costs, reliability, and availability. As is the case with many social science-based models, the domain being modeled, a deregulated electric power market, did not exist, so comparison with real-world behavior was not an option. Thus, their task was to "establish an argument that the model produces sound insights and sound data based on a wide range of tests and criteria that 'stand in' for comparing model results to data from the real system" (p. 3). They constructed a validation framework and process, providing a set of resources that can be used to counter objections to the validity of the model and the results it produces. The framework comprises the following elements: data validation; independent subject matter expert (SME) judgment; participatory simulation; model-to-model comparison; critical tests and key indicators; comprehensive test cases; and invalidation exercises.

Noting that data gaps or inconsistencies can invalidate model results and destroy the model's credibility, Macal and North [17] used a variety of iterative analysis techniques to validate the data inputs to the model. These included:

- Mapping and cross-referencing data definitions from different databases and converting data to common units to ensure consistency
- Assuring currency of the data
- Engaging third-parties (stakeholders) in data verification
- Including proprietary data provided by stakeholders
- Using data visualization techniques to identify data anomalies that were not readily apparent through inspection

Macal and North [17] used a workshop format to engage SMEs in evaluating the model, model assumptions, and agent behavior. SMEs were instrumental in identifying test cases that would reveal model weaknesses during the invalidation exercises. In these exercises, multiple models runs were executed in systematic attempts to have the model exhibit unexpected behaviors, thereby invalidating the model assumptions. Unexpected behaviors became focal points for more in-depth analysis and model refinement. Macal and North note that the systematic design of the invalidation exercises was critical to avoiding validation bias, or the tendency to perform only those validation tests that are likely to validate the model.

Among the most novel of Macal and North's [17] validation methods was the use of "participatory simulation" (p. 4) in which real people played the roles of agents in the energy market. Results of the simulation were compared to the model results. The EMCAS results closely matched the results of the participatory simulation, validating model assumptions about agent strategies and behaviors under deregulation.

Using more traditional V&V methods, Macal and North [17] also generated test cases to evaluate model performance in replicating known system behaviors, such as replicating results for the regulated energy market; compared EMCAS against two other validated models representing special cases of EMCAS's operational parameters; and performed comprehensive testing across the full spectrum of plausible agent strategies (bounded by the assumption of rationality) and parameter settings for variables such as price, quantity, and generating capacity.

Macal and North [17] concluded that use of the validation process described above did result in the model being better accepted as a tool for answering important questions with respect to electricity deregulation, and that the process they developed is a generalizable and practical framework for agent-based model validation.

3.2 *Innovations Using Epistemological Decomposition for V&V*

Ruvinsky et al. [16] have proposed a very different V&V methodology, which is based on analysis of the epistemological structure of the model rather than on the operationalization of the model per se. They challenge the assumption that V&V methods for social science-based models should move to the left-hand side of the objectivity continuum and claim that their method "move[s] beyond viewing verification and validation as limited by empirical testing by providing mechanisms and techniques to verify and validate all aspects of knowledge/information that a model uses or produces" (p. 6595). Essentially, their goal is to develop a more robust construct validation method as an alternative to more traditional "operational" V&V.

Ruvinsky et al. [16] methodology takes what they call the focal V&V point of view, focusing assessment on "how well a model explains the phenomenon for which it was designed" (p. 6598). The first step in their method is the decomposition of the model into its hierarchical epistemological elements (element definitions are from [16], p. 6596). At the conceptual level, the elements are a social ontology—the set of background entities and beliefs about the world that pertain to and characterize a basic structure of reality and a paradigm or conceptual repertoire—a grand scheme of worldview that brings to bear the basic concepts prior to claims about specific domains. At the conceptual level, the elements are theories—abstract statements about reality describing relationships between or among concepts; social model—a representation of real world system behavior based upon theories and concepts; and hypotheses—conjectures within a theory regarding the relationships of two concepts to be explored in the social model. Finally, at the operational level, the elements are the application model—a description of how the social model will be refined and the concepts operationalized; the implementation model—the equations, parameter settings and coding rules to enable execution of the application model and manipulation of the raw data into model parameters; and data—selection of specific data bases and methods to access specific data sets from them.

For each tier, there are computational model and data artifacts that can be examined during V&V. For example, at the Data Level the data sources and raw data that serve as input into model parameters would be the artifacts associated with the computational model and the data model, respectively.

Evaluators independently assess the model against a series of "metadata elements" presented in questionnaire format, assigning a numerical or letter grade to each and noting discrepancies found. Reviewers' results are analyzed for similarities and differences, and scores and evaluator comments are used to make judgments about verification and validation at the various levels of the hierarchy. Ruvinsky et al. [16] assert that this ability to assess, for example, the V&V of a hypothesis separate from the V&V of the specific application model of the social construct from which it is derived, not only helps in managing the V&V effort but also enables model improvements by isolating the faulty parts of the model's epistemology.

4 LANL Experiences with V&V of Extensible Logic Models

Shevitz and O'Brien [4] have described a modeling methodology, called Extensible Logic Models (ELM), which they and their research group have used to model a wide variety of counter-terrorism scenarios for which there is a desire to prioritize a highly varied technology portfolio based on the expected effectiveness of the strategy or technology in countering the threat. The threat scenarios can be generically described by the lower activity sequence shown in Fig. 2. The objective of counter-terrorism models is to address the threat from beginning to end, as shown in the upper activity sequence. ELM models developed to represent the threat sequence include both human factors, such as group characteristics and pre-incident rhetoric and behavior, and operational and technical factors, such as time of day, terrain characteristics, the types of materials needed, and the ability to obtain materials through theft or purchase. When ELM models are used for technology assessment and prioritization, the technologies of interest are mapped into the counter-activity sequence; the factors considered include the ease and cost of developing the technology, the ease and cost of deployment, and the effectiveness of the technology at addressing the threat.

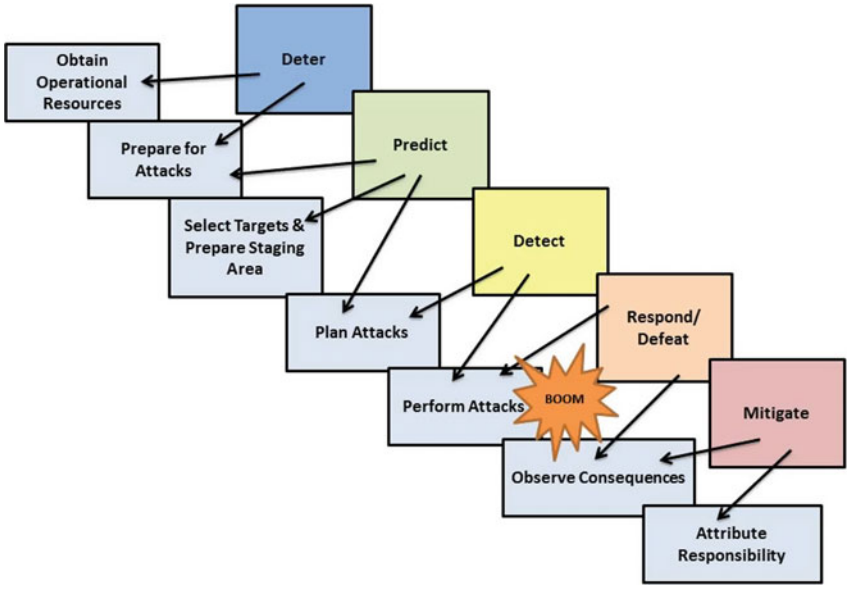


Fig. 2 Terrorist attack sequence with counter-activity overlay (adapted from a Department of Homeland Security figure in [18])

4.1 ELM Description [4, 18]

Extensible Logic Models are "programmable outlines" that enable analysts and decision-makers to manage hierarchical complexity through the use of fault and event tree-based logic diagrams, i.e. hierarchical models composed of "and" and "or" relationships, coupled with decision metrics utilizing either numerical or linguistic input data or both. The nodes of the tree are individually and independently programmable, allowing for complicated and customizable solutions. The decision metric produced by ELM can be risk or other values of interest, such as cost or consequence.

ELM analysis is an iterative process consisting of the following steps: (1) author the tree logic, creating the "and" and "or" gates and associated narrative; (2) assemble the data to be used in the analysis; data can be numbers, probabilities, linguistics (binned quantities such as "highly capable" or "incompetent"), results of other simulations, etc.; (3) program the ELM model; this entails creating and setting "path states" or user-specified pieces of information attached to the nodes, creating look-ups to databases containing the data to be used in the analysis, and using probabilistic risk tools and evidentiary reasoning to calculate decision metrics; (4) solve the model, generating all traversals of the model consistent with the program logic; and (5) post-process the results, aggregating the results into useful summary statistics, risk-ranked solutions, etc.

One of the strengths of ELM is that allows for integration of various data types and supports a variety of data analysis methods. Probability inputs may be simulation- or data-based or be the result of expert elicitation. Linguistic expression of uncertainty, in which analysts assign "degrees of belief" to different alternatives, and use of rule bases to "roll up" degrees of belief to infer an aggregate result support approximate reasoning for cases where probabilities cannot be evaluated in a straightforward manner due to lack of sufficient or reliable data or because likelihood inferences involve a complicated set of interrelated factors. Node computations may also be based on physical data (database look-ups) or physics-based models; any algorithm that is programmable can be incorporated into the model.

4.2 ELM Validation

As is the case in several previously cited instances, V&V of ELM collapses into the more general problem of validation, and specifically, of establishing construct validity, as it is described in Fig. 1. Validation of ELM models used for technology assessment takes place at multiple levels, including validation of: the generic threat and counter-activity sequences; details related to a specific threat; assignment of technologies to assessment factor levels; and technology prioritization results.

Recall that construct validity comprises translation and internal validity. Translation validity focuses on whether the model operationalization is a good

reflection of the underlying constructs [9]; it is established as a function of face and/or content validity. For the ELM models developed for counter-terrorism threat analysis, content validation of the generic activity sequences is possible by inspection (by subject matter experts [SMEs]) because consensus has been built regarding the general form of both the threat- and counter-activity sequences, as published in many venues including [18]. This speaks to the provenance of the constructs reflected in the models, which is one of the criteria discussed by Ruvinsky et al. [16] for validating models at their conceptual level. Generally, such content validation of ELM models is done during a series of table-top exercises using SME judgment, and is similar to the validation workshops described by Macal and North [17].

Content validation of the threat scenarios, and the theories underlying them, becomes more difficult when the level of analysis drops down to the specific threat scenario. For example, the theory of the process of radicalization for Group A may be very different than that for Group B, and there are most likely limits to the ability to test the assumptions or understandings that underpin the theory. This may be further complicated by the need to compartmentalize information for security purposes, such that the data that support the theoretical constructs are not available for inspection by independent SMEs.

Verification of the correct implementation of the technologies being assessed is also a matter of SME judgment. This has generally been solicited in an interview format, with interviews conducted by members of the V&V team rather than the modelers themselves. In some cases, there is objective evidence related to the performance of similar or precursor technologies that can be incorporated into the models; this evidence is subjected to data validation checking, as described below.

Validation at the operational level of Ruvinsky et al. [16] epistemological hierarchy is probably the most important aspect of translational validation for ELM. Specifically, they question whether the policy-relevant findings are substantively useful (p. 6600). Here, the evidence for ELM's validity is in its use by decision-makers within the sponsor community, which has been extensive [4].

Internal validity, or the extent to which the relationships between variables are represented correctly in the model [8], is concerned with the data and logic used in the model. This is the area in which ELM shines. Since ELM models are logic trees, the embedded logic is transparent, and can essentially be read just as one would read an outline, while also being an executable program, facilitating logic checking by inspection. Further, the software contains multiple mechanisms for "pruning" nonsensical or inconsistent possibilities, which helps manage the combinatorial explosion of possibility generation [18].

Although data collection is typically the hardest part of a modeling problem, ELM assists in data specification by making explicit what information is needed and suggesting ways to proxy missing data based on what information the modeler does know [18]. Further, allowing for the linguistic expression of variables facilitates the ability to fill in gaps in hard data since it is much easier and more defensible to assign "degrees of belief" and compute over binned quantities than to provide definitive quantities for metrics like cost (actual dollars vs. "very

expensive") and risk (% likelihood vs. "unlikely") as long as there is a common understanding of the anchors associated with the bins.

Data validation for ELM follows the analysis techniques described by Macal and North [17], specifically:

- Ensuring consistency of data definitions and units from different data sources
- Assuring currency of the data
- Incorporating stakeholder-provided data
- Engaging stakeholders in data verification

4.3 Observations on ELM Validation Relative to Macal and North [17] and Ruvinsky et al. [16]

As can be seen from the foregoing discussion, members of the ELM V&V team did rely on the literature when developing the ELM validation approach. For the most part, Macal and North's [17] approach, which builds upon traditional V&V techniques and emphasizes internal validation, was more understandable to, and thus better accepted by, stakeholder groups, including the modelers themselves. Inclusion of stakeholders in data specification and verification was key to this result.

The approach advocated by Ruvinsky et al. [16] was of interest to the ELM V&V team because it encouraged thinking through the verification and validation issues at a variety of levels. It proved difficult to implement, however, because, in some cases, the relationships of the "metatags" to the hierarchical levels were unclear. A literature search of publications by these authors produced no new insights into the actual implementation of the methodology. A case study example would be a helpful contribution for a methodology that seems to provide a quite promising construct.

5 Conclusions and Future Directions

To be accepted by decision makers, theory-based models, including social science models, must credible. Model validation is essential to credibility. Traditional, empirically-based V&V methods, however, are often not suitable for validation of social-science based models, either because there is no real world system behavior to compare the model results to or because real world data collection is intractable. Thus, researchers have been seeking alternative methods and frameworks for model validation for socio-technical models.

Although there has been significant progress in developing models for understanding, detecting, predicting, and effecting change in human behavior, there is much work still to be done. Research priorities in this area have been promulgated

by both the Department of Defense [19] and the Office of Naval Research [20]. Among the most interesting of these, from the Los Alamos perspective, is researching and engineering a "social radar" which would be a global and persistent indications and warning capability for detecting an monitoring relevant sociocultural behavior signatures, and integrating analytics of sociocultural behaviors with conventional and geospatial data [20]. Pabian [21], a LANL analyst, has described how geospatial tools have been used operationally to gather value-added follow-up information on open source reports of a former clandestine Chinese plutonium production complex, with human analysts correlating the data. Integrating a socio-cultural model with the geospatial tool set could enable the detection of contextually-dependent "suspicious behavior" and support decision-makers in determining the best course of action to deal with the observed situation.

Building a rigorous foundation for social science-based computational models must include establishing and building consensus on V&V methods [20]. Whether any of the approaches described above brings the community to such consensus remains to be seen, particularly because demonstrating that the models validated using these methodologies are "more credible"—which both Macal and North [17] and Ruvinsky et al. [16] hold out as their goal—may not be very satisfying. Rather, the challenge may be to move beyond thinking of model validation as primarily to establish "credibility" of the model and to develop metrics that begin to answer the question of whether models validated using these methods are "better" on a variety of dimensions—are the data more accurate, is the logic more internally consistent, are the models better predictors of whatever system behavior they purport to represent, etc.—to provide evidence that social science-based behavioral models are on par with the "goodness" of the science-based Standard Models.

References

1. Department of Defense, http://vva.msco.mil/Key/key-pr.pdf
2. Lustick, I., Tubin, M.: Verification as a form of validation: deepening theory to broaden application of dod protocols to the social sciences. In: Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics, pp. 6583–6592, AHFE International, San Francisco (2012)
3. Hahn, H.: The conundrum of verification and validation of social science-based models. In: Paredis, C., Bishop, C., Bodner, D. (eds.) Procedia Computer Science, vol. 16, pp. 878–887. Elsevier, Atlanta (2013)
4. Shevitz, D., O'Brien, D.: https://www.orau.gov/dhssummit/2009/presentations/march18/plenary/shevitz_mar18.pdf
5. Macal, C.: Model Verification and Validation. The University of Chicago and Argonne National Laboratory, Chicago (2005)
6. Sargent, R.: Validation and verification of simulation models. In: Ingalls, R., Rossetti, M., Smith, J., Peters, B. (eds.) Proceedings of the 2004 Winter Simulation Conference, pp. 17–28, Winter Simulation Conference Board of Directors, Washington, DC (2004)
7. King, G., Keohane, R., Verba, S.: Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton University Press, Princeton (1994)

8. Thomas, C.: Verification revisited. In: MORS, Military Modeling for Decision Making, pp. 333–350, Military Operations Research Society, Alexandria, VA, (1997)
9. Trochim, M., http://www.socialresearchmethods.net/kb/measval.php
10. Walden, D.D., Roedler, G.J., Forsberg, K.J., Hameline, R.D., Shortell, T.M. (eds.): Systems Engineering Handbook A Guide for System Life Cycle Processes and Activities. International Council on Systems Engineering, San Diego (2015)
11. Friedman, G.J.: Constraint Theory: Multidimensional Mathematical Model Management. Springer Science + Business Media Inc, New York (2005)
12. Robinson, S.: Simulation model verification and validation: increasing the users' confidence. In: Proceedings of the 1997 Winter Simulation Conference, Winter Simulation Conference Board of Directors, Atlanta (1997)
13. European Organization for Nuclear Research (CERN), http://home.cern/about/physics/standard-model
14. Peebles, P.J.E, https://ned.ipac.caltech.edu/level5/Peebles1/frames.html
15. Leadership-Central.com, http://www.leadership-central.com/great-man-theory.html#axzz3wJynLynK
16. Ruvinsky, A.I., Wedgwood, J.E., Welsh, J.J.: Establishing bounds of responsible operational use of social science models via innovations in verification and validation. In: Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics. Springer-Verlag, Heidelberg (2012)
17. Macal, C.M., North, M.J., http://www2.econ.iastate.edu/tesfatsi/EmpValidACE.MacalNorth.ElectricPower.pdf
18. Shevitz, D. Chavez, G., Key, B., Zerkle, D.: Application of the ELM Methology to C-IED Problems. Technical Report LA-UR-08-6378, Los Alamos National Laboratory (2008)
19. Office of the Secretary of Defense, http://www.dtic.mil/biosys/files/SBRE2011.pdf
20. Office of Naval Research, https://www.fbo.gov/utils/view?id=ea459ff2dde3362048607957ccc830a0
21. Pabian, F.V.: The New Geospatial Tools: Global Transparency Enhancing Safeguards Verification, Symposium on International Safeguards: Preparing for Future Verification Challenges. International Atomic Energy Agency, Vienna (2010)

Socio-cultural Modeling and Dynamic Decision Making

Orélie Desfriches Doria, Peter Svenmarck and Christophe Fagot

Abstract In the field of Cross-Cultural Decision Making, our work is oriented towards supporting decision making process by providing a dynamic and interactive visualisation interface. The idea with this approach is that users should be able to explore a problem space using visualisation that facilitate quick perceptions and understanding. Vesta-Cosy application, developed in the frame of a research project, therefore combines paraconsistent reasoning and works on graphic semiology. We first present the currently used approaches of decision making in Vesta Cosy. Thereafter, the Vesta Cosy application is described in details, especially the different models and their connection with visual and algorithmic units. We also present the theoretical framework which we use for a specific use case scenario. This theoretical framework aims at enabling the consideration of socio-cultural factors in Decision Making.

Keywords Dynamic decision making · Socio-cultural factors · Visualisation interface · User interaction · Modeling · Simulation

1 Introduction

Inspired by the systemic approach [2, 8, 9, 18], by complex systems engineering [19] and expert system engineering [10, 11], the Vesta Cosy research project, funded by "Direction Générale de l'Armement" (DGA) (French Ministry of

O.D. Doria (✉)
Ensci-Les Ateliers, 48 Rue Saint Sabin, 75011 Paris, France
e-mail: orelie.desfriches-doria@ensci.com

P. Svenmarck
Swedish Defence Research Agency (FOI), 164 90 Stockholm, Sweden
e-mail: peter.svenmarck@foi.se

C. Fagot
Intactile Design, 20 Rue du Carré du Roi, 34000 Montpellier, France
e-mail: fagot.c@intactile.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_24

Defense procurement agency), is developed in collaboration by Intactile Design and some researchers in Information and Communication Sciences.

One of the aims of the application developed in this project is to support decision-making process, by means of visualizing some aspects of the decision, and by providing the ability to describe some properties about these graphic elements. This work is based on the graphic semiology from Bertin [1], and combined with paraconsistent reasoning approaches [5, 6, 7, 21].

