QuEST
Simulation based representation tenets

March 2020

Cap

Air Force Research Laboratory

QuEST – Cognitive Exoskeleton

Kabrisky Memorial Lecture 2020

10 January 2020

Dr. Steve 'Cap' Rogers
AF Sr Scientist Autonomy

Integrity ★ Service ★ Excellence

QuEST for Consciousness "What are the tenets for machine representations (artificial qualia?) that enable flexible behaviors?"

Some material from Nagel 1974

Notes by Capt Amerika from discussion with Special K

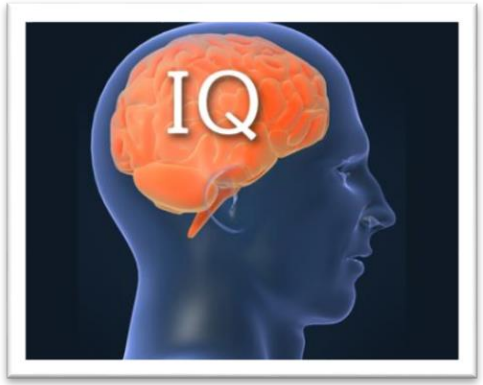
Edited April 2019 by cap

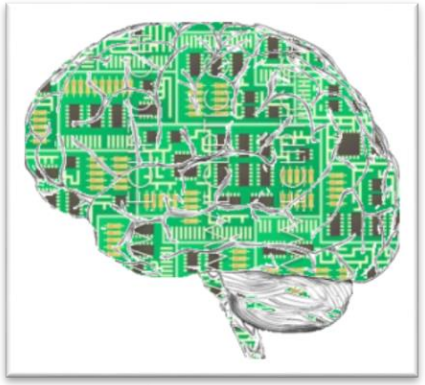
Post Ancient Mike consciousness presentation

In preparations of upcoming 'Walk with Marvin' MIT workshop

Intelligence & Artificial Intelligence

- Intelligence is the ability of an agent to gather observations, create knowledge, and appropriately apply that knowledge to accomplish tasks
- Artificial Intelligence (AI) is a machine that possesses intelligence





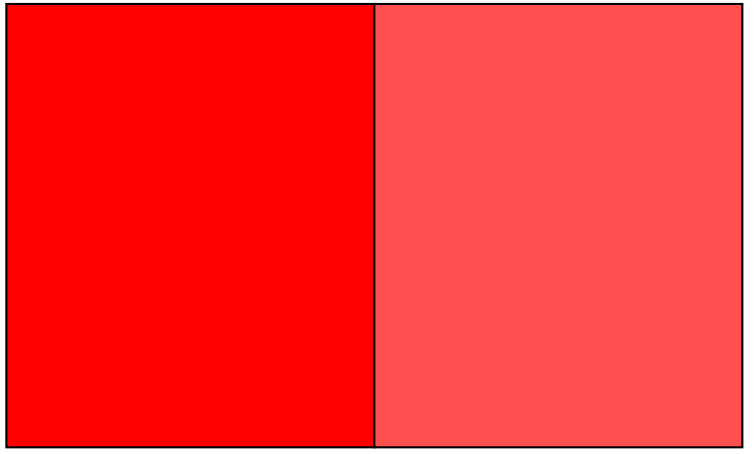
Recently heard another definition that I liked – ability to perform acceptably across a variety of tasks

Autonomous Horizons v2: https://www.airuniversity.af.edu/AUPress/Display/Article/1787830/autonomous-horizons-the-way-forward/

Examples of Conscious Representation

There is something it is like to a human to be stabbed in the hand – or see the red or pink square or hear a chord or taste a pizza!

'perception' is subjectively acceptable

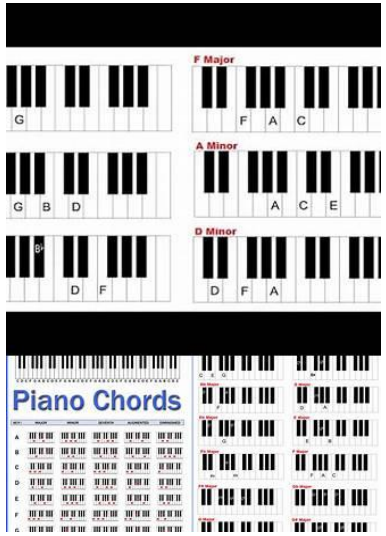


What it is like to undergo an experience?

Why do you 'feel' pain?

Individualized!

- Pheung – individualized basis **(SOUND)**
  - Chords a great example – what you hear is evoked in your mind it isn't just the notes! Headphones!



Why do you 'see' red?

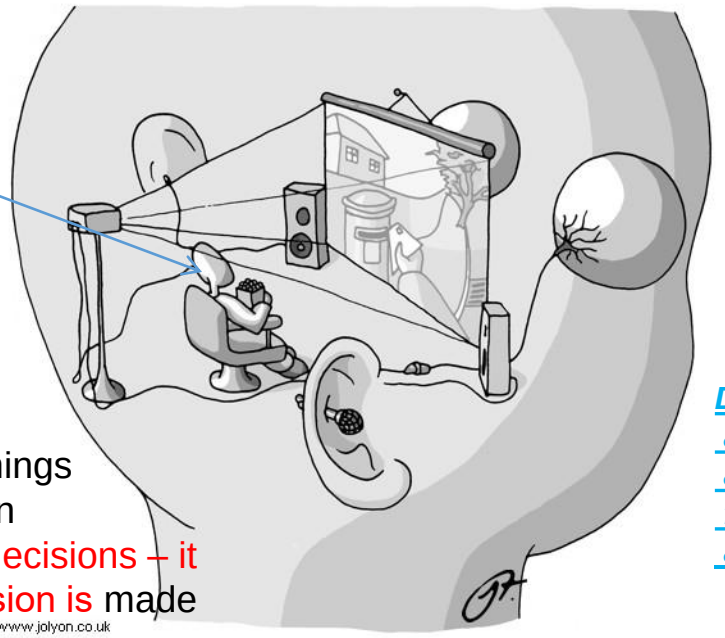
Qualia Theory of Relativity.

- ***Pizza*** doesn't have a ***taste*** until you eat it.

Illusion of Cartesian Theater – What is a Quale?
Working Memory (cognitive decoupled / simulation)

The 'subject' in the subjectively acceptable representation

Thoughts



Qualia associated with representation of the physical environment and to represent the thoughts you are having all in a stable, consistent and useful way

Introspection

Discernible aspects can be used internally to 'think' or communicated via language to an aligned agent -

Pretending to be in charge of things beyond its control – user illusion

Although claims to be making decisions – it isn't even there when the decision is made

Discernible meant to capture that there is a difference between one state and the alternative (blue versus brown) and the fact that aspects are introspectively available

For the purpose of this presentation I will ***define qualia*** as ***any discernible aspect of the illusory Cartesian theater*** = any aspect of your world model that you are aware of (meaning you know is part of, meaning you experience that aspect) as being part of that world model, the fact you can 'see' the redness of a car means that red attribute of your world model is a quale red – any sound you hear (the attributes of the sound that evoked JND aspects to them are each a quale at that moment) – any thought you have at that moment you are thinking it is in your world model as a thought in your mind so there is a quale of thought associated with it AND thoughts in fact are composed of qualia (the primitives of all thought)

Model that suggest multiple representations: Dual process agent

- Dual-Process Theories of Higher Cognition: Advancing the Debate, Perspectives on Psychological Science 8(3) 223–241 © The Author(s) 2013
  - Evans and Stanovich
- Dual Process Theories,
  - Betram Gawronski, Laura A. Creighton, in D.E. Carlson (Ed.) (2013) the Oxford Handbook of Social Cognition, pp 282-312, Oxford University Press, Ny Ny

We posit that there are lessons in how nature creates its 'conscious' representation that are key to making autonomous systems that have flexible behaviours

Meaning will be the changes in both of these processes as a result of observation being processed with knowledge

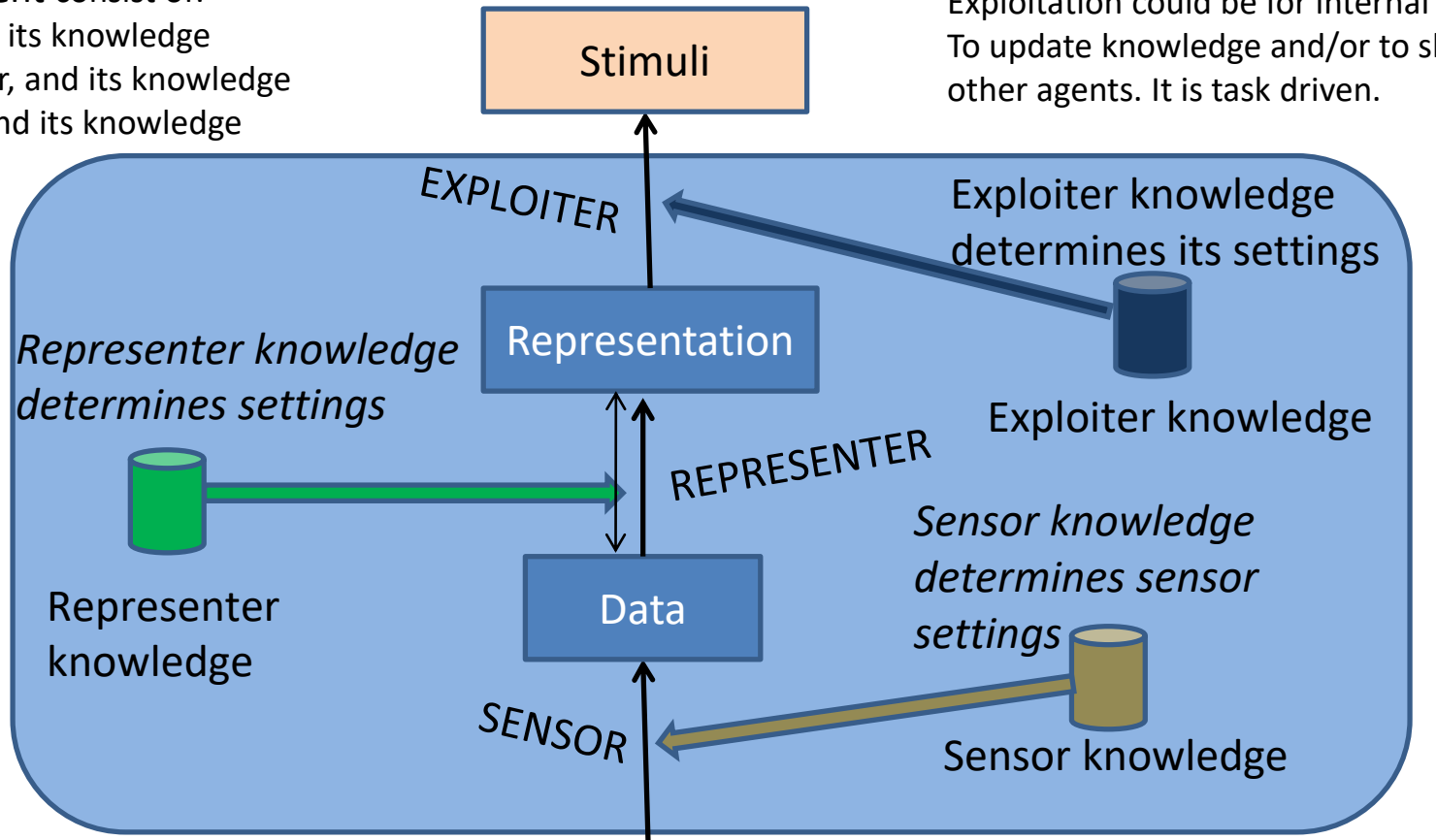
Work/interactions of Benjamin Libet initially led us here

If you model the two systems as disparate agents – do they together exemplify the flexibilities – example Peer flexibility between sys1 / sys2

Atomic Agent, Stimuli, Data, Information, Knowledge and Query

Atomic Agent consist of:
- Sensor, and its knowledge
- Representer, and its knowledge
- Exploiter, and its knowledge

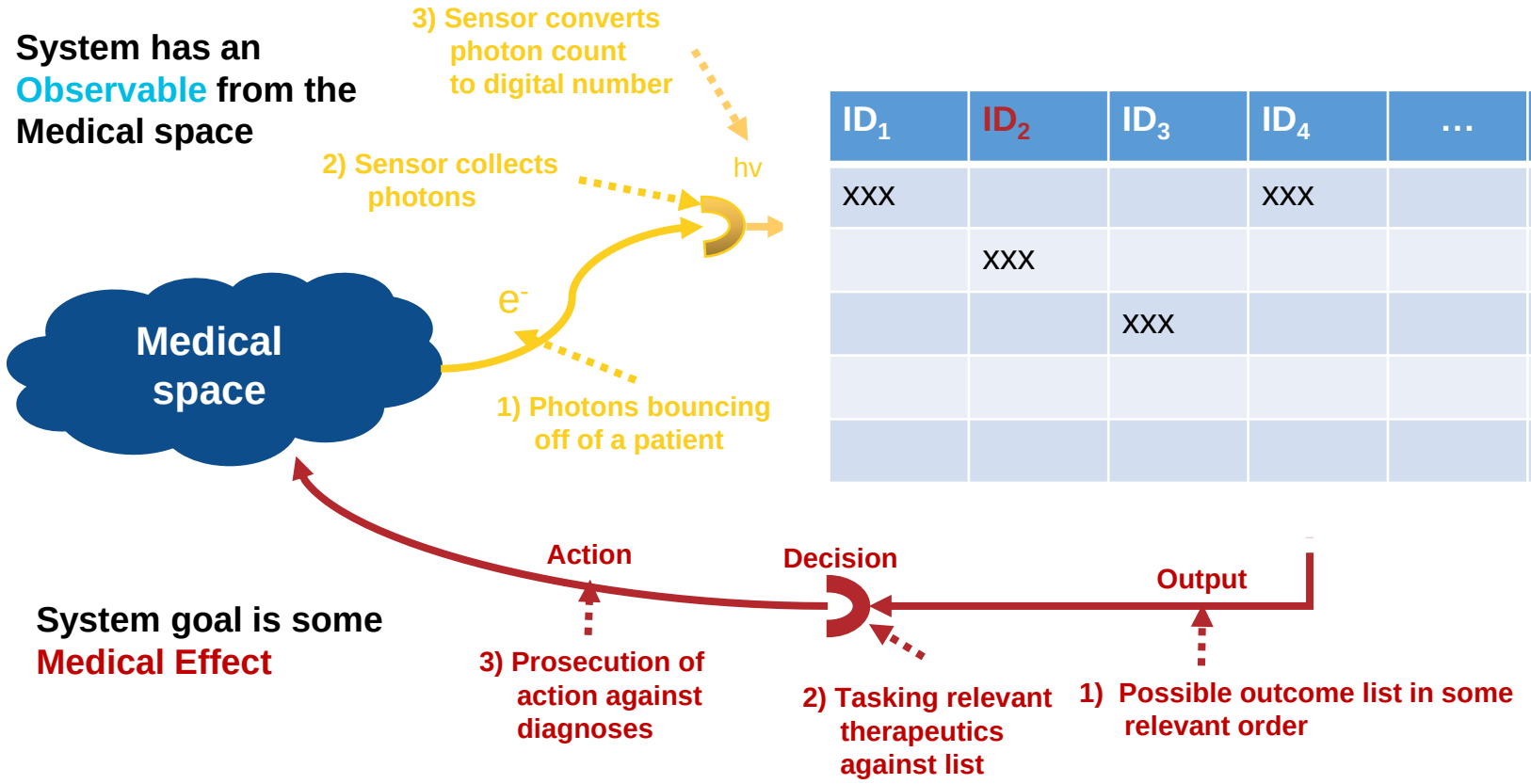
Exploitation could be for internal use
To update knowledge and/or to share with other agents. It is task driven.



Atomic Agent = (S, K_S, R, K_R, X, K_X)

A query is defined as **the act of a stimulus being provided to an agent and the agent responds.**

Example 1: Detection & Treatment



Example 2: Detection & Treatment

System has an Observable from the Medical space

3) Sensor converts photon count to digital number

2) Sensor collects photons

hv

e⁻

Medical space

1) Photons bouncing off of patient

AI System

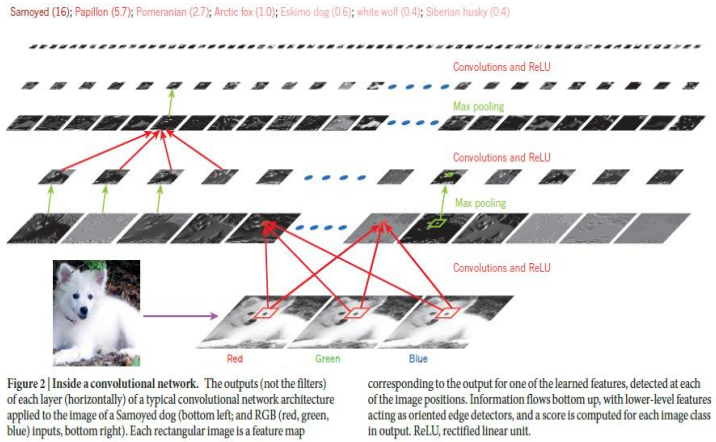


Figure 2 | Inside a convolutional network. The outputs (not the filters) of each layer (horizontally) of a typical convolutional network architecture applied to the image of a Samoyed dog (bottom left; and RGB (red, green, blue) inputs, bottom right). Each rectangular image is a feature map corresponding to the output for one of the learned features, detected at each of the image positions. Information flows bottom up, with lower-level features acting as oriented edge detectors, and a score is computed for each image class in output. ReLU, rectified linear unit.

