OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

Knowledge Representation in Neural Systems (KRNS) Program: Proposers' Day

LEADING INTELLIGENCE INTEGRATION

Dr. Brad Minnery
IARPA, Incisive Analysis Office
12 July 2012

INTELLIGENCE ADVANCED RESEARCH PROJECTS ACTIVITY (IARPA)

NOTE: As a result of feedback from Proposers' Day, minor changes have been made to these slides to improve clarity.

There may be further differences between the information contained in these slides and the BAA.

Disclaimer

- This presentation is provided solely for information and planning purposes
- The Proposers' Day does not constitute a formal solicitation for proposals or proposal abstracts
- Nothing said at Proposers' Day changes the requirements set forth in a BAA
- BAA supersedes anything presented or said at the Proposers' Day by IARPA

Today's Goals

- Familiarize participants with IARPA's interest in KRNS – Please ask questions & provide feedback; this is your chance to alter the course of events.
- Foster discussion of complementary capabilities among potential program participants, AKA teaming. Take a chance, someone might have a missing piece of your puzzle.

Schedule

- Full Proposals will be due 60 days after the BAA is published
- Once BAA is released, questions can only be answered in writing on the program website

Today's Topics

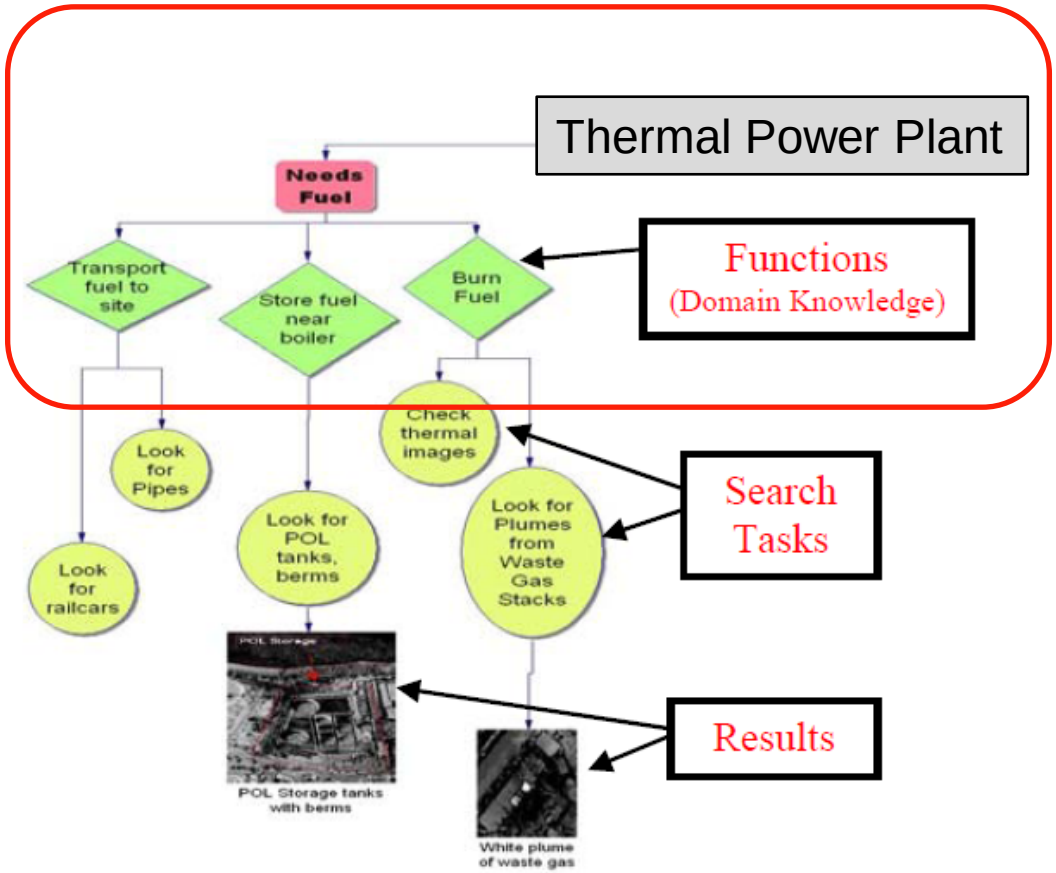
- Program Overview
  - Background, Goals & Definitions
  - Program Structure
  - Milestones and Metrics
- Award Information
- Eligibility Information
- Application Review Information

Program Overview

Background, Goals & Definitions

Background

- Making sense of intelligence data requires rich repertoires of conceptual knowledge in order to:
  - resolve ambiguities
  - make inferences
  - draw conclusions
- Understanding how the human brain represents conceptual knowledge may lead to:
  - new analysis tools that acquire, organize and wield knowledge with unprecedented proficiency
  - novel techniques for training analysts and linguists



