Overview

1. The 3 triplets
2. World models
3. Mind models
4. Signs & Language
5. Conclusion
My Background

- 5 year diploma (MEng) in EECS, unfortunately mainly EE. [AUTH]
- BSc in Math Sciences (+music) [british OU]
- MS in EE/Control [UCLA]
- HiScore – Face recognition [ITI]
- MIT Media Lab courses:
  - Patrec, Deb’s computational semantics, Cognitive Science core, Cognitive Architectures, Human Intelligence Enterprise
  - [i.e.: Never had any formal linguistics courses!]
- MIT Media Lab research: Ripley’s Mental Model, GSM’s, Resolver…
1. The three triplets
The three triplets

- Areas:
  - Main, Technical, Contextual... >2K pages

- Papers
  - Minimal GSM design for “Token Test”
  - Extending HAL (& sensory-motor UG)
  - Cross-species intention recognition

- The Big Picture
Triplet 1: Areas

- MAIN: Language and Meaning
  - Information & Meaning
  - Computational Linguistics
  - Speech Acts & Dialog

- TECH: Computer Vision and Action Reps
  - Vision (active & 3D shape / motion)
  - Action Generation
  - Action Recognition

- CTXT: Representational forms & reasoning
  - Sensible vs. Physical Reality
  - Representations: Categorization, Marr etc.
  - Reasoning: Mental Models, diagram. reason., naïve physics
Triplet 2: Papers

□ MAIN:
- A near-minimal grounded situation model design for a conversational robot that can pass the “Token Test”
  [Also AAAI-05 MCAI Paper: Grounded Situation Models for Robots: Bridging Language, Perception, and Action]

□ TECH:
- Pathways towards studying human action and the proposal of the “Human Activity Language”

□ CTXT:
- Constraints for cross-species intention recognition and feeling / desire communication.
  [Also long unofficial draft: Initial exploration of the problem of communication among substantially different beings]
Triplet 3: The Big Picture

- Almost …The Semiotic Triangle

- MIND (concepts)
  - Acting/Sensing
  - Producing/Detecting

- WORLD (objects)
  - Referring

- SIGN (words)

Models of worlds, minds, languages…
Realism, Mentalism
World: contains signs & minds
Mind: affects/affected by both, contains reps of both
Signs: Directly about mind, indirectly about world [intentional, not natural]
2. World Models
Models of the World

- The thing vs. description. Different Knowers (completeness, observer/actor, scientist)
- Informational lattice – “granularity” [nesting in Dretske]
- Missing knowledge, Possible Alternatives, Probabilistic views
- Static versus Dynamical views
- Physical law (static/dynamic), predict what given what?, determinism, reversibility
- Plato’s “chora”: deriving spacetime (indexicals) & dimensions
- Sensible vs. Physical Reality [Hayek, Planck] – observer-independence
- Individuation: objects, events [BC Smith], affordances of objects [Gibson]
- Properties: Apparent / Objective [BC Smith’s “stabilisation”]
- Quantization: What are the right categories? [Roche, Horvitz]

SOME MODELS:
- Newtonian Mechanics
- Modern Physics
- Model-theoretic semantics
- Possible Worlds [Appelt]

Possible Worlds:
In all the worlds that P, also Q. If ~P, either Q or ~Q
Living Beings in the World

- Self-preserving, replicating Unities
- Teleology of Life & Shared Goal Hierarchy
- Physical law + evolutionary continuity:
  Similar bodies / functions in similar habitats, complex structures

EFFECT ON MODELS OF WORLD:

- Process vs. Teleology
- Agents vs. Objects, Animacy
- Analytic / Nomic / Conventional Constraints
- Dichotomies: Physical/Mental, Body/Soul, Brains and Minds…
- From societies to Plato’s All-inclusive being w Anima Mundi
3. Mind Models
Models of Minds

- **Purpose** of Minds:
  - Ultimate goal: Serve Teleology of Life / Designer. Act accordingly!

- **Interface** to World:
  - Sensing / Acting: [models for recognition/generation – Technical Area]

- **Internal Representation**: [Hayek, Laird, Dretske]
  - None in very simple [ATP]
  - Why? Classification, Dealing with space-time detachment [BC Smith]
    Causal-Predictive cycle [Roy]. Ultimate reason: driving actions
  - Situation Models (versus imagery and propositions)
  - Handling uncertainty, multiple levels of specificity (topo), rep. the past
  - Problem: flexible self-modifying reps, paradigm shifts [Kuhn, Aloim.]