Digital numbers processed/stored in a specific way enabling object extraction and labeling

System achieves some Medical Effect

Action

Decision

Output

3) No action taken, collateral damage avoided

2) Do nothing, object is not treatable

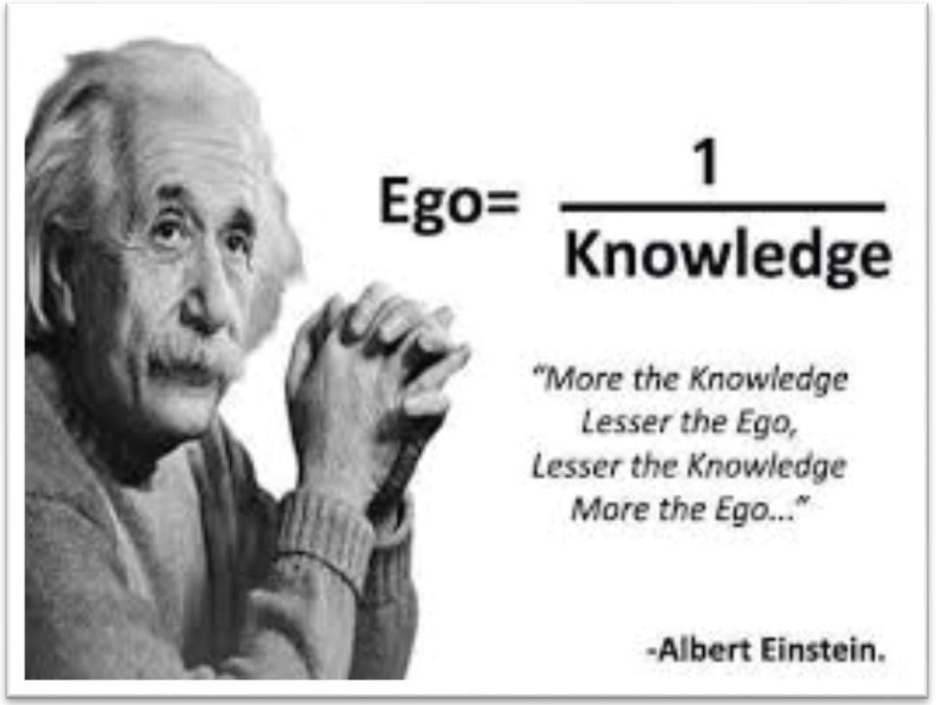
1) System Determines Disease Category ~ Dog Type = Samoyed

Representation and knowledge

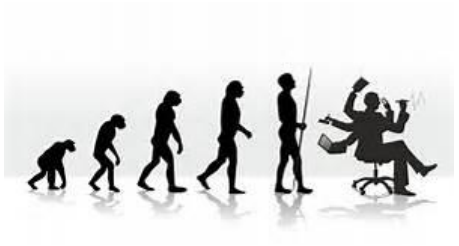
- Representation is how an agent structures its knowledge
- Knowledge is what an agent uses to generate meaning (knowledge includes the representation and the processes of how to generate that meaning using that representation)
- Example: we have experienced a probabilistic characterization of the occurrence of a particular event – that is knowledge – we decide to represent that knowledge in a PDF (probability density function) and then use a particular instantiation inside the computer based on relative frequency - maybe some parzen windows for a representation of that knowledge.
- Example2: we have experienced seeing lots of pictures with labels ‘cat’, there exists in that set of pictures what could be used by an agent as knowledge about ‘what is a cat’ – the agent has to create the knowledge it will use to decide if a picture is of a cat from that data and how it will represent that knowledge and what processes will be used to generate subsequent meaning of pictures

What is knowledge?

Knowledge is what is used to generate the meaning of the observable



Where does Knowledge come from?



Evolution

Experience

Culture

Machine Generated

Speed at which knowledge is generated

10^0 10^1 10^2 10^3

- Each source of knowledge is an order of magnitude quicker than the previous
- Most future knowledge on Earth will come from machines extracting it from the environment –machine generation of knowledge key for the future of Medicine

Pedro Domingos book: The Master Algorithm

There is more to your internal representation than what you are conscious of – Type 1

'What is done by what is called myself is, I feel, done by something greater than myself in me' – James Clerk Maxwell on his deathbed, 1879 (user illusion)

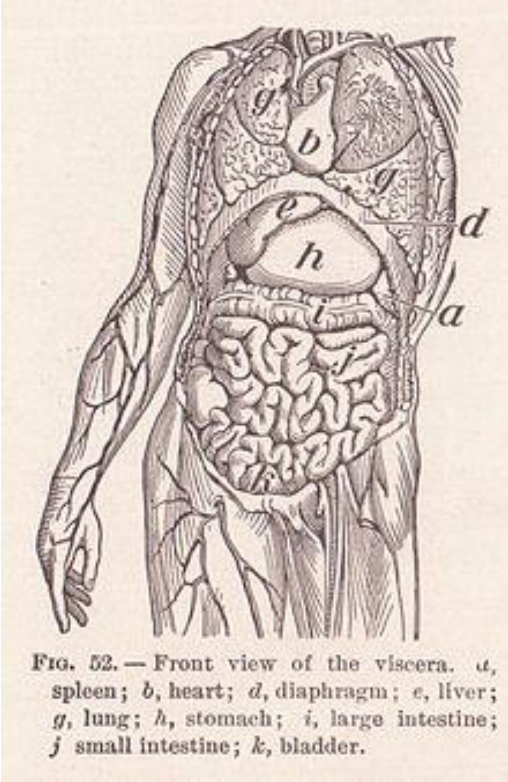


FIG. 52. — Front view of the viscera. *a*, spleen; *b*, heart; *d*, diaphragm; *e*, liver; *g*, lung; *h*, stomach; *i*, large intestine; *j* small intestine; *k*, bladder.

Defining gut (intuition or hunch) feelings:

Appears quickly in consciousness (we would say the results of the calculation gets posted to consciousness)

Whose **underlying reasons** we are not fully conscious of - Is strong enough to **act** upon

John Rollwagen, a colleague for many years, tells the story of a French scientist who visited Cray's home in Chippewa Falls. Asked what were the secrets of his success, Cray said "Well, we have elves here, and they help me". Cray subsequently showed his visitor a tunnel he had built under his house, explaining that when he reached an impasse in his computer design, he would retire to the tunnel to dig. "While I'm digging in the tunnel, the elves will often come to me with solutions to my problem", he said.

Below the level of consciousness does NOT imply lack of importance to the system's solution! Should Cognitive Engineering Design include type 1 information?

Benjamin Libet – Harvard University Press

Theory of Consciousness

- **THE ULTIMATE GOAL of a theory of consciousness is a simple and elegant set of fundamental laws, analogous to the fundamental laws of physics.**
- We provide the QuEST tenets – they are unlikely to be the right answer to this challenge.
  - **Structural Coherence (interaction to ensure stable, consistent and useful representation)**
  - **Situation based processing (situations as variables) – fundamental unit of conscious cognition (narratives)**
  - **Conscious representation of situations are done via simulation (cognitively decoupled – imagined past, present and future in the form of a cohesive narrative)**

Consciousness is **Stable, consistent and useful** ALL SOURCE **situated simulation that is structurally coherent**

Consciousness Tenets Summary

1. **Structural Coherence**
    1. Interaction – enough fidelity with reality (bits of awareness info) facilitate conscious driven interaction (situations are learned via interactions)
    2. There is a similarity measure applicable for the conscious representation – (color wheel example) – the similarity measure is between situations/qualia/chunks
2. **Situation** based processing (situations as variables) – fundamental unit of conscious cognition
    1. Links (types of links, possibly capture similarity, meaning as what links are evoked – source of exformation – entities (situations) are defined based on how they are situated that is their meaning)
    2. Gists as key part of representation–low bandwidth representation –what is situated/simulated
    3. Multimodal – integrates multiple sensors representation into common framework – part of situated
    4. One quale at a time – for any aspect of the illusory cartesean theater (example invertible illusions)
    5. Qualia Theory of Relativity – only value (meaning) is in the relationships between (dictionary)
    6. Narrative based representation–situated in time/space/multiple modalities (plausible narratives compete) Stream of consciousness is a cohesive narrative
    7. TD/BU – means to do context - rapid high level first –
    8. Types of Qualia types of situations - time as a quale, Affect as a quale, types of speech, ToM – (Evolving not static), aha and negative aha (means to know what is known and what is NOT known by the agent)
        1. Self – special type of qualia/situation (qualia self interacts with Continuity, unity, embodiment, sense of free will, reflection)
3. Conscious representation of situations are done via **simulation** (cognitively decoupled)
    1. simulation is an organized body of knowledge that *produces* specific *simulations of a situation's instances*
    2. Imagined past, imagined present, imagined future – cognitively decoupled
    3. Exformation (pattern completion inferring mechanism)
    4. Compression (infinite number of stimuli into a single quale, low bandwidth 50 bits/sec)
    5.) Ability to generate meaning / situated simulation of a new concept – the unexpected query

Other decks covered structural coherence and situated representations

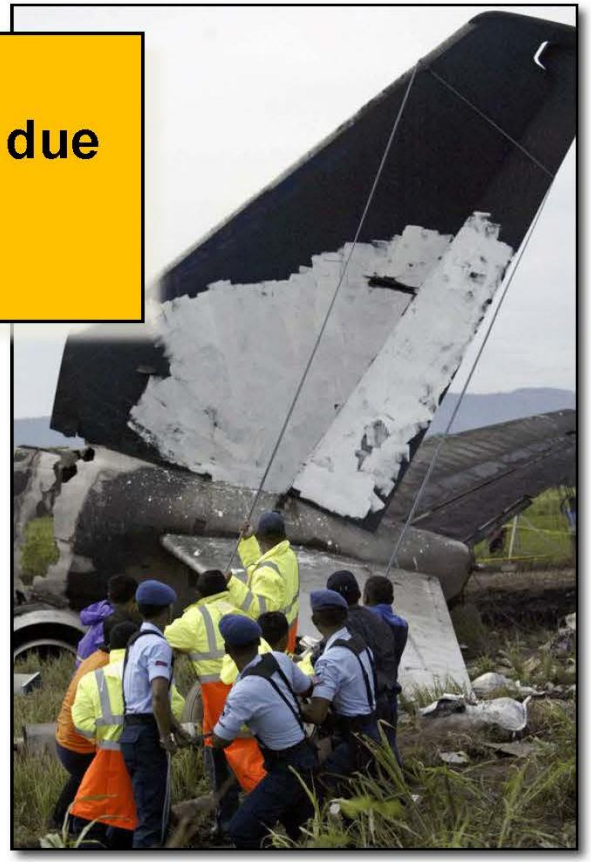
Consciousness Tenets Summary

simulation

1. **Structural Coherence**
    1. Interaction – enough fidelity with reality (bits of awareness info) facilitate conscious driven interaction (situations are learned via interactions)
    2. There is a similarity measure applicable for the conscious representation – (color wheel example) – the similarity measure is between situations/qualia/chunks
2. **Situation** based processing (situations as variables) – fundamental unit of conscious cognition
    1. Links (types of links, possibly capture similarity, meaning as what links are evoked – source of exformation – entities (situations) are defined based on how they are situated that is their meaning)
    2. Gists as key part of representation–low bandwidth representation –what is situated/simulated
    3. Multimodal – integrates multiple sensors representation into common framework – part of situated
    4. One quale at a time – for any aspect of the illusory cartesean theater (example invertible illusions)
    5. Qualia Theory of Relativity – only value (meaning) is in the relationships between (dictionary)
    6. Narrative based representation–situated in time/space/multiple modalities (plausible narratives compete) Stream of consciousness is a cohesive narrative
    7. TD/BU – means to do context - rapid high level first –
    8. Types of Qualia types of situations - time as a quale, Affect as a quale, types of speech, ToM – (Evolving not static), aha and negative aha (means to know what is known and what is NOT known by the agent)
        1. Self – special type of qualia/situation (qualia self interacts with Continuity, unity, embodiment, sense of free will, reflection)
3. Conscious representation of situations are done via **simulation** (cognitively decoupled)
    1. simulation is an organized body of knowledge that *produces* specific *simulations of a situation's instances*
    2. Imagined past, imagined present, imagined future – cognitively decoupled
    3. Exformation (pattern completion inferring mechanism)
    4. Compression (infinite number of stimuli into a single quale, low bandwidth 50 bits/sec)
    5.) Ability to generate meaning / situated simulation of a new concept – the unexpected query

One way to estimate meaning making accuracy is to analyze decision errors

As much as 88% of human error in aviation is due to problems with situation awareness



Levels of situation awareness

Projection at its Core is an Imagined Future

Note imagination is all-source, integrated, situated and simulated possibly key to getting information and understanding its meaning also

Gists and Links

Sys2, conscious, qualia as a simulation

Qualia Theory of Relativity - simulation based Sys2

Some locations in simulation that generates the Cartesean Theater devoted to perception – some to memory some to imagination each has 'tag' = feeling to distinguish

Complete with a ***pattern completion inference mechanism***

Imagined present / past / future

Memory, Perception, Imagination - simulation based Sys2

Can such an implementation satisfy the 'conscious' characteristics? Continuity, Unity, …

A good perceptual system has to go beyond the information given; **it has to 'invent' things**.

Your brain sees more than what your eye senses – **NOTE there is nothing in the definition of a situation that would require pattern completion inferencing – but is key to qualia Representation – there is nothing in the definition of a situation that requires simulation**

See Larry Barsalou – Emory

Simulation

Sounds like a case

- Consider a simulator that represents the concept of bicycle. Across encounters with different instances, *visual information about how bicycles look becomes integrated in the simulator, along with auditory information about how they sound, somatosensory information about how they feel, motor sequences for interacting with them, affective responses to experiencing them and so forth.*
- The **result is a distributed system throughout the brain's feature and association areas** that *accumulates and integrates modal content processed for the category*.
- As Barsalou (2003a) describes, **many additional simulators develop to represent properties, relations, events and mental states relevant to bicycles** (e.g. spokes, mesh, pedal, effort).

All Qualia are situated – not all the aspects mentioned here become qualia

Simulators we would say develop to represent anything you attend to in sys2 – sys1 provides stimuli for the simulation to account for – when it captures attention the simulators must generate a stable, consistent and useful simulation – that then is potentially stored for later use

Modification of ideas from Barsalou

- Across encounters with different instances, *visual information about how bicycles look becomes integrated in the simulator, along with auditory information about how they sound, somatosensory information about how they feel, motor sequences for interacting with them, affective responses to experiencing them and so forth.*
- *--- this statement suggest that the perceptual system captures what the bike looks like – I don't think that is required at all – the simulation has to incorporate characteristics of the visual sensed data and the audio and somatosensory characteristics that create a stable, consistent and useful representation that does NOT imply captures what the bike really looks like or sounds like or feels like ~ Hoffman*

****

BEHAVIORAL AND BRAIN SCIENCES (1999) **22**, 577–660
Printed in the United States of America

Perceptual symbol systems

Lawrence W. Barsalou
Department of Psychology, Emory University, Atlanta, GA 30322
barsalou@emory.edu **userwww.service.emory.edu/~barsalou/**

Notes by Capt Amerika – v2

April 2010

****

Summary Points

- Representation of symbols and their relationship to what they 'refer' to and their perceptual representation -
- **Functional conceptual system requirements**
- Importance of 'simulators'
- Importance of 'interactions' to bind concepts to referents
- **Key is to have perceptual representation the same as what is used for cognition NOT separate distinct systems**

Perception and cognition systems are the same!

Primary goal of learning is to establish simulators – once can simulate to a subjectively acceptable accuracy = adequate understanding of it!

****

Summary: Perceptual symbols defined

- *Perceptual symbols are modal and analogical.* They are modal because they are represented in the same systems as the perceptual states that produced them.
  - The **neural systems that represent color in perception, for example, also represent the colors of objects in perceptual symbols, at least to a significant extent**.
- On this view, a ***common representational system underlies perception and cognition, not independent systems***.
  - Because perceptual symbols are modal, they are also analogical. The ***structure of a perceptual symbol corresponds, at least somewhat, to the perceptual state that produced it.1***

Could it be the perceptual Sys1 representation is really modal – but once conscious it is amodal – conscious perception is amodal!

****

Summary: Amodal symbols defined

- *Whereas earlier schemes assumed that **cognitive representations** utilize perceptual representations (Fig. 1), the newer schemes assumed that cognitive and perceptual representations constitute separate systems that work according to different principles* see that the symbols in these systems are amodal and arbitrary.
- They are amodal because their **internal *structures bear no correspondence to the perceptual states that produced them***.
  - The *amodal symbols that represent the colors of objects in their absence **reside in a different neural system** from the representations of these colors during perception itself*.
  - In addition, **these two systems use *different representational schemes and operate according to different principles***

This could be the key failure of current amodal approaches – they ignore the real problem which lies in perception being cognitive! Amodal representations result in the S-Q gap!

****

I'm sure there is considerable work on perceived similarities of color – the question could be asked – are our perceptual symbols 'arbitrary' with respect to similarity of physics? The color wheel suggest they are not separate

Summary: Distance not preserved for amodal!

- As a consequence, **similarities** between amodal symbols are not related systematically to similarities between their perceptual states, which is again analogous to how **similarities between words are *not related systematically to similarities between their referents***.
- Just as the words "blue" and "green" are not necessarily more similar than the words "blue" and "red," the amodal symbols for *blue and green are not necessarily more similar than the* amodal symbols for *blue and red.2*

Is distance preserved in the qualia space – since it doesn't make sense to talk about similarities of perceived color we would suggest even though the physics might make adjacent wavelengths more similar – once you get any 'distance' from a wavelength you lose the preservation of distance in the PERCEPTION SPACE – where he points it out for the symbol space

******key deficiency in amodal is lack of ability to represent spatio-temporal knowledge***

****

Summary: Amodal systems face issues

- Symbol grounding issue – complement of transduction issue

Since we link the perceptual state to our symbols we don't have to link back to perceptual states – BUT – as in all work the mapping to entities in the world is always an issue – ***this may be why we need interaction in the grounding of symbols – why structural coherence!***

In our world – imagination = perception = memory, all are grounded by reference to the same machinery originally grounded with physical referents via interactions.