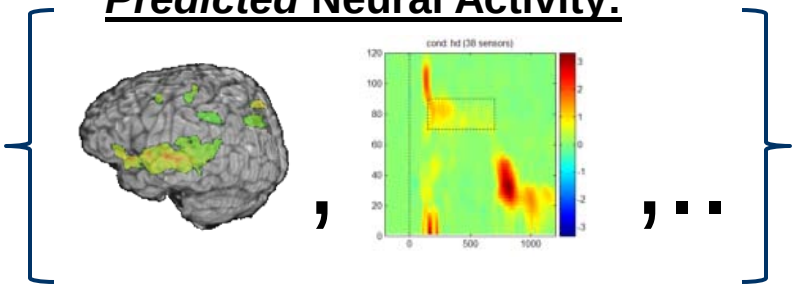
Kurland et al., MITRE Technical Paper, 2005

http://www.mitre.org/work/tech_papers/tech_papers_05/05_1365/index.html

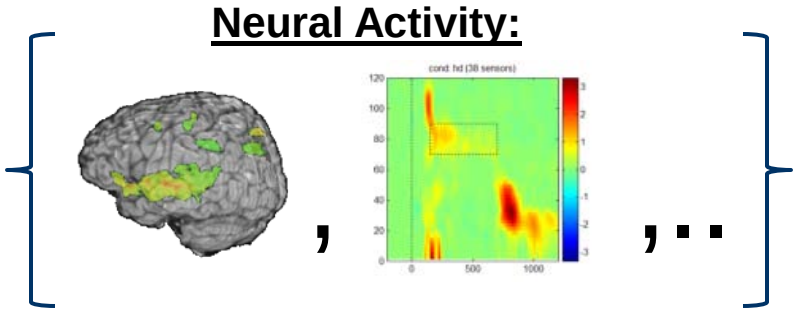
Program Goals

- *Develop* predictive **theories** that explain how the brain represents multiple types of conceptual knowledge

Concept(s):
e.g., "The doctor drove the car"



- Rigorously *assess* these theories based (in part) on their ability to **interpret** concepts from patterns of neural activity



Interpreted Concept(s):
e.g., "Doctor, drive, car, road, hospital…"

Definitions

- ***Conceptual knowledge,*** a.k.a. semantic knowledge*,* refers to knowledge of entities and their properties, and of relationships among them.
  - Conceptual knowledge is distinct from episodic memory and procedural knowledge.
  - Types of concepts of interest in KRNS include:
    - **Objects** (both animate and inanimate); e.g., “animal,” “crowd”
    - **Simple actions***;* e.g., “kick,” “laugh,” “fall”
    - **Spatial and temporal settings**; e.g., “office,” “plaza,” “winter”
    - **Human social roles and relationships**; e.g., “athlete,” “victim,” “friend”
    - **States, properties, conditions, and emotions**; e.g., “dry,” “red,” “damaged,” “sad”
    - **Events and activities**; e.g., “wedding,” “tornado,” “protest”

The term “concept” in KRNS refers to both *individual* concepts (typically represented as single words, e.g., “apple”) and to meaningful *combinations* of individual concepts (represented as simple* sentences, e.g., “The man ate the apple.”)