We first present the existing and currently used approaches of decision making in Vesta Cosy. Thereafter, the Vesta Cosy application is described in details, especially the different models and their connection with visual and algorithmic units. We also present the theoretical framework inspired by Schwartz [20], which we use in the frame of a use case scenario SOCUMOD 2, for a specific research project coordinated by the European Defense Agency (EDA).

2 Decision Making

About systems which aim at supporting the decision making process, we suggest, that these systems are designed according to the vision of the decision making process itself [17]. These authors demonstrate that the definition of uncertainty is subjective. Moreover it depends on the decision model adopted.

Among different types of uncertainty definition, in decision-making process, we focus on Weick's approaches [24]. He mentions that a situation can give rise to a diversity of interpretations. Another interesting assumption about uncertainty is provided by the mental models theory, based on the following hypothesis first formulated in 1943 by the Scottish psychologist Kenneth Craik: "for understanding and predict reality, human mind needs to mentally shape models for simulation of its functioning" [23]. Moreover, when a multiple option situation is implied, the decision making process is more difficult [15]. Vesta Cosy's main claim is to give users the ability to explore a problem space using visualization to facilitate quick perceptions and understanding. This semiotic medium is designed to support the mental modeling, to handle multiple options situations easier, and to support the decision process to make the best decisions possible.

Naturalistic Decision Making approaches describe the operational decision making situation in context, including pressure, emergency, ill-defined goals, changing goals, information inconsistency, and pre-existing expert knowledge [12, 13]. For these authors, "Naturalistic Decision Making (NDM) is the study of how people make decisions in the workplace and in their personal lives. Researchers in NDM try to understand and describe the strategies people use in diagnosing a situation or in choosing a course of action. The focus of interest is on actual situations, because decisions are made in context, and the features of context need to be taken into account to understand the decisions" [12]. However, we do not completely agree with their ideas, especially because of their assumption that people usually choose the first correct solution, which we consider to be not

necessarily the best one. Indeed, according to Le Moigne [16], this sort of choice as opposed to the optimum choice, are only assessed by means of efficiency criteria. The decision making process, as we define it, is a complex cognitive process. It can imply several satisfactory solutions assessed with effectiveness criteria. Effectiveness is here understood as the multidimensional relation between the system performance and its purposes [16]. The system in this context is considered as the socio-technical complex that includes human factors, interfaces, and their interactions.

3 Vesta-Cosy

Although there are many successful applications based on paraconsistent logic, any system that is intended for end users most also have a good user interface. Intactile Design's Vesta-Cosy therefore combines paraconsistent reasoning with ideas from Jacques Bertin's work on graphic semiology in his search for meaning [1]. The idea with this approach is that users should be able to explore a problem space using visualization that facilitates quick perceptions and understanding. The interface therefore uses expressive polymorphic objects that can change size, transparency, blur, fine drawing, and mutate to evoke a change in substance or nature. Further, the network of relationships between objects is also considered a polymorphic element that may express membership, contact, influence, a statement, or uncertainty.

The Vesta-Cosy interface provides:

- Quick perception of problem in information dimensions, such as consistency, volume, quality, certainty, completeness, dangerousness, etc.
- Information unification into a single workspace of information that is used during the analysis, as well as of information from external sources, such as Internet, databases, decision support systems, or other users.
- Production and exploration of arguments relating to the purpose of a particular data or knowledge.
- Collaboration with domain experts by visualization of joint work in progress, presentation of other users' perceptions, and information sharing.

In practice, users set some entities from knowledge models on the symbolic map, and also define their relations, to get a schematic and symbolic representation of the situations.

The complex network that characterizes the situation can exist implicitly. The user may interact with it by setting entities from the knowledge model on the map, but also by visualizing different viewpoints. Even if all existing connections between entities are represented, the visualization may not present all of them at the same time, in order to enable reading and reasoning only with the necessary materials. Thus, some information and interconnected actors can be represented on the symbolic map as a graph, but some additional information may be presented as

tag clouds, or as notifications. The knowledge model also supports the investigation because it includes tradecraft rules.

The situation represented on the Vesta Cosy's maps can be analyzed dynamically and in reference to key aspects of the problem at hand. In this perspective, data visualization is considered as a process of decreasing the complexity of components for a logical qualitative and dynamic graph.

The next section is dedicated to a detailed description of the software components and their interactions.

4 Description and Interactions of Vesta Cosy's Components

Vesta Cosy includes an interface for modeling complex situations, through a graphic and symbolic schematization. This visualization is based on domain models including relevant components, for assessing, simulating, or predicting the evolution of the handled situation, and achieving the best possible decision, in context. It also implies the ability to simulate the way of thinking of influential actors in the handled situation. These aspects will be presented further in this paper. For these purposes, several models are available and are interacting in Vesta Cosy. The following exposition is designed in reference to a theoretical framework for characterizing influential actors [20]. We go into this theory presentation in depth further in this paper.

4.1 Domain Knowledge Models

This type of model is providing all types of entities related to the domain or to the tradecraft, and their properties, but also some specified objects for a contextual situation. Moreover, it contains the known connections between entities.

The domain or tradecraft is here understood as, for example, the medicine domain for doctors, or the environmental domain about a subject focused on a controversy on shale gas. Thus, a model about food additives could be used for analyzing questions about aspartame or glutamate.

The role of this kind of knowledge model consists in providing to the user a semantic support to add entities on the symbolic map, by using existing entities in the model, or by creating new entities based on the existing types in the model. It also includes a set of inference rules connected to the domain, which enable to produce on the fly:

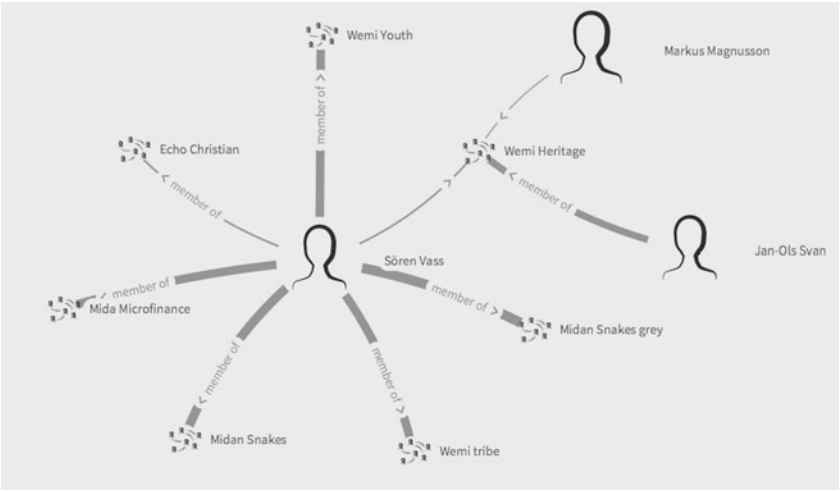


Fig. 1 Symbolic map about an actor

- new information or connections between entities in a contextual situation
- warn the user that some required data are missing, according to the professional standard
- detect some potential disagreements between the user and the inferred information, or inconsistencies between different tradecraft rules (the rules engine is presented below).

In the Fig. 1, a person is specified by the user with the label "Sören Vass", and other entities are also available on the symbolic map. The latter are social groups, as "Wemi Tribe" or "Echo Christian". This entity "Sören Vass" has already been specified in the model, so the user just has to drag and drop it on the symbolic map, from the model. Each entity is also connected to others entities with a labeled link which allows to represent the nature of their relation. In reference to this map above, we can also notice that some of the actors' icons are bigger or smaller, depending of their influence in context. Some of the links connecting different entities are stronger or weaker as well, depending of the frequency of these actors' interactions, for example.

In reference to the Actor Network Theory [3, 14], all the entities included in the maps are considered as agents, human or non-human ones. Thus an entity labeled as "Risk" could be mentioned on the map and considered as an agent. All agent types in Vesta Cosy are described with a set of properties, that the user can fill in or modify if necessary. All the entities' properties are presented in information forms. The Fig. 2, shows the information form about the actor previously mentioned "Sören Vass" on the right part of the figure. All the information forms are fully semantically described, in RDF format, and can contain numerical or textual data, GPS data, or personal information, such as a phone number.

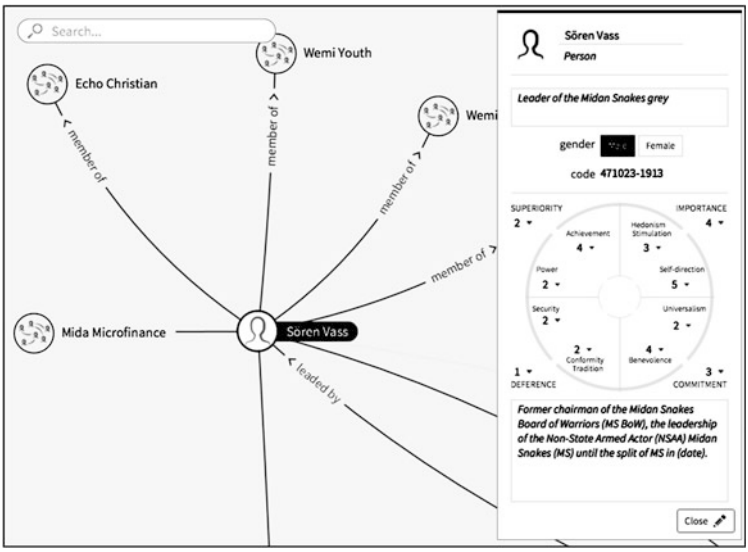


Fig. 2 Information forms about a human agent

4.2 Arguments Model

This model uses the same principles to represent arguments. It could be considered as a domain model about reasoning and debate skills. It aims at generally implementing the functioning of arguments by representing them on symbolic maps to enable their understanding and connections in the contextual situation.

This argument model is at the same level as knowledge models, but can be commonly used in all user cases. As the domain knowledge models, it includes reasoning rules and general decision rules. Actually, we consider arguments as agents too. In this respect, arguments can be described with some properties, and included in information forms as well. Some research work is still in process about these assumptions, especially concerning the graphic representation for the properties of entities in this model. This research is inspired by some sociologists' propositions [4, 22].

Indeed, when using Vesta Cosy for analyzing a situation in a decision making process, these two models are applied, the domain knowledge model and the argument model. The lining up of the two models is supported by the rules engine, which is presented below.

4.3 The Symbolic Maps

The symbolic maps that can be created in Vesta Cosy have previously been introduced in Figs. 1 and 2. This application supports the specification of the knowledge model in reference to a contextual situation that needs to be represented for analysis or simulation. The role of the symbolic map interface consists in providing a graphic representation of viewpoints, which composition arises from the knowledge model, and/or from added entities based on professional standards. The map is used to stage a situation in a contextual way. It is composed by entities coming from the knowledge model and by viewpoints, defined as a consistent set of arguments about which some actors would agree. The map can thus be considered as a pivot between the specified entities coming directly from the knowledge model in reference to a domain, and required missing entities in this knowledge model, according to the decision context.

4.4 Semantic Zoom Feature

Available on the maps, the semantic zoom feature allows the user to interact with symbolic representations. It aims is to provide the ability to switch gradually from a detailed view of the specified model for analyzing a contextual situation to a more general view. Granularity levels management is thus embedded in this feature: according to the selected zoom level, the information and detail levels displayed varies, including some descriptive elements about entities (tag cloud, notifications, annotations, etc.). Thus, when a user is zooming in, by means of mouse scroll wheel or stretching out on a touch screen, he not only wants the enlargement of entities on the symbolic map, but he wants more detailed information about these entities next to the selected area too.

By contrast, zooming out by the same means, doesn't only indicate a request to reduce the size of graphic elements, it also indicate that the user wants less information details displayed about the present entities, and that he wants to switch to a global visualization by mass entities layout. The different detail and information levels depend on their semantic description and on zoom level. It constitutes on of the major contribution of designers in this project. The Fig. 3 shows varying visual degrees of the semantic zoom for an entity labeled "Paris".

4.5 Web Crawling Component

The web crawling component is an optional functionality the user can choose to activate or not. It consists of automatic information or data retrieval about one or more entities specified in a symbolic map. These queries can be connected to the

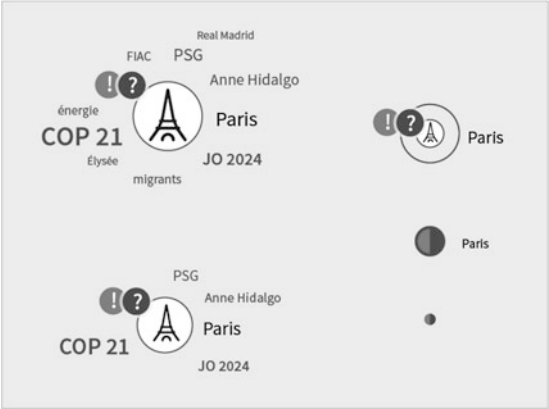


Fig. 3 Presentation of visual variations for an entity with the semantic zoom

Web environment and eventually to specialized databases. This component has the ability to weight information in accordance to its significance regarding the concerned entity and other retrieved information.

This information feedback can be displayed on the symbolic map in various ways, for example, a tag cloud can reveal the actual trend on the Web about an actor specified on the map. Consistently with the semantic zoom, this tag cloud layout depends on the detail degree selected in the semantic zoom. Thus, the user can switch from a very detailed zoom where visualization of information feedback is displayed as a tag cloud, toward a less detailed zoom, where a circle around the edge of the entity icon is coloured.

4.6 Paraconsistent Component

This paraconsistent component's role is focused on steering the graphic interface and produce inferences based on both the knowledge model and the argument model. It's carrying out the interactions between these two models and guarantees the reactivity of the knowledge model by bringing out information related to this model's rules on the symbolic map.

The Fig. 4 is showing a knowledge model's rule in relation with a specific domain for simulating social roles and identification modes to social groups. The specificities of Schwartz's theory [20] for these purposes are presented in Sect. 5 of this paper, and also in [21].

This rule engine is almost similar to existing ones (Apache, Jena, Pellet, Drools). Nevertheless, it presents two distinctive features. First, it is used in real time, so it produces updated inferences in the course of the addition or of the remove of information by the user about entities specified on the symbolic map (either in the entities' information forms or in the connections between entities). Secondly, it can handle inconsistencies, which implies a few consequences:

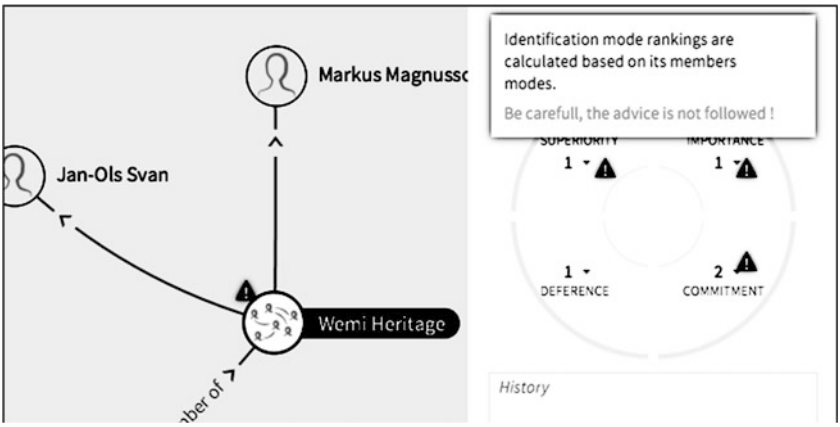


Fig. 4 Rules engine running for social domain simulation

- Tradecraft rules used by this rule engine can infer opposed information (this sort of case happens in real situations where experts cooperate, in domain such as defense, medicine, legal issues, taxation);
- The user can disagree with the information produced by the rules run in the rules engine and when this case happens, the user's opinion takes priority over the inferences that could result from the process (and not information derived from professional standards)

5 Vesta Cosy for Simulation About Actor's Social Roles and Memberships

As mentioned above, the Actor Network Theory [3, 14], allows us to consider all entities included in the maps as agents, human or non-human ones. In this perspective, human agents, as well as non-human entities, can be described with a set of properties. There is a variety of ways to describe human agents, from physical properties to more holistic approaches that would consider the following properties consistent for describing a human agent: social position, general and personal goals related to a situation, social memberships, definition of individual and general interests, type of concern or motivation related to a situation, epistemic culture, emotional register, etc.

In this perspective, the theoretical framework of Schwartz [20] approaches is especially interesting to qualify and understand actors' motives and intentions, the distribution of political and economic power among actors, relationships among the actors, socio-cultural dynamics, as well as effects of own and others' actions.

5.1 Schwartz's Theory About Social Values

Within the area that commanders are responsible for, there may be many types of actors or groups that each have their own motives and intentions for supporting and protecting the development of their own group.

In Schwartz's theory [20], an actor or group is considered as a number of individuals with shared preferences, routines, and convictions that have an essentially cooperative attitude towards each other. Influential actors, in turn, are representatives for a group that have larger impact on or control over their group compared to other individuals within the group.

The assumption is that the large variety of actors and influential actors can be explained with a few basic mechanisms of human behavior. Such mechanisms include how values form from social and contextual factors, personality traits, and evolutionary processes that govern group values and identification with groups [21]. Schwartz [20] describe how values support social functioning by guiding what is considered appropriate behavior, as well as actions that may benefit the group or the individual.

5.2 Value Structures from Contextual and Social Factors

The values that groups cultivate are an important driver for how people identify with the group. Values often develop adaptively as a response to conditions and factors that are important for the group. How groups prioritize values therefore gives some insight into the groups' context. According to Schwartz theory, values support social functioning by:

- Promoting cooperative relations among group members;
- Providing motivation for work;
- Legitimizing gratification for individual needs and desires that does not conflict with group goals.

Schwartz describes how ten values suffice to define most groups where [20]:

1. Benevolence refers to the importance of preserving and enhancing the welfare of other members in groups that the person identifies with. Benevolence is necessary for groups to function and also support the need for belonging to a group. Benevolence is therefore an internal motivation that promotes cooperative and supportive social relations.
2. Universalism refers to the importance of preserving and enhancing the welfare of the society and world. Universalism is necessary to avoid destruction of scarce resources and dangerous conflicts.
3. Self-direction refers to the importance of having control and independence. Self-direction therefore promotes freedom and own choice of goals.

4. Security is essential both for individuals and groups. Without an appropriate level of security, a group may not be sustainable over time.
5. Conformism refers to the importance of avoiding actions that may disrupt group interaction and functioning. Conformism therefore regulate the amount of conflicts and what type of conflicts that are acceptable within groups.
6. Hedonism refers to the importance of experiencing pleasure and gratification for oneself. Some form of personal satisfaction is important for many types of social contributions.
7. Achievement refers to the importance of personal gain from adhering to social standards. For example, the gain may be access to additional resources and/or social recognition.
8. Tradition refers to the importance of a group's practices, symbols and ideas. Tradition symbolizes solidarity and contributes to the group's survival.
9. Stimulation refers to the importance of variety for a positive activation rather than a negative activation from threats. Stimulation creates excitement from facing new challenges.
10. Power refers to the importance of social status and control over people and resources. The differentiation of social status is necessary for the functioning of social institutions, whereas dominance and control may be important for survival.

5.3 Applying Schwartz's Theory in Vesta Cosy

In the frame of a specific use case scenario evoked in Sect. 1, the theoretical framework previously mentioned is applied in Vesta Cosy content, understood as information forms about entities in the maps. That allows us to apply a generic framework for describing and analyzing actor's sources of motivation and intentions, in simulation about strategic decisions in operational contexts for defense or economics. Indeed decision making in security context, as in economic environment, is a complex cognitive process, which involves facts, beliefs, rules, situational elements, strategies, uncertainty, inconsistency, risks, and availability of information.

According to Schwartz [20], values are beliefs, connected to emotions. They constitute sources of motivation, and transcend specific actions or situations. They compose an upper common frame of reference for actors in decision making and can be applied in various contexts. They compose a common culture that may represent a standard and affect the way people evaluate, justify, or blame some behaviors. As they are ordered by importance for each individual, the implementation of this system of values in Vesta Cosy, allows the user to order in various ways different values concerning one or more human entities in the maps. By doing this, it becomes possible to simulate some varying scenarios, in the context of the

Fig. 5 Schwartz's theory application in Vesta Cosy's Information forms

handled decision situation, depending on the value priorities attributed to some actors or social groups.