****

Summary: Use of symbols whether peceptual or amodal

- supports all of the higher ***cognitive functions, including memory, knowledge, language, and thought***
- *categorization, concepts, attention, working memory, long term memory, language, problem solving, decision making, skill, reasoning, and formal symbol manipulation*

These cognitive functions are accomplished using symbols – memory, knowledge, language and thought

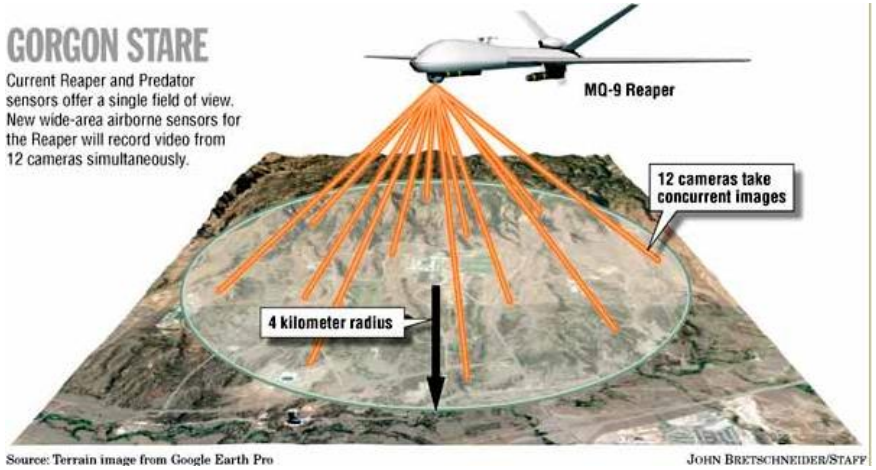
Selective attention controls memory

Summary: Focus on high level architecture

- Because this target *article focuses on the high level architecture of perceptual symbol systems*, it leaves many details unspecified.
  - The theory **does not** specify the **features of perception**, or **why attention focuses** on some features but not others.
- The theory **does not address how** the cognitive system divides the world into **categories, or how abstraction processes establish categorical knowledge**.
  - The theory ***does not*** *explain how the* ***fit between one representation and another is computed****, or how* ***constraints control the combination of concepts***.

Places where Capt Dube or Mr. Derriso can focus on breaking new ground!

Why we care Summary: Layered Sensing challenge



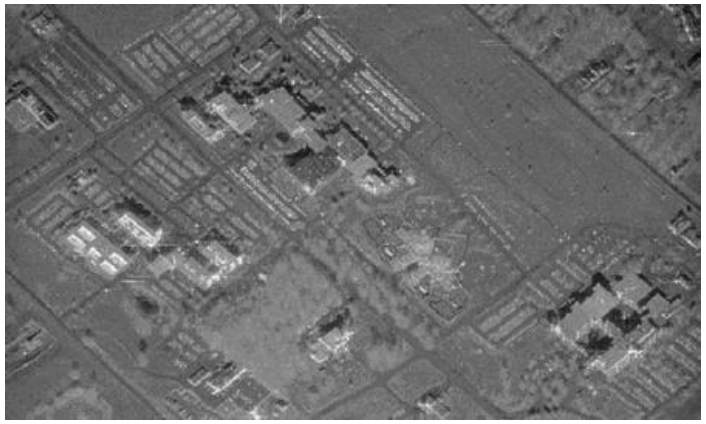
http://www.wired.com/images_blogs/dangerroom/images/2009/02/17/gorgon_stare.jpg

The system is initially expected to ***provide a main full-motion video and 12 pre-programmable sub-views***. Air Force officials have previously said that their goal is to provide up to 30 sub-views in future generations of the sensor.

Concept encoding is a wicked problem because the data is so variable and the sensors never exactly replicated – point being the specs are changing

Air Force Chief of Staff Gen. Norton Schwartz challenged the ISR agency to use its in-depth expertise to solve the problem of dealing with the deluge of data coming into ISR systems from the increasing number of sensors and ISR sorties – May 4 2010 Sensor Rally.

"***We are going to be swimming in sensors and drowning in data,***" Gen Deptula told reporters at a July 7, 2009 Pentagon briefing.



http://www.armytimes.com/xml/news/2007/10/gns_spytechnology_071002/100207gotcha_800.JPG

Processing this data ***requires "abstraction" -> data to concepts***

- current approaches process the raw sensory data
- How to map the # ***data to concepts***? **********
- How to do that in a general way without having to predefine in every detail (which results in loss of flexibility and robustness) –

- Close the Loop! Query to sensing to hypothesis to new query

a concept is equivalent to a simulator ****

Why we care Summary: Recording systems versus conceptual systems

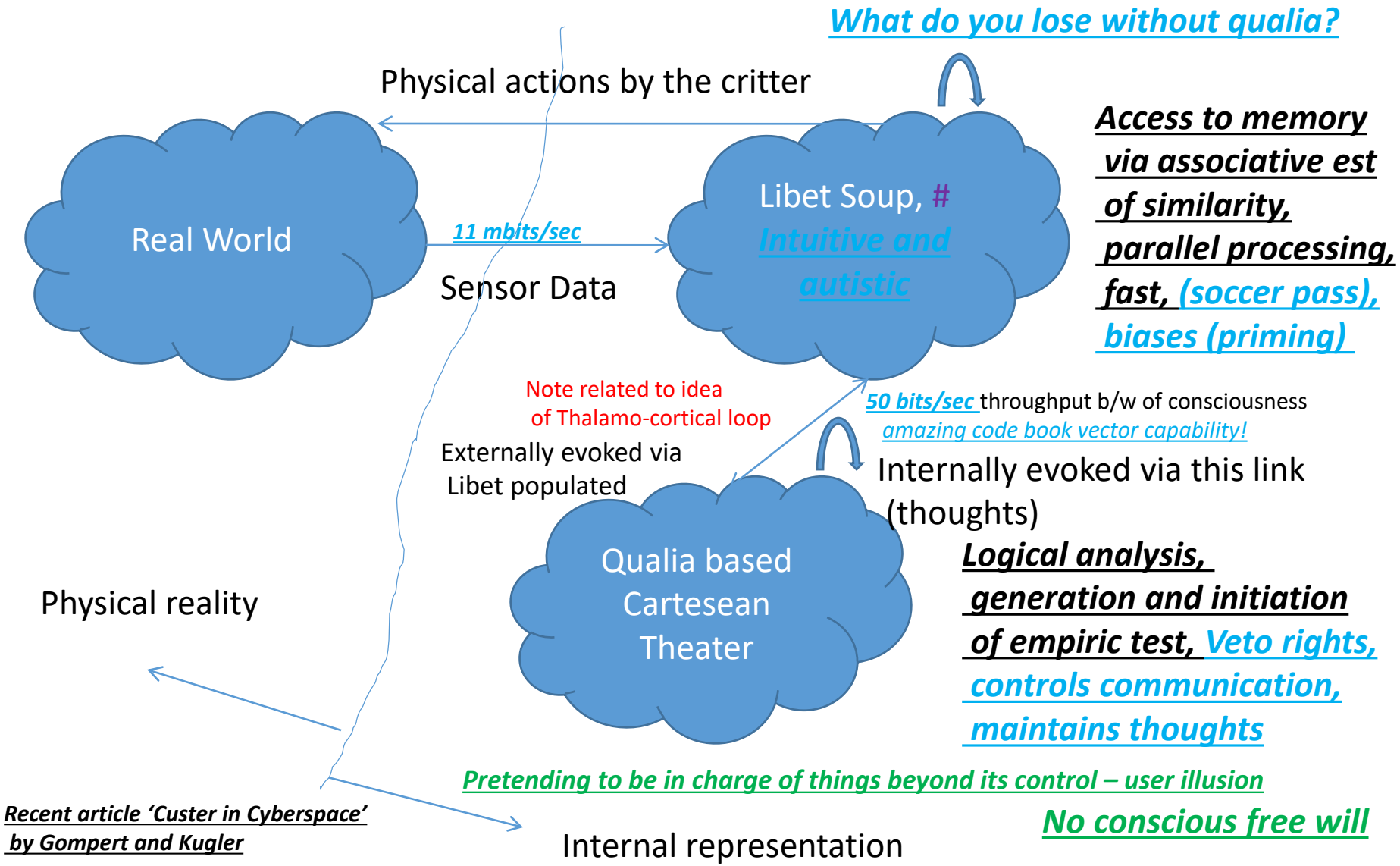
- It is widely believed that *perceptually based theories* of knowledge *do not have sufficient expressive power to implement a fully functional conceptual system*.
  - As described earlier (sect. 1.2.1), a **fully functional conceptual system represents both types and tokens, it produces categorical inferences, it combines symbols productively to produce limitless conceptual structures, it produces propositions by binding types to tokens, and it represents abstract concepts**.
- The **primary purpose of this target article is to demonstrate that perceptual symbol systems can implement these functions naturally and powerfully.**
  - *To accomplish* this, the *conceptual system* ***binds specific tokens in perception (i.e., individuals) to knowledge for general types of things in memory (i.e., concepts)***.
  - Clearly, a system that only records perceptual experience cannot construe individuals in this manner – it only **records them in the holistic context of an undifferentiated event**.

This is why we care – generating the data won't solve the problem – we need a fully functional conceptual system to solve the Gen Schwatz challenge. – see layered sensing slide

Why we care Summary: Layered Sensing challenge

****

We have proposed a key part of the concept encoding solution is the simulation level of the representation = Qualia? ********



****

Why we care Summary: Layered Sensing challenge

Theory of knowledge ****

- Solution must be a fully function concept encoding system ****
  - Key aspect of solution is 'simulators' - ***simulators implement a basic conceptual system*** *that represents types, supports categorization, and produces categorical inferences.* These **simulators further support productivity, propositions, and abstract concepts, thereby implementing a fully functional conceptual system!**

BUT concepts that are 'amodal' – like feature lists, frames, … are inherently limited – they miss the power of perceptual symbol systems – perception is the key – perception is cognitive as we said in our ooda article! The internal representation for perception HAS to be the same as for language, memory and thought to be able to process intelligently = MODAL.

Fully functional conceptual systems require:

- ***Simulators that can:***
  - *represents types (***binding a concept (type) to an individual (token) in a manner that is true or false**.),*
  - *supports categorization, and*
  - *produces categorical inferences.*

Slide ~70 has 4 additional properties of a conceptual system

These **simulators further support:**

productivity (*from integrating simulators combinatorially and recursively to produce complex simulations)***,**

propositions (*result from binding simulators to perceived individuals to represent type-token relations)***, and**

abstract concepts (grounded in complex simulations of combined physical and introspective events).

****

Summary: Categorization, categorical inferences, and affordances.

- Tracking a category successfully requires that its members be categorized correctly when they appear. Viewing concepts as simulators suggests a different way of thinking about *categorization*.
- ***Whereas many theories assume*** **that relatively static, amodal structures determine category membership (e.g., definitions, prototypes, exemplars, theories), simulators suggest a more dynamic, embodied approach: if the simulator for a category can produce a satisfactory simulation of a perceived entity, the entity belongs in the category.** If the **simulator cannot produce a satisfactory simulation, the entity is not a category member.**12

Similar to our breakthroughs in speech – HMMs all about 'simulating' for categorization

****

Summary Introspection

- Relative to sensory-motor processing in the brain, introspective processing is poorly understood.
- Functionally, **three types of introspective experience appear especially important: representational states, cognitive operations, and emotional states.**
  - **Representational states** include the representation of an entity or event in its absence, as well as construing a perceived entity as belonging to a category.
  - **Cognitive operations** include rehearsal, elaboration, search, retrieval, comparison, and transformation.
  - **Emotional states** include emotions, moods, and affects.

What sort of processing for the conscious representation?

Exactly the problem with modern sensing/exploitation approaches – recording not concept encoded processing

****

Abstract

- Prior to the twentieth century, **theories of knowledge** were inherently perceptual. Since then, *developments in logic, statistics, and programming languages have inspired amodal theories* that *rest on principles fundamentally different from those underlying perception*. In addition, perceptual approaches have become widely viewed as untenable because they are assumed to implement recording systems, not conceptual systems. A **perceptual theory of knowledge is developed here in the context of current cognitive science and neuroscience.** During perceptual experience, association areas in the brain capture bottom-up patterns of activation in sensory-motor areas.

Perceptual systems often discounted since not inherently conceptual often thought of as recording system

This paper proposes a perceptual theory of knowledge

Activation in sensory motor areas = perceptual representation = what barsalou calls perceptual experience – he does not equate to conscious experience like we do – he includes unconscious experience in the phrase perceptual experience (includes libet).

Quest is all about a Theory of Knowledge – what a system can know. The fundamental question is the language of that knowledge – does it consist of perceptual symbols or in some other 'amodal' form (like logic, statistics, …)?

This idea is very consistent with out tenets description of a quest architecture – you perceive the feedback! Top-down

****

Abstract

- Later, in a *top-down manner, association areas partially reactivate sensory-motor areas to implement perceptual symbols*. The **storage and reactivation of perceptual symbols operates at the level of perceptual components – not at the level of holistic perceptual experiences**.

***

- Through the *use of selective attention, schematic representations of perceptual components are extracted from experience and stored in memory* (e.g., **individual memories of *green, purr, hot***).

***

Perceptual components – idea is at the shape, texture, edge level = components, he suggests that is the level of perceptual symbols

****

Hierarchical Quest Architecture
'Chunking' for complex problems (not BU alone)

Infraconscious path is the hierarchy of instances and their links

Sensor outputs – most recent temporal data from which will extract relationships

- Doppler
- Micro-Doppler
- Impedence

Respiration rate is a feed forward representation temporal plot

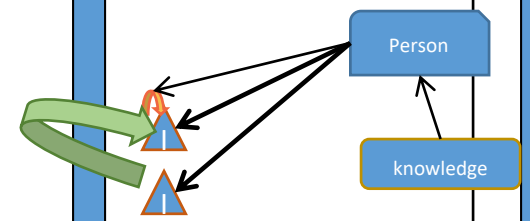
Concepts that might be for example a codebook vectorization in the feed forward path

Concepts that might be speaking or hyperventilating

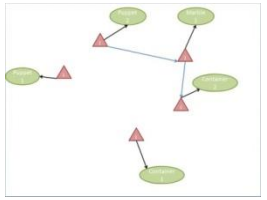
Layer 1 feed forward

Layer 2 feed forward

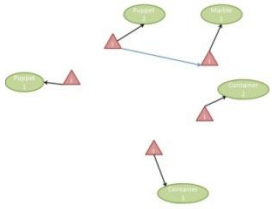
Layer 3



Instances and labels

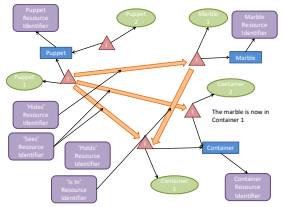


Who is speaking? Who is hyperventilating?



'Scared' connection

Modified and perceived sensed data!



Perception

Experiences "speaking"

Versus feed forward senses

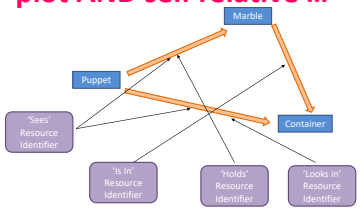
Layer 1 feed back

Layer 2 feed back

Constraints being applied at layer 1! Results in a qualiarized form of temporal plot AND self relative …

Constraints being applied at layer 2! Results in a qualiarized form of the codebook vector view

Constraints being applied at layer 3! Qualiarized view of those higher concepts (talking, sleeping, hyperventilating, …)



Example of one of Feed Back parallel paths!
As they say in HTM – reliable predictability of sequences
Feedback can be means to overcome ambiguity or noise in feed forward concept formation.

I like the approach of linking together conceptualization – whether it is for sensory data or … the use of a simulation approach is also nice – this is key to our approach (LSOC) – below discussed as simulators are required *to be able to do all that a full conceptual solution has to be able to do!* ****

Abstract

- As *memories of the same component become organized around a common frame*, they implement a simulator that produces limitless simulations of the component (e.g., **simulations of *purr)***. *Not only do* such *simulators develop for aspects of sensory experience*, they also *develop for aspects of proprioception* (e.g., ***lift, run***) *and introspection* (e.g., ***compare, memory, happy, hungry***). *Once established,* ***these simulators implement a basic conceptual system*** *that represents types, supports categorization, and produces categorical inferences.* These **simulators further support productivity, propositions, and abstract concepts, thereby implementing a fully functional conceptual system**. *Productivity results from integrating simulators combinatorially and recursively to produce complex simulations.*

Where we disagree – is that we suggest that we don't perceive then simulate – we perceive in the simulator space! – note he distinguishes out of perceptual experiences proprioception and introspection

Slide ~69 has 6 core properties of a conceptual system

****

Fully functional conceptual systems require:

- ***Simulators that can:***
  - *represents types (***binding a concept (type) to an individual (token) in a manner that is true or false**.),*
  - *supports categorization, and*
  - *produces categorical inferences.*

Slide ~ 70 has 4 additional properties of a conceptual system

These **simulators further support:**

productivity (*from integrating simulators combinatorially and recursively to produce complex simulations)***,**

propositions (*result from binding simulators to perceived individuals to represent type-token relations)***, and**

abstract concepts (grounded in complex simulations of combined physical and introspective events).

