MTR140409
MITRE TECHNICAL REPORT

MITRE

Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS):

Overview of Test and Evaluation Materials

Kevin Burns
Craig Bonaceto
Michael Fine
Carsten Oertel

November, 2014

This publication is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) ICArUS program, BAA number IARPA-BAA-10-04, via contract 2009-0917826-016, and is subject to the Rights in Data-General Clause 52.227-14, Alt. IV (DEC 2007). Any 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 ODNI, IARPA or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

McLean, VA

Abstract

The IARPA (Intelligence Advanced Research Projects Activity) program ICArUS (Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking) developed and tested brain-based computational models of "sensemaking" – a cognitive component of intelligence analysis. MITRE's role was in Test and Evaluation (T&E) of the neural-computational models developed by several teams of performers in two phases of the program, which began in December, 2010 and ended in June, 2014. This overview document summarizes the major T&E deliverables, providing an integrated introduction to more detailed documents available at http://www.mitre.org/publications and software/data available at: http://www.mitre.org/research/technology-transfer.

Table of Contents

1 Introduction 2
2 A Computational Basis for ICArUS Challenge Problem Design 3
3 Challenge Problem Design and Test Specification 3
4 Challenge Problem Walkthrough 5
5 Test and Evaluation Development Guide 5
6 Challenge Problem Software and Human Behavioral Data 6
7 References 9

1 Introduction

This document summarizes materials developed by MITRE in Test & Evaluation (T&E) of ICArUS: Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking, a program funded by the Intelligence Advanced Research Projects Activity (IARPA). Details of the ICArUS program itself are provided in the Broad Agency Announcement (BAA, 2010) available at: http://www.iarpa.gov/index.php/research-programs/icarus/baa.

As stated in the BAA (2010): *"The goal of the ICArUS Program is to construct brain-based computational models of the process known as sensemaking. Sensemaking, a core human cognitive ability, underlies intelligence analysts' ability to recognize and explain relationships among sparse and ambiguous data. By shedding light on the fundamental mechanisms of sensemaking, ICArUS models will enable the Intelligence Community to better predict human-related strengths and failure modes in the intelligence analysis process and will point to new strategies for enhancing analytic tools and methods."*

The ICArUS program included two phases over a 3.5-year period. In Phase 1, which ran from December, 2010 through December, 2012, IARPA funded three performer teams led by Hughes Research Laboratory, Raytheon/BBN, and Lockheed-Martin. In Phase 2, which ran from January, 2013 through June, 2014, IARPA funded two performer teams led by Hughes Research Laboratory and Raytheon/BBN. In both phases, performers developed neural-computational models that MITRE assessed in T&E.

In accordance with the BAA (2010), Phases 1 and 2 differed in the scope of laboratory challenge problems as well as in the performance scores that models were required to meet on those problems in qualitative and quantitative assessments by T&E. For Phase 1, the challenge problem involved spatial sensemaking in which the underlying probabilities of events varied in space but were constant in time. For Phase 2, the challenge problem involved spatial-temporal sensemaking in which probabilities of events were changing in both space and time. For each phase, performers' models were required to meet pre-established success criteria (more stringent for Phase 2 than for Phase 1) in three components of T&E: a qualitative Neural Fidelity Assessment (NFA), a quantitative Cognitive Fidelity Assessment (CFA), and a quantitative Comparative Performance Assessment (CPA).

For each phase, T&E includes: designing a challenge problem that poses cognitive task demands prototypical of geospatial sensemaking; collecting behavioral data to measure human performance on the challenge problem; and assessing the extent to which neural-computational models developed by performers can explain, predict, and emulate human sensemaking on the challenge problem. In accomplishing these T&E activities, MITRE developed a number of products that that are publically available, including documents available at http://www.mitre.org/publications and software/data available at http://www.mitre.org/research/technology-transfer. These products are summarized in the remaining sections of this overview document.