The Fig. 5 is an extract of an Information form about an actor from a map. This information form thus includes a model for representing the value system containing the ten previously presented values, from Schwartz's theory. For each value, the user can change the dedicated weight. On each corner of this figure, and still according to Schwartz's theory, the following elements "importance", "commitment", "superiority", "deference" show modes of identification to a group.

These are weighted up as function of the weight of the ten values. Based on theses two types of description, the user can analyze first, the match between a person and a group, secondly, the balance or instability of an actor according to the configuration of his individual values.

6 Conclusion and Perspectives

As we include the model of prioritization human actors' values, arising from Schwartz's theory, we define a general frame for describing human actor properties, in the context of a specific project. By changing weight on the values' priorities, and by means of the rules engine running in Vesta Cosy, one can obtain dynamic simulation scenarios. This case represent one of the ways this application can support dynamic decision making, by providing on the fly informed rules and information. It is also because Vesta Cosy allows the user to interact with all entities in the model and the maps, and also with the information form formats, it can be considered as a solution to support dynamic decision making and dynamic

simulation process. To conclude, applying the abilities and flexibility of Vesta Cosy to embed a sociologic theory in the interface and the model, is one way to define how interdisciplinarity can be achieved for decision making support.

References

1. Bertin, J.: Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes, Les Ré-impressions des Editions de l'EHESS (1999), Paris (1973)
2. Bertin, E., Gandrillon, O., Beslon, G., Grauwin, S., Jensen, P., Schabanel, N.: Les complexités: point de vue d'un institut des systèmes complexes. Hermès, La Revue **60**, 145–150 (2011)
3. Callon, M.: Actor-network theory—the market test. Sociol. Rev. **47**, 181–195 (1999)
4. Chateauraynaud, F.: La contrainte argumentative. Les formes de l'argumentation entre cadres délibératifs et puissances d'expression politiques, Revue européenne des sciences sociales XLV-136 (2007)
5. da Costa, N.C.A., Krause, D., Bueno, O.: Paraconsistent logics and paraconsistency. In: Jacquette, D. (eds.) Philosophy of Logic (Handbook of the Philosophy of Science), pp. 791–912. North-Holland (2007)
6. Da Silva Filho, J.I.: Treatment of uncertainties with algorithms of the paraconsistent annotated logic. J. Intell. Learn. Syst. Appl. **4**(02), 144–153 (2012)
7. Da Silva Filho, J.I., Lambert-Torres, G., Ferrara, L.F.P., Mário, M.C., dos Santos, M.R., Onuki, A.S., Rocco, A.: Paraconsistent algorithm extractor of contradiction effects-paraextrctrctr. J. Softw. Eng. Appl. **4**(10), 579–584 (2011)
8. Donnadieu, G.: Systémique et science des systèmes: quelques repères historiques. Association Française de Systémique (AFSCET). Quelques repères historiques - Afscet (2004)
9. Donnadieu, G., Durand, D., Neel, D., Nunez, E., Saint-Paul, L.: L'Approche systémique: de quoi s'agit-il? L'Approche systémique : de quoi s'agit-il - Afscet (2003)
10. Feigenbaum, E.A., Buchanan, B.G., Lederberg, J.: On generality and problem solving: a case study using the DENDRAL program. Mach. Intell. **6**, 165–190. Edinburgh University Press, Edinburgh (1971)
11. Hoffman, R.R., Lintern, G.: Eliciting and Representing the Knowledge of Experts. Cambridge Handbook of Expertise and Expert Performance, pp. 203–222. Cambridge University Press, New York (2006)
12. Klein, G.: Naturalistic decision making: implications for design. Klein Associates Inc Fairborn OH (1993)
13. Klein, G., Klinger, D.: Naturalistic decision making. Human Systems IAC GATEWAY, vol. XI, 3, (1991)
14. Latour, B.: On actor-network theory: a few clarifications. Soziale Welt. **47**(4), 369–381 (1996)
15. Legrenzi, P., Girotto, V.: Mental models in reasoning and decision making processes. Mental models in cognitive science, pp. 95–118. Taylor and Francis, Psychology Press, Hove (1996)
16. Le Moigne, J.L.: La modélisation des systèmes complexes. Bordas, Dunot, Paris (1990)
17. Lipshitz, R., Strauss, O.: Coping with uncertainty: a naturalistic decision-making analysis. Organ. Behav. Hum. Decis. Process **69**(2), 149–163 (1997)
18. Morin, E.: Introduction à la pensée complexe. Seuil (2005)
19. Rechtin, E., Maier, M.W.: The Art of Systems Architecting. CRC Press (2010)
20. Schwartz, S.H.: An overview of the Schwartz theory of basic values. Online Readings in Psychology and Culture, vol. 2, issue 1, (2012)
21. Svenmarck, P., Sjöberg, E., Fagot, C.: A planning tool for interaction with influential actors based on paraconsistent logic. In: AHFE 2016, 27–31 July, 2016, Orlando, Florida (2016)

22. Thévenot, L.: La connaissance dans l'action, L'action au pluriel, Paris, La Découverte, TAP/Politique et société (2006)
23. Thevenot, C., Perret, P.: Le développement du raisonnement dans la résolution de problèmes: l'apport de la théorie des modèles mentaux. Développements **2**(2), 49–56 (2009)
24. Weick, K.E.: The Social Psychology of Organizing. Adisson Wesley, Reading (1979)

Examining the, Ideological, Sociopolitical, and Contextual Factors Underlying the Appeal of Extremism

Grace-Rose Williams, Michael L. Bernard and Robert F. Jeffers

Abstract This paper discusses and seeks to synthesize theories regarding the role of ideology and psychosocial contextual factors in shaping motivations and behaviors of individuals within violent extremist movements. To better understand how these factors give birth to and nurture extremist social movements, theory from a multitude of disciplines was incorporated into a conceptual model of the drivers associated with terrorist behaviors. This model draws upon empirically supported theoretical notions, such as the violation of socioeconomic and geopolitical expectations, the concept of perceived threat, one's mental construction of the world and group polarization. It also draws upon the importance of one's social identity, sense of belonging, and the perceived "glamour" associated with extremist group behaviors.

Keywords Extremism · Religious ideology · Terrorist behavior · Violent extremist organizations

1 Introduction

The 'War on Terror,' 9/11, and more recently the rise of Islamic State (IS), has generated a vast amount of discussion pertaining to the motivations of terrorists' actions and the allure of violent extremist organizations. However, much of this has focused on explanations of behavior that emphasize religious ideology. The intent of this paper is to challenge this assumption as the sole or even the root cause as to

G.-R. Williams (✉)
Defence Science Technology Laboratory, Porton Down, Wiltshire, UK
e-mail: grwilliams1@dstl.gov.uk

M.L. Bernard (✉) · R.F. Jeffers (✉)
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: mlberna@sandia.gov

R.F. Jeffers
e-mail: rfjeffe@sandia.gov

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_25

why Muslims become terrorists. It is argued here that ideology is often used to explain actions, particularly in extremist organizations, that are in fact influenced by broader socio-cultural-economic factors. That is, ideology may not be best understood as a fixed feature, but rather a fluid set of beliefs that interact with factors such as economic stress, political dysfunction, and inequality. With this in mind, this paper will discuss how socio-political and economic factors interact with beliefs and attitudes and other psychological and sociological factors to influence individuals' movement towards violent extremism. This work is based on an effort being undertaken by the US Department of Energy's Sandia National Laboratories (SNL) in collaboration with the UK Ministry of Defence's Defence Science and Technology Laboratory (Dstl).

Considering the factors mentioned above, a multi-disciplinary approach is taken to explore common factors associated with the rise of violent extremist organizations (VEO) and how they can transform and re-align as a result of internal and external pressures. In doing so, we have drawn upon and sought to integrate theory from anthropology, behavioral economics, political science, psychology, and sociology to better understand how these factors give birth to and nurture extremist social movements. This discussion draws upon empirically supported theoretical notions, such as the violation of socioeconomic and geopolitical expectations, the concept of perceived threat, one's mental construction of the world and group polarization. It also draws upon the importance of one's social identity, sense of belonging, and the perceived "glamour" associated with VEOs. These notions are influenced by broader external factors that can affect the support, and thus the strength and influence of extremist organizations. Ultimately, this paper seeks to provide a greater theoretical understanding of the terrorist phenomenon by integrating the concepts discussed above into a more unified conceptual model.

2 The Role of Ideology in Extremist Behaviors

As stated previously, the influence of certain ideologies, specifically certain religious ideologies, has been attributed to the rise in militant extremism of specific individuals. According to the Oxford dictionary, an ideology is defined as "a system of ideas and ideals, especially one that forms the basis of economic or political theory and policy" where context is defined as "the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood and assessed" [1]. An ideology can also be thought of as a specific way of conceptualizing beliefs and attitudes, which can be shared within societies. Thus, in a societal sense, ideology can underlie a common belief about the world and how individuals should behave in it. Ideologies have also been associated with the underlying attitudes of individuals towards other, non-conforming ideologies and those who hold it [2].

Within societies, a dominant ideology can serve as a common point of reference where other ideologies are compared, often negatively. Consequently, a dominant

ideology can have the effect of marginalizing other, less dominant ideologies. This is especially true in societies that are less accustom to differing views. Although competing ideologies is often cited as a key factor in the rise of extremism [3], it can also simplify a complex debate and ignore the broader socio-cultural factors affecting individual involvement in VEOs. That is, although ideology can inform our perception of the world, it is also shaped and influenced by external dynamics which can alter our beliefs, attitudes and view of others. One example of this is how countries use ideology and norms to forge cultural and political boundaries between followers of different groups [2]. The identification of groups whose beliefs and attitudes may differ from the dominant ideology enables majoritarian discourse to stigmatize minority groups as problem communities. The 'otherisation' of minority populations is not only a divisive mechanism which contributes to binary constructions of 'us' and 'them,' but can also marginalize these populations from wider society. This is a general notion that is applicable across many societies.

It is argued here that these types of external dynamics, such as institutionally divisive practices, more often than ideology on its own, influences individuals to join extremist organizations. For example, McDermott found in his study of the September 11 terrorists, that the ringleader, Atta, came from "an ambitious, not overtly religious middle-class household in Egypt." Hani Hanjour, the Saudi who piloted the American Airlines flight 77 into the Pentagon, "had lived in the United States off and on throughout the 1990 s, mostly in Arizona." He was considered "intelligent, friendly, and 'very courteous.'" Ziad Jarrah, the Palestinian who piloted the plane that crashed in Pennsylvania, was "the only son of an industrious, middle-class family in Beirut," a "secular Muslim" family that "was easygoing" [4]. These and other examples suggest that while religious ideology can play an important role in shaping certain beliefs and practices, other factors can also have a prominent role in influencing the actual commission of extremist behaviors [5].

3 The Confluence of Psychosocial and Structural Factors

3.1 Identity and Belonging

Within a global context of heightened concern about Islamic extremism and terrorism more broadly, discursive constructions of identity and belonging, particularly discourses of race, religion, and nation permeate societal institutions, beliefs, and behavior. This is problematic as it brings citizenship, identity, and belonging to the forefront of the debate which in turn can push individuals towards VEOs due to feelings of exclusion. Citizenship expresses universal human rights and duties, while identity implies particularism and group membership. Although it has been argued that citizenship and identity are exclusionary categories, this paper emphasizes the relationship between them as well as how they both can generate a sense of belonging. Brubaker has asserted that citizenship is about inclusion and

exclusion, for individuals who are citizens, their status is ascribed, they are 'insiders,' but for many ethnic minorities, despite having an official citizenship status they are considered 'outsiders' [6]. Citizenship is considered central to an individual's self-understanding and assertions of who they are [7]. This sense of difference as an outsider is enforced by racism and discrimination, which in turn can have an important influence on a young person's identity construction, self-identification and sense of belonging.

The significant role of ideology and religion in the construction of one's identity and sense of belonging brings into question how individuals within minority populations negotiate and respond to being seen as fundamentally 'different' to the majority. For example, the ascription of non-western ways of life to minorities, particularly Muslims, as something inherently negative, defines what is considered an acceptable difference (disability, sexuality) and an unacceptable difference (Islam) [8]. The focus on difference has turned religious experience into political categories, 'good Muslim;' moderate Islam, the genuine Islam and 'bad Muslims;' political 'extremists' [9]. As a result, religious belonging has become a symbol of racial difference. The focus on 'manufacturing homogeneity' and 'managing difference' in order to regulate the threat of heterogeneity is central to the idea of maintaining a dominant ideology [10]. Differences in beliefs and attitudes within the context of religion and racial identity are often interpreted as a 'clash of identity' [11]. As a result normative Muslim thought, behavior, and culture are increasingly challenged and questioned as being opposed to Western lifestyle and values. This has led to Muslim identity being defined in terms of negativity, disadvantage, and alienation, serving only to stigmatize the Muslim population further [12]. An individual's interpretation of their identity, citizenship and sense of belonging within a social context that is perceived as unjust, unfair and/or marginalizing can influence their decision making to seek out groups (including VEOs) that can alleviate this.

The environment that an individual is situated in will also help frame the socio-cultural context of that individual and in turn their identity and sense of belonging. These context effects can be both highly salient and/or very subtle, and are highly subjective, given the history of the individual. That is, contexts are "mental constructs of participants; they are individually variable interpretations of the ongoing social situation. Thus, they may be biased, feature personal opinions, and for these reasons also embody the opinions of the participants as members of groups" [13, p. 7]. As a consequence, these effects can affect the individual's perceived quality of life and sense of fairness and trust. According to Johns, context can be perceived as "situational opportunities and constraints that affect the occurrence and meaning of organizational behavior as well as functional relationships between variables" [14, p. 386]. That is, the environment, in which an individual is situated will help frame the socio-cultural context of that individual. Examples of contextual effects could be the rise in social strains caused by changes in the status quo, the more apparent unequal distribution of wealth, and a lack of accountability by the government and marginalization of society, as well as many other factors.

3.2 The Sense of Loss and Grievance

The feeling of social dissatisfaction typically begins with a perceived sense of loss or some form of grievance. The sense of loss can include recent events or events that might have occurred generations ago. These events can be, and often are, distorted through time to favor some form of narrative. Examples include loss of territory, sovereignty (e.g., perceived Palestinian loss of territory), hegemony and lives (e.g., massacres against Jewish people). Grievances may stem from these losses and are considered to be a root cause for collective political action [15]. A grievance is "an individual's belief that he or she (or a group or organization) is entitled to a resource which someone else may grant or deny" [16, p. 52]. A dispute exists when the purported perpetrator of the grievance rejects this claim. Grievances are typically associated with perceptions of inequality, relative deprivation, injustice, or some form of moral indignation [17]. As the number of people who share the same grievance grows, a sense of social solidarity can grow as well, which not only multiplies the effect of the grievance, but also strengthens one's sense of identity and belonging to that leader/group [18]. Thus, the strength of the perceived grievance can be leveraged for a leader and/or group to cause blame and frame the situation to incite a social movement, then serves as a catalyst for a social uprising.

3.3 The Violation of Expectations

When one thinks of the causes of social uprisings, a common assertion is that the people associated with the uprising are those who have little to lose, those on the margins of society. However, in examining social movements, and leaders associated with them, one typically finds that the frustration that is felt begins with people who can address both their basic physiological needs and have time (perhaps unemployed) to generate the type of frustration that leads to actionable behaviors. This could also stem from such things as a perceived sense of obligation towards those who are less able to defend themselves. Thus, a person who is weakened by extreme poverty is less likely to think about higher ideological needs [19]. For example, studies have found that terrorists tend to be from higher educated and wealthier families than the average population [20]. Accordingly, frustration tends to come less from an absolute standard of deprivation than from the perception of deprivation in comparison to an ideal [21]. The frustration with one's situation can explain the multitude of individuals from different socio-economic and political backgrounds within VEOs. In these cases, a VEO can apparently provide the 'ideal' (which can vary from individual to individual) that they cannot achieve in their current situation. Interestingly, the discordance between a perceived ideal and reality can come at a time when, in comparison to the past, there is a general rise in the socio-economic and/or political condition of that group, but the expectation of the group rises faster than the perceived rise in change. The idea that collective

discontent can develop if there is a significant gap between expected and achieved welfare of the group is outlined in the theory of relative deprivation [22].

Relative deprivation theory refers to the idea that the perception of deprivation and discontent occurs as a group negatively compares their perceived situation to a desired point of reference, such as with other groups and, societies. That is, when a group believes its expectations are legitimate and are being blocked within their society, or by other societies, relative deprivation can occur. This is particularly true for discontent arising from the status of an entire group as compared to a similar, referent group. To achieve greater social satisfaction, members of that group will attempt to reduce this deprivation, often by using actions that highlight their deprivation and discontent. This type of behavior is generally considered to be a chief factor in explaining the desire for and the actions associate within social movements [22]. This deprivation also tends to strengthen a group's collective identity, making them more cohesive [23]. For example, relative deprivation can be perceived between those representing the populous and the government (e.g., Egyptian government dissenting groups with each other and with the various Egyptian government administrations), between religious sects (e.g., Shiite vs. Sunni), and between different socio-religious societies (e.g., Israelis vs. Palestinians) or more broadly, the Middle East vs. the West. In each case, one group compares its standing against the other. This is particularly true for groups that have long-standing conflicts with each other. As with all humans, negative comparisons are more psychologically salient than positive ones [24]. Thus, in comparison to another group, any deprivation that the group perceives will be more profound than any positive comparison. This is particularly true if the comparison group is perceived to be a threat. In the examples mentioned above, each group could consider the other group as a threat.

3.4 The Concept of Threat

The perception of threat by some external group can have a strong and lasting effect on both the attitudes and ultimately, behaviors of an internal group. According to social identity theory, group members are motivated to develop and maintain biased intragroup comparisons in order to promote a positive social identity [25]. This may be particularly true if there is a high degree of comparison between groups, which can spawn greater chauvinism between them. Moreover, studies have suggested that changes in the relationship between groups can strengthen this form of chauvinism.

Research suggests there are two major types of threats that can influence attitudes towards an external group. The first is the concept of realistic threat. Realistic threats refers to a perceived threat by an external group that has the potential to significantly affect one's own power, resources, and general welfare [26]. This can take the form of military, economic, and/or other physical or material threats to the group. For example, the rise in prosperity among some states in the Middle East,

such as Iran, will influence its relative power within the region, potentially being perceived as a greater realistic threat among states that consider it to be an adversary. This could be offset by an increase in military spending by an adversary state. Of course, this has the potential for a tit-for-tat response, thereby increasing the perceived threat by both states.

The second type of threat, called symbolic threat, concerns the threat to a group's honor, religion, values, belief system, ideology, philosophy or morality by another group. Here, out-groups that are perceived as having a different worldview can be seen as threatening the cultural identity of the in-group. This threat is particularly strong if the out-group is dominant, which can lead to a heightened fear that the out-group's culture will override the in-group's way of life. These can be perceived threats to a group's religion, belief system, honor, or worldview. The realization of this threat is the loss of the in-group's social identity and honor. Proponents of the concept of symbolic threat have suggested that prejudice is often a result of conflicting values and beliefs—even more so than from material threats [26]. For example, perceived threats to an in-group's values by foreigners were related to increases in negative attitudes toward immigrants [27]. Studies that have measured both realistic and symbolic threats have shown that both types of threats can account for different portions of the variance in attitudes toward out-groups [28]. Concerns around the realistic threat from terrorism has reinforced perceptions of a symbolic threat from Muslim communities, resulting in control practices which reproduce and maintain in-group hegemonic power and perpetuate negative stereotypes and prejudice. Moreover, Riek and colleagues found that in-group identification had a significant impact on realistic and symbolic threat but the impact was stronger for symbolic threat than realistic threat [29]. In turn, the stronger the identification with the in-group, the stronger the reactions to group esteem threats [30].

3.5 Attraction to the Perceived 'Glamour' of Violent Organizations

It is argued here that the (perceived) 'glamour' of VEOs can be a key driving force in the appeal of VEOs. The Oxford English Dictionary defines glamour as, "the attractive and exciting quality that makes a person, a job or a place seem special, often because of wealth or status" [1]. Thus, glamour can be thought of as not only something (a quality, commodity, lifestyle, etc.) which is desirable, but also something which can provide power, status, and respect. The 'glamour' of groups is a relatively undeveloped concept, having mainly been used to explore the appeal of sub-cultures, specifically street gangs [31–33]. Within this context, it is asserted that the perceived external image of a gang being powerful, rich, and glamorous can be an attractive lifestyle to those who have limited access to legitimate means of social and economic success [31].