****

This suggests the architecture we have talked about – a LSOC like virtual simulation anchored in physical detections

Abstract

- ***Propositions*** *result from binding simulators to perceived individuals to represent type-token relations*. ***Abstract concepts*** **are grounded in complex simulations of combined physical and introspective events**.
- Thus, a *perceptual theory of knowledge can implement a fully functional conceptual system while avoiding problems associated with amodal symbol systems*. Implications for cognition, neuroscience, evolution, development, and artificial intelligence are explored.

Idea of a perceptual conceptual system overcomes the gap between current machine learning approaches based on statistical theory and real world learning

***Quest connections

- Most current approaches to engineer solutions required for computational intelligence solutions for problems like ISHM, cyber warfare, ATR, layered sensing etc. – are based on abstractions from the perceptual experiences (using Barsalou ideas – abstractions from the sensory data representation) – he discusses here some of the issues with perceptual symbol systems – with respect to concept encoding.

We have tied together in our OODA paper the ideas of perception has to be cognitive – so we certainly agree fields have to talk

****

Introduction

- For the last several decades, the **fields of cognition and perception**
- **have diverged.** Researchers in these two areas know
- ever less about each other's work, and their discoveries have
- had diminishing influence on each other. In many universities,
- researchers in these two areas are in different programs,
- and sometimes in different departments, buildings, and university
- divisions. ***One might conclude from this lack of contact***
- ***that perception and cognition reflect independent or***
- ***modular systems in the brain. ********

Argues as we have in quest the binding together of perception and cognition – just be careful – we are not in total agreement – in the sense that he suggest that perception Includes both unconscious and conscious representations. We reserve use of the perception word just for the conscious part of the internal representation.

Where we disagree is not that cognition is inherently perceptual – we believe that perception is cognitive! And although he is right that we need to tie perception to cognition they both are done in an amodal manner!

Introduction

****

Keep in mind his use of the word perceptual also include Libet

- Perceptual systems pick up information from the environment and pass it on to *separate systems* that support the various cognitive functions, such as *language, memory, and thought*. I will **argue that this *view is fundamentally wrong*.**
- **Instead, *cognition is inherently perceptual*, sharing systems with perception at both the cognitive and the neural levels**. I *will further suggest that the divergence between cognition and perception reflects the widespread assumption that cognitive representations are inherently nonperceptual, or what I will call amodal.*

***

Cognitive functions of language, memory and thought – he argues that you can't break these into separate systems distinct from the perceptual system – he defines amodal – representation that is inherently non-perceptual

****

Architecture question****

- Really interesting architecture question – does perception use a distinct system separated from language, memory and thought?
- We would suggest NOT.

****

Grounding cognition in perception

- In contrast to modern views, it is *relatively straightforward to imagine how cognition could be inherently perceptual*.
- As Figure 1 illustrates, this view begins by assuming that perceptual states arise in sensory-motor systems. As discussed in more detail later (sect. 2.1), a *perceptual state can contain two components*: an **unconscious neural representation of physical input, and an optional conscious experience**.

****

Suggests a sys1 and sys2 view – but links the two – suggests an architecture where they don't operate independently – a perceptual state consists of both components

We would suggest he has missed some steps here – the perceptual neural activation results in a sys 1 **** activation – that is NOT conscious yet – that leads to the perceptual representation that is the conscious experience – which of course also is achieved through neural activation

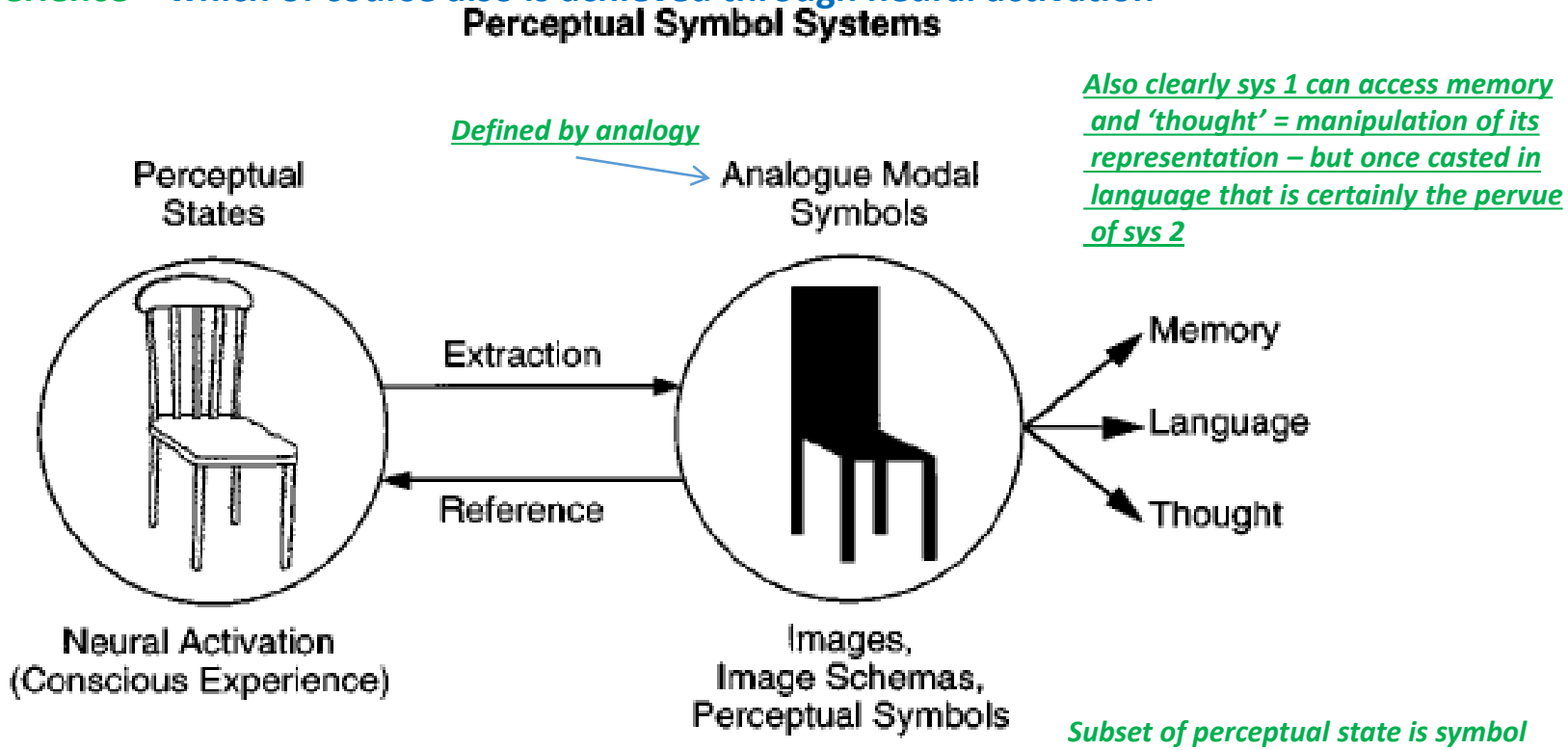
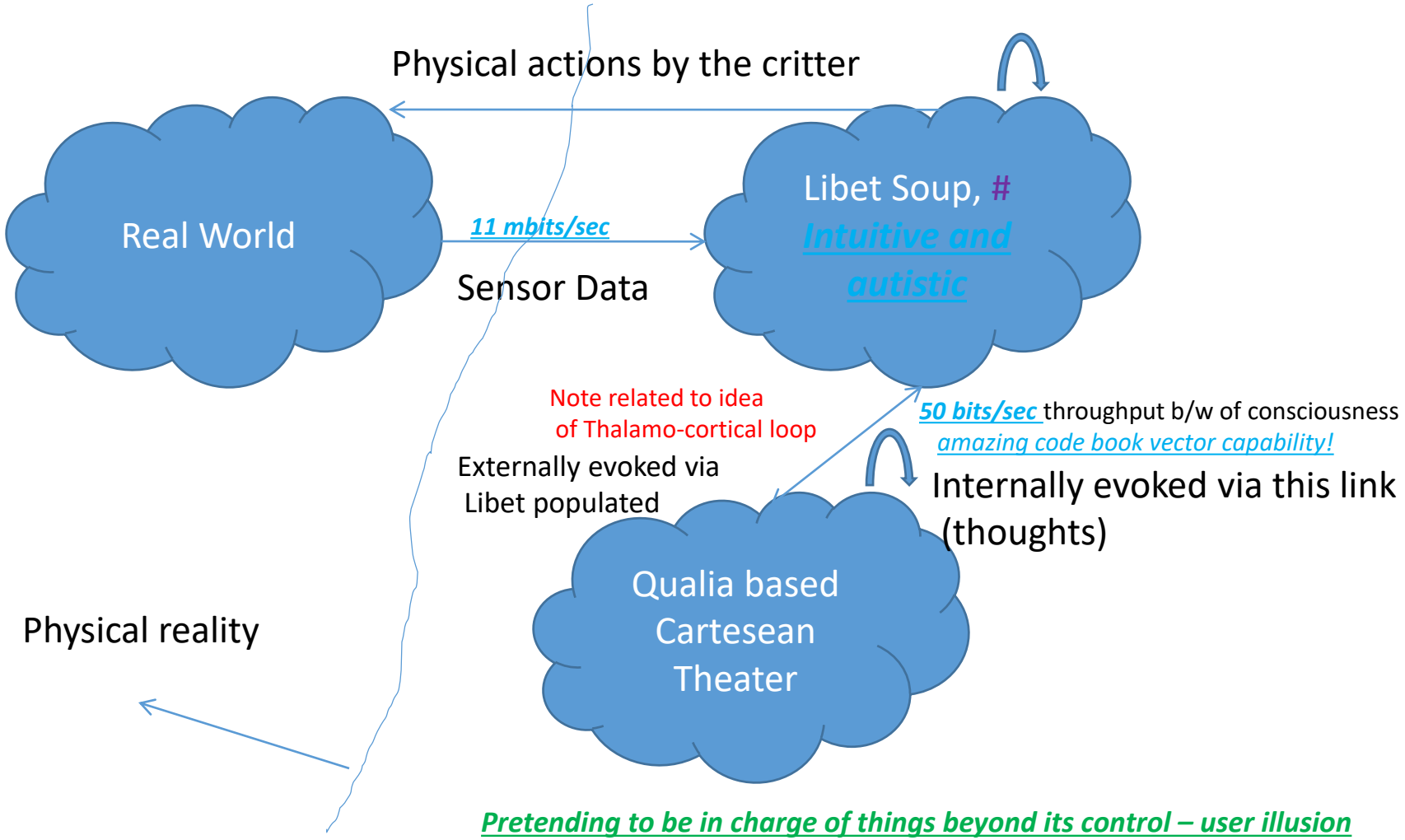


Figure 1. The basic assumption underlying perceptual symbol systems: Subsets of perceptual states in sensory-motor systems are extracted and stored in long-term memory to function as symbols. As a result, the internal structure of these symbols is modal, and they are analogically related to the perceptual states that produced them.

The subset that is extracted is associated with attention and then stored in LTM.

****

QUEST view of symbol set



We have taken the idea that the 'subset' is really a GIST – like a LF Fourier version – certainly this is in perceptual space where that includes Libet ****

Grounding cognition in perception

- Once a perceptual state arises, a *subset of it is extracted via selective attention and stored permanently in long-term memory*. On later retrievals, this perceptual memory can function symbolically, standing for referents in the world, and entering into symbol manipulation. As ***collections of perceptual symbols develop, they constitute the representations that underlie cognition**.

***

***

This is where we have been suggesting qualia are the perceptual symbols – they form the representations that underlie cognition – HIS PERCEPTUAL SYMBOLS ARE OUR QUALIA! except he also allows perceptual symbols in the unconscious processing - Libet

****

**** perceptual symbols and Libet

- We could allow the same symbols in the Libet soup and the Qualia Cartesean Theater – this would make us consistent with Barsalou – we would have to answer the question of why we have Qualia then – the answer might lie in the types of manipulation of the symbols are different in the cartesean ‘simulator’ environment
- Or the point that using the perceptual symbols versus the sensory data is one level of abstraction

His definition of modal – represented in the same systems (perception systems) I can't disagree with – but if the perception system is already qualia we may be different

****

Grounding cognition in perception

- *Perceptual symbols are modal and anal*ogical. They are
- modal because they are represented in the same systems as
- the perceptual states that produced them. The **neural systems**
- **that represent color in perception, for example, also**
- **represent the colors of objects in perceptual symbols, at**
- **least to a significant extent**. On this view, a *common representational*
- *system underlies perception and cognition, not*
- *independent systems*. Because perceptual symbols are
- modal, they are also analogical. The *structure of a perceptual*
- *symbol corresponds, at least somewhat, to the perceptual*
- *state that produced it.1*

*** *completely agree – but we would suggest perceive in symbol space - qualia*

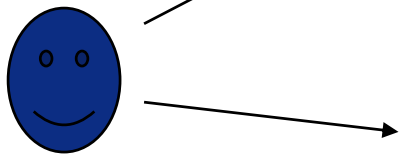
*** again we agree – but where we emphasize that the perceptual state is in the symbol space not in the sensory space ***

Let's be clear – we do also believe that the generation, maintenance and manipulation of qualia all require the exercising of the same pieces of meat used during perception – so by his definition of perceptual (uses the same system – I can't disagree with the statement that cognition is perceptual) – but the quest point is perception is already in the 'symbol' space

- Given how reasonable this perceptually based view of
- cognition might seem, why has it not enjoyed widespread
- acceptance? Why is it not in serious contention as a theory
- of representation? Actually, this view dominated theories
- of mind for most of recorded history. For more than
- 2,000 years, theorists viewed higher cognition as inherently
- perceptual. Since Aristotle (4th century BC/1961) and Epicurus
- (4th century BC/1994), theorists saw the representations
- that underlie cognition as imagistic.

Problem space and representation

What are the characteristics of the representations being generated by pipelines doing tasks – that could be used to determine when decision improvement could be obtained for your task – when should I attempt to bring in ‘privileged information’



We could use a system to extract a caption for a snippet of video from that world

We could use a system to extract a track of an individual a HVT from that world

There is a world and a person in the world doing stuff – moving about …

Exformation

This is the idea I was pushing on using a virtual environment as the scaffolding upon which we place sensations (computer representation of sensor data) for quest solns!

- exformation is 'context' - it is the reduction of uncertainty in the qualia representation not including the mental counterparts for sensory data = sensation ('reduction of uncertainty = meaning that values for some qualia are being set and not left undefined - although not to be confused with idea that 'reduction of uncertainty' implies matching to reality reducing the uncertainty of the representation with respect to its fidelity with respect to accurately capturing the physical reality) -
-
- Some of the qualia are associated with the internal representation at the introspectively available (Cartesean theater level) of the sensory captured stimuli = data, but there is much more represented in the Cartesean theater than the direct mental counterparts of the sensory data (sensation) and that is generated internally to provide a complete theater - wasn't sensed - it was *created internally to provide context for the sensed data* - that generated internally qualia we call exformation -
-
- It is the result of generating an accepted plausible narrative where much of it being composed of 'imagined' qualia not directly evoked from sensory data.

Exformation is the reduction of uncertainty resulting from the process of ‘thinking’ – the manipulation of the qualia results in setting values for qualia without additional sensation – that is exformation

Exformation additional thoughts

- I want to propose a change in our view of ‘context’ – I want input to an agent ALWAYS be through the sensors and from the stimuli pool – some of the stimuli pool is from the external to the colony of agents and some is from other agents – it is the job of the sensors and the sensor knowledge to be able to process the stimuli into data – thus I will say some data can be extracted from the outputs from other agents –
-
- Context on the other hand will be the ‘information’ that is generated as a result of ‘thinking’ – as a result of manipulation of the current set of qualia – to distinguish this source of reduction of uncertainty we might use the term ‘exformation’ to be consistent with the view in the book the ‘user illusion’ – or we might coin a new term as per the ox comment
-
- Exformation is the reduction of uncertainty resulting from the process of ‘thinking’ –
- the manipulation of the qualia results in setting values for qualia without additional
- sensation –
-
- the sensation (mental counterpart in the publisher’s representation when received the character ‘?’) – resulted in lots of qualia in addition to the representation of the character – all the rest of that qualia being set reduces the uncertainty of those qualia from many potential values to some specific settings – that reduction of uncertainty is a special type of ‘information’ that we have been calling ‘exformation’

Prior view of Exformation, Context and Agents

- With respect to alignment – our old view is that when two agents (be that between people or between a human and a computer) are able to directly incorporate the output from the other agent into the reduction of uncertainty of some aspect of its internal representation we will say that aspect of their internal representations are 'aligned' - we have called the input mechanism to an agent for this the 'context' input - context inputs have to be aligned - so the Knowledge of other agent = Ka allows the saving of the normal work required to transduce sensor inputs to information -
-
- the idea is that reduction of uncertainty in the internal representation through the context input is exformation

WE WON'T USE THIS VIEW ANYMORE – CONTEXT IS RECEIVED FROM OUTSIDE AN AGENT BUT IS GENERATED AS A RESULT OF INTERNAL MANIPULATION OF THE QUALIA – IT IS THE RESULT OF THE PROCESS OF THINKING

Exformation and Libet

- Since we've defined 'information' as the reduction of uncertainty in the internal representation – 'information' can be generated as a result of processing the data using knowledge – where our agent discussion has focused
-
- OR
-
- it can also be generated via 'thoughts' – since this information (reduction of uncertainty of the internal representation) didn't come through a sensory channel but was created out of manipulation of the existing qualia using prior experiences (memory) and knowledge we wanted to call this out as distinct and thus called it 'exformation' (context)
-
- what about Libet
-
- the issue comes about that Libet is certainly part of the internal representation – it is not introspectively available but is critical to the critter – sensory data can result in information at the Libet level – Do we want to allow exformation at this level also – my suggestion is NO– prior experiences (memory) and knowledge can complement the sensory data for a richer internal representation even at the Libet level – BUT WE WON'T ALLOW LIBET TO STORE IN STM – WON'T HAVE THE QUALIA REPRESENTATION AVAILABLE FOR THINKING – THE WHOLE REASON FOR QUALIA IS FOR THE GENERATION OF EXFORMATION

Biederman Problem – clear we need a new approach

- Children in first 6 years:
  - vocabulary approximately 10,000 words
  - master visual recognition of approximately 3,000 “entry-level” (shape-based) categories
  - master visual recognition of approximately 30,000 “perceptual-level” categories ***** 10 bits *****
- Approximately one perception-level category per waking hour **(about 50k hours in 6 years)**
- To make upper bound on error (using approximations like the VC dimension) – how many training samples to solve Biederman’s Problem – 10^2,400,000 – need **Yotta-samples**

Biederman, I. (1995). Visual object recognition. In S. F. Kosslyn and D. N. Osherson (Eds.). An Invitation to Cognitive Science, 2nd edition, Volume 2., Visual Cognition. MIT Press. Chapter 4, 121-165.