2 A Computational Basis for ICArUS Challenge Problem Design

Per the BAA (2010), ICArUS challenge problems are designed to address core processes of a notional framework by Klein, et al. (2007), known as the data-frame theory of sensemaking. The data-frame theory offers a conceptual description of sensemaking, but does not provide a computational specification of functional processes or knowledge representations – as needed for quantitative design and assessment of ICArUS challenge problems. In particular, the ICArUS BAA (2010) requires that T&E include two types of quantitative assessments, namely: Comparative Performance Assessment (CPA), using a numerical percentage to measure how well a neural model matches human sensemaking performance; and Cognitive Fidelity Assessment (CFA), using normative (Bayesian) solutions as benchmarks for measuring whether neural models exhibit cognitive biases like those of human subjects.

To meet these needs, MITRE developed a Bayesian-computational framework that models sensemaking in a cycle of eight stages, dubbed the *Octaloop*. This model served as the computational basis for designing ICArUS challenge problems that address all core sensemaking processes, and for assessing the performance of humans and models in CFA and CPA.

The Octaloop model was derived from a real-world story of sensemaking described by Klein, et al. (2007), and the same model applies beyond ICArUS experiments to cases of real-world intelligence analysis. Formulation of this Bayesian-computational Octaloop, its application to ICArUS challenge problem design, and discussion of potential transition to techniques, training, and tools of real-world intelligence analysis, are all provided in the following document:

> Burns, K. (2014a). *A Computational Basis for ICArUS Challenge Problem Design*. MITRE Technical Report, MTR140415.

This document is available at: http://www.mitre.org/publications.

3 Challenge Problem Design and Test Specification

Besides requirements for CPA and CFA, noted above, the BAA (2010) also imposed other constraints on the design of ICArUS challenge problems. One important constraint that applied across Phases 1 and 2 was to minimize the role of rich and sophisticated knowledge representations (RASKRs) held by human subjects, because it is currently infeasible to endow neural models with comparable knowledge for use in sensemaking.

Additional design constraints were specific to each phase of the program. In particular, Phase 1 was to focus on spatial sensemaking processes, without a temporal component, whereas Phase 2 was to address spatial-temporal sensemaking processes. Also the T&E requirements for CPA, CFA, and a qualitative Neural Fidelity Assessment (NFA) differed between the two phases. Phase 2 had more stringent success criteria in terms of the modeling scope (e.g., how many functional brain areas are addressed in NFA; how many cognitive biases are addressed in CFA) and performance scores (e.g., how good a match to cognitive biases in CFA; how good a match to human sensemaking in CPA). For both phases, the ICArUS challenge problem comprised a suite of tasks (also called "missions"), each with multiple trials (and multiple stages per trial), as needed to address sensemaking processes that included analytical *inferencing*, operational *decision-making*, and informational *foraging* (to support inferencing and decision-making).

In both phases, but especially Phase 2, the challenge problem was designed to pose cognitive demands that were prototypical of real-world sensemaking (within constraints of the BAA, 2010). This was accomplished by reviewing dozens of case studies collected via interviews with practicing analysts, and by relating those case studies directly to task demands of the challenge problem.

Also in both phases, an important aspect of challenge problem design was to compute normative (Bayesian) solutions as benchmarks for assessing cognitive biases. This was a difficult requirement for T&E to satisfy, as discussed in *A Computational Basis for ICArUS Challenge Problem Design* (Burns, 2014a). As a result, the task demands of missions in Phase 1 and Phase 2 challenge problems were shaped largely by the cognitive biases (relative to normative solutions) that were specified by the BAA (2010) for CFA in each phase.

There were four biases addressed in Phase 1, namely: *Anchoring and Adjustment*; *Confirmation Bias*; *Representativeness*; and *Probability Matching*. There were four more biases (for a total of eight biases) addressed in Phase 2, namely: *Satisfaction of Search*; *Change Blindness*; *Availability*; and *Persistence of Discredited Evidence*.

Details of the ICArUS challenge problems are documented in two reports, one for each phase. Each document presents the underlying design rationale as well as the associated test specification developed for use in quantitative assessments of neural-computational models, including both CFA and CPA. The qualitative approach to Neural Fidelity Assessment (NFA) was similar for both phases and is described in the Phase 1 document.

The Phase 1 document is as follows:

> Burns, K., Greenwald, H., & Fine, M. (2014). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 1 Challenge Problem Design and Test Specification*. MITRE Technical Report, MTR140410.

The Phase 2 document is as follows:

> Burns, K. (2014b). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Challenge Problem Design and Test Specification*. MITRE Technical Report, MTR140412.