*Description of sentence structure and format will be provided in the BAA

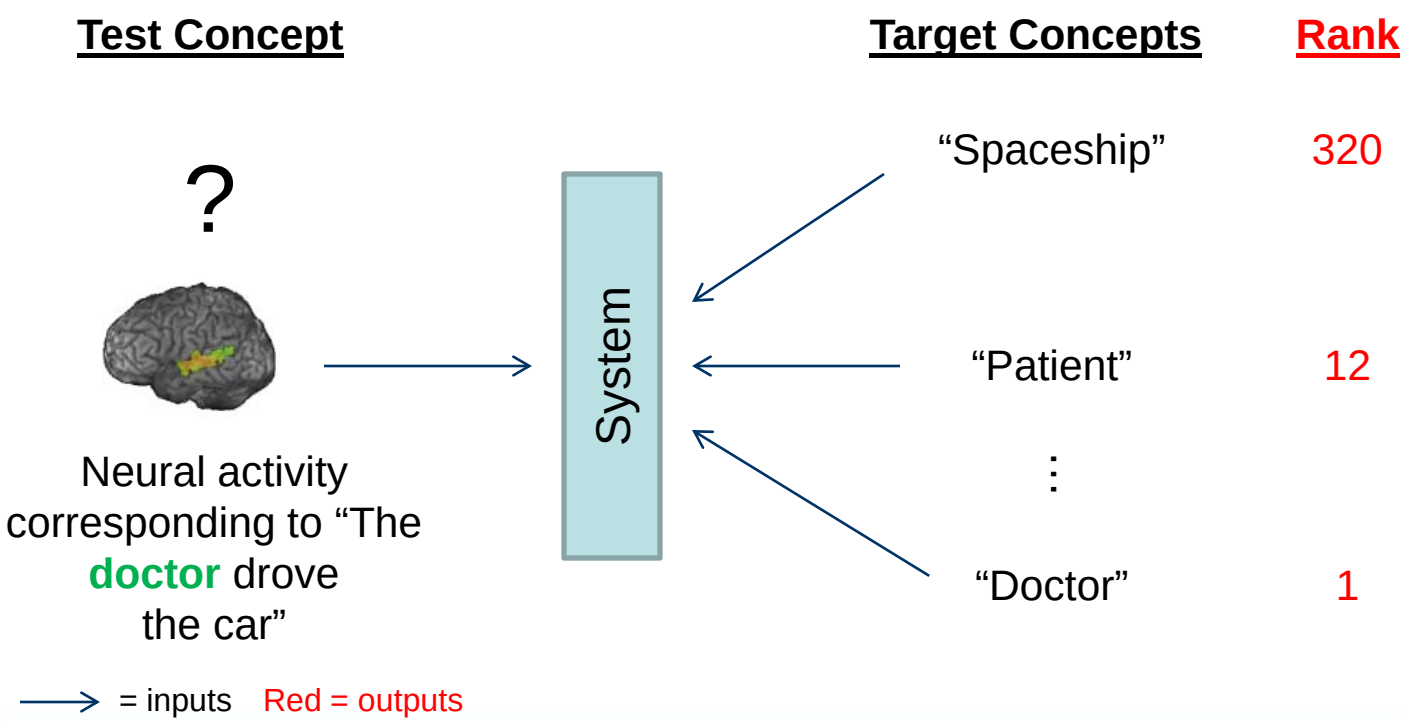
Definitions

- ***Theory*** of conceptual knowledge representation:
  - Specifies one or more *representation schemes* by which the brain represents concepts
  - Identifies which aspects of a representation scheme(s) are common/variable across individuals
  - Identifies the specific brain systems (and networks of systems) involved in the representation of specific types/aspects/features of conceptual knowledge
  - Explains the manner and extent to which the neural representation of an individual concept varies as a function of its semantic context
  - Explains how the brain represents combinations of individual concepts
  - Explains how diverse types of conceptual knowledge are represented in measureable patterns of neural activity
  - Is plausible based on existing neural, cognitive, and behavioral science research

In KRNS, the theory will be assessed both qualitatively and via the ability of a software system *based on that theory* to interpret concepts from patterns of neural activity.