Yotta is the largest unit prefix in the International System of Units (SI), denoting a factor of 10^{24} or 1000000000000000000000000. It has the unit symbol **Y**.
The prefix name is derived from the Greek οκτώ (*októ*), meaning *eight*, because it is equal to 1000^8. It was added to the SI in 1991.[1]

Hinton point from Yann

- 10^14 connections – live only 10^9 sec – m ore parameters than data – can’t do the labelled data approach to solving

Biederman Problem

- Conclusion nature's solution is something else – not enough time ontogenically **(the entire sequence of events involved in the development of an individual organism)** to get to what we are able to do or even phylogenically **(as a species capturing the concepts)**
- Conclusion – we don't learn representations of the stimulus data – ***we learn parameters of simulation***
  - (we **would say we learn qualia** – we think we are conscious of a simulation of the world NOT the world itself and NOT the sensory data)

The conscious meaning of the stimulus is in the simulation!
We would suggest the 'conscious' part of the meaning of a stimulus to a person is the aspects of the simulation attended to – the qualia. But begs the question how are the qualia represented – how can we represent them in a computer

Stable consistent and useful objective function

- Idea that the objective function used by nature in generating and updating the qualia vocabulary is stability, consistency and usefulness –
  - this last attribute tied to 'structural coherence':
    - If you have multiple learning systems that use the same objective function for optimizing and are using basically the same machinery (sensors and cognitive approaches) – what can you say about the resulting representations? With respect to what aspects have to be identical or what aspects have to be mutually consistent between those multiple systems (multiple agents)? Seems to me this is an idea if followed could answer how agents that 'align' have to have representations that are 'consistent' in some way?
- How do we define an objective function that captures structural coherence to be used in current machine learning systems – associated with the Embarassingly simple approach to zero shot learning article –it provides an attribute based approach – We've seen a related system that used mechanical turk to extract constraints on a representation of concepts (which are closer and which are farther apart …) and then sets up a transformation from what is learned to that new representation space – the thought was to use those constraints as an objective function when extracting the features …

Challenge to jared / ox / Scott

- If you have multiple learning systems that use the same objective function for optimizing and are using basically the same machinery (sensors and cognitive approaches) – what can you say about the resulting representations?
  - With respect to what aspects have to be identical or what aspects have to be mutually consistent between those multiple systems (multiple agents)? Seems to me this is an idea if followed could answer how agents that 'align' have to have representations that are 'consistent' in some way?
- Example to start with: take two systems that use the same data set and same algorithmic approach say deep learning – but let the two systems start with different random weights and different orders of presentations and different parsing into training / test / validation sets – what can we say is in common with the resulting representations?

Of course real goal is to define what has to be constrained for two agents representations to be alignable

How to find the right representation

- Mean square error for classification as the objective function does not necessarily generate a representation that replicates the characteristics we consider critical for consciousness – we believe the constraints used by nature are
  - Stability
  - Consistency
  - Usefulness
- We think we can achieve the first by using streaming video and enforcing the features extracted to not be bouncing / varying too much (assuming stabilized video) – unsupervised learning
- We think we can achieve the second by doing video of similar locations and reward consistency – supervised learning
- We think effectors / interactions with the environment – we have to close the loop around the representation to achieve the last – reinforcement learning

Knowledge and representation

- How about this – back to knowledge and representation then heading towards simulation:
- The knowledge includes the representation and the processes that facilitates manipulation of that representation – for example the representation in a DL system is the architecture and the weights but there are also the processes that I have to use to exercise that representation to generate any meaning as well as processes that make the deduction that the highest output node indicates something of value to be used in a particular fashion outside the system – all that together is the knowledge – so far more than the representation – but the representation places constraints on what can be done suitably with the knowledge
- Now if my agent uses simulation as part of its cognition – then as you say the parameters of the simulation are part of the knowledge as are the details of executing the simulation – the process of executing the simulation
- With respect to mutual information – we've often suggested that the idea of 'cognitive decoupling' is key – the key idea is to in fact be able to generate a simulation that is stable, consistent and useful with the least possible mutual information – thus not relying on the noise but the essence of the information –

- To generalize this across past/present/future, we probably should distinguish between the "simulation" -- i.e., the process that is a particular type of "thinking" and the result of the simulation -- i.e., the meaning that is generated by the simulation that has a spatio-temporal framework and low mutual information with the "world."
- The word "simulation" connotes "not-realness" (and I think that this is intuitively how we have used the word) -- so it seems to me that the essence is in this low mutual information with reality, not the past/present/future distinction.
- I also think of the parameters of the simulation as just another part of the knowledge (knowledge seems to be the catch-all for everything modeling the agent).
- Jared

Cognitively decoupled

- Proving as problem solving: The role of cognitive decoupling : Abstract
- This paper discusses the process of proving from a novel theoretical perspective, imported from cognitive psychology research. This perspective highlights the role of hypothetical thinking, mental representations and working memory capacity in proving, in particular the effortful mechanism of *cognitive decoupling*: problem solvers need to form in their working memory two closely related models of the problem situation – the so-called *primary* and *secondary* representations – and to keep the two models *decoupled*, that is, keep the first fixed while performing various transformations on the second, while constantly struggling to protect the primary representation from being "contaminated" by the secondary one. We first illustrate the framework by analyzing a common scenario of introducing complex numbers to college-level students. The main part of the paper consists of re-analyzing, from the perspective of cognitive decoupling, previously published data of students searching for a non-trivial proof of a theorem in geometry. We suggest alternative (or additional) explanations for some well-documented phenomena, such as the appearance of cycles in repeated proving attempts, and the use of multiple drawings

Cognitive decoupling

- Cognitive decoupling (see Fig. 3) is an alternative to serial associative cognition. It consists in simulating situations that are counterfactual with respect to the system's current situation (as represented by the activity of the Reactive Level's *Knowledge agents*). When the Reflective Level launches cognitive decoupling (see below for the conditions in which the Reflective level will do so), *SecondaryRepresentation agents* are initialized using the reduced representation sent by the *UpdateStatus* agents (SU in Fig. 3). As described by Stanovich, decoupling is a cognitively costly operation. Through the action of a *SimAct agent*, cognitive decoupling recruits *RequestStatus* (RS2 in Fig. 3) and *Control agents* by rerouting the messages they initially sent to or received from the Reactive level to *SecondaryRepresentation agents* (SR in Fig. 3). The Algorithmic level is then focused on carrying out the cognitive decoupling operation. Instead of using goals to control the Reactive Level activity of the agent, the system will use them while cognitive decoupling is in effect to control a simulation of a possible world. The result of this operation may affect goal selection at the Reflective level. Cognitive decoupling is carried out by agents at the Algorithmic Level, but initiated by Reflective Level agents. While it is focused on carrying out the cognitive decoupling operation, the Algorithmic Level is still "keeping an eye" on the Reactive Level activity. In the current version of the system, when cognitive decoupling is launched, there still is one RequestStatus agent, one *Control* agent (C2 in Fig. 3) and one *UpdateStatus* agent assigned to monitor activity at the Reactive Level. This number of agents is not sufficient to carry out complete serial associative cognition, but it allows the system to be able to interrupt cognitive decoupling and go back to serial associative cognition if and when something important happens in the system's environment. When the system is in an unstable environment, it needs to be able to interrupt cognitive decoupling and go back to full serial associative cognition. If, for example, current Reactive Level activity start to involve many *Knowledge agents* associated with "Fire" (or some other danger) while the system is carrying out a cognitive decoupling operation, this will lead to an high emotional response that will influence the frequency at which the *Control agents* and the two status agents (*RequestStatus* and *UpdateStatus*) send messages (for more on the influences of artificial emotions on the different functional structures of the system, see Larue, Poirier, & Nkambou (2012)). Increasing the message sending frequency of these agents will lead to an increase in the number of messages sent to all of the *Goal agents*. If the number of messages sent is sufficient to make a *Goal agent* match a reduced representation coming from the Reactive Level, cognitive decoupling is interrupted and the system goes back to full serial associative cognition mode.

Hypothetical-thinking based on cognitive decoupling and thinking dispositions in a dual cognitive agent
LaRue - Biologically Inspired Cognitive Architectures
Volume 6, October 2013, Pages 67–75

Subject: dynamic vocabulary

- Sorry for the size of the email - cathy can you post on the vdl for this week -
- Jared / ox - one idea that I've been thinking about this week is application of the ideas from the Octopus article to cyberspace - that is often we speak of things like cyber defense in terms of a central smart brain taking in all the measurements of activity that then deduces situation awareness and plans actions -
- With the amorphous body in cyberspace - the tentacles of the network are not unlike those of an octopus - and maybe that provides a better architecture -
- Now to the point - assuming each 'limb' is smart - each can taste / touch the world and experience it independently - since there is some top down control from the brain also - I would assume there is a common vocabulary that the system uses for interactions -
- Cap

This exacerbates the issue of alignment – there is no way for systems to stay alignable if something like the last slide is NOT true

Jared response

- Cap,
- I know that there has been a lot of work in decentralized decision making, although what I'm aware of has mostly focused on things like consensus across a network.
- How does the octopus response at the limbs compare to human reflexive responses? I mean, how "smart" does a subsystem need to be before you start thinking of it as a separate "agent" with its own vocab? Or maybe this is the wrong test -- maybe it is more (as I think you suggest) about the ability to "interact" using some sort of vocabulary with the assumption being that the more spatially dispersed and limited bandwidth communications would require more sophistication in this interaction.
- In other words, what is the difference between a decentralized networked swarm of ISR agents (with limited comms) and the cyber problem (or the octopus)?
- Jared

- That last point is where I've been all week - even in the human mind we get a glimpse of the same issue - if a person gets an anti epileptic seizure surgery to severe their corpus collosum they often have one side of their body do something and the other side attempt to prevent - examples include one hand pulling the pants up and the other attempting to pull them down -
- That brings me back to the challenge I gave you/ox - my assumption is this issue of 'aligning' representations is critical - and in the case of a single critter with multiple 'smart' limbs and the similar (in my mind) case of two critters of the same species - I'm going down the path that there is some genetic predisposition for the generation of the sensory information and an 'objective' function (like make the representation stable, consistent, useful for this critter to survive and procreate) - those two facets together result in representations that have something in common (are the same? Or are there characteristics that can be proven relatable by the two facets?)
- Cap

How to have bots communicate

- Representation is how an agent structures its knowledge (what it uses to generate its meaning from a stimuli) and how its meaning is structured and how it generates its understanding and comprehension (this is the key - how it generates and maintains its situations - relationships between situations and how situations can interact with each other - this is the key aspect of the meaning and what is consistent between agents that can align) - this last point is what stays consistent between two agents - it is what is the same between your red and my red - if our respective representations maintain this characteristic (consistent relationships and consistent interactions between represented situations) then we can use those concepts consistently and can thus communicate and coordinate action

Cap shot at formalizing

- How do we formalize this problem – assume there is a world – there is physics – there are photons – they have characteristics – observables by an agent's sensors – the agent uses those observables to generate a representation to capture some aspect of the physical world – that transformation from sensory observations to an internal representation is guided by an objective function – the mechanism used by the agent to accomplish that (to generate the representation) may be in common with some other agents – and the sensors are similar to other agents – and the objective function is in common - what can we say?

Jared comments

- This line of thinking makes me wish I knew more about model theory. Here's a crude attempt to formalize some of these ideas:
- A task (or set of tasks) defines the semantics of a problem (for instance, a categorization task might give logical statements about whether a particular image contains a given object of interest). In this color problem, the task seems to be to not only categorize colors, but also determine similarity, opposition, etc. We can then come up with a set of statements in those semantics (a theory) and ask whether or not those statements are satisfiable in a particular model (mathematical formalism, such as the color-space model for human color perception), which includes a mapping of the semantics into that model.
- Your question then takes the form of saying that you have several models derived using similar machinery, and we can ask whether the knowledge of the variation in the generation of those models allows us to say anything about whether the general theory is satisfiable in those multiple models -- if so, we can say that we have "alignment" (at least up to whichever statements are jointly satisfied).
- For example, if you go to the simplest example of a light switch for which the task is determining whether the light is "on" or "off," then the semantics just contains two statements (which are negations of each other) and so any model that can handle a binary state would be theoretically alignable -- e.g. I can align to that light switch because in both of our models, we can compute satisfiability of the statement "The light in on," even though I might call it something different.
- For our deep net classification problem, this would entail thinking about the mapping of the semantics in the network (i.e., how do you make decisions given a trained network) -- and then using our understanding of things like variation in training data, backprop, and regularization to argue about the satisfiability of the semantic statements. This seems hard in general.
- I'm sure that someone who actually knows something about model theory beyond a vague understanding of these words could probably say something more useful (and more correct). I would be shocked if some philosophers have not gone down a similar road with human-to-human model alignment at some point -- the hard part to me is how to say something concrete about a particular set of models.

- Ok I read through your email – and I think I can ‘align’ the two – but I want to make a couple of other points first – the why first – making QuEST AI bots:
- From Microsoft:
- CEO Satya Nadella calls it "Conversations as a Platform", and it’s all about chatty bots.
- “Bots are the new apps,” Nadella told Microsoft’s annual Build developer conference this week, giving the wannabe platform pride of place. We’re told that conversational bots will marry the “power of natural human language with advanced machine intelligence”. There’s a Bot Framework for developers. Ambitiously, he wants corner shops to develop their own bots.
- “People-to-people conversations, people-to-digital assistants, people-to-bots and even digital assistants-to-bots. That’s the world you’re going to get to see in the years to come.”
- So from a quest perspective the issue is NOT how to make this one central enormously complex AI solution not to make a single conscious computer to solve all our problems – but how to make the simplest possible set of conscious agents that can align together to formulate interesting solutions – and going down this path screams to me we need to focus on the respective representations – and thus our fundamental open question –
- Assume an architecture for QuEST agents – assume some objective function (possibly based on stability, consistency and usefulness ….) – turning them loose in the world to get exposed to stimuli – they generate ‘knowledge’ (means to impact later meaning making by this agent when exposed to related stimuli) – but they also encounter other agents that they can align with and thus can acquire knowledge via culture (agent to agent communication) – so this last aspect is what I seek to understand – under what circumstances can we have ‘culture’ – where one agent can impact knowledge of another –
- Cap

- But I want to remind all how I'm using the words – knowledge and representation –
- Knowledge is what an agent uses to generate meaning –
- Representation: how the agent structures what it knows about the world – so for example its knowledge (what it uses to generate meaning of an observable)
- Reasoning: how the agent can change its representation – the manipulation of the representation for example for the generation of meaning
- Understanding: application of relevant parts of the representation to complete a task – the meaning generated by an agent relevant to accomplishing a task in the estimation of an evaluating agent
- So the one of the keys to this thread is we use our favorite learning approach and turn loose a bot into the wild – BUT – we have also release a series of other bots – initially they all have the same knowledge – this is the 'evolutionary' knowledge that is programmed in at bot birth – one current best practice here is supervised learning
- Now the bots go out and they gather observations – and interact with the world – one current best practice here is reinforcement learning – this second means of knowledge acquisition is 'experience' – Scott in your case of multi-sensor fusion I could imagine although the bots start out similar they are exposed to their respective sensing modalities thus their respective knowledge is becoming somewhat distinct –

- But now I allow my bots to interact with each other – ‘culture’ – I want the bots to be able to share their knowledge to each others benefit – (by the way although I’m using the term bot I also want my bots to be able to share their knowledge with humans – so instead of bots I could say agents) ---- this is where this thread is at – what do I have to do to ensure that the representations will facilitate culture based knowledge transfer – if you will transfer learning –
- I envision from the transfer learning literature we can extract some very demanding requirements that would also reduce the flexibility of when the knowledge could be transferred and what knowledge could be transferred and how to transfer the knowledge – but we want to abstract this out –
- Taking that view back to the ‘is your red my red issue’ – let’s assume that the only reason to have language is ‘culture’ – I need to do transfer learning – what can I say has to be mutually consistent between the respective representations to do that – AND – what was done by nature to ensure that happened (same objective function?, same machinery?) – but note I can do transfer learning with my dog – he teaches me to notice certain things (big leap in logic there but I think important to consider agent interactions between drastically different types of agents)
- Cap

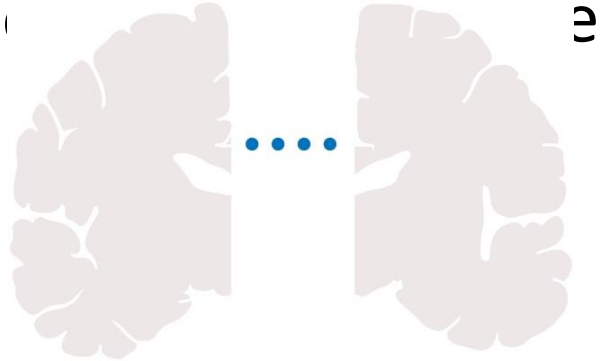
- Cap,
-
- It seems to me that turning these bots loose on the environment independently will make incorporating culture "later" much more difficult. It would seem better to embed the need to align (indirectly?) into the objective function from the start -- so that the bots learn to align as they gain (perhaps vastly different) experiences.
-
- If you could put a metric on efficient communication from the start (maybe after some initial supervised learning?), that might help ensure that the reps are coherent.
-
- Jared

scott

- I know we've talked about it before but, there's literature describing deep-learning experiments wherein an inter-agent communication channel was allowed. Consequently, the system learned a communication protocol that improved the performance of the "society" on some tasks as compared to those of the individual agents without communication. https://arxiv.org/abs/1605.07736
- While this work corroborates the idea that communication is useful and that our current algorithms can learn it, it doesn't directly address the efficacy in the timing, i.e. it's better to communicate as you go vs. collect information and communicate later. Might be a means of exploring the idea, though. This is the direction I'm leaning in for the fusion experiments.
- Even though this system learns to communicate amongst agents, it wouldn't be that helpful to communicate with people since it was "designed" in the training process - can't really expect a person to take part in that. Not now, anyway.
- ==========
- Given my last statement, I think this implies Jared's point about incorporating culture after the fact would be more difficult; not because the human couldn't interact with the training mechanism but, because a "new culture" has to be learned. It might just be that a culture learned during exploration would be more efficient somehow, e.g. somewhat dubious claims about the Himba people and the shades of green (http://languagelog.ldc.upenn.edu/nll/?p=17970).
- The other thing I wanted to mention was that since these learning mechanisms are probabilistic, maybe looking at metrics in a similar way (e.g., on the space of events related to the task) would be helpful. As we ran into in our hallway discussion, we currently characterize the systems empirically. Maybe we should start with a simple learning approach that can be described somewhat analytically, e.g. Naïve Bayes?