The concept of 'glamour' with respect to VEOs, can be framed within Agnew's general strain theory [34]. Agnew's general strain theory argues that individuals engage in 'deviant' behavior (defined as actions or behavior which violates social norms [35]) when they experience goal blockage (access to socially approved goals) which results in strain. Specific strains include; failure to achieve positively valued goals (e.g., money or goods), the removal of positively valued stimuli (e.g., loss of valued possession) and the presentation of negatively valued stimuli (e.g., physical abuse). Agnew argued that conditions under which strain may lead to crime are when they are (1) seen as unjust, (2) high in magnitude, (3) associated with low social control and (4) create some incentive to engage in criminal coping [34]. Strain is particularly highlighted when individuals experience long-term unemployment, poverty, marginalization and a comparison to those in higher economic positions. It could be argued that for individuals who experience strain, VEOs may offer an alternative lifestyle that provides an escape from the frustration of their current position. For example, young Muslims in the West who are treated with mistrust, whose voices are rarely listened to, and are subject to racism and labels of 'terrorist.' Being part of a VEO allows young people to actively control and win space (in contrast to 'places' which are fixed and stable, 'spaces' are fluid, they are created by relationships and interactions with our environment [36]) for themselves, while gaining a position within that space [37]. The external image of a VEO as being 'glamorous,' rich, and fighting for a cause is also often tied in with notions power and freedom. The VEO becomes something to aspire to, an entity that provides a place to belong, an identity, thus giving individuals the social agency to obtain status, respect and a source of empowerment free from societal constraints and misrepresentations. However, by joining a VEO, it can also result in their ultimate marginalization through arrest, prison or deportation.

3.6 One's Mental Construction of the World

To make sense of one's world, individuals create mental models regarding such things as how societies should and do behave, how their world is ordered, the nature and role of justice, as well as the nature and role of men and women. These mental models, often called schemas, provide continuity and predictableness to the world [38]. A schema is a type of heuristic that helps to cognitively construct and organize one's perception of the world. A schema is developed over time and can be very resistant to change. That is, schema inconsistent information tends to be forgotten easily or simply ignored; whereas schema consistent information is typically remembered more easily and incorporated into the schema via assimilation. With regard to ideology, schemas play a large and important role. One aspect is the role of religion in one's schema. According to McIntosh, "religion is more than a cognitive organization of beliefs. Religion is broader in that it exists outside the person in the form of text, symbols, and traditions, and it is narrower in that it appears in the form of individuals' rites, habits, and other behaviors" [39, p. 1].

Regarding beliefs, one's schema about God might include the existence, purpose. and degree of guidance. Associated with these beliefs are the written texts that describe the directions established by God [39].

Moreover, schemas can play a powerful role in the attention and behavior of individuals toward male and female roles. That is, individuals with a high masculine gender schema tend to attend to more masculine behaviors that support their schema [40]. They also tend to react negatively to violations to their gender schema. In highly conservative societies, such as the Middle East, this violation might be the perceived blending of traditional male versus female roles and/or the wearing of western clothing (i.e., not wearing a hijab) for women (which could also violate one's religious schema). The degree to which one reacts negatively to a schema violation is typically a function of one's culture, which often affects how "schematic" (i.e., the degree to which an individual rigidity follows their schema) a person might be across a number of contexts. Thus, behaviors are affected by schemas, which are, in turn, affected by ones culture and context. For example, in the Middle East, certain ideologies have a very set view of the world.

3.7 The Polarization of Groups

In addition to the mental construction of one's social-religious attitudes toward oneself and others, developed via a schema, interactions with others and societal institutions can strengthen this. This concept is called group-induced attitude polarization [41]. Group polarization is said to occur when an initial tendency of group members' attitudes toward a given direction is enhanced following group interactions [42]. This can result in more extreme positions in the same attitude direction over time. For example, Myers and Bishop found that groups with prejudice-leaning individuals became more prejudiced, as a group, over time, while groups with less-prejudice leaning individuals became less prejudice over time [41].

A typically cited reason for this phenomenon involves the idea of information exchange and social comparison. Specifically, when individual group members exchange concurring information, the information can serve to both strengthen and add to each member's beliefs about a specific topic. Also, through dialogue with other members, each member can discern the general group orientation towards the topic and can support the group, and bolster one's position within the group, by taking on positions that further push the position of the group in the same direction [43]. In fact, Myers and Lamm found that the degree to which moderate fundamentalist ideas can morph to more extremist fundamentalist ideas is at least partly due to the group polarization effect [44]. Cultural difference, as discussed above, can also provide a basis for group polarization and can lead to the development and expansion of ethnic and religious boundaries between groups [45].

4 Developing an Integrated Conceptual Model

In attempting to bring the concepts discussed above into a more comprehensive understanding with regard to extremist behavior, a conceptual integrative model was developed (shown in Fig. 1). This conceptual model illustrates the role of societal and religious institutions influencing and being influenced by religious and other schemas (such as gender schemas). Societal institutions and schemas provide context for the construction of one's identity as well as informing ones expectations and the violations of those expectations. Aided by societal/religious institutions and one's schema, these factors can intensify the perception of threat from external groups. This can induce a drive towards some type of threat response.

The perceived 'glamour' associated with VEO can help facilitate this response. In fact, VEOs can then prey on these types of individuals to draw them into their organization or at least carry out behaviors that are aligned with the VEO's objectives. If individuals decide to join a VEO, the normative pressures will typically be to increase their level of violence. Alternatively, if individuals receives positive (non-violence) counseling and attention, the behaviors could potentially shift towards less violent actions. However, as discussed above, many factors have the potential to ultimately affect the behavior of these individuals.

The above model demonstrates not only the complexity of issues that need to be taken into consideration when analyzing the appeal of extremisms, but also the importance of a multi-disciplinary approach. Societal and religious institutions play an important role in ordering individuals' lives, internalizing norms and modes of behavior as well as developing ones identity. These institutions also provide socially accepted expectations and goals for citizens, for example, the American dream promises wealth under the premise of meritocracy. However, when

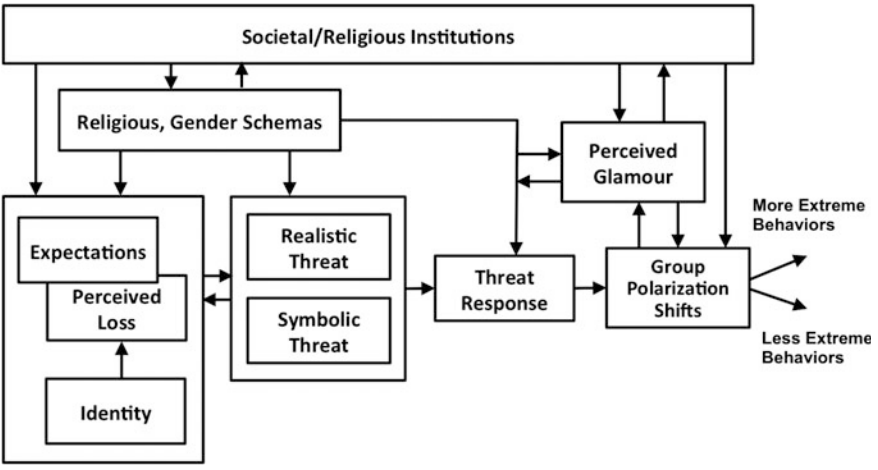


Fig. 1 The proposed integrated conceptual model of extremist behavior

individuals cannot achieve this wealth (often due to deep social inequalities), actionable behavior can result from the frustration of having a lack of opportunities. That is, when ones expectations are being blocked or denied, these can be interpreted as losses and/or grievances, particularly when in comparison to other groups standing. If a number of people share the same grievances, solidarity can grow, strengthening ones identity and sense of belonging to that group. As well as societal institutions, schemas (which are also affected by broader societal factors) also inform ones perception of the world, including how to behave, how society is ordered and ones relationship to other groups.

An individual's identity is created by group perceptions of who we are and how we define ourselves, they are socially bestowed, socially maintained and socially transformed [33]. Thus, once individuals join a VEO, the organization becomes central to their lives, which heavily influences their identity and in turn values, norms, and behaviors. Initial behaviors are internalized so they become part of an individual's self-perception and in turn transform their identity, whereby the group is an extension of the individual and the individual is an extension of the group [46]. This is also important within the context of both realistic and symbolic threat, whereby the threat to one's group not only reinforces ones identification with the group, but also informs their expectations and losses compared to another group.

The perceived 'glamour' of groups has both a push and pull effect on individuals' decision to join a VEO. The 'glamour' of a group can be perceived as reputation, status, respect and/or power, which can provide individuals with a sense of fulfillment and access to success (even if this is illegitimate). Societal representations of 'glamour' permeate individual's lives in different guises, for example, wealth, celebrity culture, commodities etc. For those who have been unable to achieve legitimate forms of success, a VEO can offer something to aspire to, either as an entity that can provide wealth, power and/or status or as an entity that allows them to reject the societal value of glamour. Particularly for those who are disenfranchised and/or have experienced perceived losses, the 'glamour' of a VEO can offer an alternative means of success to which mainstream society may have denied them. Yet, despite offering an alternative empowering lifestyle, joining a VEO can also lead to their ultimate marginalization.

The conceptual model aims to provide a deeper theoretical understanding of the movement of individuals towards VEO beyond explanations, which emphasize religious ideology. Taking into account the broader socio-political and structural factors, this paper used a multi-disciplinary approach to identify and interrogate how influencers such as perceived glamour, threat and ones expectations affect the appeal of VEOs and individual behavior. It aims to demonstrate how societal institutions and individual schemas dynamically influence and inform individual decision making and perception of the world. This model could be used as a foundation to build a more complete picture of the myriad of factors, which influence individuals' decision making to join a VEO.

Acknowledgments This research was possible, in part by, by funding from the UK's Ministry of Defence and the US's Department of Defense. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly-owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Approved for unlimited release: SAND2016-2131 C.

References

1. The Oxford Dictionaries. Oxford University Press. http://www.oxforddictionaries.com (2016)
2. Abercrombie, N., Turner, B.S.: The Dominant Ideology Thesis, pp. 396–414. Macmillan Education, UK (1982)
3. Huntington, S.P.: The Clash of Civilizations. Simon & Schuster, London (1997)
4. McDermott, T.: Perfect soldiers: The 9/11 hijackers: Who they were, why they did it. Politico's (2005)
5. Weinberg, L., Pedahzur, A: Suicide Terrorism. Religion Compass 4/4, pp. 234–244 (2010)
6. Brubaker, R.: Citizenship and Nationhood in France and Germany. Harvard University Press, Cambridge (1992)
7. Hussain, Y., Bagguley, P.: Citizenship, ethnicity and identity: British Pakistanis after the 2001 'riots'. Sociology **39**(3), 407–425 (2005)
8. Kalra, V., Kapoor, N.: Interrogating segregation, integration and the community cohesion agenda. J. Ethn. Migr. Stud. **35**(9), 1397–1415 (2009)
9. Mamdani, M.: Good muslim, bad muslim: a political perspective on culture and terrorism. Am. Anthropol. **104**(3), 766–775 (2002)
10. Goldberg, D.T.: The Racial State. Blackwell, Oxford (2002)
11. Kapoor, N.: On the North West 10(12): postcoloniality, the British Racial State and the War on Terror. Ident. Global Stud. Cult. Power. **20**(1), 61–76 (2013)
12. Alexander, C.: The Asian Gang. Berg, Oxford (2000)
13. Van Dijk, T.A.: Discourse, ideology and context. Folia Linguistica **35**(1–2), 11–40 (2001)
14. Johns, G.: The essential impact of context on organizational behavior. Acad. Manag. Rev. **31** (2), 386–408 (2006)
15. Klandermans, B.: The Demand and Supply of Participation: Social-Psychological Correlates of Participation in Social Movements. The Blackwell Companion to Social Movements, pp. 360–379. Blackwell, Oxford (2004)
16. Miller, R.E., Sarat, A.: Grievances, claims, and disputes: assessing the adversary culture. Law Soc. Rev. **15**, 525–566 (1980)
17. Klandermans, B.: The Social Psychology Of Protest. Blackwell, Oxford (1997)
18. Useem, B.: Solidarity model, breakdown model, and the Boston anti-busing movement. Am. Soc. Rev. **45**, 357–369 (1980)
19. Maslow, A.H., Lewis, K.J.: Maslow's hierarchy of needs. Salenger Incorporated (1987)
20. Krueger, A.B., Laitin, D.D., Kto K.: A cross-country study of the origins and targets of terrorism. Terrorism, economic development, and political openness, pp. 148–173 (2008)
21. Gurr, T.R.: Why Men Rebel. Princeton University Press, Princeton (1970)
22. Morrison, D.E.: Some notes toward theory on relative deprivation, social movements, and social change. Am. Behav. Sci. **14**(5), 675 (1971)

23. Singer, M.S.: The application of relative deprivation theory to justice perception of preferential selection. Curr. Psychiatry **11**(2), 128–144 (1992)
24. Kahneman, D., Tversky, A.; Prospect theory: an analysis of decision under risk. Econom. J. Econ. Soc. 263–291 (1979)
25. Tajfel, H., Turner, J.C.: The social identity theory of intergroup behavior. In: Worchel, S., Austin, W.G. (eds.) The Psychology of Intergroup Relations, pp. 7–24. Nelson-Hall, Chicago (1986)
26. Kinder, D.R., Sears, D.O.: Prejudice and politics: symbolic racism versus racial threats to the good life. J. Pers. Soc. Psychol. **40**(3), 414 (1981)
27. Esses, V.M., Hodson, G., Dovidio, J.F.: Public attitudes toward immigrants and immigration: determinants and policy implications. In: Beach, C.M., Green, A.G., Jeffrey, G.R. (eds.) Canadian Immigration Policy for the 21st Century, pp. 507–535. John Deutsch Institute, Queen's University, Kingston (2003)
28. McLaren, L.M.: Immigration and the new politics of inclusion and exclusion in the European Union: the effect of elites and the eu on individual-level opinions regarding European and non-European immigrants. Eur. J. Pol. Res. 39(1), 81–108 (2001)
29. Riek, B.M., Mania, E.W., Gaertner, S.L.: Intergroup threat and outgroup attitudes: a meta-analytic review. Pers. Soc. Psychol. Rev. 10(4), 336–353 (2006)
30. Branscombe, N.R., Ellemers, N., Spears, R., Doosje, B.: The Context and Content of Social Identity Threat (1999)
31. Sauma, J.: Street encounters: betrayal and belonging in youth gangs. Public Policy Res. **15**(1), 32–35 (2008)
32. Zdun, S.: Violence in street culture: cross cultural comparison of youth groups and criminal gangs. New Dir. Youth Dev. **119**, 39–54 (2008)
33. Stretesky, P.B., Pogrebin, M.R.: Gang-related gun violence socialization, identity, and self. J. Contemp. Ethnogr. **36**(1), 85–114 (2007)
34. Agnew, R.: Foundation for a general strain theory of crime and delinquency. Criminology **30**, 47–87 (1992)
35. Giddens, A.: Sociology. Polity, London (2012)
36. Bell, M.M.: The ghosts of place. Theor. Soc. **26**, 813–836 (1997)
37. Barker, C.: Cultural Studies. Theory and Practice. Sage, London (2008)
38. Fiske, S.T., Linville, P.W.: What does the schema concept buy us? Pers. Soc. Psychol. Bull. **6** (4), 543–557 (1980)
39. McIntosh, D.N.: Religion-as-schema, with implications for the relation between religion and coping. Int. J. Psychol. Relig. **5**(1), 1–16 (1995)
40. Markus, H., Smith, J., Moreland, R.L.: Role of the self-concept in the perception of others. J. Pers. Soc. Psychol. **49**(6), 1494 (1985)
41. Myers, D.G., Bishop, G.D.: Discussion effects on racial attitudes. Science **169**(3947), 778–779 (1970)
42. Isenberg, D.J.: Group polarization: a critical review and meta-analysis. J. Pers. Soc. Psychol. **50**(6), 1141 (1986)
43. Abrams, Dominic, et al.: Knowing what to think by knowing who you are: self categorization and the nature of norm formation, conformity and group polarization. Br. J. Soc. Psychol. **29** (2), 97–119 (1990)
44. Myers, D.G., Lamm, H.: The group polarization phenomenon. Psychol. Bull. **83**(4), 602 (1976)
45. Kunovich, R.M., Hodson, R.: Conflict, religious identity, and ethnic intolerance in croatia. Soc. Forces **78**(2), 643–668 (1999)
46. Kanter, R.: Commitment and Community. Harvard University Press, London (1972)

Using Computational Modeling to Examine Shifts Towards Extremist Behaviors in European Diaspora Communities

Asmeret Bier Naugle and Michael L. Bernard

Abstract We created a simulation model to investigate potential links between the actions of violent extremist organizations (VEOs), people in the VEO's home country, and diaspora communities from that country living in the West. We created this model using the DYMATICA framework, which uses a hybrid cognitive-system dynamics modeling strategy to simulate behaviors based on psycho-social theory. Initial results of the model are given, focusing on increases to VEO funding and recruiting resulting from an invasion of the VEO's home country. Western intervention, prejudice, and economic drivers are also considered.

Keywords Diaspora · European diaspora · Extremism · Violent extremist organizations

1 Introduction

Middle Eastern diaspora communities residing in the West have recently received a great deal of media attention regarding the potential for some individuals to become radicalized. In extreme cases, this can lead to a some of these individuals becoming terrorists within their own communities, as well as a possibility that radicalized individuals will join international violent extremist organizations (VEO) in conflict with Western nations. The road to radicalization is typically a complex one, consisting of variables such as local pressures to fit in, perceived social, economic, and institutional prejudice and harm, desire for affiliation, social isolation, and ideological justifications. This type of environment can help frame the socio-cultural context for an individual, which can be highly salient and highly subjective, dependent on the history of the individual. A question then arises: to what degree

A.B. Naugle (✉) · M.L. Bernard
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: abier@sandia.gov

M.L. Bernard
e-mail: mlberna@sandia.gov

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*,
Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_26

do these types of variables affect the individual's propensity to either affiliate within European society or to resist European societal norms? Even further, which conditions are likely to lead to an individual actively seeking to commit violent acts of terrorism? Accordingly, the intent of this work is to explore how various diaspora communities understand their realities and investigate why they choose to either support or resist their host societies over time.

Using engineering and social science techniques, this effort is quantifiably assessing current events and choice options ("what-if" queries) concerning geopolitical inter-group/regional dynamics within diaspora communities. More specifically, this effort is considering how events, perceptions, and attitudes affect radicalization likelihood for diaspora communities in Western countries, and how radicalization may induce behavioral shifts pushing individuals toward interaction with VEOs. The computational model assesses both moderate diaspora communities and more militant diaspora groups. It addresses the conditions that attract more moderate members towards militant groups within the host country, as well as the conditions that may attract more militant individuals toward international VEOs. Variables that are included within the model include perceived Western prejudice towards Islam, VEO recruitment activities, economic opportunities, social isolation, and military conflict within or near the home country.

The preliminary model has so far been used to look at the effects of military activity in the diaspora's home country on funding and recruitment to VEOs. Early results indicate that such military activity is likely to increase political alignment with the VEO, thus increasing recruitment. Remittances to families and organizations are likely to increase, as is funding to the VEO from diaspora communities. The VEO thus gains overall strength from the military activity, even while losing some funding and personnel due directly to the same military activity. The model is still in preliminary form, and results should be considered illustrative rather than predictive.

2 Diaspora Communities Within Europe, and Recent Challenges

Europe has a long history of accepting migrants into its borders. The flow of immigrants from other parts of the globe has recently accelerated, mostly due to migrants from the Middle East and Africa. For example, Europe (particularly Germany and France) has recently taken in large waves of immigrants from Syria, Afghanistan, Iraq, and Eritrea [1]. This is expected to put a strain not only on government services and the general society within these areas, but also on existing diaspora communities. More specifically, existing diaspora (and other) communities will have to share a limited pool of government services and resources with the incoming diaspora. New migrants may also bring with them the psychological

aftermath of violent conflict and poverty, potentially adding to the need for greater services and resources [2]. In addition, most of the recent immigrants have been young [3, 4] and male. While there is potential for this to benefit Europe by contributing to economic growth and slowing overall population aging, this demographic, along with changes to sex ratios caused by the relative homogeneity of the migrant group, tends to contribute disproportionately to instability, violence, and insurgence [5].