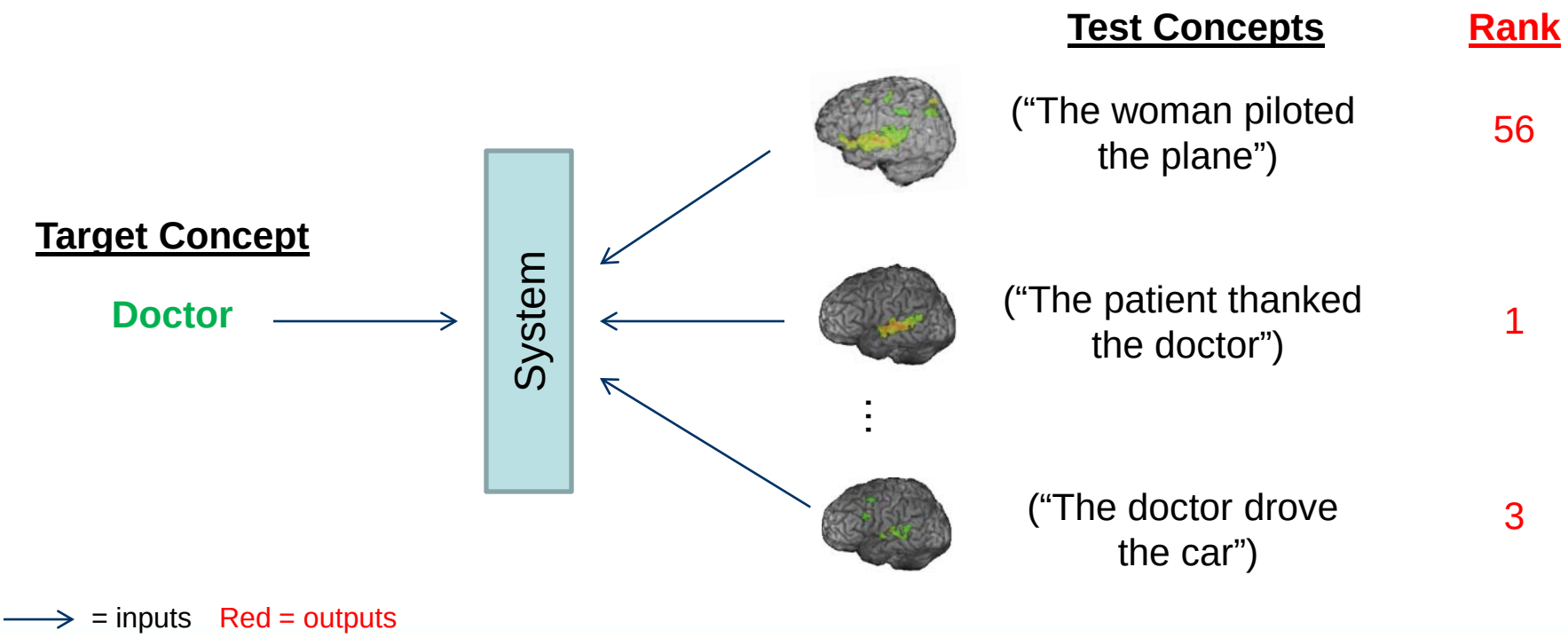
Definitions

- ***Interpretation*** refers to a system's ability to:
  1. Determine which concepts (among a set of "Target Concepts") are best represented in a specified pattern of neural activity (corresponding to an unspecified "Test Concept")



Definitions

- ***Interpretation*** refers to a system's ability to:
  1. Determine which concepts (among a set of Target Concepts) are best represented in a specified pattern of neural activity (corresponding to an unspecified Test Concept)
  2. Determine which pattern of neural activity (among a set of patterns, each corresponding to an unspecified Test Concept) best represents a specified Target Concept



Out of Scope

- Theories that focus on neural *processing/computation* as opposed to *representation*
- Theories (and aspects of theories) that are not relevant to the task of *interpretation* as defined in KRNS
- *Non*-theory-based ('black box') approaches to interpretation
- Invasive neural imaging/measurement methods of all types (e.g., microelectrodes)
- Deception detection research
- Development of new types of neuroimaging hardware
- Direct manipulation of neural activity via pharmacological or other methods (e.g., transcranial magnetic stimulation)
- Behavioral studies (NOTE: Limited behavioral studies may be permitted if they are narrow in scope and aimed at informing specific aspects of the theory.)

State of the Science

- Much research has focused on how the brain represents sensory and motor information, but less has focused on conceptual knowledge
- Studies to date have focused on a limited number of coarsely defined concept classes (e.g., faces and places), but a general predictive theory of the neural basis of conceptual knowledge remains elusive
- Various approaches for interpreting neural activity have been demonstrated:
  - “Black box” classifiers
  - Inverting cortical models of sensory encoding (e.g., of static images, or video)
  - Identifying correspondence between observed neural activity and that predicted by “encoding models” utilizing semantic “basis functions”

Desired Research

KRNS seeks innovation in the following areas:

1) Neural **theories** of conceptual knowledge representation
2) Theory-based **algorithms** (implemented as a software system) for interpreting concepts from neural activity
3) Concept **elicitation protocols**
   - Novel protocols for eliciting concepts such that interpretation accuracy is maximized. Includes:
     - Instructions given to subjects prior to and during neural activity measurements
     - Stimuli used to evoke concepts (e.g., verbal, still image, video, or combination thereof)

Desired Research (cont)

4) **Neuroimaging protocols**
   - Type of *non-invasive* neural imaging system(s), e.g., functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and parameters for data collection
   - Secondary physiological measures, e.g., heart rate variability, eye-tracking (gaze and pupil dilation)

IMPORTANT POINTS:
• KRNS performers will collect the neural activity data against which their own theories and algorithms will be assessed. Neural data will *not* be provided by the Government or the Test & Evaluation (T&E) Team.
• Although the Test/Target Concepts will be given to performers by the T&E Team in the form of text, performers will decide how these concepts will be elicited in subjects (e.g., via text, image, video, etc.) as well as how neural activity will be measured.