- **Action** Selection:
  - All the way from reflexive to partial planning and reflective
    (spatiotemporal horizon increases too)

- **Prediction**: [includes Reasoning – Ctxt area]
  - Why? Drive action selection towards goals:
    NextState=f(CurrentState, Chosen Action) (possibly NO ACTION)
    *Maximize Utility* of NextState towards Goal.
Some Existing Mind Models

- **Simplistic:**
  - believe(Agt, Prop), want(Agt, Prop), or dynamical systems

- **SOAR:**
  - Production system, symbolic, also cogn. model. [Newell / Rosenbloom]

- **EM-1:**
  - Frames-based, reflective, multi-level mental critics [P. Singh]

- **Anigrafs:**
  - Multi-agent, topology and action selection-centric [W. Richards]

- **Semiotic Schemas:**
  - Component schemas, focus on language, not modelling goals [D. Roy]

- **GSM-based Agent Models:**
  - Situation-model centric, sets of processes/reps, 3-layer [N. Mavridis]

Very different purposes, focuses, and levels of detail! [WR MM wshop]
GSM-Based Agent Models

**Representations:**
R1) Present Situation model: known world (not only physical necessarily)
R2) Internal state (goals, affect)
R3) Past states/situations (includes events)
R4) Future predictions
R5) Action list

**Processes:**
P1) Sensor-to-SituationModel process
P2) SituationModel-to-InternalState proc.
P3) Storage/Maintenance of Past process (includes event recognition)
P4) Future Prediction Generation process
P5) Action Selection Process (SituationModel-to-action table or planner)

Also: further breakdown! And more arrows…
GSM-Based Agent Models

Sensor-to-Situation Model Process (P1)
Example: Color Property

P1.1) Sensor
[Deb’s sensor projection]
For example, vision, proprioception etc.
Let’s say color camera

P1.2) Individuator
[~senseRegion!=senseRegionLocation]
Finds object-relevant information in all sensory streams
Visual Segmenter with region-object resolution

P1.3) Property estimator
[“Measure color” transformer]
Estimates property of Object
Color estimator: (110,12, 21)

P1.4) Categorizer
[“colorCat” categorizer]
Categorizes property
Trained probabilistic classifier: (Class 3 = “red”)

L2 Continuous Layer = [“Color” Analog Belief], L3 Categorical Layer = [“COLOR” Categorical Belief]
Another application: CLASSIFY POSSIBLE DIFFERENCES ACROSS WORLDVIEWS, try to align!
A Hierarchy of Minds

Level-0: Reflexive (here-and-now stimulus-response)

Level-1: Reflexive w internal state

Level-2: Conditionable Reflexive

Level-3: Fixed planners

Level-4: Flexible planners (w models of others – “proxies”)

Main Axis:

Action selection, temporal horizon, models of others
The World within the Mind

- Situation Models: [Johnson-Laird, Zwaan]
  Language-to-model, Not senses-to-model.
  Main dimensions: space, time, agency,

- Grounded Situation Models: [Mavridis, Roy]
  Specify connection to senses
  Can handle levels of specificity & confidence
  Effectively bridge Perception, Action & Language
  Multimodal, Uniform speech/motor act handling
  Also handle the past through Moments/Events

First conversational robot to: [AAAI-05 paper]
1) Handle inform speech acts and blend w senses
2) Talk about the past
3) Verbalize uncertainty
Soon: First robot to pass the “Token Test!”
Example Contents of GSM

At time = 137 (present)

- Agent #1 (“I”)
  - Physical:
    - Compound Object with 5 objects (body parts)
    - Object #1: (“head”)
      - Property #1: (“position”)
        - Layer #1 (Stochastic): Circular Gauss. Mix G1\{m=(1.5,1.89,2), s=2\}, G2{…}
        - Layer #2 (Continuous): (1.6, 1.85, 2.05)
        - Layer #3 (Categorical): #7 (“At the center”)
    - Property #2: (“pose”)
      ...
  - Object #2: (“neck”)
    ...
  - Interface:
    - Viewpoint: R eye center (1.4, 1.89, 2.1), viewdir (-14deg, 5deg, 0deg)
    - Mover: Head Motor: attachment point (1.5, 1.89, 1.8), rot axis: (…)
    ...
  - Mental:
    - Goals:
      - Estimated GSM: (what Robot believes Robot’s GSM contents are)
        ...