On top of these challenges, Western governments and societies must determine how to deal with the small but concerning portion of the diaspora population (both existing and new) that has been receptive to the lure of VEOs. For example, there are instances of men and women in their early 20s quietly leaving their families and friends to join VEOs, such as ISIS in Syria and Iraq, Boko Haram in Nigeria, or al-Shabaab in Somalia [6, 7]. Moreover, it is known that at least some of the funding for terrorist activities in these regions has been routed through remittances sent by diaspora communities in Western countries [8].

This leads to the question of what circumstances drive individuals, who migrate from the Middle East to Europe, to either assimilate within the European culture or reject it in favor of extremist behaviors. Better understanding these circumstances can enable Western governments to better prepare and equip themselves with policies and programs that reduce the likelihood of a shift towards extremist behaviors.

2.1 Questions of Interest

Our modeling and assessment seeks to address the following questions. The assessment described here is an initial version, and may not yet address all of these. However, the model structure is designed to allow policy evaluation and other assessment activities that will contribute to answering these questions.

(1) What factors are likely to contribute to determining VEO recruitment and funding from diaspora communities?
(2) How might Western actions influence perceived Western prejudice toward Islam, thus affecting VEO recruitment and funding from diaspora communities?
(3) How might Western intervention in the Middle East/North Africa affect VEO recruitment and funding from diaspora communities?
(4) How does the balance of remittances from diaspora communities to families and non-violent organizations within the original homeland versus VEOs affect VEO recruitment?

2.2 Initial Scenario

This assessment is based on a scenario intended to focus the work by providing a context and initiating event. The scenario involves occupation by foreign military troops in the area of interest (the home country of the associated diaspora in the model). A 10-year time horizon was selected in order to capture both short-term and longer-term interactions of VEOs with the diaspora, influenced by both exogenous and endogenous variables.

2.3 General Motivations of Diaspora Communities Integrated into European Society

We consider two general groups of diaspora communities: one that is relatively integrated in its host society, and one that remains relatively isolated. While these groups are not necessarily well defined, the concept is useful for identifying motivations and behavior patterns that tend to be associated with one of these mindsets. There are certainly cases where motivations attributed more strongly to one group are also of concern to the other, or where individuals could be considered as belonging to either group, but considering the groups separately allows for more useful discussion and simulation of the situation.

For many in the more-integrated diaspora community, providing financial support for themselves and their families is a strong behavioral motivator. For example, the Somali diaspora within Europe alone provide at least $1.6B in annual remittances to families within Somalia. This is in line with the traditional obligation to their extended family group. Doing this provides critical support to the Somali society [9]. For diaspora communities that are more integrated within the larger European society, another motivation can be to appropriately blend Western culture and Islamic religion. Often, these communities will attempt to embrace what is most valued in both worlds. This can lead to a sense of confusion regarding what culture and lifestyle to embrace, and to publically express, at any giving time [10]. It also can cause conflict between families and generations regarding what to embrace. A sense of honor and dignity is also typically very important within these communities. A perception of widespread prejudice and discrimination can thus have a pronounced affect on the behaviors of the community, particularly its youth [11].

2.4 General Motivations of Diaspora Communities Isolated from European Society

More isolated diaspora communities will typically follow the obligation to their extended family groups, and will uphold their sense of honor and resistance to

prejudice and discrimination. However, because of their isolation, their contact with Western society can be minimal. Instead, they tend to have a greater desire for isolation from and resistance to the West and its perceived anti-Islam, anti-Middle East foreign policies. This includes policies and secularist laws within Europe that are perceived as discriminatory and insulting toward Islam and Muslims. This can be particularly true for individuals who feel marginalized by society due to unemployment or underemployment. This perception can be particularly strong among Muslim youth, where there is a general belief that Islam can and should guide the important behaviors of society. Another common belief is that the function of government should be to promote socially appropriate behaviors. This notion is most prominent in the teachings of political Islam. The main belief underlying political Islam is that Muslims can only truly fulfill their religious obligations when public (Sharia) law sanctions and encourages pious behavior. What this actually means is often a function of the local culture. Thus, for those who follow a more strict interpretation of the Koran, the secular laws of the West can be thought of as promoting a corrupt and immoral society [12]. Consequently, those who strongly espouse traditional political Islam commonly see Western laws as a potential threat to Islamic society and their communities.

More isolated, conservative groups may also see other, less strict interpretations and practices of Islam as potential a threat by their perceived appeasement to Western values, and potentially as a greater threat than the West itself. That is, the perceived acceptance by Islamic communities to Western, secular values can cause a spiritual and moral decay that would be more detrimental to Islamic communities than any direct confrontation through the laws and general power of Western governments.

Another factor influencing the decision making, and ultimately the behaviors, of individuals are contextual situations that help frame one's environment. For example, in diaspora communities factors such as poverty and discrimination can inflame attitudes and encourage more militant behaviors. As stated by Hroub [13], "in understanding and explaining Islamist movements, are we better served by relying on an understanding of their context or an analysis of their ideology? The easy answer, of course is 'both,' because these two undertakings should not be mutually exclusive." Accordingly, the combination of societal distrust, perceived discrimination, unemployment or underemployment, and youths propensity to congregate can make for a potentially dangerous situation of militant extremism.

2.5 The DYMATICA Simulation Framework

To model and simulate this system, we used a cognition-oriented system dynamics approach called the Dynamic Multi-scale Assessment Tool for Integrated Cognitive-behavioral Actions (DYMATICA). The theoretical framework of DYMATICA is based on well-established psychological, social, and economic

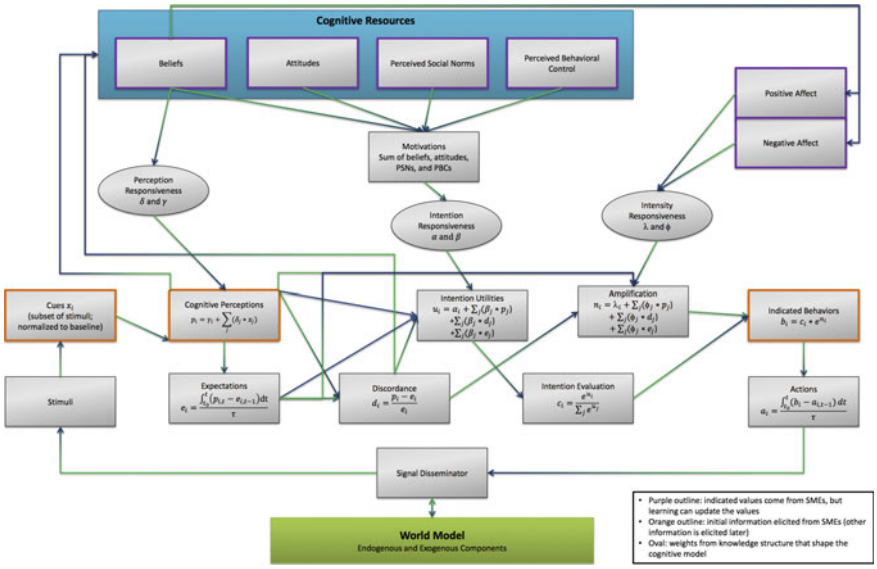


Fig. 1 Computational structure of DYMATICA

theories that have been incorporated into a single structure (Fig. 2) that is both self-consistent and dynamic. Cognitive models are implemented using system dynamics and embedded into an encompassing system dynamics model, which simulates interactions between people, groups, and physical, economic, or other system components (Fig. 1).

The cognitive portion of DYMATICA begins with individuals or groups being exposed to cues (stimuli relevant to the decision-maker). These cues are processed to create cognitive perceptions, the decision-maker's assessment of the world or situation. Over time, cognitive perceptions become expectations, which are compared to cognitive perceptions to determine discordance with the current situation. Discordance and cognitive perception affect beliefs, a category of cognitive processes that includes the components of the theory of planned behavior (attitudes, social norms, perceived behavioral control) [14] and affect. Intentions are calculated using utility functions. A multinomial logit function [15] compares intentions to determine realized behaviors, and over time those behaviors become physical realized actions. One of these cognitive models is populated for each individual or group being included in the system. These cognitive models are connected to each other and to a world model sector using system dynamics. The world model sector includes all of the non-cognitive components of the system of interest, including physical systems, economics, etc. Outputs from the world model and the cognitive models act as inputs, or stimuli, for the cognitive model in subsequent time steps.

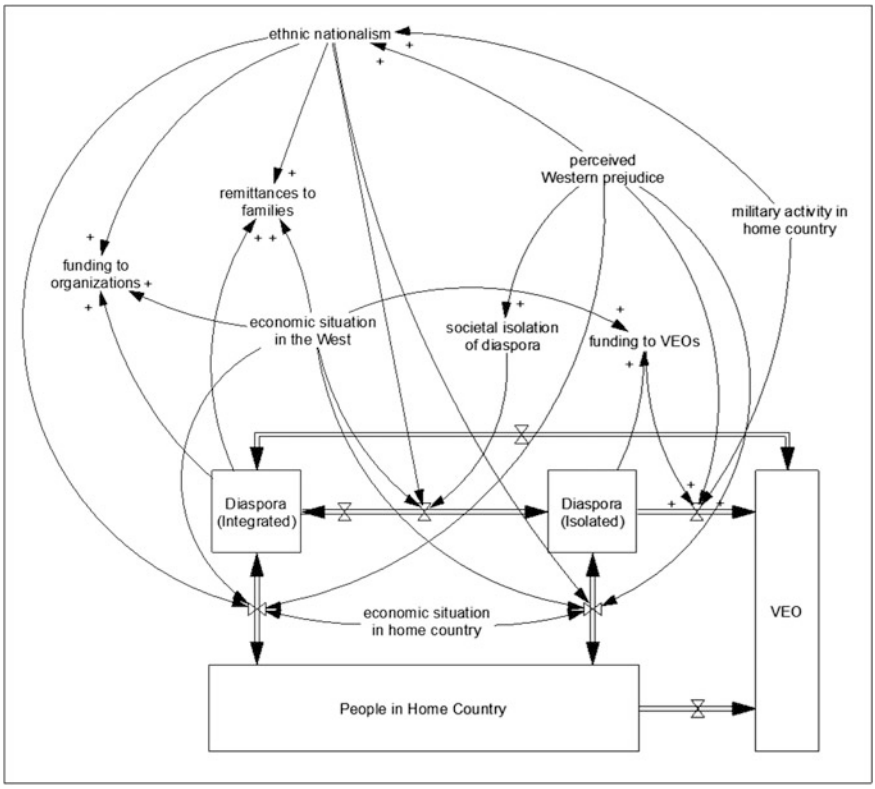


Fig. 2 Basic structure of the diaspora model

2.6 The Diaspora Model

We used the DYMATICA framework to model and simulate interactions between a VEO, the isolated and integrated diaspora communities, and people in the home country of interest. We assume that people in the home country can move to either diaspora community, and vice versa. These moves will be affected by economic situations in the west and the home country, as well as ethnic nationalism and perceived Western prejudice. People in the more integrated and more isolated diaspora communities can also move between those communities, behavior largely determined by ethnic nationalism, societal isolation of the diaspora in general, and the economic situation in the West. People in the home country may choose to join the VEO, as may people in the isolated diaspora community. These shifts will be determined by perceived Western prejudice, VEO funding, and military activity in the home country. The diaspora communities have a large affect on the situation at home through sending remittances to their families, funding to organizations, and funding to VEOs.

2.7 *Initial Model Results—Base Case Scenario*

Initial results of a base case run of the model are given here. These results should be seen as illustrative rather than predictive and are meant to give insight into potential outcomes of the system under the assumptions of this particular model structure. Future work for this project will include further scenario analysis, as well as uncertainty quantification and assessment of the resulting outcomes.

The results of this scenario are driven by an invasion of the diaspora's home country. This invasion begins 12 months into the 10 year time horizon and lasts for two years. Because of this invasion, nationalism throughout the diaspora communities and the home country community increases, leading some of the people in each population to increase their support of the VEO, which is seen as fighting against the invasion (Fig. 1). The VEO can also be seen as the most effective, or at least the most destructive, force against the invading force. This perception can last to some degree over the course of the invasion and beyond, especially if the VEO is seen as successful or the home country's military is perceived as incompetent. In this case, he more initially aligned with the VEO, the greater and longer the support will be for the VEO (Fig. 3).

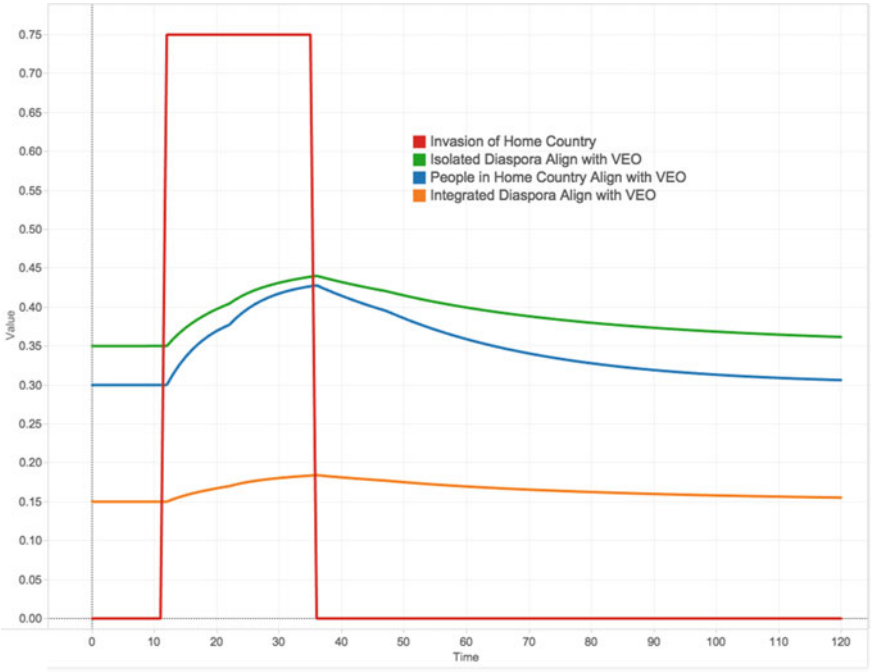


Fig. 3 Invasion of the country of interest and resulting alignment with the VEO among diaspora and home country communities

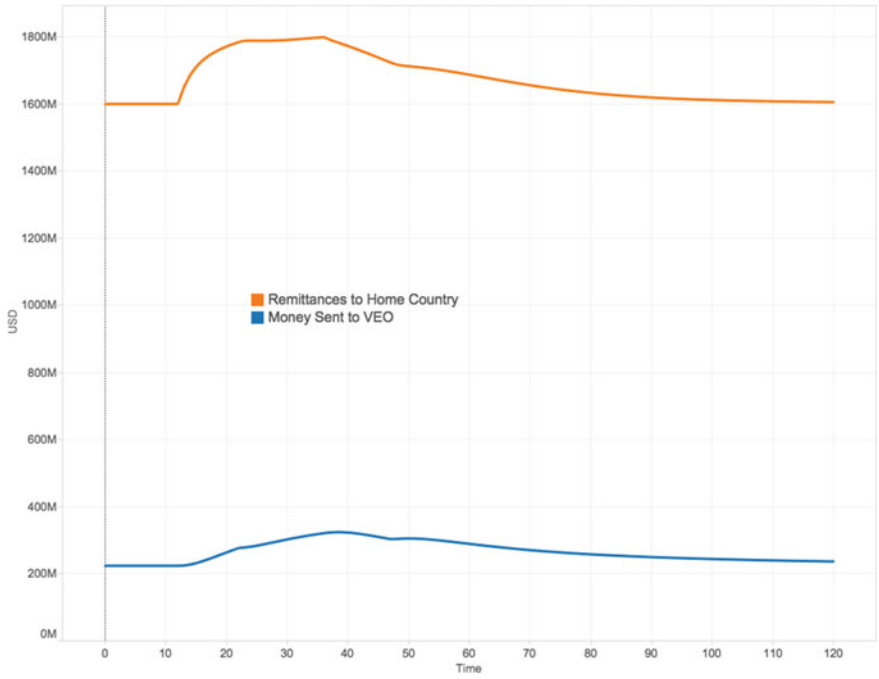


Fig. 4 Funding sent from diaspora to the home country

The invasion spurs the diaspora populations in Western countries to send more money to the home country (Fig. 4). Remittances grow over the time period of the invasion, as do funds sent specifically to the VEO. The diaspora populations see the VEO as helping to counter the invasion of the home country. Since sending funds is easier and safer than participating in a conflict, a rise in remittances may reflect a greater age range of individuals supporting the VEO than participating in the conflict.

Some of the people from each community will also choose to join the VEO. VEO recruiting is spurred by increased funding (Fig. 4) as well as by the perception that it is helping to counter the invasion. Nationalism, which also increases in response to the invasion, also encourages some to join the VEO. Figure 5 shows a stacked graph of VEO members coming from the home country, isolated diaspora in Western countries, and integrated diaspora in Western countries. Conditions in the home country and in Europe affect recruitment. The isolated diaspora community might tend to focus more than the integrated diaspora on events in the home country, and thus be more easily drawn into volunteering to participate in the conflict.

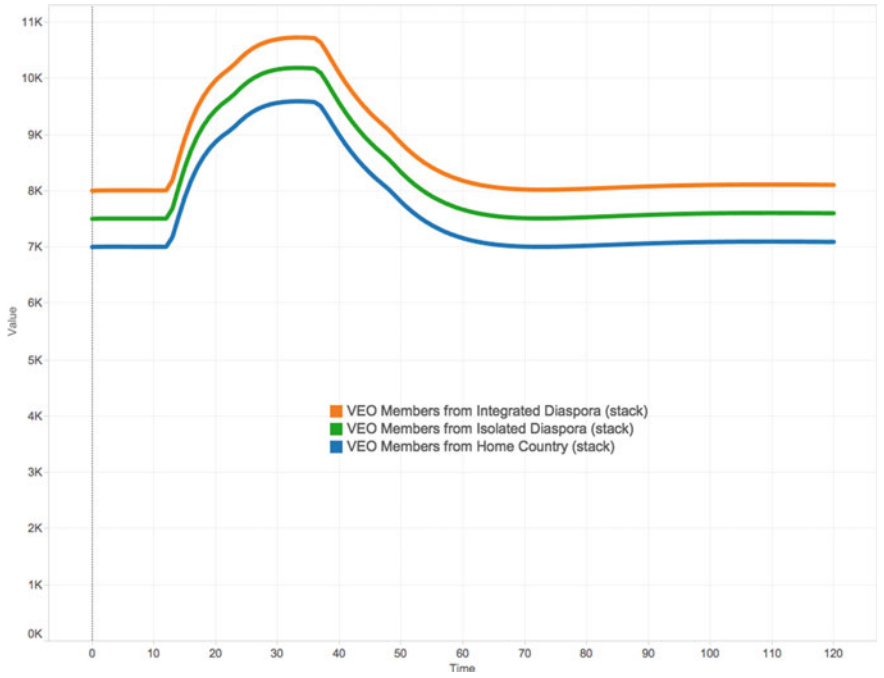


Fig. 5 VEO members from the home country and diaspora communities

Increased membership and funding cause public perception of the VEO's strength to grow during the invasion (Fig. 6, pink line). However, as this perception of strength increases, Western countries have more incentive to take action against the VEO's growing power in the home country. Once the perception of the VEO's strength reaches a threshold, Western countries take action against the VEO (Fig. 6, blue line). This decreases the VEO's overall strength and thus the perception of its strength among the general populations of the home country and Western countries (Fig. 6, red line). The overall perception of VEO strength is thus checked by Western action, and drops after the invasion ends as diaspora and home country support weaken, leading to lower levels of recruitment and funding for the VEO.