Split brain patients

- Split-brain is a lay term to describe the result when the corpus callosum connecting the two hemispheres of the brain is severed to some degree. It is an association of symptoms produced by disruption of or interference with the connection between the hemispheres of the brain
- Split-brain surgery, or corpus calloscotomy, is a drastic way of alleviating epileptic s ence of sporadic electrical storms in the brain



The split brain: A tale of two halves

- Since the 1960s, researchers have been scrutinizing a handful of patients who underwent a radical kind of brain surgery. The cohort has been a boon to neuroscience — but soon it will be gone
- In the first months after her surgery, shopping for groceries was infuriating. Standing in the supermarket aisle, Vicki would look at an item on the shelf and know that she wanted to place it in her trolley — but she couldn't. “I'd reach with my right for the thing I wanted, but the left would come in and they'd kind of fight,” she says. “Almost like repelling magnets.” Picking out food for the week was a two-, sometimes three-hour ordeal. Getting dressed posed a similar challenge: Vicki couldn't reconcile what she wanted to put on with what her hands were doing. Sometimes she ended up wearing three outfits at once. “I'd have to dump all the clothes on the bed, catch my breath and start again

http://www.nature.com/news/the-split-brain-a-tale-of-two-halves-1.10213

Exformation and meaning continued

- The meaning of both were unmistakable by either party. Measured in Shannon bits (30 odd characters or approx 5 bits each) there were only about 10 bits exchanged over the communication channel. The correspondence in reality refers to a plethora of 'information' otherwise it would have had no meaning to either.
- Hugo 'discarded' a mass of information that was consuming him and he in a very real way was referring to it without including it. Explicit discarding of all that other stuff. Explicitly discarded information = 'exformation' – user illusion.
- The measurement of how much 'exformation' a given message implies – only the context can tell you. We would say it can only be estimated by the 'alignment' between the agents involved. It is the meaning to those agents!

Many computer agents only process information they can ‘see’

- **Word-based algorithms are limited by the fact that they can process only the information that they can ‘see’.**
- As **human text processors, we do not have such limitations** as every word we see activates a cascade of semantically related concepts, relevant episodes, and sensory experiences, all of which enable the completion of complex tasks (such as **word-sense disambiguation, textual textual entailment, and semantic role l**abeling) in a quick and effortless way.

Recall the idea of Exformation – key is the link game -

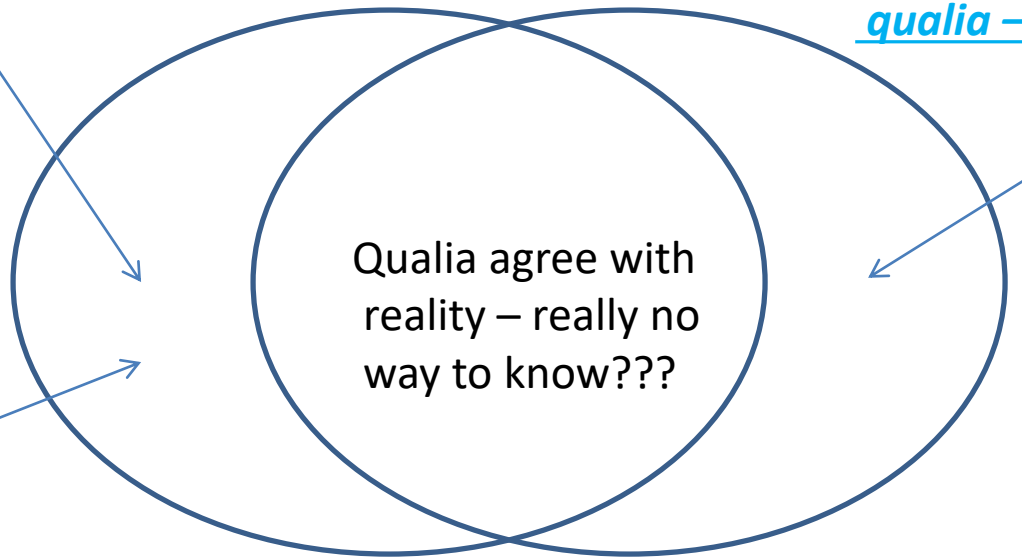
Projection at its Core is an Imagined Future

Note imagination is all-source, integrated, situated and simulated possibly key to getting information and understanding its meaning also

Venn Diagram Awareness vs. Consciousness

Qualia don't always agree with reality = delusion
BIIDs, AHS - generating a 'not part of self' quale that isn't correct -> low awareness of leg being yours

Qualia don't agree with reality = ex blindsight – Libet visual system is aware and the person can respond appropriately but no visual qualia –> no visual consciousness



Dreaming!

consciousness

awareness

Conscious but not aware

Aware but not Conscious
Sys1 Aware but Sys2 NOT

**Libet system – sys1 is about awareness
The conscious sys2 is about generating a stable/consistent/useful representation**

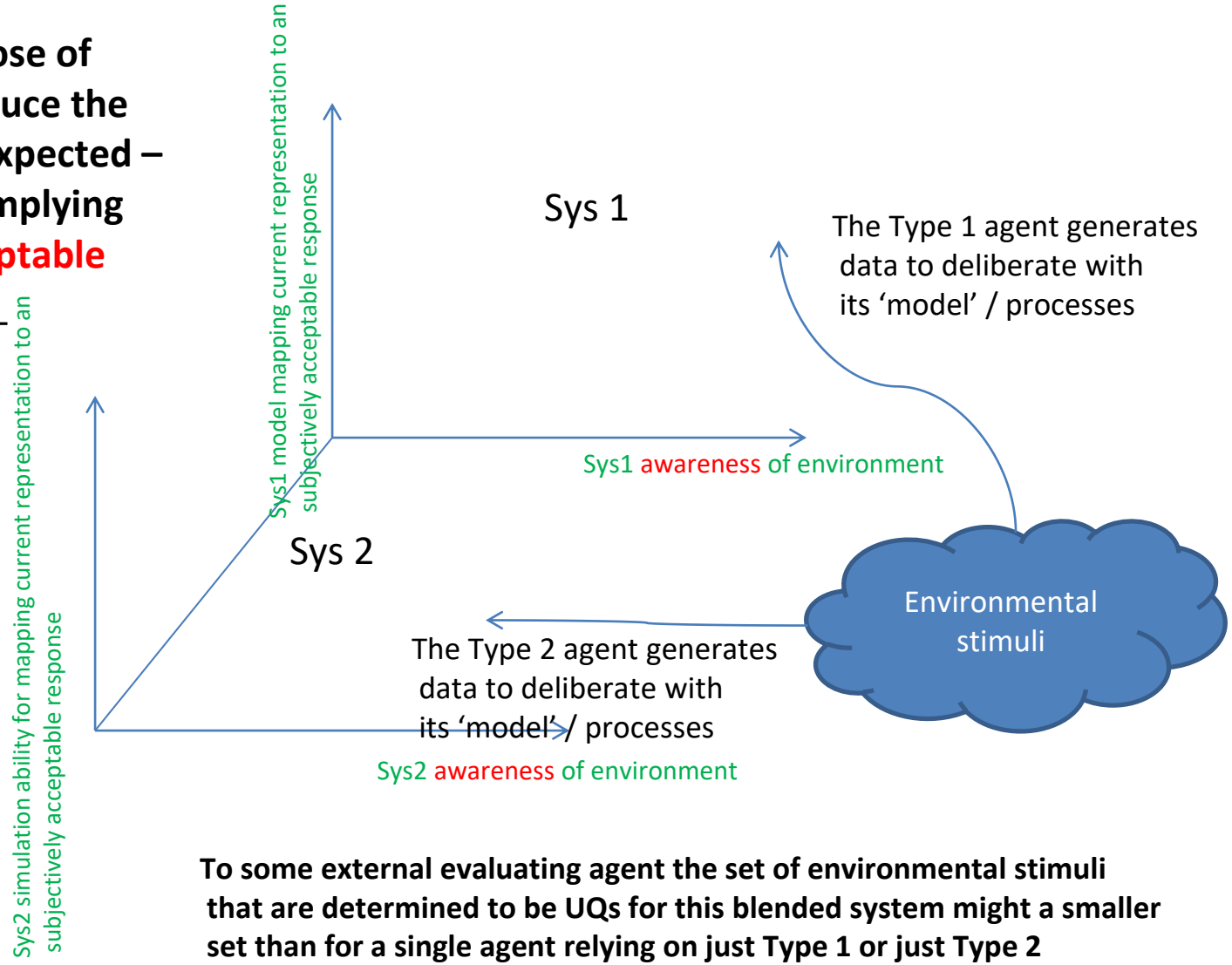
Framework for letting conscious or subconscious dominate decision making

- Along these lines, **Hammond et al. (1997)** presented a framework for analyzing how the surface and deep characteristics of various tasks can serve to induce the corresponding cognitive styles of decision making and reasoning. These characteristics include the following:
  - a. *number of cues—intuitive-inducing tasks possess a large number of cues,* whereas analytic-inducing tasks possess a small number of cues;
  - b. *measurement of cues—intuitive-inducing tasks involve perceptual measurement,* whereas analytic-inducing tasks entail objective measurement;
  - c. *redundancy among cues—intuitive-inducing tasks involve high redundancy;* analytic-inducing tasks entail low redundancy;
  - d. *degree of task certainty—intuitive-inducing tasks involve low certainty; analytic* tasks entail high certainty;
  - e. *display of cues—intuitive-inducing tasks involve simultaneous display; analytic-* inducing tasks entail sequential display; and
  - f. *time period—intuitive-inducing tasks involve brief time periods while analytic inducing* tasks entail long time periods.

Blending

Extended View of UQ for Blended Multi-Representational Agents

Implication is the purpose of consciousness is to reduce the trade space of the unexpected – where unexpected is implying can't generate an acceptable response! --- or maybe refers to the amount of effort it takes to get to an acceptable response ---



To some external evaluating agent the set of environmental stimuli that are determined to be UQs for this blended system might a smaller set than for a single agent relying on just Type 1 or just Type 2

xAI

- Goal line tech - no more accurate but people believe and in tennis

Information is not independent of the processor

- This idea is not new to us – for example in the reference below –
- It is our view that we will use the word information to be some reduction of uncertainty in a processors representation thus clearly tied to the processor that is generating it

J Evol Econ (2004) 14: 43–67
DOI: 10.1007/s00191-003-0181-9

Journal of Evolutionary Economics

Idea that information is out there and our job is to go get it is flawed To use the word information we propose you have to associate with it the agent that is using in its representation

Data, information and knowledge: have we got it right?

Max Boisot[1] **and Agustí Canals**[2]

[1] IN3, Universitat Oberta de Catalunya, Barcelona, Catalonia, Spain, and Sol Snider Center for Entrepreneurial Research, The Wharton School, University of Pennsylvania, Philadelphia, USA (e-mail: boisot@attglobal.net)
[2] IN3, Universitat Oberta de Catalunya, Barcelona, Catalonia, Spain, and ESADE Business School, Universitat Ramon Llull, Barcelona, Catalonia, Spain (e-mail: acanalsp@uoc.edu)

QUEST Information and Shannon

- We will need to have our view of information tied to the meaning of the information to the processor that is generating it (Shannon didn't focus on 'meaning' – to Shannon the amount of self information in a given symbol is inversely proportional to the probability of occurrence
- We will thus have to address meaning (not normally addressed in 'information')

Cryptography example

Cryptography example – data itself is public – freely available – only those with the key can get the 'information' – this authors suggest that you can't extract the information without the key – thus cryptography exploits deep differences between data and information --- *we originally suggested that decrypting the data doesn't make it information – it only makes it the transmitted data – it doesn't become information until it reduces the uncertainty of some aspect of a representation of the world (**** note there is a reduction in the uncertainty of which words were transmitted so in that sense the key does allow the meaningless data bits translation into which 'words' were transmitted – but we could take that even further – capturing of the bits reduces the uncertainty of what bits were transmitted so at that level that is also information ***) bottom line is there are levels of data and at each of those layers of abstraction there is associated information – thus *** next slide ****

Information defined

- **information as a relation between in-coming data and a given agent – *by this definition information is 'subjective'***
- ***Information is the reduction of uncertainty in any aspect of an Agent's representation of its world***
- ***It is generated via a process that has as its input data and uses knowledge to reduce the uncertainty in its representation***

Knowledge

- ***Knowledge is set of expectations held by agents and modified (reduction in uncertainty) by the arrival of the data and its conversion into information --- straight from the economics article***
- ***We would say knowledge is what is used in the process to map data into information***
- ***We would include the 'knowledge' used in the sensor to convert stimuli into data***

****

Wolfram problem

The amount of knowledge will have to do with the representation used by Wolfram Alpha by specification of the number of possible states – thus how to quantify the impact of the Alpha knowledge

- We're approaching 7 million lines of mathematic code. It's a little hard to quantify. I can tell you how much code is inside. If you ask ***how much knowledge*** is there, there is no particularly good way to quantify that *** this is something I think our discussion is about – if we define knowledge as the means we transform data into information- ***we should be able to quantify how much knowledge is in alpha*** ***. One of the big projects is to curate data and make it computable. The challenge there is to find really reliable sources of data. Going from the raw table of numbers to the full thing — that is, setting it up so that it can be computed — is a big challenge. About 98% of the time the data you get from WolframAlpha is internal to our computer systems. There are a few cases where it's real time data. Stock quotes are one example
- ***--- I might be willing to say 'to know something means that it has evoked in you an 'aha' (as per matt) – a very specific quale – thus tying to know to being aware of something – thus currently only within a critter'***

Note how to know something does not mean that it is correct – just means that it was experienced – it generated an aha quale in the critter – thus the critter knew it to be so

Agent

A(s,r,k) – an agent is defined by its sensors, its representors (function that map the data to a modification of its representation) and its knowledge that it uses to choose what subset of stimuli to make into data and how to map that data to a change in its representation

- We will use the word 'Agent' to denote the entity that captures some stimuli (that subset of stimuli captured by the Agent's sensors - we will call ***data*** for that agent and what subset it captures is determined by its ***knowledge*** encoded in its sensors) and uses that data along with its ***knowledge*** to modify its internal representation (that reduction of uncertainty of its representation we call ***information***)

PREMISE OF PIECE IS THAT ALAN TURING HAD AN APPROACH TO PROBLEM SOLVING - HE USED THAT APPROACH TO CRACK THE NAZI CODE – HE USED THAT APPROACH TO INVENTING THE IMITATION GAME – THROUGH THAT ASSOCIATION WE WILL GET BETTER INSIGHT INTO WHAT THE IMPLICATION OF THE IMITATION GAME MEANING IS TO COMING UP WITH A BETTER CAPTCHA, BETTER APPROACH TO 'TRUST' AND AN AUTISM DETECTION SCHEME – and a unique approach to Intent from activity (malware)

****

Turing and code breaking and the Imitation Game

Notes by Capt Amerika, ideas from Special K, Ox, Adam

Oct 2009

As ox suggested maybe the 'deceit' aspect is also universal in his thoughts – he lived a life based on trying NOT to let people break his personal 'code' – his sexuality

Meaning – considerations for definition

- Note the meaning could be a recalled memory (the memory can serve as data that with knowledge can be converted to information {reduce the uncertainty of some aspect of the Agent's world model})– can be a projection (using imagination) {some olfactory data stimulates a memory that can be processed into information – a particular internal representation in my world model – but it reminds me (memory) of something – thus it 'means' something to me associated with that memory.

Relation between information and knowledge

- Relation with information – a symbol (used the word symbol to capture both qualia level information and Libet level information) disambiguates some aspect of the agent's representation of the world. The impact of the disambiguation to the entirety of the agent's world model is its meaning to that agent.