- Agent #1 (“I”)
  - Physical:…
  - Interface:…
  - Mental:
    - …Estimated GSM: (what Robot believes human’s GSM contents are)
GSM-Based Agent Models

Ripley’s GSM Modular Implementation Architecture
The GSM in operation
Deb’s Semiotic Schemas


Fig. 12. Schema for a tangible (touchable, graspable, moveable, visible) object such as a cup.
Whitman’s Anigrafs

1 organism = many agents. Focus on action selection (CondorcetTally). Very neat graph formalism, enables finding concrete conditions. Scales all the way up to societies. Higher-level view.

“Mating game”

Notice “proxies”

Also: Julian Jaynes and “hallucination”
4. Signs & Language
Signs and Language

Communication-related disciplines:
- Engineering: info-theory, signals, RF, cryptology
- Physics: transmission of waves etc.
- Humanities: linguistics, psychology, philosophy
- Also: animal training, journalism, performing arts and more!

The “why”, the “what” and the “how”
- Two shared anchors enable communication:
  - External reality & Teleology of life
  - Why (purpose): should serve life
  - What (content): mental entities, caused by external reality
  - How (medium, method): through external reality
How do intentional signs arise?

- Evolution of “Explicitly Communicative” Actions
- Deb’s classification: Natural, Intentional, Indexical
- A predator-prey toy model: Early warning for flee! (“guard” service)
  - 2 stimuli: Hear predator, Hear shout (radii), 2 actions: Flee, Shout
  - What is the why, the what and the how here? Inform/Request? “Intention” exists?
  - Population diversification & Lies/Cheating.
Basic Information Theory

Shannon 1940’s
Focus on “quantity”, not content!

Suppose events \{x_k\} with \(P(x_k)\)
Self-Info of particular \(x_k\):
\[
I(x_k) = \log \frac{1}{P(x_k)}
\]

Now suppose observable signs \{y_i\}
Info of \(y_i\) about \(x_k\)
\[
I(x_k; y_i) = \log \frac{P(x_k | y_i)}{P(x_k)}
\]

Extends to encoding with many chars:
\[
I(x_k; y_i z_j) = I(x_k; z_j) + I(x_k; y_i | z_j)
\]

But then very quickly…
Ensemble averaging enters the picture ☹

Focused on average info of source,
not info of particular message!

Avg. Mutual Infos:
\[
I(X; y_i) = \sum_x P(x | y_i)I(x; y_i)
\]
\[
I(X; Y) = \sum_y P(y)I(X; y) \geq 0
\]

Noise/Equivocation Equation:
\[
I(X; Y) = H(X) - H(X | Y) = H(Y) - H(Y | X)
\]
Where:
\(I(X;Y)\) = Avg Rxed Info about source
\(H(X)\) = Avg Txed Info
\(H(X|Y)\) = Info lost due to equivocation
\(H(Y)\) = Avg Rxed Info (source + noise)
\(H(Y|X)\) = Noise Part of Rxed info

Then, theory extends to:
Channel capacity
Optimal Average code length & code
Sequences and Markovian sources
And much more…
Dretske’s View

- Information pre-exists; meaning is what is made out of it by organisms
- Q: How to connect \{Info\}(cogsc/CS) w \{evidence, belief, knowledge\}(\Phi)?
- Proposes:
  - Supplement Shannon to get Semantic theory of Info! Get “rigid” defs of \Phi terms.
  - Focus on “surprisal”, i.e. self-info of particular messages
  - Then Comm Theory: gives us an “upper bound” on what info a signal might carry about source (not receiver-dependent, but objective)

- Definitions:
  - Info content: A sign r carries the info that (s is F) ⇔
    ⇔ P(s=F | r and k) = 1 but P(s=F | k only) < 1 [k=receiver’s existing knowledge]
  - Nesting: The info that (t is G) is nested in (s is F) ⇔
    ⇔ (s is F) carries the info (t is G) [analytical/nomic]
  - Knowledge: K knows that (s is F) ⇔
    ⇔ K’s belief that (s is F) is caused by the info that (s is F) [ONLY P=1! Not less]
  - Sensory/Cognitive significance: (s is F) is in digital form if it carries no additional info about s, while analog form if (s is F) is one of the internally nested cycles of the signal.