3 Conclusions

The intent of this effort is to investigate potential links between the actions of VEOs, people in the VEO's home country, and diaspora from that country living in the West. Our base case involves an initiating event focused on an invasion of the home country. Funding and recruitment to the VEO from diaspora populations are investigated, and Western intervention against the VEO is also included. We

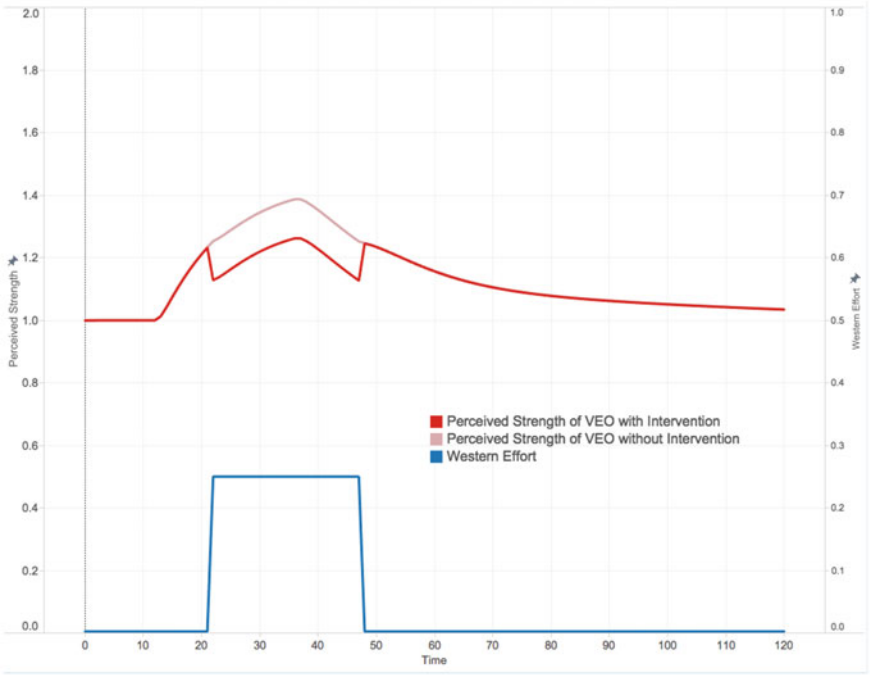


Fig. 6 Perception of VEO's strength and Western intervention

believe that with some alteration, this general model might apply to a variety of situations in which VEOs interact with home country and diaspora populations. This can give some insight into radicalization of diaspora populations, a topic which has recently been of great interest, especially considering recent influxes of migrants into Europe.

This is an initial study looking at diaspora populations and how they might be spurred to join or support VEOs. We plan to extend this by incorporating this model with two others. The first focuses on how VEOs might gain power among a population that is not ideologically inclined to support the VEO. The last model will focus on interactions between VEOs that want the support of diaspora and home country populations. We hope that by considering these three aspects of VEO interactions, we can gain insight into how these groups gain strength, membership, and money.

Acknowledgments This research was possible in part by funding from its internal Laboratory Directed Research and Development program. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly-owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Approved for unlimited release: SAND2016-2216 C.

References

1. Council on Foreign Relations: CFR Backgrounders: Europe's Migration Crisis (2015). http://www.cfr.org/migration/europes-migration-crisis/p32874. Accessed 23 September 2015
2. Staub, E.: Reconciliation after genocide, mass killing, or intractable conflict: understanding the roots of violence, psychological recovery, and steps toward a general theory. Political Psychol. **27**(6), 867–894 (2006)
3. Eurostat: Migration and migrant population statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics. Accessed 1 March 2016
4. Pew Research Center: Refugee Surge Brings Youth to an Aging Europe (2015). http://www.pewresearch.org/fact-tank/2015/10/08/refugee-surge-brings-youth-to-an-aging-europe/. Accessed 8 October 2015
5. Politico: Europe's Man Problem: Migrants to Europe Skew Heavily Male—and How That's Dangerous. http://www.politico.com/magazine/story/2016/01/europe-refugees-migrant-crisis-men-213500. Accessed 5 January 2016
6. Lehman, S.: ISIS Lures 3 Teenage Girls: Reporter's Notebook. The New York Times. http://www.nytimes.com/2015/08/18/insider/isis-lures-3-teenage-girls-reporters-notebook.html. Accessed 18 August 2015
7. Shane, S.: From Minneapolis to ISIS: An American's Path to Jihad. The New York Times. http://www.nytimes.com/2015/03/22/world/middleeast/from-minneapolis-to-isis-an-americans-path-to-jihad.html. Accessed 21 March 2015
8. Sipus, M.: Support for al-Shabaab through the diaspora. Forced Migrat. Rev. (37), 29. Vancouver (2011)
9. Welsh, T.: Remittance Cutoffs Threaten Somalis. U.S. News & World Report. http://www.usnews.com/news/articles/2015/02/19/remittance-cutoffs-threaten-somalis. Accessed 19 February 2015
10. Dhami, S., Sheikh, A.: The family: predicament and promise. Caring for Muslim patients, pp. 43–56 (2000)
11. Rousseau, C., Hassan, G., Moreau, N., Thombs, B.D.: Perceived discrimination and its association with psychological distress among newly arrived immigrants before and after September 11, 2001. Am. J. Public Health **101**(5), 909–915 (2011)
12. Pew Research Center: Religion & Public Life. The World's Muslims: Religion, Politics and Society. http://www.pewforum.org/2013/04/30/the-worlds-muslims-religion-politics-society-overview. Accessed 30 April 2013
13. Hroub, K. ed.: Political Islam: Context Versus Ideology, vol. 8. Saqi (2012)
14. Ajzen, I.: The theory of planned behavior. Organ. Behav. Hum. Decis. Process. **50**, 179–211 (1991)
15. McFadden, D.: Qualitative response models. In: Hildenbrand, W. (ed.) Advances in Econometrics. Cambridge University Press, New York (1982)

A Statistical Approach to the Subnational Geolocation of Event Data

Jennifer Lautenschlager, James Starz and Ian Warfield

Abstract Geolocation is the identification of the real-world geographic location of items such as coded news events. We have developed software to geolocate high volumes of event data, often to the subnational level, by combining existing entity extraction technologies with new statistical ranking algorithms. Our three-stage pipeline consists of: (1) named-entity recognition (identifying the text strings that represent named entities from underlying text and classifying them by type); (2) entity resolution (matching location strings to specific real-world locations referenced in a gazetteer); and, (3) location determination (selecting the most appropriate location for the event). We have used this software operationally to geolocate tens of millions of events and have formally evaluated both the accuracy and specificity of our results. Our latest formal evaluation had an overall subnational accuracy of 78 %, with 85 % of all events geolocated at a subnational level.

Keywords Geolocation · Geocoding · Geospatial data · Named-Entity recognition · Entity resolution · Entity linking · Gazetteer · Data analytics · Event data

1 Introduction

Event data, capturing the interactions between actors as extracted from news stories, has been of interest to social scientists for some years. Historically these events have been grouped by country, to facilitate the exploration of socio-political

J. Lautenschlager (✉) · J. Starz · I. Warfield
Lockheed Martin Advanced Technology Laboratories, 3 Executive Campus, Cherry Hill, NJ, USA
e-mail: jennifer.lautenschlager@lmco.com

J. Starz
e-mail: james.c.starz@lmco.com

I. Warfield
e-mail: ian.warfield@lmco.com

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_27

phenomena at the country level. However, when focusing analysis at the subnational level—for example, in the case of urban crisis monitoring—a greater degree of specificity is required. Events may need to be grouped by provinces, by cities, or even by specific neighborhoods within a city.

Geolocation is the identification of the real-world geographic location of items such as coded events. Such geolocated events can then be grouped to create time series data of value to social scientists, to track trends for the purposes of Indicators and Warnings (I&W), or to develop forecasting models capable of predicting unstable political situations. We have developed software to geolocate high volumes of event data, often to the subnational level, by combining existing entity extraction technologies with new statistical ranking algorithms. Our three-stage pipeline consists of: (1) named-entity recognition (identifying the text strings that represent named entities from underlying news stories and classifying them by type); (2) entity resolution (matching location strings to specific real-world locations referenced in a gazetteer); and, (3) location determination (selecting the most appropriate location for the event). Stage 1 relies predominately on existing open source software while stages 2 and 3 use the GeoNames gazetteer and custom-developed algorithms.

Our statistical approach to entity resolution uses a lightweight production rule system to assign nominal scores to the potentially hundreds of gazetteer matches for a given location string, with higher scores indicating likelier matches. These rules take into account factors such as exactness of the string match; presence and proximity of other locational cues within the underlying text; the type of location (e.g., city, square, province); commonality of the name; the actors associated with the event; the event type; and location population sizes. The collection of resolved entities serves as input to the location determination algorithm, which iterates over the collection to retain the most appropriate valid match.

We have used this software operationally to geolocate tens of millions of events and have formally evaluated both the accuracy and specificity of our results. Our latest formal evaluation had an overall subnational accuracy of 78 %, with 85 % of all events geolocated at a subnational level. We have also identified issues specific to geolocating event data derived from news sources, such as differing Arabic transliteration techniques used by different publishers, and have developed methods to address these challenges.

In this paper, we present a background section briefly describing past work in the technologies underlying geolocation along with a review of candidate knowledge bases. We then detail the three steps in our geolocation pipeline, outlining the data, methodology, experimentation and results of each step. Finally, we conclude with our observations and thoughts on future research.

2 Related Work

Our approach to event geolocation builds off two core information extraction tasks, named-entity recognition (NER) and entity resolution (also known as entity linking). Both exist as active areas of research within the Natural Language Processing (NLP) community.

Research into named-entity recognition dates back to the 1990s, with many different approaches explored [1]. While early systems relied on learned lookup lists [2], focus quickly shifted to computational models such as finite state transducers [3] and rule-based grammars [4] augmented by lookup lists. Though such models were initially hand-crafted, machine learning techniques soon became the norm [5, 6]. Today there are many open source systems available [7–9].

Entity resolution is an extension of the earlier problem of coreference resolution, which involves finding all expressions that refer to the same entity in a text [10, 11]. In entity resolution, references not only link with one another but also into a populated knowledge base such as a gazetteer or Wikipedia. Entity resolution is a relatively new field of study, with some of the earliest publications occurring in the mid-2000s [12, 13]. Some key challenges include name variations, such as alternate spellings or acronyms, and the ambiguity that occurs when the same textual string identifies multiple entries within a knowledge base. These challenges have led to a typical two-stage approach to the problem: first identifying possible candidates within the knowledge base, and then ranking the candidates so as to determine the likeliest option [14].

Candidate identification, at its simplest, can be done via string matching, though the challenge of name variations renders this approach alone insufficient. One technique to broaden the search is query expansion, where the underlying knowledge base is mined to generate multiple identifying text strings for the same object [15]. For example, Wikipedia is a commonly used knowledge base for research purposes and has several useful features for query expansion such as redirection links, disambiguation pages, and anchor text. Another technique is query relaxation, where partial or approximate string matches are used to identify candidates [16].

Researchers have proposed numerous approaches to candidate ranking. In graph-based entity resolution, the interdependence of entities within a body of text is used to supply additional context to the ranking algorithm [17]. Other approaches involve machine learning techniques such as classifiers [18] or trained models [19]. One of the more novel approaches combines crowdsourcing with a probabilistic model to identify potentially malicious input [20].

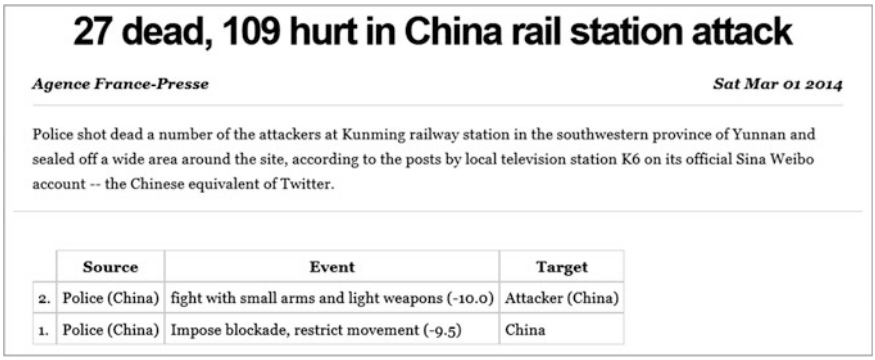
Another concern for geolocation involves which gazetteer to use as a knowledge base, and we considered several alternatives. OpenStreetMap, which advertises itself as a collaborative mapping wiki, has a web service called Nominatim for context-free lookup of location strings [21]. Another web service is the NGA GEOnet Names Server, at over 11 million locations with metadata [22]. GeoNames is a graphical database with over 8 million locations and associated metadata such as alternate spellings, latitude and longitude, and populations [23].

3 Named-Entity Recognition

Data. Our input data was a collection of roughly 14.5 million automatically coded events, in the form of "who did what to whom", along with the news stories from which they were extracted. Figure 1 shows several example events extracted from a paragraph of text. Though the text was made available in English, roughly a quarter of the news stories were translated from foreign languages which included French, Spanish, Portuguese, and Arabic. A full set of this event data, excluding the associated copyrighted source text, is available on the Harvard Dataverse in the ICEWS Coded Event Data Dataset [24].

We also needed a list of known location names as input to one of the third-party tools our algorithm used. We chose to use the GeoNames gazetteer as the basis for this list; we made this decision based on its depth of coverage and accessibility. We created our list by exporting a subset of the GeoNames gazetteer, starting with the English-language names of all 1st and 2nd order administrative districts and their capitals. To this, we added the names of any place with a recorded population greater than 5000. Next we added names of a few dozen feature types within the GeoNames gazetteer, such as hospitals, airports, and universities. Finally, we removed any entries that consisted solely of common English words or person names.

Methodology. As named-entity recognition is a well-studied field, we chose to leverage existing open source tools for this stage in our pipeline. We started with the hypothesis that by combining multiple NER tools with differing NLP approaches, we would increase our identification of location strings while decreasing our false positive rate (i.e., the percent of strings identified as indicating locations when they do not). We based this hypothesis on the belief that by combining multiple tools, we could use the strengths of one tool to counter-balance weaknesses in another.



27 dead, 109 hurt in China rail station attack

Agence France-Presse *Sat Mar 01 2014*

Police shot dead a number of the attackers at Kunming railway station in the southwestern province of Yunnan and sealed off a wide area around the site, according to the posts by local television station K6 on its official Sina Weibo account -- the Chinese equivalent of Twitter.

	Source	Event	Target
2.	Police (China)	fight with small arms and light weapons (-10.0)	Attacker (China)
1.	Police (China)	Impose blockade, restrict movement (-9.5)	China

Fig. 1 Two sample events extracted from the lead paragraph of a news story

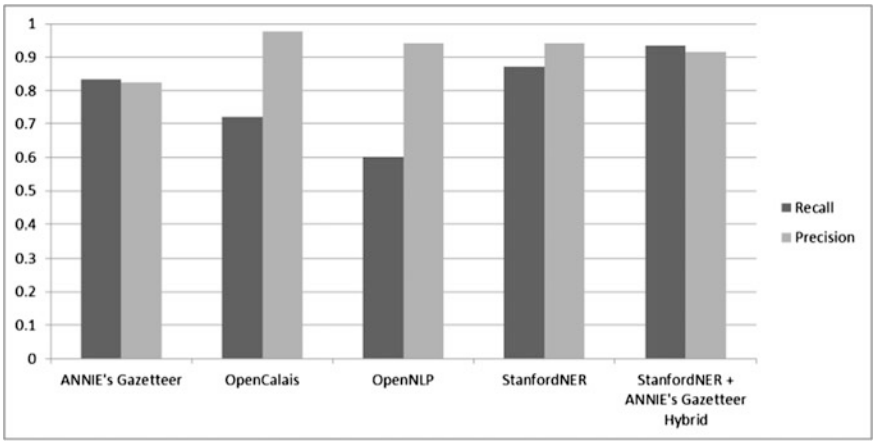


Fig. 2 A comparison of several open source NER alternatives

Experimentation. For our experiment, we created a hybrid algorithm based on two previously studied NER tools. The first was A Nearly-New Information Extraction (ANNIE) system, part of the General Architecture for Text Engineering (GATE) tool suite [25]. This component identifies entity names in text based on user-defined lookup lists and fast string-matching algorithms; we used our previously created list of location strings for its lookup list. The second was the Stanford NER tool, which is a Java implementation of a Named Entity Recognizer [11]. It is a CRFClassifier, which provides a general implementation of arbitrary order linear chain Conditional Random Field (CRF) sequence models.

We ran both tools over our sentence of interest, and then took the superset of their output as our set of location strings. However, given that ANNIE's Gazetteer operates without any knowledge of context, it would sometimes incorrectly identify a string as a location when it was actually part of the name for a person or organization (e.g., 'Jonestown' the city versus 'Paul Jonestown' the person). To compensate for this, we removed any string that the Stanford NER system tagged as being a person or organization.

Results. We compared our hybrid entity-name recognizer with several open source alternatives—ANNIE's Gazetteer, OpenNLP [8], Stanford NER, and Open-Calais [26]—by running them over the same corpus of news stories. Figure 2 shows the results of our analysis. We did not fully prove our hypothesis as the hybrid precision is slightly lower than that of the best performer. However, we improved our recall rate which, for our purposes, was an acceptable trade-off.

4 Entity Resolution

Data. In the entity resolution stage, we start with our list of previously identified location strings and match them, where possible, to entries in a knowledge base. We used a modification of the GeoNames gazetteer for our knowledge base. Our modifications were a combination of metadata augmentation and data cleaning. We have encapsulated them in an SQL script that we re-apply whenever we update the underlying GeoNames data.

Our metadata augmentation adds several pieces of information to each gazetteer entry, which we later use for ranking purposes in our resolution algorithm. Typically this information is already embedded in the gazetteer, and we merely extract it during a pre-processing step and then store it in a more explicit form. We record three Boolean flags per entry, as follows:

1. **`isCapital`**: Denotes a nation capital, based on the `feature_code` attribute within GeoNames.
2. **`isTopCity`**: Denotes a large-population city, based on the `population` attribute within GeoNames.
3. **`isCommon`**: Denotes a commonly mentioned city that is neither a capital nor a 'top city.' We set this attribute manually following periodic evaluations when we see many correct references to a city lacking other flags.

One final value we set on all gazetteer entries is a *specificity level*, which is an integer value indicating the relative size of a geographic region. We derive these values from the `feature_code` of the GeoNames entry. Smaller values indicate more specific locations, such as cities or buildings within a city, while larger values are used to indicate administrative districts, nations, or continents.

Data cleaning of the gazetteer is a mixture of automated and manually identified changes. On the automated side, we either remove or modify any names that are common English words or proper names. We choose to remove when the name matches either a minimally populated city or any spot within the city. Those left we modify, by both prepending a short identifier and appending a country name (e.g., "town of Douglas", "Douglas, South Africa"). Another automated change involves rules to rename foreign administrative districts to their most common English name (e.g., "Hiroshima-shi" -> "Hiroshima") and to deconflict airport names from the names of the city in which they are found (e.g., "Ghazni" -> "Ghazni Airport").

Manual cleaning results from erroneously identified geolocations that are flagged either by end users or in periodic evaluations. In cases where the error occurred due to the underlying gazetteer data, we follow a similar "remove or modify" strategy on a case-by-case basis.

Methodology. As with many entity resolution algorithms implementations, we adopted a two-step approach of first identifying all possible matches within our gazetteer and then ranking them. Our belief was that we could effectively identify possible matches through a combination of two factors: (1) leveraging the 'alternate names' metadata in GeoNames; and, (2) performing query expansion focused on

alternative spellings of Arabic names. For ranking our possible matches we calculated a numerical score based on a combination of location metadata and event context, and explicitly ignored the challenge of NIL detection (i.e., determining when a string corresponds to no entry in the gazetteer). Our hypothesis was that we could ignore the additional computational complexity related to NIL detection because, in the subsequent stage of location determination, any erroneously resolved locations would rank lower than the correct one and become effectively irrelevant.

Experimentation. The first step, querying for matching entries, is based on text matches between the location string and all names in the gazetteer. The GeoNames gazetteer provides multiple names (known as 'alternate names') for a given entry to account for transliteration and spelling variants, and we make use of all English language alternatives in our search. If no matches are found, we extend our search via query expansion by generating simple Arabic transliterations of our location string and querying for those within Arabic-speaking countries. We chose to focus only on Arabic transliteration as it is those spelling variants that we found most often missing in the GeoNames gazetteer.

In the second step, we use a simple rules system to calculate a *likelihood score* for each candidate entry. Examples, shown in Table 1, include rules like "add 15 points to the score of entries matching these criteria." These rules have been hand-weighted based on early experimental work and involved no machine learning. Inputs into the system include metadata about the candidate entry along with contextual information about the sentence and event from which we originally derived the location string. Information about the latter includes the country affiliations of the source and target of the event, the type of event, and geographic terms from the sentence (e.g., 'city', 'district', 'state'). We then choose the result with the highest score, with ties broken at random. Table 2 gives an example of scoring a set of candidate entries.