Example

- My Porsche is red. There is information in its redness to me. Its redness also has meaning to me. Of all the uncertainty I have available to my world model to account for the hue/saturation/brightness of my car I reduce that uncertainty when I associate with my car the symbol red = information = generation of a specific symbolic representation from the input data (note there is the word symbol information ‘red’ AND also the quale of redness level of information generated). In addition to the information associated with reduction of uncertainty to account for the data my senses received there are implications to my world model by the fact that I’ve associated with my car the color red. Red ‘means’ to me that the car has that ‘arrest me’ quale to any cop that might see me driving. Note the meaning has links to implications to my world model that aren’t always associated with the information generated from similar data. Sometimes that same hue/saturation/brightness has a different meaning – ‘firetruckness’.

Matt comments:

- I think that our model of links+quale is a functional definition of "meaning". To prove to someone else that you "know" the meaning of some stimulus requires only that you produce a link list that is reasonably congruent to that of the person you are speaking to.
- The **internal AHA that you get, to indicate that you know the meaning**, is your private realization that you have your links to the quale that the stimulus generates. **Dictionaries are lists of links that some lexicographer knew.**
- The links idea dismisses the philosophers' trick question: "What is the meaning of meaning?" They didn't have any clue about links+quale.

Matt comments

- I get uncomfortable with our attempts to generate ***airtight*** labels, [made out of words, which are (qualia+links)], to "define" data, information, knowledge, meaning...
- I also get uncomfortable with some of the proposed specific definitions like: "Information is: The result of the reduction of uncertainty". That's useful for me only when I'm standing in Shannon Land, where the uncertainty is the choice of one of a few possible trivial events in the presence of added Gaussian noise
- We all know that the **only thing information is**, in the qualia mind of the recipient of the data stimulus, is **qualia+links**. If there is **something we might like to call information in a machine, its nature is unlike the information in my qualia mind.**

Matt comments

- When a stimulus generates a qualia+links feeling in my brain, that's information to me. If **you want me to tell you what it feels like, I will describe the link structure it made me feel** (and impersonate a dictionary). If **my link structure make sense to you, we have aligned ourselves on that stimulus.** My links structure is somewhat congruent to yours, partner.
- We know that sensory stimuli are mapped reasonably conformally from the sensor hardware (like retinas or cochleas) onto primary cortex. This has to be the where the Libet soup computation begins, and the sign of what that might be, is suggested by the densely interconnected cortical association connectivity and interconnection, that follows sensory mapping. If Libet is a correlator, searching through sheets of stored links, no EE or Opticker would be surprised by the extant cortico-cortical connectivity systems.

Capt Amerika reply

- I like your point - the definitions we are generating cannot be for the purpose of 'airtight' in court defense -
- The second point - 'reduction of uncertainty' - only useful when there are relatively few alternative choices worries me - and I certainly don't want to have to assume Gaussian noise - my hope was that we would define ***groupings of related qualia (like color, emotions, ...)*** - and within those groupings we could then use a set of math to describe the 'knowledge and data' necessary to disambiguate specific compound qualia - like 'tank' --- and sneak up on a theory of pattern recognition - something that would quantify the questions like --- you will never be able to break the 10% Netflix challenge if you only use data that represents these qualia axes

Adam Comments

- where we bring value in our model is attempting to optimize the quality of the information given a theory of mind that the transmitting agent has of the receiving agent, allowing him to modify his output to make it more consumable and more useful for the receiver.

Adam comments

- i worry that 'meaning' as we are trying to define it may still be another form of 'information', a subjective piece of our model evoked by data using our own internal knowledge set ***(**** I agree with this statement – meaning can't have meaning unless it reduces some aspect of uncertainty of the Agent's world model – thus it is information ****)***. it is as if the information that was generated by incoming stimulus/data is ***then re-used to stimulate*** the same agent and generate another level of information given that agent's knowledge set. ***(**** I again agree – so it begs the question of why differentiate the two – I was chasing this as insight into a CAPTCHA – determine the meaning of the symbols to the Agent using them – we might be able to cast similarly finding the information this symbol has for that Agent that generated it ****)***
- in your porsche example, photons with a certain wavelength enter the retina and serve as the stimulus. our knowledge set has the ability to recognize this as a discernible difference from the rest of the stimulus at that time, so it is now 'data'. this data is then processed upon given our knowledge set to evoke a piece of information, as we have now reduced the uncertainty in our world model that the color of that car should be represented as red instead of any other available colors. this red is the resulting information from the stimulus. **** completely agree ***

Adam comments

- now this is where you and i might differ. i think that what may be happening after this point is that **'redness' is then subsequently processed again, now as data, inside of the same agent, in this case your brain. the relevant links that we have come to associate with a piece of information may be nothing else than the results of processing that piece of information as data, a level up in a hierarchy**. ******* I agree **** maybe I should have said the meaning of a piece of data is all the information that is generated by a given Agent as a result of processing that piece of data ******
- we are comfortable in saying that what is information to one agent may be data to another, i think that the two also need to be defined differently depending on where in the processing path they play. **and if we are trying to capture the differences between why humans can make 'meaning' of things by generating a link set from a piece of information, it could be because of the supplemental processing steps that continue to work on a piece of evoked information**, where these steps are non-existent inside of a computer because their representation can only perform processing on a certain kind of data, once it has been transformed to information it CANT be processed on anymore from the same knowledge pool.

Walkaway definition of meaning

- ***maybe I should have said the meaning of a piece of data to a given agent (recall it can't be termed data without specifying the agent that is consuming it) is all the information that is generated by that given Agent as a result of processing that piece of data into an infon (piece of information) --- all the links / qualia***
-
- ***that is information may begot information since it can become data to subsequent processing – a given piece of data that gets converted to information by an agent can use that information as data in subsequent processing to generate additional information – the sum total of the impact to the representation of an agent by a piece of data is what the piece of data 'means' to that Agent***

Meaning – considerations for definition

- Note the meaning could be a recalled memory (the memory can serve as data that with knowledge can be converted to information {reduce the uncertainty of some aspect of the Agent's world model})– can be a projection (using imagination) {some olfactory data stimulates a memory that can be processed into information – a particular internal representation in my world model – but it reminds me (memory) of something – thus it 'means' something to me associated with that memory.

Relation between information and knowledge

- Relation with information – a symbol (used the word symbol to capture both qualia level information and Libet level information) disambiguates some aspect of the agent's representation of the world. The impact of the disambiguation to the entirety of the agent's world model is its meaning to that agent.

Example

- My Porsche is red. There is information in its redness to me. Its redness also has meaning to me. Of all the uncertainty I have available to my world model to account for the hue/saturation/brightness of my car I reduce that uncertainty when I associate with my car the symbol red = information = generation of a specific symbolic representation from the input data (note there is the word symbol information ‘red’ AND also the quale of redness level of information generated). In addition to the information associated with reduction of uncertainty to account for the data my senses received there are implications to my world model by the fact that I’ve associated with my car the color red. Red ‘means’ to me that the car has that ‘arrest me’ quale to any cop that might see me driving. Note the meaning has links to implications to my world model that aren’t always associated with the information generated from similar data. Sometimes that same hue/saturation/brightness has a different meaning – ‘firetruckness’.

Matt comments:

- I think that our model of links+quale is a functional definition of "meaning". To prove to someone else that you "know" the meaning of some stimulus requires only that you produce a link list that is reasonably congruent to that of the person you are speaking to.
- The **internal AHA that you get, to indicate that you know the meaning**, is your private realization that you have your links to the quale that the stimulus generates. **Dictionaries are lists of links that some lexicographer knew.**
- The links idea dismisses the philosophers' trick question: "What is the meaning of meaning?" They didn't have any clue about links+quale.

Matt comments

- I get uncomfortable with our attempts to generate ***airtight*** labels, [made out of words, which are (qualia+links)], to "define" data, information, knowledge, meaning...
- I also get uncomfortable with some of the proposed specific definitions like: "Information is: The result of the reduction of uncertainty". That's useful for me only when I'm standing in Shannon Land, where the uncertainty is the choice of one of a few possible trivial events in the presence of added Gaussian noise
- We all know that the **only thing information is**, in the qualia mind of the recipient of the data stimulus, is **qualia+links**. If there is **something we might like to call information in a machine, its nature is unlike the information in my qualia mind.**

Matt comments

- When a stimulus generates a qualia+links feeling in my brain, that's information to me. If **you want me to tell you what it feels like, I will describe the link structure it made me feel** (and impersonate a dictionary). If **my link structure make sense to you, we have aligned ourselves on that stimulus.** My links structure is somewhat congruent to yours, partner.
- We know that sensory stimuli are mapped reasonably conformally from the sensor hardware (like retinas or cochleas) onto primary cortex. This has to be the where the Libet soup computation begins, and the sign of what that might be, is suggested by the densely interconnected cortical association connectivity and interconnection, that follows sensory mapping. If Libet is a correlator, searching through sheets of stored links, no EE or Opticker would be surprised by the extant cortico-cortical connectivity systems.

Capt Amerika reply

- I like your point - the definitions we are generating cannot be for the purpose of 'airtight' in court defense -
- The second point - 'reduction of uncertainty' - only useful when there are relatively few alternative choices worries me - and I certainly don't want to have to assume Gaussian noise - my hope was that we would define ***groupings of related qualia (like color, emotions, ...)*** - and within those groupings we could then use a set of math to describe the 'knowledge and data' necessary to disambiguate specific compound qualia - like 'tank' --- and sneak up on a theory of pattern recognition - something that would quantify the questions like --- you will never be able to break the 10% Netflix challenge if you only use data that represents these qualia axes

Adam Comments

- where we bring value in our model is attempting to optimize the quality of the information given a theory of mind that the transmitting agent has of the receiving agent, allowing him to modify his output to make it more consumable and more useful for the receiver.

Adam comments

- i worry that 'meaning' as we are trying to define it may still be another form of 'information', a subjective piece of our model evoked by data using our own internal knowledge set ***(**** I agree with this statement – meaning can't have meaning unless it reduces some aspect of uncertainty of the Agent's world model – thus it is information ****)***. it is as if the information that was generated by incoming stimulus/data is ***then re-used to stimulate*** the same agent and generate another level of information given that agent's knowledge set. ***(**** I again agree – so it begs the question of why differentiate the two – I was chasing this as insight into a CAPTCHA – determine the meaning of the symbols to the Agent using them – we might be able to cast similarly finding the information this symbol has for that Agent that generated it ****)***
- in your porsche example, photons with a certain wavelength enter the retina and serve as the stimulus. our knowledge set has the ability to recognize this as a discernible difference from the rest of the stimulus at that time, so it is now 'data'. this data is then processed upon given our knowledge set to evoke a piece of information, as we have now reduced the uncertainty in our world model that the color of that car should be represented as red instead of any other available colors. this red is the resulting information from the stimulus. **** completely agree ***

Adam comments

- now this is where you and i might differ. i think that what may be happening after this point is that **'redness' is then subsequently processed again, now as data, inside of the same agent, in this case your brain. the relevant links that we have come to associate with a piece of information may be nothing else than the results of processing that piece of information as data, a level up in a hierarchy**. ******* I agree **** maybe I should have said the meaning of a piece of data is all the information that is generated by a given Agent as a result of processing that piece of data ******
- we are comfortable in saying that what is information to one agent may be data to another, i think that the two also need to be defined differently depending on where in the processing path they play. **and if we are trying to capture the differences between why humans can make 'meaning' of things by generating a link set from a piece of information, it could be because of the supplemental processing steps that continue to work on a piece of evoked information**, where these steps are non-existent inside of a computer because their representation can only perform processing on a certain kind of data, once it has been transformed to information it CANT be processed on anymore from the same knowledge pool.

Walkaway definition of meaning

- ***maybe I should have said the meaning of a piece of data to a given agent (recall it can't be termed data without specifying the agent that is consuming it) is all the information that is generated by that given Agent as a result of processing that piece of data into an infon (piece of information) --- all the links / qualia***
-
- ***that is information may begot information since it can become data to subsequent processing – a given piece of data that gets converted to information by an agent can use that information as data in subsequent processing to generate additional information – the sum total of the impact to the representation of an agent by a piece of data is what the piece of data 'means' to that Agent***

Wiltshire and 50 bits/sec

- He evidently has eidetic memory even though an adult; some children have it up to about 8 years of age and then lose it. If you go to Wikipedia, you can see a good article
- about it, including mention of Wiltshire. Some folks claim to have it because they THINK they can remember precise details of many things, but have never been properly
- checked; try NOW to draw the dashboard of your car and you will see how low- pass your memory is.

- So it looks like W., even as an adult, retains access to a real eidetic memory incedental to his other strange intellectal abilites and short comings.

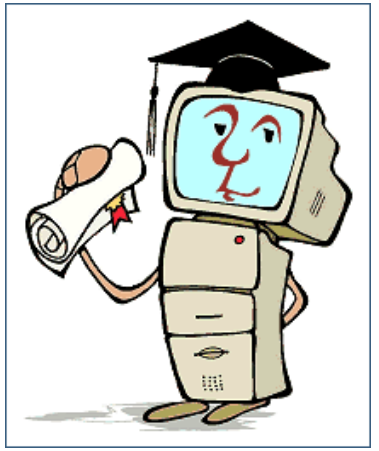
- matt

- --- "Rogers, Steven K CIV USAF AFRL/SN"
- <Steven.Rogers@WPAFB.AF.MIL> wrote:

- > Adam - asks the question - does he violate the 50 bits/sec rule?
- >
- >
- >
- > I suggested he doesn't based on watching him draw - he doesn't do it
- > that fast that made me wonder - but I wanted to think about it
- >
- >

Alternatives to solve problem

- Lots of people – even if we could resource the people they can't handle the data flow (***50 bits/sec throughput***)
- Autonomous computer solutions – after 50 years we have failed in automating conceptualization and processing in data space won't work
- Answer: ***integrated human/computer*** solution!



http://www.etproductions.com/images/ai.gif

Cyber, Automatic Target Recognition, Integrated Systems Health Monitoring, Nurse shift facilitator

50 bits / sec bandwidth of consciousness

- Igor was asking how to convert the human cognition type 2 processing to bits/sec in a Shannon sense - below is a cut/paste from an old discussion Matt /Adam and I had some time back when we were adding the **50 bits/sec tenet** (now a sub tenet in the simulation tenet)
- I think that that 50 bits/sec comes from requiring a subject to read unfamiliar text (like a newspaper article) as fast as he can. At about **2.5 bits/letter => about 7 or 8 bits/five letter word and about 300 words/minute**, you get around **40 bits/second**. If you **memorize the text, you can speak faster than that, but then the listener has trouble understanding what has been said (sounds like the lawyer-speak at the end of a TV contest offer)**. I used to extend this the vision too, by making an assumption about the number of pictures any person could identify, and rate at which he could do it; it also comes out at about 50 bits/sec.
- Suppose that the **receiver (in Shannon's formal channel) is a qualia decoder** (like the human visual system is) and is therefore looking for only a VERY small subset of all the possible signals (formally, an infinite number of possible world events).
- I think that **this channel, which consists of the real world as a transmitter** [of photons] to the receiver [which is the 50 bit/sec visual system] turns out to have an extremely high information transmission rate for the things that it cares about. In this way, the HVS evades Shannon rate limits (so much for that physics stuff).
- For QUEST to work this way and **exploit the power of qualia matching as a detector**, it will have to have some **efficient way of selecting what the qualia need to be for any specific task**.

QuEST view of creating the qualia vocabulary - **efficient way of selecting what the qualia need to be for any specific task**

- How can a 50 bit/second comm channel (like the human visual channel) **enable construction of an exquisitely detailed model of the real world**, in real time (with a slight 200 ms delay), inside the mind?
- Only because hardly any of the sensed world data are needed to cue up the already stored internal qualia out of which the world model gets constructed. **ONLY A QUALIA BASED SYSTEM CAN WORK the way animal sensory channels do**. Once in a while, the wrong qualia are triggered into the mind and we get nekkered (as in Neker cube); that's small price to pay for a very fast sensory analysis system.
- We spend the first years of our lives generating all the qualia we will use to internally compose the Cartesean theatre in our mind for the rest of our lives.
- What that means for QUEST is that we must be able to construct a set of internal qualia sufficient to span the entire set of things we expect to have to identify (make a list of statements about). Notice that fovea based visual systems avoid having to generate lots of possible qualia (that would be needed to compensate for PREDICTABLE variations in the real world, namely scale and rotation transformations), by building log r/theta hardware.
- I don't think the web 3.0 folks have the least idea of things like this; it would be like us trying to do PR in pixel space. Their approaches will never scale - will never be able to handle the Biederman issues
- Capt amerika

Low Bandwidth Sys2 characteristic: 50 bits / sec

What if the 50 bits/sec are what is the goal of consciousness to generate – and use to provide 'context' for sys1 decisions that it closes with respect to the environment


The New York Times

Inspired by the Traditions of Africa
But Baked by a Contemporary Spirit

Read a newspaper article as fast as you can. At about 2.5 bits/letter => about 7 or 8 bits/five letter word and about 300 words/minute, you get around 40 bits/second.

Bottom line: Human Brain is limiting – key to ability is finding a useful representation of world – World Model that can be consistently generated and efficiently exploited to respond appropriately to stimuli.

Are qualia required for exformation? Victor Hugo example?

Bits of awareness

- Since we define information as the reduction of uncertainty in the representation – the resolution along each distinct axis in a critter's world model defines that critter's 'bits of awareness' for that quale (some might cause this different levels of awareness {the fact a dog has more resolution in the olfactory representation than a human} – we suggest that is a different level of fidelity with representing 'reality')
- We define **'exformation'** as the reduction in uncertainty (generation of qualia) not associated with 'sensation' - it didn't come through the sensors - but is part of cartesean theater - you might say it is 'imagination' - provides the context scaffolding upon which you can anchor with sensory data mental counterparts - sensation

This is the idea I was pushing on using a virtual environment as the scaffolding upon which we place sensations (computer representation of sensor data) for quest solns!