Program Structure

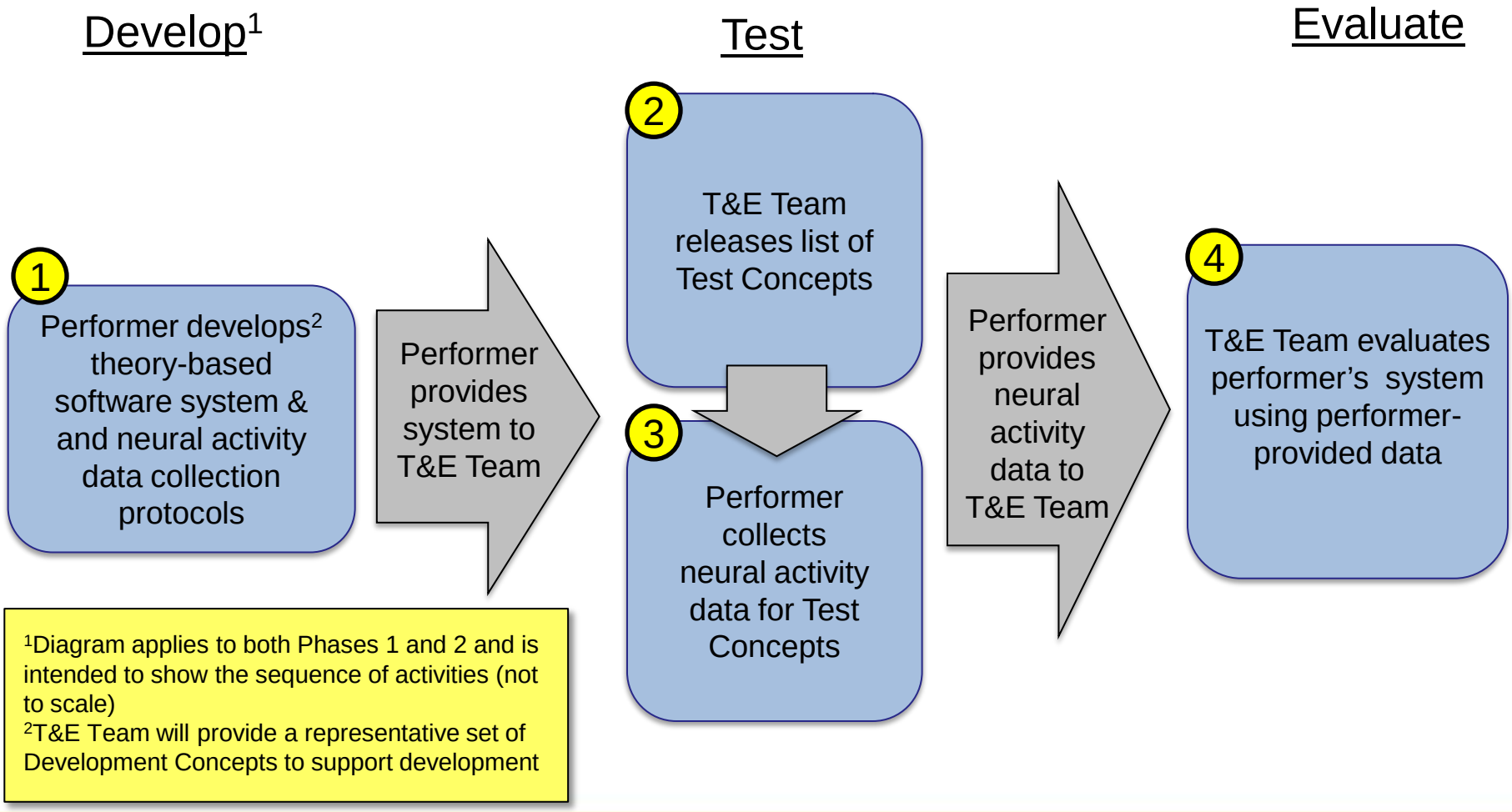
Program Phases

Two phases spanning 36 months

- **Phase 1:** 22 months
  - Phase 1a: 14-month Base Period
  - Phase 1b: 8-month Option Period
  - **Focus:** *Single* concepts in *variable* contexts
- **Phase 2:** 14 months (single 14-month Option Period)
  - **Focus:** *Multiple* concepts in *specific* contexts

The KRNS BAA will solicit proposals for *both* Phases 1 & 2

General outline of activities



[1]Diagram applies to both Phases 1 and 2 and is intended to show the sequence of activities (not to scale)
[2]T&E Team will provide a representative set of Development Concepts to support development

Phase 1

- **Key Question:** How and to what extent does the brain's representation of a *single* concept vary as a function of the *context* in which it is embedded?