Table 1 Example rules for calculating the likelihood score of a candidate entry

ID	Priority	Condition	Type	Value
4	1	specificity == 1 OR specificity == 0	Additive	15
21	2	specificity == 2 AND containsProvinceTerm	Additive	40
31	3	isCapital	Additive	35
32	3	isTopCity	Additive	25
41	4	isLargeCity	Additive	30
42	4	isMediumCity	Additive	5
43	4	isTinyCity	Additive	−5
53	5	isTargetBiasedEvent AND isInSourceCountry	Multiplicative	3
54	5	isTargetBiasedEvent AND isInTargetCountry	Multiplicative	5.5
56	5	isTargetBiasedEvent AND neighborsTargetCountry	Multiplicative	2.5

Table 2 Subset of sample scores for evaluation the location string "Tripoli" in the event triple ("Jeffrey D. Feltman", "Makes a visit", "Omar Karami")

ID	City	Province	Country	Rules Matched	Score
266826	Tripoli	Liban-Nord	Lebanon	4, 32, 42, 54	247.5
2210247	Tripoli	Tarabulus	Libya	4, 31, 41	80
4879018	Tripoli	Iowa	United States	4, 43, 53	30
252601	Tripoli	Peloponnese	Greece	4	15

Results. No formal analysis was done at this stage on our entity resolution, though we did review our results periodically to ensure that they seemed plausible for the inputs given.

5 Location Determination

Data. Our input into the location determination algorithm is the set of resolved entities from the previous stage of the processing pipeline, along with the event we are geolocating and the news story from which it was extracted. Related to the event is the concept of *affiliated countries*, which are countries with which the source actor and target actor are affiliated. For example, an event about Vladimir Putin making a visit to meet Hassan Rouhani would have a source actor of 'Vladimir Putin' and a target actor of 'Hassan Rouhani'. The two affiliated countries would, therefore, be Russia and Iran.

Methodology. Our hypothesis was that the location of events was highly correlated to the country affiliations of its involved actors. Furthermore, we believed that the type of the event provided strong clues as to which country affiliation (source or target) was more likely correct. For example, an event about a Russian politician making a visit to meet with an Iranian politician will probably occur somewhere in Iran, while an event about Turkish police arresting Syrian refugees will most likely occur somewhere in Turkey.

Experimentation. We enter our geolocation algorithm with the previously discovered set of resolved locations and the list of actor-affiliated countries described above. We start by removing any unsuitable locations from our candidate pool. We consider a location to be unsuitable **unless** it meets one of the following criteria: (1) it lies within an affiliated country; (2) it lies within a country that borders an affiliated country; (3) its specificity level indicates that it is a province, country, or region; (4) it is considered to be a top city of any country; or, (5) it is a national capital.

We then use this set of suitable locations as input into a rules engine, similar as we did with location strings during the entity resolution stage. Our rules are hand-crafted to give precedence to more specific locations (e.g., city vs. province), locations within our list of affiliated countries, and locations following prepositions

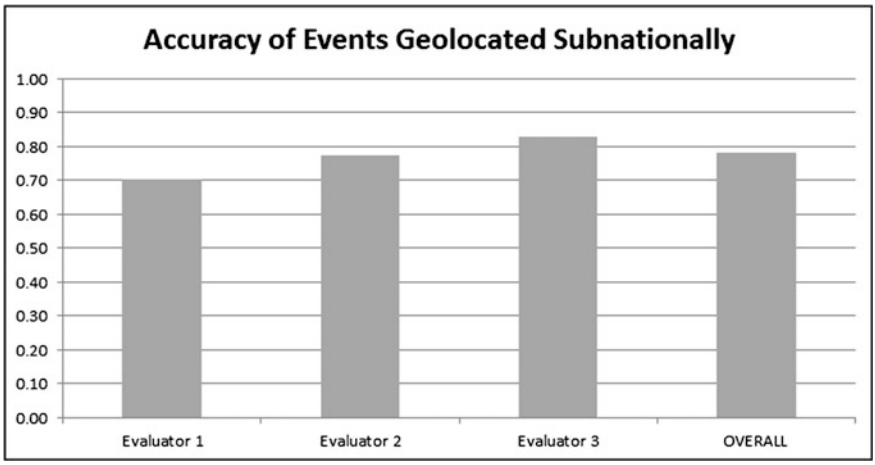


Fig. 3 Accuracy of a random set of events geolocated at the subnational level. Results are accurate at the 95 % confidence level ±5 % points

of place (e.g., at, in, near) in a sentence. We geolocate the event by choosing the location with the highest score, with ties broken at random.

The failure to identify a valid location, as occurs when we discover no suitable candidates for a sentence, causes the entire processing pipeline to repeat. We begin once again with the named-entity recognition stage and use the surrounding sentences as input. Repeated failures will eventually halt the overall process. At this point, we set the event's location to the most probable actor-affiliated country or, lacking any identified country affiliations, to nothing at all.

Results. We have used our geolocation algorithms for the past 5 years on over 14.5 million events. During this time, we have made several major revisions of the system to improve our accuracy. Our latest evaluation involved three trained evaluators and a set of 500 events randomly selected from the most recent year. The results of our analysis are shown in Fig. 3. We also randomly selected half a million events from the entire dataset and analyzed the specificity of the geolocations. We found that roughly 85 % of them were geolocated below the country level, with two-thirds of them geolocated to a specific city.

6 Conclusion and Future Research

The accurate and specific geolocation of coded event data is of value to researchers in numerous fields. We have developed and fielded software to geolocate a high volume of event data—roughly 14.5 million events spanning a 5-year period—with a three-stage pipeline. The first stage combines multiple open source NER tools, leveraging their strengths while offsetting their weaknesses, to identify location

strings in the underlying text. We then resolve these strings to actual gazetteer entries by statistically ranking candidate matches via a rule-based system. A similar approach is used to select the location of the event from one of the multiple locations found in the underlying news story. Our latest formal evaluation had an overall subnational accuracy of 78 %, with 85 % of all events geolocated at a subnational level.

One area of future research involves tuning the system rules, which we developed through manual means, with machine learning algorithms. Our existing collection of geolocated events provides a large training set for this purpose. Other potential research areas include the challenge of transliterated location names outside of the Arabic language (e.g., Chinese, Hebrew) and how better to disambiguate between multiple same-country locations with identical likelihood scores during the entity resolution step.

This work was funded under the Office of Naval Research (ONR) contract number N00014-14-C-0387.

References

1. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvist. Invest. **30**(1), 3–26 (2007)
2. Palmer, D.D., Day, D.S.: A statistical profile of the named entity task. In: Proceedings of the Fifth Conference on Applied natural language processing. Association for Computational Linguistics (1997)
3. Appelt, D.E., et al: SRI International FASTUS system: MUC-6 test results and analysis. In: Proceedings of the 6th Conference on Message Understanding. Association for Computational Linguistics (1995)
4. Mikheev, A., Moens, M., Grover, C.: Named entity recognition without gazetteers. In: Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics, pp. 1–8. Association for Computational Linguistic (1999)
5. Florian, R., et al.: Named entity recognition through classifier combination. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL. Association for Computational Linguistics (2003)
6. Liao, W., Veeramachaneni, S.: A simple semi-supervised algorithm for named entity recognition. In: Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing. Association for Computational Linguistics (2009)
7. Finkel, J.R., Grenager, T, Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics (2005)
8. OpenNLP, Apache. Welcome to Apache OpenNLP (2012). http://opennlp.apache.org
9. Agerri Gascón, R., Cuadros Oller, M., Gaines, S., Rigau Claramunt, G.: Opener: open polarity enhanced named entity recognition (2013)
10. Soon, W.M., Ng, H.T., Lim, D.C.Y.: A machine learning approach to coreference resolution of noun phrases. Comput. Linguist. **27**(4), 521–544 (2001)
11. Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M., Jurafsky, D.: Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In: Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, pp. 28–34. Association for Computational Linguistics (2011)

12. Bunescu, R.C., Pasca, M.: Using encyclopedic knowledge for named entity disambiguation. EACL **6**, 9–16 (2006)
13. Cucerzan, S.: Large-scale named entity disambiguation based on Wikipedia data. EMNLP-CoNLL **7**, 708–716 (2007)
14. Ji, H., Grishman, R.: Knowledge base population: successful approaches and challenges. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 1148–1158. Association for Computational Linguistics (2011)
15. Zhang, W., Su, J., Tan, C.L., Wang, W.T.: Entity linking leveraging: automatically generated annotation. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 1290–1298. Association for Computational Linguistics (2010)
16. Rao, D., McNamee, P., Dredze, M.: Entity linking: finding extracted entities in a knowledge base. In: Multi-source, Multilingual Information Extraction and Summarization, pp. 93–115. Springer, Berlin (2013)
17. Han, X., Sun, L., Zhao, J.: Collective entity linking in web text: a graph-based method. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 765–774. ACM (2011)
18. Lehmann, J., Monahan, S., Nezda, L., Jung, A., Shi, Y.: LCC approaches to knowledge base population at TAC 2010. In: Proceedings of the TAC 2010 Workshop (2010)
19. McNamee, P.: HLTCOE efforts in entity linking at TAC KBP 2010. In: Proceedings of the TAC 2010 Workshop (2010)
20. Demartini, G., Difallah, D. E., Cudré-Mauroux, P.: ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In: Proceedings of the 21st International Conference on World Wide Web, pp. 469–478. ACM (2012)
21. Haklay, M., Weber, P.: Openstreetmap: User-generated street maps. IEEE Pervasive Comput. **7**(4), 12–18 (2008)
22. NGA GEOnet Names Server (GNS): National Geospatial-Intelligent Agency (NGA) (2012). http://geonames.nga.mil/gns/html/
23. Wick, M., Vatant, B.: The geonames geographical database. http://geonames.org
24. Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., Ward, M.: ICEWS Coded Event Data. Harvard Dataverse (2015). doi:10.7910/DVN/28075
25. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: a framework and graphical development environment for robust NLP tools and applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (2002)
26. Reuters, T.: OpenCalais (2008). http://www.opencalais.com/

Tweet Sentiment Analysis with Pronoun Choice Reveals Online Community Dynamics in Response to Crisis Events

Samira Shaikh, Laurie Beth Feldman, Eliza Barach and Yousri Marzouki

Abstract We describe the emergence of an online community from naturally occurring social media data. Our method uses patterns of word choice in an online social platform to characterize how a community forms in response to adverse events such as a terrorist attack. Our focus is English Twitter messages after the Charlie Hebdo terrorist attack in Paris in January 2015). We examined the text to find lexical variation associated with measures of valence, arousal and concreteness. We also examine the patterns of language use of the most prolific twitter users (top 2 % by number of tweets) and the most frequent tweets in our collection (top 2 % by number of retweets). Differences between users and tweets based on frequency are revealing about how lexical variation in tweeting behavior reflects evolution of a community in reaction to crisis events on an international scale.

Keywords Human factors · Sentiment analysis · Social media · Natural language · Big data

S. Shaikh (✉) · L.B. Feldman · E. Barach
State University of New York, University at Albany, 1400 Washington Avenue, Albany, NY 12222, USA
e-mail: sshaikh@albany.edu

L.B. Feldman
e-mail: lfeldman@albany.edu

E. Barach
e-mail: eliza.barach@gmail.com

Y. Marzouki
Qatar University and Aix-Marseille University, Doha, Qatar
e-mail: yousri.marzouki@gmail.com

Y. Marzouki
Aix-Marseille University, Marseille, France

S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making*, Advances in Intelligent Systems and Computing 480,
DOI 10.1007/978-3-319-41636-6_28

1 Introduction

Variety in online communication is now a popular subject of investigation. These new communication formats range from synchronous to asynchronous, from private to public and from spontaneous to triggered by specific events [10]. These communications often have no specific addressee and yet they can be very revealing about the sender and those with whom she/he communicates. Further, because the roles of sender and recipient continually alternate, the distinction between roles typically loses importance. We use patterns of interaction at the level of groups of users rather than individuals to define community and subcommunity and this is our focus of study.

Dubois and Gaffney [7] and Xu et al. [13] showed that opinion leaders and influencers could be quantitatively identified in Twitter-mediated political discussions. Others differentiate between tweeters who are at the epicenter and those who are more peripheral to the focal event and emphasize how the synergy between them scales up the visibility of an event and helps in disseminating information [3]. Giglietto and Lee [8] report that retweets are particularly characteristic of emergency events when compared to ceremonial events (e.g., royal birth) that elicit more original tweets. Tweet popularity within a set of hashtags provides a means to detect whether tweeters have a bias to expose themselves selectively to those with similar ideological positions or whether the tweeting environment is more consistent with cross-ideological dissemination [2].

Patterns of lexical variation within a community including the use of emoticons intermixed with words within a bilingual work environment have been informative about language background [1]. The choice of personal pronouns in behavioral descriptions collected across diverse corpora has been revealing about group dynamics and psychological states [6]. Sometimes, relative to self-excluding pronouns (HE, SHE, THEY), news reports with self-including pronouns (I, WE) are associated with more positively valenced accounts [9].

Using Twitter data as a corpus to investigate the conversational characteristics of users is a timely scientific endeavor. For example, Burns and Stieglitz [5] highlighted the importance of studying a shared data set that is built upon one hashtag and stated that hashtags can be considered as "shared conversational markers", because they are inclusive without requiring a uniformity of opinion. In addition, they scale up faster than other types of discourse.

The following study aimed at extending the work conducted by Giglietto and Lee [8] about the twitter hashtags as behavioral indicators of online discourse and counter-discourse dynamics when users are facing a crisis event. Giglietto and Lee [8] used network co-occurrence patterns and semantic cluster analyses of content by hashtags to show that tweets associated with the shooting attack by Islamist gunmen at the office of the Charle Hebdo publication served as a genuine vehicle to exhibit various characteristics of the users' online discussions and reactions. Comparative analyses across hashtags #JeSuisPasCharlie and #JeNeSuisPasCharlie revealed interesting hidden patterns such as activism, grief, resistance, ethnocentrism and

islamophobia in the 74,074 posted tweets collected between 7th and 11th of January 2015.

The current study analyzed a more sizable sample than the one used by Giglietto and Lee [8] and within a more stable time window, covering the period from January 14th to February 9th. Moreover, we used many diverse hashtags to cover a sample with different emotional reactions. More importantly, instead of tagging the tweets associated with one opposing hashtag, we opted for a more fine-grained analysis of lexical variation, by differentiating among tweets according to pronoun use in natural human discourse. Our goal was to reveal nuanced discrepancies of group identities originating from the target hashtag #JeSuisCharlie. Indeed, the "Je" in the hashtag refers to the first person pronoun in French (I), so tracking the dynamics of the discourse and counter-discourse based on pronouns and the key words with which they occur is particularly illuminating. We have hypothesized that the presence of the "Je" in the key hashtag of the whole event is a trigger and a prime for an identity issue that a close analysis will help us decipher and understand.

2 Data Collection and Method

The data we collected consisted of all the tweets that contained at least one of 16 hashtags pertaining to the Charlie Hebdo attacks collected via Twitter streaming API. The final dataset consisted of 404,918 tweets from about 190,000 unique twitter ids that were shared between January 14th and February 9th 2015. The unique hashtags across the corpus are listed in Table 1.

3 Data Analysis

To ascertain that lexical variation is associated with pronoun choice, we extracted all tweets with pronouns and classified them as to whether a pronoun was present or absent and then whether the pronoun was in first (I, ME, MINE, etc.), second (YOU, YOUR, etc.) or third (HE, HIM, HIS, SHE, HER, HERS, THEY, etc.) person. For first person pronouns, we separated them into singular and plural so that we could track individual and collective perspectives [9].

Table 1 List of unique hashtags and their frequency in our tweet collection

#JeSuisCharlie	#IamCharlie	#MarcheRepublicaine	#JeSuisAhmed
#ParisShooting	#NousSommesCharlie	#LaFranceEstCharlie	#jesuisfranck
#lemondeestcharlie	#ThanksTheWorldFromFrance	#jenesuispascharlie	#parisattacks
#parisattacks	#parisattack	#ParisEstCharlie	#CharlieHebdo

In this article, we analyzed three sets of data for valence, arousal and concreteness measures. The first set is all of the tweets containing first-person singular pronouns, the second is the set of all tweets containing first-person plural pronouns and the last is the set of all tweets containing no first- or second- person pronouns. We defined the last set as the baseline set. In future work, we aim to include tweets containing second-person pronouns (YOU, YOUR etc.) in our pronoun analyses.

Each of the three sets was pre-processed as follows: We first removed stop-words (non-content words such as THE, A etc.) from the tweets. Next, we converted all words to their lower-case equivalents. In the last step, we calculated the frequencies of all words that co-occur with our target pronoun set (first-person singular, first-person plural and baseline). Thus, we obtained a list of words, ranked by their frequencies that co-occurred with the sets of pronouns we were interested in. We then determined the valence, arousal and concreteness scores of any word that was listed in our corpus by consulting databases collected by researchers [4, 12]. Psycholinguistic measures [4, 12] included valence (the pleasantness of a given word), arousal (intensity of evoked emotion, terrified > grief) and concreteness (the degree to which the concept is perceptible to one of the senses). We calculated weighted means for arousal valence and concreteness for each word using the following formula: multiplying the arousal, valence and concreteness score of a word by the number of times that word appeared in the corpus. When considering the valence, arousal and concreteness scores, we focused on those words that exhibited the most extreme scores—which are the 25 % most negative words and 25 % most positive words for valence, and similarly for arousal and concreteness, the top most 25 % words from the respective databases were used.

4 Results

We provide descriptive statistics of the data in Sect. 4.1. In Sects. 4.2–4.4, we present the analysis of tweets that contrasts the measures of valence, arousal and concreteness across the three different sets defined by pronoun use (first person singular, first person plural and no first person pronouns: the baseline). In Sect. 4.5, we contrast the language use patterns for the subcommunity of super-tweeters (top 2 %) and the remainder of the community. In Sect. 4.6, we contrast the language use patterns that characterize the super-tweets (top 2 %) versus the remainder of our tweet collection.

4.1 Distribution of Tweets and Users

Figure 1 shows the distribution of tweets in our data collection. We note that the rate of tweets was higher in the days immediately after the event, than as time passed.

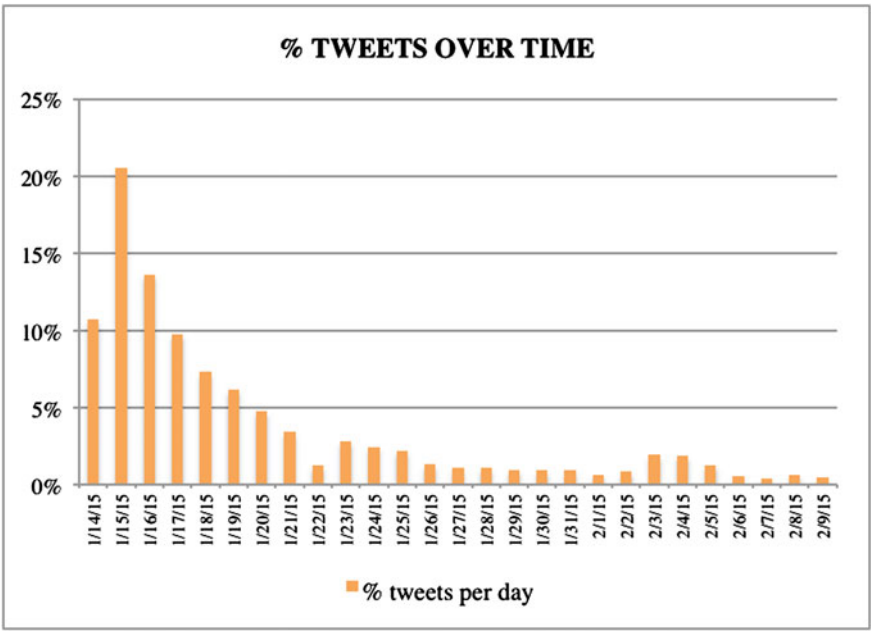


Fig. 1 Histogram of tweets rate over time in our collection

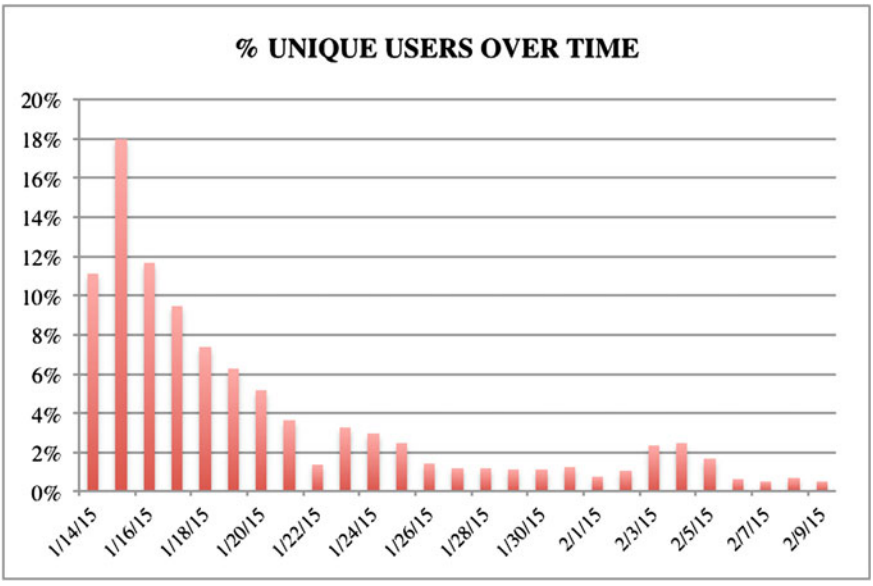


Fig. 2 Histogram of unique twitter users over time

Fig. 3 Word cloud showing the most common words in the collection

Figure 2 shows the histogram of unique twitter users per day ascertained by the number of unique twitter IDs per day. In Fig. 3, we show a word cloud with the most frequently appearing words in our collection.