- However, no real “semantic value of information” definition attempted…

One could propose:
   Expected Utility of agent’s modified behavior after reception of sign
Gorin’s Auto Lang Acquisition

Widely deployed, real-world systems that learn and adapt!
Gets input (speech + other), makes action, user provides feedback

Five Clear Principles:
1) Purpose of language = convey meaning, i.e. lang. acquisition = decoding meaning
2) Language acquisition through interaction with environment, which provides feedback
3) Meaning grounded in device’s interaction with environment
   meaning = {action or internal state change} OR {associations to I/O periphery}
   Uses info-theoretic nets using mutual information, map:(word/phrase) to (action)
   Extend with environmental input: map: (lang, environment) to (action)

   **Semantic Distance** between two words: KL-dist of \(<P(c_k|v_1)>,<P(c_k|v_2)>\)
   **Salience** of word: semantic distance from null word (i.e. word with zero I/O assoc.)

   Finally, a “Semantic Information Quantity” definition! (not value, though)

   But – for a VERY SIMPLE mind… (conditionable stimulus-response level-0, at most 2)

4) Specialized, not homogeneous NN’s (need innate structure, reflects environment)
   Product networks, semantic/sensory primitive subnets

5) Language acquisition proceeds in developmental stages
B&P’s Situations

- Basics:
  - Individuals (might have parts), Properties, Relations & Locations: All are “uniformities”—they appear again and again across situations
  - These build situations; some of the abstract situations, are actual (correspond to reality). Abstract situations = locations + situation types, i.e. at I: barks, Molly, yes; Situation types are partial; they leave things open.
  - Course of events = function from locations to situation types, i.e.:
    - In e: at I: barks, Molly; yes. at I’: shouts at, Mr. Levine, Molly; yes
    - Situations are abstract set-theory types; semanticists use them to classify real situations.

- The Relation Theory of Meaning:
  - “Smoke means fire” (nomic), “Kissing means touching” (analytical-part of), “Ringing bell means class is over” (conventional; potential mis-information), “COOKIE means cookie” (Hearing “COOKIE” is a uniformity across situations, so is the existence in a situation including an object having the property cookie).
  - Event meaning (particular event) vs. event-type meaning: “efficiency” of lang.
  - Meaning of declarative = relation between utterances and described situations. Can invert!
    - Interpretation of declarative sent. (on a specific occasion) = described situation
    - \( u[I \ AM \ SITTING]e \leftrightarrow \exists \ location \ l, \ individual \ a : in \ u : at \ l : speaks, \ a; \ yes. \ in \ e : at \ l : sits, \ a; \ yes. \)
  - Extend to: d,c[BITING MOLLY]s,e
    - Discourse Situation (d): speaker, addressee, discourse-location, expression (utt)
    - Speaker Connections (c): function from referring words to referents
    - Setting (s): elements provided by fragment that can be used through anaphora (p.129)
Computational Linguistics

[Jurafsky/Martin]: A reference textbook

**WORDS** The fundamental building blocks

**SYNTAX** The skeleton – the formal relationships between words

**SEMANTICS** The literal meaning of utterances, and how it is composed

**PRAGMATICS** The relation of language and context-of-use
Words

- Phonology (phonetics etc.)
- Morphology (inflections etc. root/lemma)
- Regular Expressions
- N-grams and smoothing
Syntax

Word classes
- Dionysius Thrax 8; now up to 146!
- Def: co-occurrence (“distributional props”) / similar inflections [not semantic ;-)]
- Closed class / open class

POS-tagging
- Rule-based / stochastic

Grammars and Parsing
- Constituency: 1 UNIT as they appear in similar envir., but individual words cannot repl. whole
- CFG Grammars and rules: A -> B C
- Agreement: Does you have a flight? Subcategorization: permissible verb complements!
- Parsers: top-down, bottom-up, Earley, probabilistic…
- Human Parsing? Modularist / Anti-modularist (whether semant./pragmat. intervenes)

Language and Complexity
- Chomsky Hierarchy: Regular (A -> xB, A -> x) [FSA], CFG’s (no ε) (A -> γ) [PSG], CSG (A -> γ / α-β ctxt) [Tree- adjoining G], Type 0 (α -> β) ATN’s… Is NL ctxt-free? (Swiss G)
Semantics

Representing Meaning

- Meaning = state of affairs of situation
- Some ways: FOPC, Semantic Nets, Frames... everything translatable to FOPC
- “I have a car”
- Semantic analysis: “We take linguistic inputs and construct meaning reps that are made up of the same kind of stuff that is used to rep this kind of everyday commonsense knowledge of the world”...