Examples of Test Concepts containing the Target Concept "doctor"

"The **doctor** removed the tumor"

"The patient thanked the **doctor**"

"The **doctor** was wealthy"

"The **doctor** scrubbed the floor"

Example dimensions of contextual variation:
- Semantic congruence/incongruence
- Properties emphasized (e.g., wealth)
- Role of Target Concept in sentence (e.g., agent vs recipient)

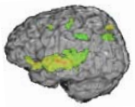
- Theory must specify major dimensions of context-related variability of neural activation patterns
- Success will be assessed in part on system's ability to identify individual Target Concepts embedded in variable contexts (Test Concepts)

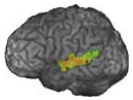
Phase 2

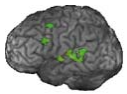
- **Key Question:** How does the brain represent combinations of *multiple* concepts in *specific* contexts?

In other words, if:

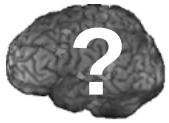
Multiple individual concepts

"nurse" = 

"worked" = 

"night" = 

Then:

"The **nurse** **worked** at **night**" = 

specific context

- Theory must specify how neural representations of composite concepts (e.g., simple sentences) are composed from the representations of individual concepts (e.g., single words)
- Success will be assessed in part on system's ability to match a specified Target Concept (sentence) with its corresponding pattern of neural activity

Program Structure Summary

	Phase 1	Phase 2
	22 Months	14 Months
Focus	*Single* concepts in *variable* contexts	*Multiple* concepts in a *specific* context
Target Concepts	**Single words** (e.g., “Doctor,” “Scare,” “Game,” “Aggressive”)	**Simple sentences**[1] (e.g., “The boy gave the carpenter a hammer.”)
Test Concepts	**Simple sentences (containing Target Concepts in variable contexts)** “The hospital hired the doctor.” “The doctor drove the car.” “The patient coughed.” “The thunderstorm scared the child.” “The dog was aggressive.” “The team won the basketball game.” “The boat left the harbor.” “The leaves are red.”	**Simple sentences (including combinations of Target Concepts)** “The boy gave the carpenter a hammer.” “The carpenter held a screwdriver.” “The man climbed the ladder.” “The plumber fixed the faucet.” “The boy played checkers.” “The boat left the harbor.” “The wagon is in the shed.”

[1]Simple sentence = one subject, one predicate (verb phrase). Each simple sentence will contain at least two concept types and may contain three or more concepts. Additional descriptions of sentence structure will be provided in the BAA.

Program Milestones & Metrics

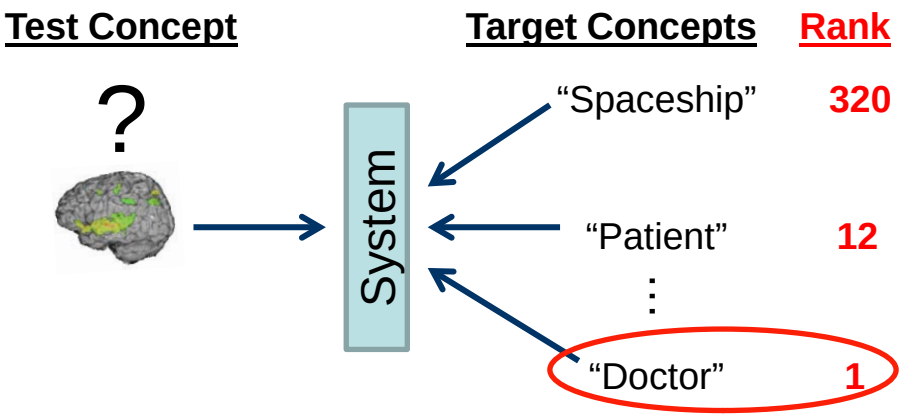
Two Types of Metrics: Quantitative and Qualitative

- Both seek to assess quality and robustness of performers' theories
  - *Quantitative*: Assessment of theory based on (theory-derived) system's ability to *interpret* conceptual content from non-invasively measured neural activity
  - *Qualitative:* Assessment of theory by independent Panel of Experts