Both documents are available at: http://www.mitre.org/publications.

4 Challenge Problem Walkthrough

The *Challenge Problem Design and Test Specification* documents (discussed above) are technical and comprehensive, as required for use by T&E and performers on the ICArUS program. In order to provide a more accessible introduction to the challenge problems for others outside the program, MITRE prepared two non-technical "walkthrough" documents.

Each walkthrough is a collection of screen shots from the tutorial instructions that human participants viewed via Graphical User Interface (GUI) before and during ICArUS experiments.

The Phase 1 walkthrough appears in the following document:

> Burns, K., Fine, M., Bonaceto, C., & Beltz, B. (2014). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 1 Challenge Problem Walkthrough*. MITRE Technical Report, MTR140413.

The Phase 2 walkthrough appears in the following document:

> Burns, K., & Bonaceto, C. (2014). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Challenge Problem Walkthrough*. MITRE Technical Report, MTR140414.

Both documents are available at: http://www.mitre.org/publications.

5 Test and Evaluation Development Guide

As another companion document to the *Challenge Problem Design and Test Specification*, MITRE also developed a *Test and Evaluation Development Guide* for each phase of the program. These development guides were written for use by software developers and specify the Extensible Markup Language (XML) formats for the Phase 1 and Phase 2 challenge problems.

Each development guide contains detailed descriptions and examples of the input and output formats for each stage of each trial of each mission posed by the challenge problem. The input format specifies each trial in a challenge problem "exam", including the "feature vectors" defining geospatial elements and intelligence data. These inputs are processed by neural models developed by performer teams, as well as by the Graphical User Interface (GUI) developed by T&E to present the challenge problem to human participants. The outputs, which represent responses to trials in the exam, are recorded using the same format for neural models and human subjects.

The Phase 1 development guide appears in the following document:

Bonaceto, C., & Fine, M. (2014a). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 1 Test and Evaluation Development Guide*. MITRE Technical Report, MTR130652.

The Phase 2 development guide appears in the following document:

Bonaceto, C., & Fine, M. (2014b). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Test and Evaluation Development Guide*. MITRE Technical Report, MTR140472.

Both documents are available at: http://www.mitre.org/publications.

6 Challenge Problem Software and Human Behavioral Data

A final T&E product is the complete Java source code used in human experiments and model assessments, which is packaged along with the following materials: XML schemas defining the challenge problem formats; XML exam and feature vector files for each experiment; all behavioral data that were collected in experiments; and Java and MATLAB software for analyzing behavioral data.

The Java software, used for performing ICArUS experiments, reads and validates XML-based challenge problem exam documents and feature vectors as specified in the *Test and Evaluation Development Guides* discussed above. The same software presents the challenge problems in a Graphical User Interface (GUI) suitable for collecting sensemaking data from human participants, and records the human responses at each stage of each trial of each mission during experiments. In addition, the Java software supports calculation of normative solutions, and allows model developers to interact with the test harness developed by T&E via Hyper Text Transfer Protocol (HTTP).

The Java and MATLAB software, used for analyzing behavioral data collected in ICArUS experiments, makes numerous plots and graphs of human and model performance. This software also computes average human performance across all participants in an experiment, and scores neural models (relative to average human performance) in Comparative Performance Assessment (CPA) and Cognitive Fidelity Assessment (CFA) as described in *Challenge Problem Design and Test Specification* documents.

Use of the challenge problem software for collecting human behavioral data was coordinated by members of the faculty and staff of the Pennsylvania State University (PSU), as well as by members of MITRE's T&E team. Participant recruitment was the sole responsibility of PSU, and data collection was performed under protocols approved by the PSU Institutional Review Board (IRB) for research with human subjects, as well as by MITRE's IRB. Participants are identified only by a sequentially-assigned ID, and the human data contain no information that would reveal the identity of any individual.

The Phase 1 and Phase 2 experiments each employed different challenge problems, implemented in different GUI software. Software and data for both phases are available in a large zip file containing a number of README files that explain what is provided and how it functions.

These README files include:

- README.txt: Contains an overview description of all software and data, describing the contents of other README files.