4.2 Differences in Valence Scores Across Sets of Tweets with Different Pronouns

The analysis across days of the valence of words in tweets containing first-person pronouns indicated that they were more strongly valenced than baseline tweets (without first-person pronouns). The scatter plot and regression lines are shown in Fig. 4. Beta values were −0.0042, 0.0020 and 0.0077 for first-person singular, first-person plural and baseline, respectively. Importantly, slopes are marginally different between first-person singular and first-person plural. Tweets using first-person singular were more negative that tweets using first-person plural and the changes in that difference over time were marginally significant ($p = 0.06537$). The presence of strongly valenced language around pronouns suggest greater diffusion of emotion when contrasted with the more impersonal baseline.

4.3 Differences in Arousal Scores Across Sets of Tweets with Different Pronouns

The analysis across days of the arousal for words indicated that tweets using first-person singular pronouns tended to have lower arousal than tweets using first-person plural or the baseline tweets. The scatter plot and regression lines are shown in Fig. 5. Means were 5.96, 6.20, and 6.15 for first-person singular, first-person plural and baseline respectively. Tweets with first-person singular pronouns failed to show a change in arousal over time. However, tweets using first-person plural pronouns and baseline showed a non-significant decrease in arousal over time (beta = 0.0044, 0.0133 and −0.0245 for first-person singular,

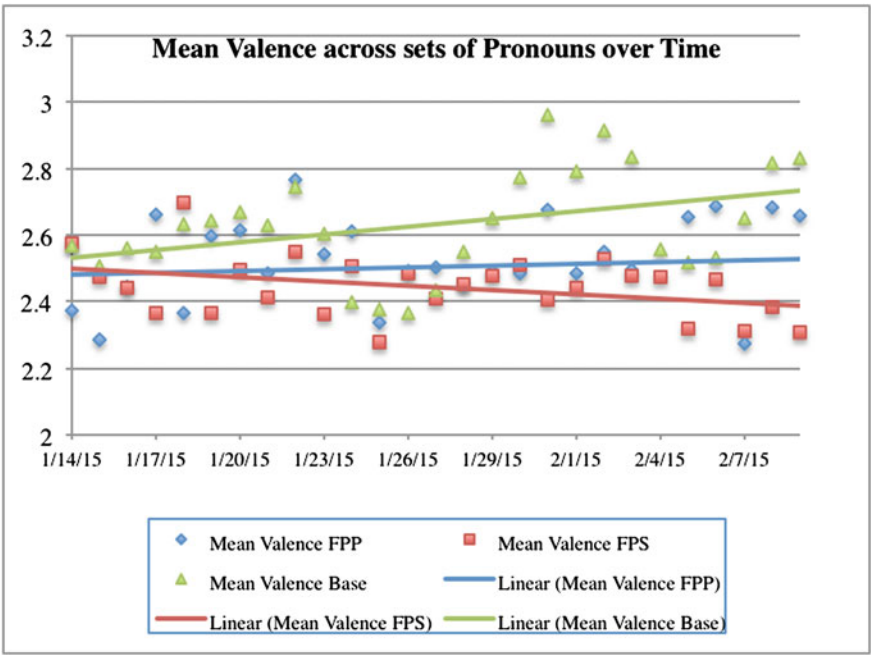


Fig. 4 Scatter diagram with regression lines over mean valence of words that co-occur with first-person singular (FPS) and first-person plural (FFP) pronouns and baseline (no first person pronouns)

first-person plural and baseline respectively; $p > 0.1$). This finding is consistent with generally high arousal within the community and minimal change with the passage of time.

4.4 Differences in Concreteness Scores Across Sets of Tweets with Different Pronouns

High concreteness is often interpreted as a measure of psychological distance [11] and tweets using first-person plural (mean = 4.34) tended to be marginally weaker in concreteness than tweets in first-person singular (mean = 4.35) or in baseline (mean = 4.38). Contributions of concreteness that vary with pronoun were negligible in the community as a whole and more obvious in the analysis of the supertweeter subcommunity described below (Fig. 6).

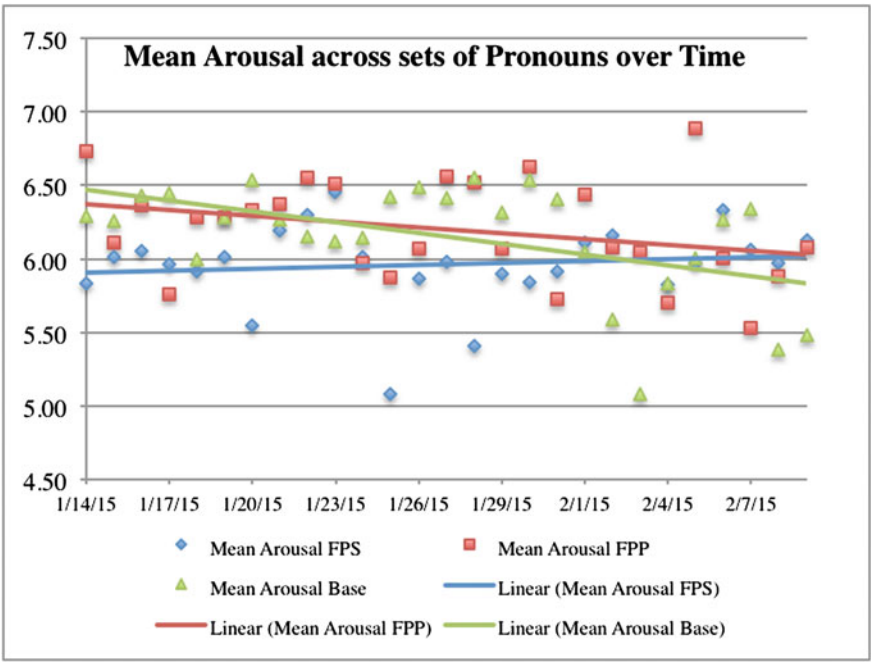


Fig. 5 Scatter diagram with regression lines over mean arousal of words that co-occur with first-person singular (FPS) and first-person plural (FFP) pronouns and baseline (no first person pronouns)

4.5 Differences in Language Use of Supertweeters Versus the Rest

Prior research suggests that supertweeters can be likened to those at the epicenter and they tend to have higher arousal than those whose activity makes them more peripheral to the focal [3]. In order to verify such findings and others, we analyzed the differences in language use of the top 2 % (by mean rate of tweeting) of the twitter users in our collection. We divided the entire collection of tweets into three separate phases of roughly equal size (in terms of number of tweets) across time. We then analyzed the lexical variation in each phase by the use of pronouns similar to our prior analyses.

Valence and Arousal Analysis (Figs. 7 and 8): Super tweeters did differ from non-supertweeters with respect to the emotional content of their tweets. T-tests showed more negative tweets (lower valence) in non-supertweeters than super-tweeters (2.66 vs. 2.78) but no difference in arousal (5.97 vs. 6.00).

Concreteness Analysis: Super tweeters did not differ from non-supertweeters in their use of more concrete words or fewer different concrete key words overall. These are words whose concreteness rating is in the extreme 25 % of the

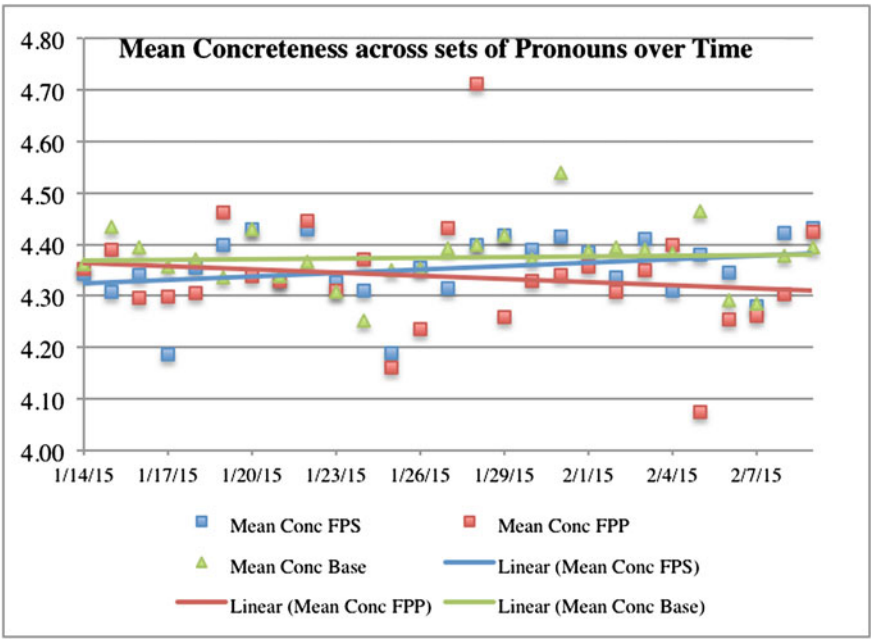


Fig. 6 Scatter diagram with regression lines over mean concreteness of words that co-occur with first-person singular (FPS) and first-person plural (FFP) pronouns and baseline (no first person pronouns)

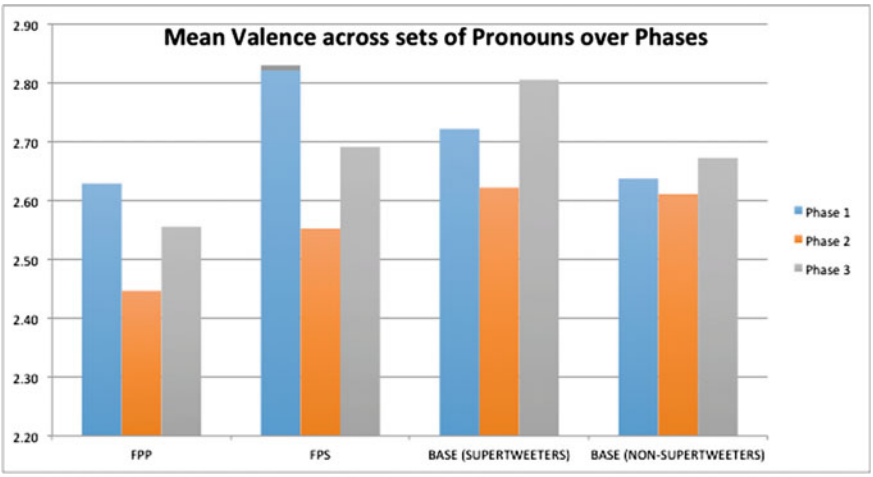


Fig. 7 Analysis of valenced language use by supertweeters across three phases and pronouns (first-person singular (FPS), first-person plural (FPP) and baseline) in our collection. The rightmost sets of bars in this chart compare these measures against the baseline for non-supertweeters

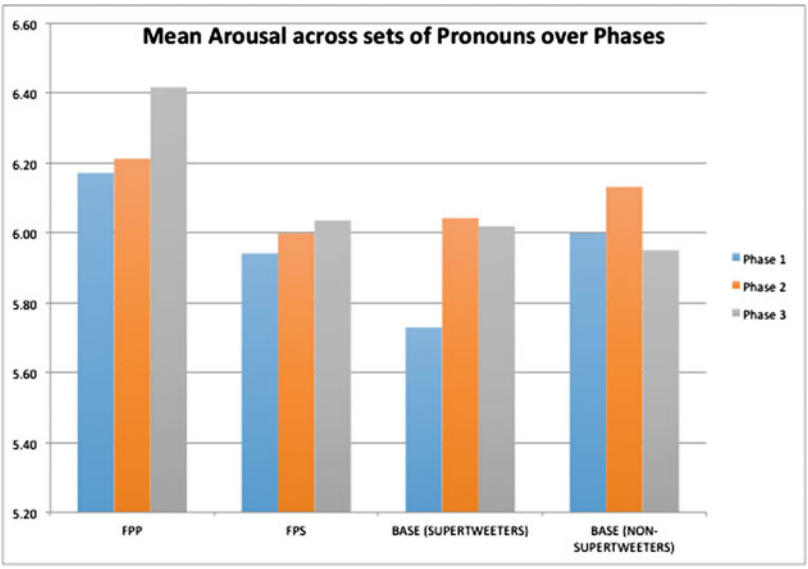


Fig. 8 Analysis of arousal language use by supertweeters across three phases and pronouns (first-person singular (FPS), first-person plural (FPP) and baseline) in our collection. The rightmost sets of bars in this chart compare these measures against the baseline for non-supertweeters

concreteness database [12]. However, among non-supertweeters, messages containing first-person singular (4.34) or first-person plural (4.29) tend to be significantly lower in concreteness than messages without pronouns (4.38). In addition, messages using first-person plural pronouns had significantly lower average concreteness than messages in first-person singular ($p < 0.005$). Among supertweeters none of these contrasts for concreteness were significant. Results suggest greatest psychological distance for tweets with no pronoun as well as greater distance with first-person plural than with first-person singular.

4.6 Differences in Language Use of Supertweets Versus the Rest

To learn about which tweets circulate most, we compare the composition of tweets based on frequency of the tweet message. We classify tweets based on those that are (re) tweeted most often (top 2 %) and the remainder. We define tweets that circulate the most (top 2 % with a retweet rate higher than 3 stdev above the mean rate of tweeting) as supertweets (N = 173) and contrast their composition with that of the remainder of tweets (N = 55954). Although the total number of tweets in our collection is over 400,000, we found that there are 56127 total unique tweets (14 %).

In line with many studies, first person pronoun usage appears to increase in the subsequent period of events associated with threat and disruption [6]. In this case, it is first person singular pronouns that increase most. Moreover, the presence of individual differences in tweeting style suggests implicit and collective leadership roles within an online community. Most interesting was supertweets containing first-person plural pronouns were significantly lower in valence (2.55) and higher in arousal (6.38) than the baseline ($p < 0.\ 05$). First-person singular supertweets did not differ from the baseline (2.69, 6.02). Like the analysis on all tweets, arousal for supertweets was significantly higher for first-person plural than for first-person singular but the difference in valence missed significance.

We examined supertweets (2 % most frequent messages) separately in each of three phases (thirds) of the data. With respect to supertweets, in each case, they were more likely to contain first-person singular than first-person plural pronouns. Moreover, the asymmetry was higher in the last phase. With respect to changes in the frequency of key words in super tweets, "truth" appeared in 3 % of tweets in the final phase but was negligible previously, whereas "insult/offense" and "hypocrisy" were close to 5 % at one of the two earlier phases but then dropped to 1 % or lower. In future work, we aim to analyze the composition of the supertweets in more detail to ascertain patterns that distinguish them from the remainder of the set.

5 Summary and Future Work

In this article, we focus on the linguistic characteristics of behavior within a network based on the diffusion of affect and other information content within tweets. We demonstrate that the tweets produced by a massive-scale network and associated with a dreadful event can be analyzed with psychological indicators such as valence, arousal and concreteness. Most consistent was that tweets with variants of the first-person plural pronoun included more high arousal words than other tweets and that arousal changed little over the course of the event. Overall, tweets with pronouns are more negatively valenced than those without and they got slightly more negative over time.

With appropriate data binning into phases, the valence of supertweeters also got more negative over time for tweets in either first-person singular or first-person plural. By contrast, in the subcommunity of supertweeters (14 %) communication without first person pronouns resembled that of nonsupertweeters, in that it got less negative over time. The valence finding captures different early affective reactions towards a dramatic event in our network. We interpret progressively more negative valence in the posts of supertweeters using first-person plural and first-person singular in conjunction with decreased negativity in the nonsupertweeters as evidence of a backlash. The increase in supertweets about truth and the decrease in supertweets about insult/offense and hypocrisy may be capturing the same phenomenon.

In summary, the relatively time constrained interactions based on individual posts could allow us to track the emergence of a sub-community that is diverging in an asynchronous manner from the community as a whole. The reversed change in the valence marks a disruption within the network that is triggered by a sub community of supertweeters whose style of tweeting deviates from that of the majority. Arousal is high throughout although patterns of lexical variation did not differ significantly between the super-tweeters and the non-supertweeters. Unlike arousal and concreteness, in our data it seems that valence better captures the strength of the disruption introduced by a sub community on the network's collective behavior. Consistent with some previous studies, we have demonstrated how quantitative analysis of sentiment in Twitter text can be immensely helpful to uncover hidden patterns and underlying processes that may provide new insights into collective behavior within a massive-scale network.

References

1. Aragon, C.R., Chen, N-C., Feldman. L.B., Kroll, J.F.: Emoticon and text production in first and second languages in informal text communication. In: Kennedy, W.G., Agarwal, N., Yang, S.J. (eds.) SBP 2014, LNCS 8393, pp. 217–222 (2014)
2. Barberá, J.T., Jost, J.N., Joshua, T., Richard, B.: Tweeting from left to right: is online political communication more than an echo chamber? Psychol. Sci. **26**(10), 1531–1542 (2015a)
3. Barberá, P., Wang, N., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J.: The critical periphery in the growth of social protests. PLoS ONE **10**(11), e0143611 (2015). doi:10.1371/journal.pone.0143611
4. Brysbaert, M., Warriner, A.B., Kuperman, V.: Concreteness ratings for 40 thousand generally known English word lemmas. Behav. Res. Methods **46**, 904–911 (2014)
5. Burns, A., Stieglitz, S.: Towards more systematic Twitter analysis: metrics for tweeting activities. Int. J. Soc. Res. Methodol. **16**, 91–108 (2013)
6. Chung, C. K., Pennebaker, J. W.: The psychological functions of function words. In: Fiedler, K. (ed.) Social Communication (pp. 343–359). Psychology Press, New York, NY (2007)
7. Dubois, E., Gaffney, D.: The Multiple facets of influence: identifying political influentials and opinion leaders on Twitter. Am. Behav. Sci. **58**(10), 1260–1277 (2014)
8. Giglietto, F., Lee, Y.: To be or not to be Charlie: Twitter hashtags as a discourse and couter-doscourse in the aftermath of the 2015 Charlie Hebdo shooting in France. In: 5th Workshop on making sense of Microposts (2015)
9. Gustafsson, S.M., Lindholm, T., Sikstrom, S.: Biases in news media as reflected by personal pronouns in evaluative contexts. Soc. Psychol. **45**(2), 103–111 (2013)
10. Stefanone, M.A., Lackaff, D., Rosen, D.: Contingencies of self-worth and socialnetworking-site behavior. Cyberpsychol. Behav. Soc. Netw. **14**, 41–49 (2011)
11. Trope, Y., Liberman, N.: Construal-level theory of psychological distance. Psychol. Rev. **117**, 440–463 (2010)
12. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav. Res. Methods **45**, 1191–1207 (2013)
13. Xu, W.W. et al.: Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election. The American behavioural scientist (2014)