Quantitative Metrics: Interpretation

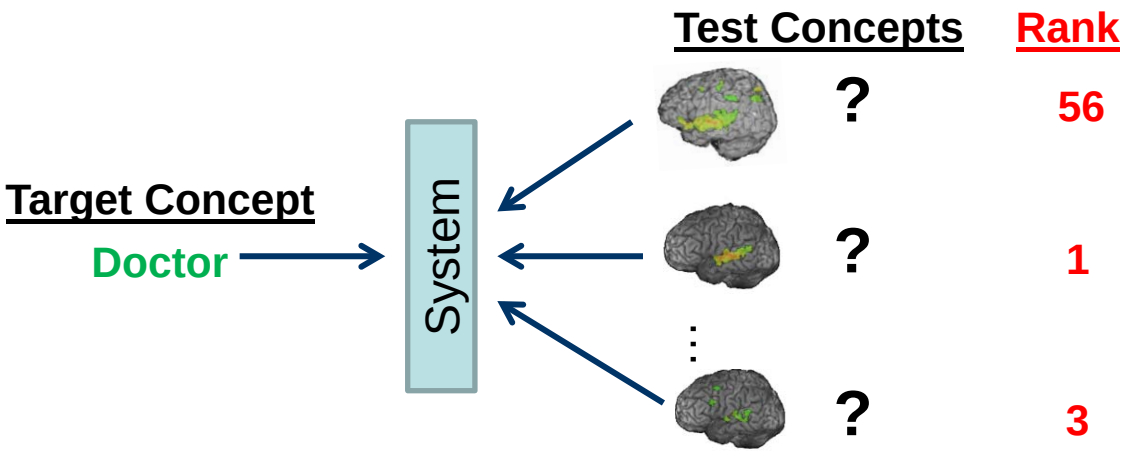
- Two interpretation tasks, each with its own metric
- Tasks/metrics are similar for each Phase, with exception that Target Concepts are words in Phase 1 and sentences in Phase 2
- The following slides illustrate tasks/metrics for Phase 1 specifically, but the same ideas apply to Phase 2

Quantitative Metrics: Interpretation



- **Task 1/ Metric 1:** For a specified pattern of neural activity (corresponding to an unknown *Test Concept*), rank a set of *Target Concepts* based on the degree to which each is reflected in the neural activity pattern
  - For each Test Concept (~500), a score will be assigned based on the rank of the correct Target Concept(s) in the sorted list
  - Note that for Phase 1, each Test Concept (sentence) contains multiple Target Concepts (words)

Quantitative Metrics: Interpretation



- **Task 2 / Metric 2:** For a specified *Target Concept*, rank a set of neural activity patterns (each corresponding to an unspecified Test Concept) based on the degree to which each pattern represents the Target Concept
  - For each Target Concept, a "semantic ranking" score will be assigned based on the correspondence between the ranked list of neural activation patterns and the "ground truth" ranked list of Test Concepts (ranked according to their semantic similarity to the Target Concept)
  - Scoring will utilize a metric that takes into account *graded* semantic similarity, such as Normalized Discounted Cumulative Gain[1] (NDCG)

[1]Jarvelin and Kekalainen, 2002

Defining Ground Truth Semantic Similarity for Metric # 2

- Semantic similarity rankings will be determined via behavioral experiments designed and conducted by the T&E Team
- In order to validate that the T&E-provided data set is a suitable proxy for KRNS subjects, each performer will obtain a limited set of behavioral rankings directly from each subject to be compared with the T&E data
  - The T&E Team will provide performers with the tools and protocols necessary for collecting these behavioral data during the course of the Program
- Performer-provided behavioral data may itself be used for evaluations depending on the scope and quality of the data sets

Exact procedures for collecting ground truth semantic similarity data will be determined in pre-program T&E pilot studies.

Summary of Quantitative Metrics

Metric	Phase 1a	Phase 1b	Phase 2
	Month 14	**Month 22**	**Month 36**
1) % rank of (correct) Target Concept(s) in sorted list	Best Four Subjects ≥ 1 Target Concept (per Test Concept) top 20th percentile All Subjects N/A	Best Four Subjects Mean of all Target Concepts (per Test Concept) top 10th percentile All Subjects ≥ 1 Target Concept (per Test Concept) top 20th percentile	Best Four Subjects All Target Concepts top 5th percentile All Subjects Mean of all concepts top 10th percentile
2) semantic ranking (algorithm > baseline*) *Methods for measuring baseline are TBD	Best Four Subjects $p \leq 0.05$ All Subjects N/A	Best Four Subjects $p \leq 0.01$ All Subjects $p \leq 0.05$	Best Four Subjects $p \leq 0.005$ All Subjects $p \leq 0.01$
* The metrics above will be computed using various quantities of neural data, up to the limits specified below			
Cumulative neuroimaging time (fMRI, EEG, etc.) per Test Concept per subject	≤ 120 s	≤ TBD	≤ TBD

Qualitative Metric: Theory

- Theories will be evaluated based on the following criteria:
  - Overall scientific value
    - Does theory significantly advance our understanding of the neural basis of conceptual knowledge representation?
    - Can theory generate novel, testable predictions?
  - Predictive accuracy
    - Comparison of predicted vs actual neural activity for Test & Target Concepts
  - Breadth
    - Does the theory address at minimum the concept types of interest in KRNS?
  - Neural plausibility
    - Is theory plausible with respect to existing literature and with respect to performer's own data?
  - Cognitive and behavioral plausibility