Pretty ad-hoc predicates, sensory deprivation, no way to handle analog / uncertainty...

- Inference / plausible reasoning with FOPC, Basic reps in FOPC

Meaning structure of NL

- Fundamental Predicate-Argument structure: “I <want> Italian Food” [pred<want>, 2 args]
- Semantic roles (subject, object etc.), Selectional restrictions (<want> <wanted restricted>)
- Principle of Compositionality: mean(A,B) = f(mean(A), mean(B), syntax)
- Substitution rule: if two fragments have same meaning, they should be substitutable:
  But then: “I believe <John smokes Marlboro>” & “John is the King of France” →
  “I believe the King of France smokes Marlboro” (Not always true!!!)

Just says: “Believe etc. are words with world creating ability” → Mental Models needed!
Semantics

Semantic Analysis
- Syntax-driven, semantic grammars, info extraction. Apps: translation, dialogue etc.

  Currently, simple info-extraction in Ripley

Lexical Semantics
- Homonymy ("bank"), Polysemy ("sperm bank"), Synonymy, Hyponymy
- Internal structure of words: Enough relations among lexems. Let's see meaning components!
  a) Thematic roles: agent, theme, experiencer, force, instrument [not standard!]
  b) Selectional restrictions: lexemes place constraints on accompanying lexemes
     "To diagonalize <a matrix> vs <a shoe?>"
  c) Primitive decomposition: everything decomposable to a set of (words) or (primitives)
  d) Semantic fields: Holistic relationship among set of words from single domain, f.e. travel

  NOT Convincing! Circularities, sensory deprivation…

  Basic lexical semantics should be grounded, and also some rules of composition learned!

  Ripley: can be seen as procedural semant. \( \text{Mean}(w) = f(\text{other w in utt, GSM state}) \) [~Gorin]

Word sense disambiguation & info retrieval
- WSD: Selectional Restriction, Robust (learning)
Pragmatics

- **Discourse** [see Grosz, too!]
  - Reference resol. (“he”, “my friend”). Discourse ctxt vs. situational ctxt. Discourse model (MM)
  - Types of ref.expr. (def/indef/pron/demonstr…). Synt/Sem constraints and preferences. Algor.
  - Text Coherence (anaphora as cohesive device). Coherence relations (sentence to next sent)
  - Hierarchical discourse structures, story grammars etc.

  *We have basic object / temporal reference resol in Ripley (extend!). No discourse as such.*

- **Dialogue**
  - Monologue to dialogue, turn taking, Transition Relevance Places, Adjacency Pair(Q/A,Greet)
  - Establishing Common Ground (Stalnaker). Speakers need not ground=ack speaker’s Utt, or else to make clear there was problem in reaching common ground (=things mutually believed)
    - 5 ways: Continued attention, contrib, acknowl, demonstr, display etc.
  - Conversational Implicature: Meaning relies on more than literal mean! S expects H to infer….
    - Gricean Maxims: {Quantity (inform), Quality(confidence), Relevance, Manner(length,ambig)}

There’s three non-stops daily => There’s only three {quantity}
When do you wanna travel? Wanna be there for meeting from 12\textsuperscript{th} to 15\textsuperscript{th} {relevance}

*We only have simple adjacency pairs in Ripley, & establish common ground through ack.*
Speech Acts

- **[Austin ‘62]**
  - Notice performatives: “I name this horse Bucephalos” (change the world! Rem. Zwaan…)  
  - Locutionary act: the act of uttering a sentence with meaning (saying: “You can’t do that!”)  
  - Illocutionary act: asking/answering/promising etc. (protesting)  
  - Perlocutionary act: production of feel/thought/act on receiver (stopping receiver from doing it)

- **[Searle ‘69]**
  - Classification of speech acts at illocutionary level: ([Cohen] thinks redundant, but useful)  
  - Assertives (INFORM): suggest, put forward, conclude, Directives (REQUEST): ask, request  
  - Commisives: promise, plan, vow, Expressives: thank, apologize, Declarations: I name, I bet  
  - Necessary and sufficient conditions for carrying out speech act:  
    Normal I/O, prep. content, preparatory, essential, force [see Cohen notes!]