- DEPLOYMENT_README.txt: Provides instructions on building the GUI software, including a desktop version of the GUI, a Java Web Start version of the GUI, and an Applet version of the GUI for use on the web. The GUI software contains functionality to: conduct human experiments; play back model performance on Phase 1 exams; and open and visualize feature vector files. Functionality to connect to the T&E test harness has been removed.

- ASSESSMENT_README.txt: Provides instructions on using the software to compute scores in Comparative Performance Assessment (CPA) and Cognitive Fidelity Assessment (CFA).

- MODEL_DEVELOPER_README.txt: Provides instructions on using the software to assist in development of neural models, including instructions on how to: read and validate XML exam and feature vector files, provide responses in XML output files, and compute normative solutions. Functionality to connect to the T&E test harness has been removed.

- MATLAB/MATLAB_README.txt: Provides instructions on using the software to create plots and graphs of human and model performance using the MATLAB source code.

Top-level folders in the zip file include:

- Data: Contains the exam and feature vector files, and all human behavioral data including responses to pre-test and post-test questionnaires.

- Distrib: Contains the built and packaged desktop version of the GUI.

- Images: Contains images and icons used in the GUI and presented to human participants.

- Lib: Contains external software dependences packaged in Java JAR files.

- MATLAB: Contains all MATLAB source code.

- Nbproject: Contains the Net Beans project files, for developers using the Net Beans Integrated Development Environment (IDE), enabling developers to open and edit the Java source code using Net Beans.

- Schemas: Contains the XML schemas defining XML exam and feature vector files.

- Src: Contains all Java source code.

- Web: Contains the built and packaged Java Web Start and Java Applet versions of the GUI, as well as example web pages with code to launch the Web Start Application or Applet.

The human behavioral data include all responses from participants in Phase 1 and Phase 2 experiments, including pilot and final experiments in each phase of the program. Pilot experiments were conducted to refine the challenge problem tasks and stimuli, as well as to collect sample human behavioral data for use in model development by performer teams. Final experiments were conducted to assess model performance in CFA and CPA, as described in the *Challenge Problem Design and Test Specification* for each phase.

The human behavioral data files include:

- Phase 1 Pilot Experiment data from N=45 participants in folder: data/Phase_1_CPD/assessment/Pilot_Exam.

- Phase 1 Final Experiment data from N=103 participants in folder: data/ Phase_1_CPD/assessment/Final_Exam.

- Phase 2 Pilot Experiment 1 (Missions 1-3 only) data from N=20 participants in folder: data/Phase_2_CPD/assessment/Sample-Exam-1.

- Phase 2 Pilot Experiment 2 data from N=30 participants in folder: data/Phase_2_CPD/assessment/Sample-Exam-2.

- Phase 2 Final Experiment data from N=123 participants in folder: data/Phase_2_CPD/assessment/Final-Exam-1.

The zip file containing all ICArUS *Challenge Problem Software and Human Behavioral Data*, as summarized above, is available at: http://www.mitre.org/research/technology-transfer.

7 References

BAA (2010). IARPA Broad Agency Announcement, *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS*). IARPA-BAA-10-04, April 1, 2010.

Bonaceto, C., & Fine, M. (2014a). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 1 Test and Evaluation Development Guide*. MITRE Technical Report, MTR130652.

Bonaceto, C., & Fine, M. (2014b). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Test and Evaluation Development Guide*. MITRE Technical Report, MTR140472.

Burns, K. (2014a). *A Computational Basis for ICArUS Challenge Problem Design*. MITRE Technical Report, MTR140415.

Burns, K. (2014b). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Challenge Problem Design and Test Specification.* MITRE Technical Report, MTR140412.

Burns, K., & Bonaceto, C. (2014). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Challenge Problem Walkthrough*. MITRE Technical Report, MTR140414.

Burns, K., Fine, M., Bonaceto, C., & Beltz, B. (2014). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 1 Challenge Problem Walkthrough*. MITRE Technical Report, MTR140413.

Burns, K., Greenwald, H., & Fine, M. (2014). *Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 1 Challenge Problem Design and Test Specification*. MITRE Technical Report, MTR140410.

Klein, G., Phillips, J., Rall, E., & Peluso, D. (2007). A data-frame theory of sensemaking. In Hoffman, R. (ed.), *Expertise Out of Context*. New York: Lawrence Erlbaum, pp. 113-155.