Independent Validation

- In each Phase, the T&E team will validate each performer's protocols and results via independent replication
- Exact duplication of results is unlikely due to differences in subject population and imaging hardware; however, substantial deviation from reported results will result in a detailed audit of protocols and data
- Replication process will be repeated at approximately six-month intervals with same subject cohort to assess longitudinal stability of interpretation algorithms

Cross-subject Evaluation of Algorithms

- Although cross-subject robustness is not a program metric *per se*, performers' theories must address the commonality/variability of representation schemes across individuals. Thus performers will be required to test and report on the cross-subject performance of their interpretation algorithms
- It is assumed that cross-subject performance may depend on multiple factors such as:
  - culture; gender; age; neuro-anatomical differences
- In order to assess the degree to which results generalize across a diverse population, performers must describe plans to recruit a diverse subject pool

Important IRB-related Waypoints

- In order for the Program to remain on schedule, it is *critical* that *all* IRB approvals (including Government approvals) be obtained **within six months** following contract award
- IRB approvals from performers' own institutions must be obtained **within two months** post award
- No IARPA funding can be used towards human subjects research until *all* approvals are granted
- To get a head start, all proposers are required to submit a complete draft IRB protocol as part of their proposal

Questions?

Award Information

Award Plan

- Three-year Program starting Q2 FY2013
  - Phase 1a – Base Period – 14 months
  - Phase 1b – Option Period – 8 months
  - Phase 2 – Option Period – 14 months
- Phase 1a performance determines participation in Phase 1b. Phase 1b performance determines participation in Phase 2.
- Multiple awards anticipated, depending upon
  - Quality of the proposals received
  - Availability of funds

Eligibility Information

Eligible Applicants

- Collaborative efforts/teaming strongly encouraged
  - Content, communications, networking, and team formation are the responsibility of proposers
- Foreign organizations and/or individuals may participate
  - Must comply with Non-Disclosure Agreements, Security Regulations, Export Control Laws, etc., as appropriate

Ineligible Organizations

- Other Government Agencies, Federally Funded Research and Development Centers (FFRDCs), University Affiliated Research Centers (UARCs), and any organizations that have a special relationship with the Government, including access to privileged and/or proprietary information, or access to Government equipment or real property, are not eligible to submit proposals under this BAA or participate as team members under proposals submitted by eligible entities

Organizational Conflict of Interest

- If a prospective offeror, or any of its proposed subcontractor teammates, believes that a potential conflict of interest exists or may exist (whether organizational or otherwise), the offeror should promptly raise the issue with IARPA and submit a waiver request by e-mail to the mailbox address for this BAA at dni-iarpa-baa-12-05@ugov.gov. A potential conflict of interest includes but is not limited to any instance where an offeror, or any of its proposed subcontractor teammates, is providing either scientific, engineering and technical assistance (SETA) or technical consultation to IARPA. In all cases, the offeror shall identify the contract under which the SETA or consultant support is being provided. Without a waiver from the IARPA Director, neither an offeror, nor its proposed subcontractor teammates, can simultaneously provide SETA support or technical consultation to IARPA and compete or perform as a Performer under this solicitation.

Application Review Information

Evaluation Criteria

Evaluation criteria in descending order of importance are:

- Overall Scientific and Technical Merit
- Effectiveness of Proposed Work Plan
- Relevance to IARPA Mission and KRNS Program Goals
- Relevant Experience and Expertise
- Cost Realism

Relevant Experience & Expertise

Successful teams must be multidisciplinary, with a variety of scientific and technical skills, such as:

- Cognitive neuroscience
- Cognitive science / Psychology
- Computational neuroscience
- Machine learning
- Functional neuroimaging
- Linguistics

Publication

- Publication of results of the research project in appropriate professional journals is encouraged as an important method of recording and reporting scientific information
- One courtesy copy of all papers and/or presentations to be presented in any public forum must be submitted to the IARPA Program Manager at least two calendar weeks prior to submission for publication
- Following publication, final copies of published papers and presentations must be submitted to the IARPA Program Manager and Contracting Officer's Representative

Point of Contact

Dr. Brandon (Brad) Minnery

Program Manager

IARPA, Incisive Analysis Office

Office of the Director of National Intelligence
Intelligence Advanced Research Projects Activity
Washington, DC 20511

Email: dni-iarpa-baa-12-05@ugov.gov

(include IARPA-BAA-12-05 in the Subject Line)

Website: www.iarpa.gov

Thank You!
Any Final Questions?