- **DAMSL**
  - Enriched catalogue of dialogue acts/conversational moves, two levels:  
    forward-looking function: like searle’s classif (f.e. propose)  
    backward-looking function: relationship to previous utt by other speaker (f.e. reject)
Speech act interp / planning

- Interpretation:
  - At first glance, simple [Imperative, declarative syntax]. Indirect Speech Acts! [“Can you give”]
  - Solutions: simplistic, plan-inferential, cue idiom-based model (words, prosody, conver. struct)
    Ripley currently uses simplistic solution with word-spotting (“who”, “there”)

- Planning:
  - [Cohen, Perrault ’79] A plan-based theory of speech acts (only Req, Inform, Question)
    Diff. Appelt: No discussion of speech-act realization w words! Both fixed goal.
    A competence theory (does not show how to create plans; only discusses ready plans)
    Belief/Goal language, Plan Operators (STRIPS-like CANDO.PR, PLAN.PR)
  - [Cohen/Levesque] A wider theory of communication from a theory of rational interaction
  - Main result: Illoc acts need not be primitive, and need not be recognized (but useful!)
    Ripley responds to requests and provides ack, does not initiate
Intention

What is Intention?
- Why bother having them? Planning takes time, and we need coordination of actions.
- Plans need to be partial, hierarchical, coordinated, & the world changes in mean time!
- DEF: Prior intentions are elements of partial hierarchical plans, and as such pose problems for solution and filter out options for deliberation. Need to be stable, change only when unantic. prob

Analyzing intention in utterances
- [Allen, Perrault ‘80] A model of cooperative behavior for an NLU system.
  Agts attempt to recognize plans of others and use this to decide what to do.
  Infer plan of others / Detect obstacles (physical/informational) / Help w obstacles
  Shown apps: Provide more info than asked. Interpret short sentence frags. Indirect SA’s
  Includes: action/plan/goal reps, plan construction, plan inference, obstacle detection etc.

Intentions, attention and the structure of discourse
  Linguistic Structure: segment of Utterance
  Intentional Structure: Purposes and their relations (DP, DiscSegmPurp, domin/satisf.prec. rel)
  Attentional State: Salient obj, prop, relations from situation model (focus space struct stack)
  Apps: interruptions (true, flashbacks, digressions),
NL Generation

- NL Generation:
  - Levels: Content selection, Lexical selection, Sentence selection, Discourse struct [compare w action generation for extended HAL!]
  - Architecture: (CommunicativeGoal + Knowledge Base) drive DiscoursePlanner
  - It produces DiscourseSpecification which SurfaceRealizer turns to NL output
  - [Appelt]: KAMP (Knowledge and Modalities Planner) – Combined speech/physical act
    4 levels: 1. Illoc acts (inform/request), Surface sp. acts (assert/command/acts), Concept Activation (describe/point), utterance acts. Opportunistic planning. Example screwdriver.

- Dialogue Managers:
  - ELIZA, SHRDLU, telephone interfaces, Ripley
  - Single initiative: FSA, with states = questions system user asks
  - Template/Production rule architecture: more convenient for user to use longer sentences (slots to be filled)
  - BDI-based with logic-based planning: True purposeful interaction!
  - Evaluation methodology: {user satisfaction, task completion cost, task completion success}
    Ripley in essence uses template-style archit, although just adjacency pairs
HAL: A Sensory-Motor Language

Human Activity Language
[Guerra-Filho, Fermuller, Aloimonos 2005]

11 actions from multiple-view motion capture system.
Translated to human stick model, w measured angles.
Clever dual-purpose visuo-motor primitives:

move(joint#, |Δφ|, sign(dφ/dt), sign(d2φ/dt2), Δt)

Phonological rules are modeled as a finite automaton.
Morphology: phonemes combined to form strings
representing activity and generate a grammar.
Visuo-motor subprograms = recurrent lexical units.
Syntax: visuo-motor sentence
subject is active joints (noun)
modified by a posture (adjective).
Verbal phrase is rep of human activity (verb)
and timing coordination among different joints (adverb).

Fig. 4: Knee angle derivatives during jog activity.

Fig. 12: A visuo-motor sentence for the jog activity.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Modifier</th>
<th>Postface</th>
<th>Modifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. Rep</td>
<td>1  4</td>
<td>B00 00</td>
<td>B00 00</td>
</tr>
<tr>
<td>R. Knee</td>
<td>1  7</td>
<td>B1 Y3 B2 G3 R4</td>
<td>Y6 B7 B6 B5</td>
</tr>
<tr>
<td>R. Anik</td>
<td>1  8</td>
<td>B6 B7 B8 Y6 B9</td>
<td>B2 G2 B1 Y3 B4</td>
</tr>
<tr>
<td>L. Rep</td>
<td>1  3</td>
<td>B2 Y4 04 Y0 Y0 Y0 Y0 Y0</td>
<td>14343343342343432</td>
</tr>
<tr>
<td>L. Knee</td>
<td>1  4</td>
<td>Y4 B0 04 G6 R4</td>
<td>Y3 B3 G2 R4</td>
</tr>
<tr>
<td>L. Anik</td>
<td>1  5</td>
<td>B1 Y3 B0 Y3 Y3 Y3</td>
<td>G2 R4 Y8 Y8 Y8</td>
</tr>
</tbody>
</table>
HAL - Extensions

- [from my Technical Area exam] – Intro w 50 more frequent verbs of English
- Expanded language metaphor at levels higher than syntax: discourse, paragraphs, as well as dialogues

On the way:
- Introduced two higher sensory-motor control levels (for the action primitives: “takewalkstep(size,duration)”/“maketurnstep(angle, duration)” and “goto(object)”).
- Discussed recognition and generation of these new primitives, and commented upon the role of egomotion estimation and visual servoing.
- Saw how “goto(object)” can be seen as a context-dependent paragraph generator, and introduced “discourse coherence”, “discourse purpose”.

Then:
- Saw how the introduction of interactions with other objects and humans creates “dialogue”, and how dancing is related to polyphonic music. We observed that the conditional-execution requirements of x-schemas align well with dialogue generators.
- Briefly commented upon the semantics of the visuo-motor language, and discussed what and where we might expect to find in visuo-motor universal grammar (shape and dynamics of body; gravity, envir. struct, purposeful tasks).
- Discussed the relation between the spoken and visuo-motor languages, and contemplated upon the possibility of a unified theory encompassing both.
Discovering the world

- Discovering your sensors/actors and simple models of space [Pierce/Kuipers]
- Discovering your body
- Discovering objects [Drescher]
- Discovering affordances [Stoytchev]
- The limits of empirical induction [Nicod]
- What Geometry can be empirically learnt [Nicod]

An open question: What is innate, what is not?

A killer app: intelligence in abstract information spaces, with geometries very different than physical space
5. Conclusion
Conclusion

- This looks like Ithaca; here we seem to end our journey of Worlds, Minds and Signs (recall!)
  And we have visited so many cities and ports of ideas and methods…
  Although, not so many people deal with the AI side of language grounding!
  [cogmac, berkeley, reiter, regier … see Deb TICS paper!]

- A final triplet of recommendations:
  - **Models of Minds:**
    You need better defined models of minds in order to get more concrete and usable models of meanings
  - **Goals of Life:**
    Have to be taken into account more seriously, as they provide the main top-down constraint
  - **Unified Language:**
    Visuo-motor language theory should be unified with natural language for a better overall view –
    - both the parts and the whole will gain!
I would like to close this presentation with a speculation of ancient origin, and a modern wish:

If for a single instant we were able to gain consciousness of the whole universe, then thousands of changes - actions and interactions would become noticeable, and would appear to the ears of our soul as a cosmic concert, of which our actions would be a small part. Every note carries with it expectations, effects, consequences of the past, predictions of the future, and remote interactions with melodies played at seemingly great spatiotemporal distances. May those of us who can use their limited knowledge towards acting in order to help what might sometimes seem to be a temporary discord resolve again into cosmic harmony.

And thus the final return to Ithaca seems as distant as ever – or anyway, maybe a little less distant…