Exformation

- exformation is 'context' - it is the reduction of uncertainty in the qualia representation not including the mental counterparts for sensory data = sensation ('reduction of uncertainty = meaning that values for some qualia are being set and not left undefined - although not to be confused with idea that 'reduction of uncertainty' implies matching to reality reducing the uncertainty of the representation with respect to its fidelity with respect to accurately capturing the physical reality) -
-
- Some of the qualia are associated with the internal representation at the introspectively available (Cartesean theater level) of the sensory captured stimuli = data, but there is much more represented in the Cartesean theater than the direct mental counterparts of the sensory data (sensation) and that is generated internally to provide a complete theater - wasn't sensed - it was *created internally to provide context for the sensed data* - that generated internally qualia we call exformation -
-
- It is the result of generating an accepted plausible narrative where much of it being composed of 'imagined' qualia not directly evoked from sensory data.

Exformation is the reduction of uncertainty resulting from the process of ‘thinking’ – the manipulation of the qualia results in setting values for qualia without additional sensation – that is exformation

Exformation additional thoughts

- I want to propose a change in our view of ‘context’ – I want input to an agent ALWAYS be through the sensors and from the stimuli pool – some of the stimuli pool is from the external to the colony of agents and some is from other agents – it is the job of the sensors and the sensor knowledge to be able to process the stimuli into data – thus I will say some data can be extracted from the outputs from other agents –
-
- Context on the other hand will be the ‘information’ that is generated as a result of ‘thinking’ – as a result of manipulation of the current set of qualia – to distinguish this source of reduction of uncertainty we might use the term ‘exformation’ to be consistent with the view in the book the ‘user illusion’ – or we might coin a new term as per the ox comment
-
- Exformation is the reduction of uncertainty resulting from the process of ‘thinking’ –
- the manipulation of the qualia results in setting values for qualia without additional
- sensation –
-
- the sensation (mental counterpart in the publisher’s representation when received the character ‘?’) – resulted in lots of qualia in addition to the representation of the character – all the rest of that qualia being set reduces the uncertainty of those qualia from many potential values to some specific settings – that reduction of uncertainty is a special type of ‘information’ that we have been calling ‘exformation’

Prior view of Exformation, Context and Agents

- With respect to alignment – our old view is that when two agents (be that between people or between a human and a computer) are able to directly incorporate the output from the other agent into the reduction of uncertainty of some aspect of its internal representation we will say that aspect of their internal representations are 'aligned' - we have called the input mechanism to an agent for this the 'context' input - context inputs have to be aligned - so the Knowledge of other agent = Ka allows the saving of the normal work required to transduce sensor inputs to information -
-
- the idea is that reduction of uncertainty in the internal representation through the context input is exformation

WE WON'T USE THIS VIEW ANYMORE – CONTEXT IS RECEIVED FROM OUTSIDE AN AGENT BUT IS GENERATED AS A RESULT OF INTERNAL MANIPULATION OF THE QUALIA – IT IS THE RESULT OF THE PROCESS OF THINKING

Exformation and Libet

- Since we've defined 'information' as the reduction of uncertainty in the internal representation – 'information' can be generated as a result of processing the data using knowledge – where our agent discussion has focused
-
- OR
-
- it can also be generated via 'thoughts' – since this information (reduction of uncertainty of the internal representation) didn't come through a sensory channel but was created out of manipulation of the existing qualia using prior experiences (memory) and knowledge we wanted to call this out as distinct and thus called it 'exformation' (context)
-
- what about Libet
-
- the issue comes about that Libet is certainly part of the internal representation – it is not introspectively available but is critical to the critter – sensory data can result in information at the Libet level – Do we want to allow exformation at this level also – my suggestion is NO– prior experiences (memory) and knowledge can complement the sensory data for a richer internal representation even at the Libet level – BUT WE WON'T ALLOW LIBET TO STORE IN STM – WON'T HAVE THE QUALIA REPRESENTATION AVAILABLE FOR THINKING – THE WHOLE REASON FOR QUALIA IS FOR THE GENERATION OF EXFORMATION

To Make Sense of the Present, Brains May Predict the Future

By J O R D A N A C E P E L E W I C Z

Notes by Cap v2

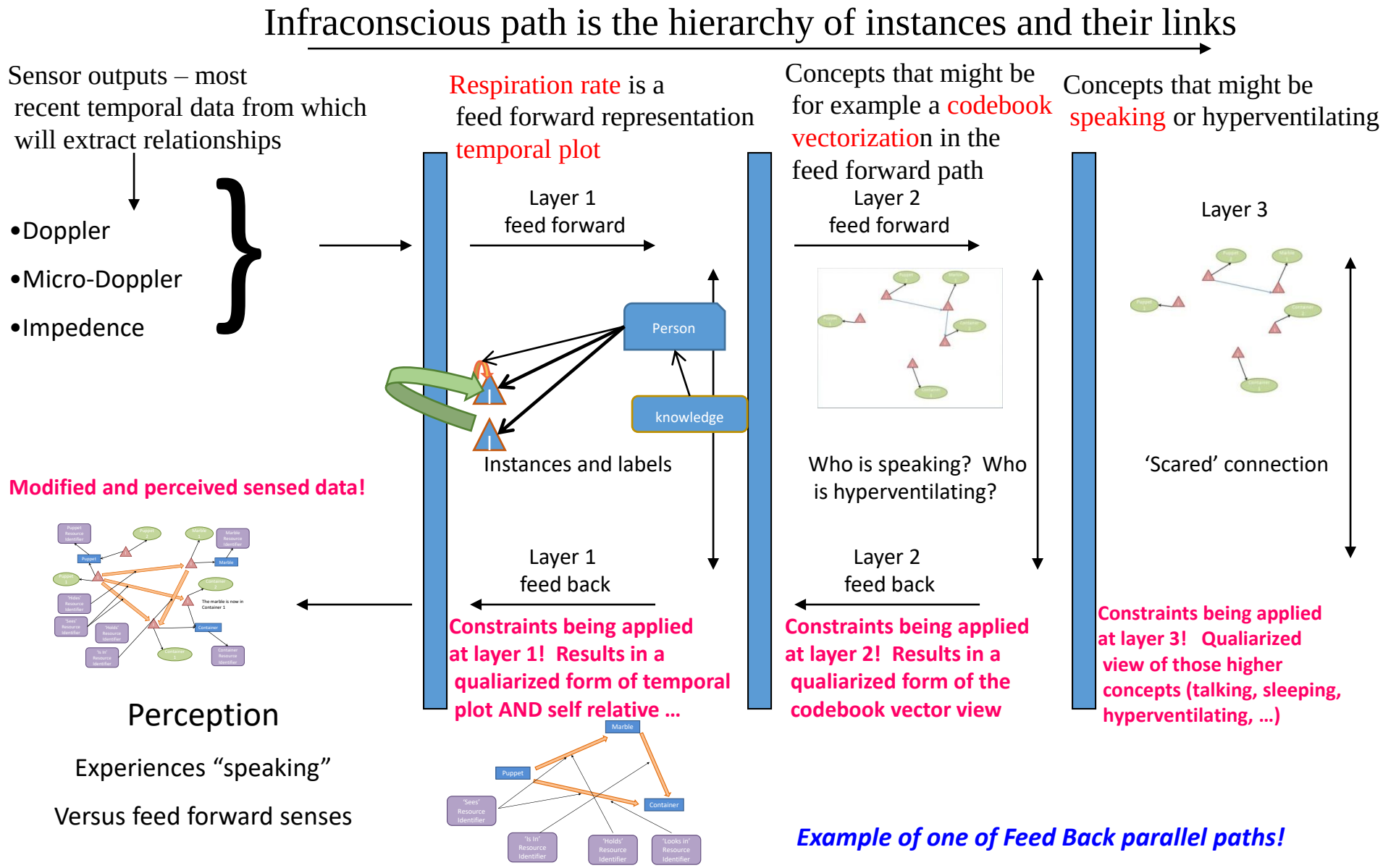
Aug 2018

https://www.quantamagazine.org/to-make-sense-of-the-present-brains-may-predict-the-future-20180710/

July 10, 2018

- *A controversial theory suggests that **perception, motor control, memory and other brain functions** all depend on comparisons between ongoing actual experiences and the brain's modeled expectations.*
- *QuEST has suggested in fact the conscious perception is in fact an expectation modulated by the sensed data – a confabulation, an 'imagined present' in the words of Edelman*

Hierarchical Quest Architecture
'Chunking' for complex problems from QuEST 2007



Perception as ‘controlled hallucination’ – QuEST compliant! For ‘conscious’ part of perception

- Some neuroscientists favor a predictive coding explanation for how the brain works, in **which perception may be thought of as a “controlled hallucination.”**
- This theory emphasizes the brain’s expectations and predictions about reality rather than the direct sensory evidence that the brain receives.

Generative Query Network, or GQN

- Last month, the artificial intelligence company DeepMind introduced new software that can take a single image of a few objects in a virtual room and, without human guidance, **infer what the three-dimensional scene looks like** from entirely new vantage points.
- Given just a handful of such pictures, the system, dubbed the **Generative Query Network, or GQN**, can successfully model the layout of a simple, video game-style maze.

Evokes in me GAN but instead of just noise and does it look real you provide viewpoint

- There are obvious technological applications for GQN, but it has also caught the eye of neuroscientists, who are particularly interested in the training algorithm it uses to learn how to perform its tasks.
- From the presented image, GQN generates predictions about what a scene should look like — **where objects should be located**, how shadows should fall against surfaces, which areas should be visible or hidden based on certain perspectives — and uses the differences between those predictions and its actual observations to improve the accuracy of the predictions it will make in the future.
  - "It was the **difference between reality and the prediction that enabled the updating of the model**," said Ali Eslami, one of the project's leaders.

Here the 'model' is the prediction model – it uses the 'error' between predictions and 'reality' to next time make a better prediction

Consistency constraint

- According to Danilo Rezende, Eslami's co-author and DeepMind colleague, "the algorithm changes the parameters of its [predictive] model in such a way that next time, when it encounters the same situation, it will be less surprised."
- Seems to be a way to do the consistency part of the QuEST stability, consistency and usefulness constraint on conscious perception

“predictive coding” theory

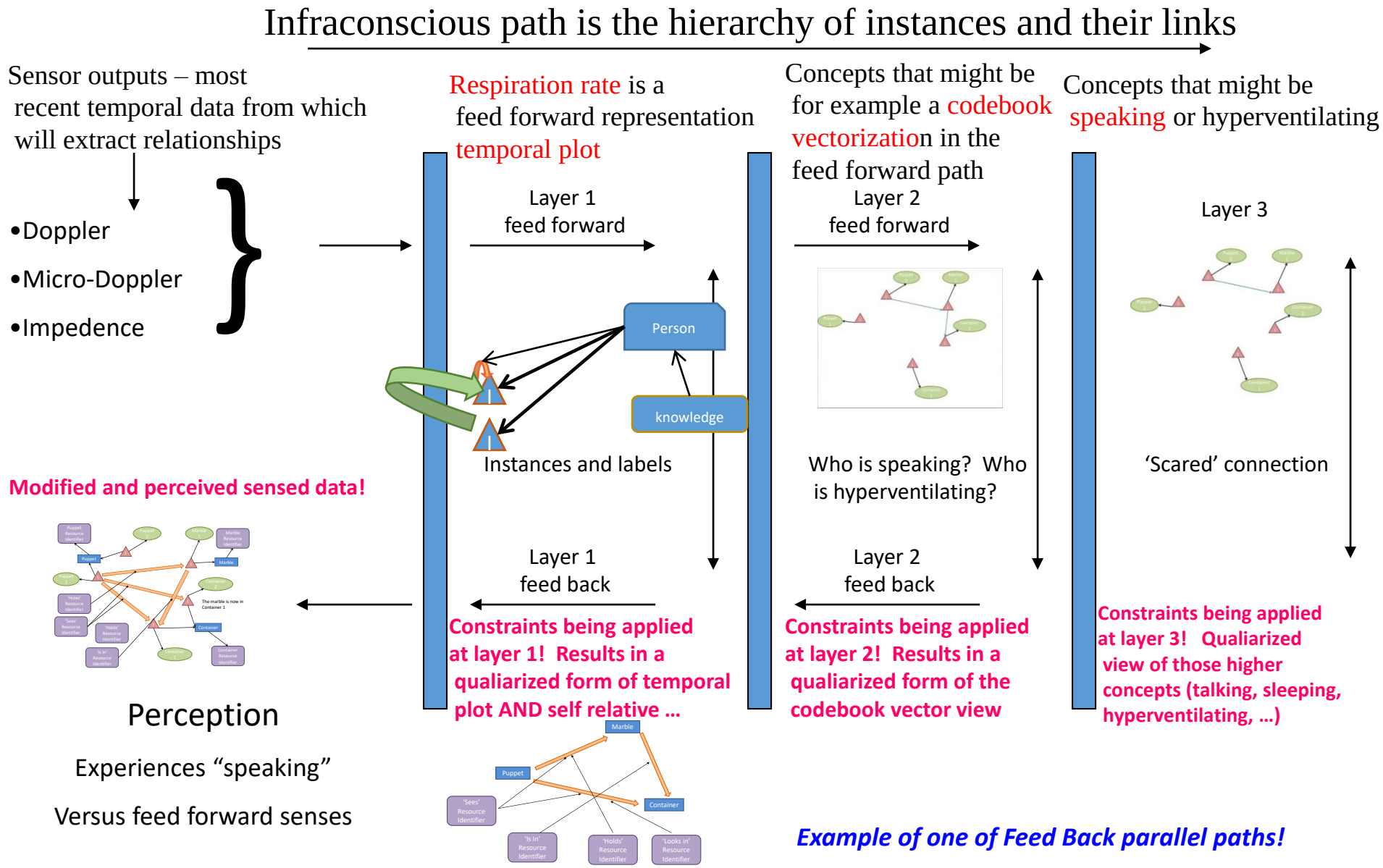
- **Neuroscientists** have long suspected that a similar mechanism drives how the brain works.
  - (Indeed, those speculations are part of what inspired the GQN team to pursue this approach.)
- According to this **“predictive coding” theory**, at each level of a cognitive process, the **brain generates models**, or beliefs, about what information it should be receiving from the level below it.
- These beliefs get translated into predictions about what should be experienced in a given situation, providing the best explanation of what’s out there so that the experience will make sense.

Recall the 10:1 ratio from cortex to lower level LGN like locations

- The **predictions then get sent down as feedback to lower-level sensory regions of the brain**.
- The brain compares its predictions with the actual sensory input it receives, "explaining away" whatever differences, or prediction errors, it **can by using its internal models to determine likely causes for the discrepancies.**
  - (For instance, we might have an internal model of a table as a flat surface supported by four legs, but we can still identify an object as a table even if something else blocks half of it from view.) **** my understanding is the representation is NOT at this symbolic level – it is GAN like ****

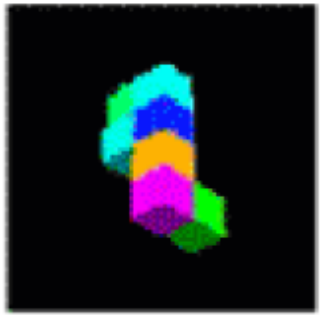
I'm not convinced there is a need to 'explain away' differences – if in fact the brain is perceiving the expectation it doesn't explain away – but does use the difference to 'learn' what to expect

Hierarchical Quest Architecture
'Chunking' for complex problems from QuEST 2007

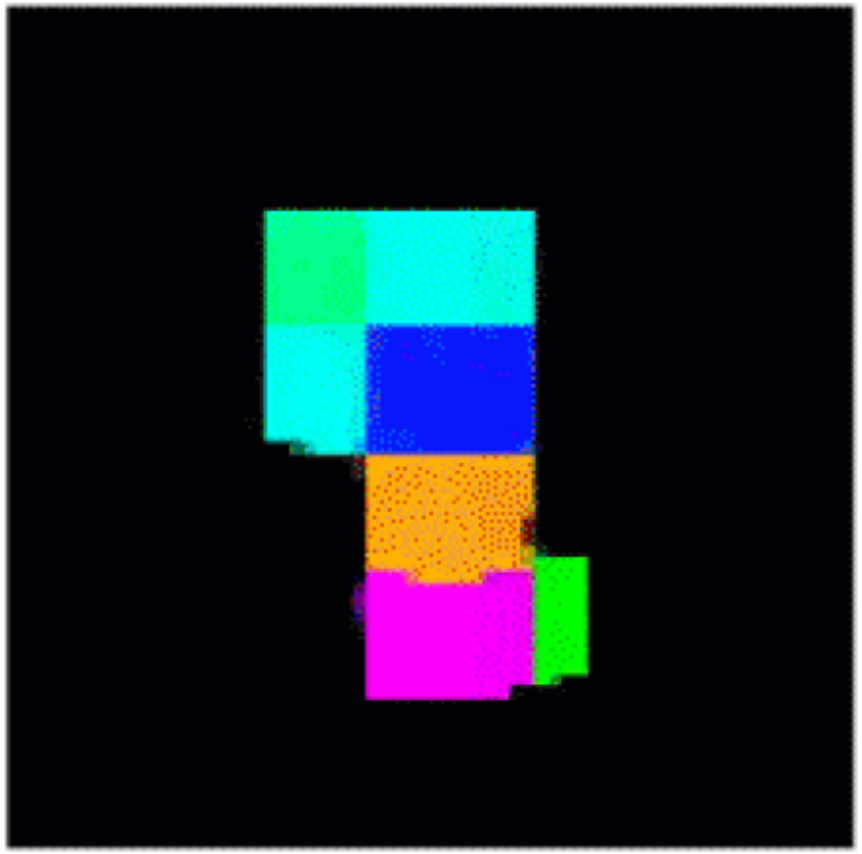


two-dimensional image of a pattern of blocks

Observation



Neural rendering



the General Query Network artificial intelligence **can infer their three-dimensional arrangement in space (right).**

neuroscience theory known as predictive coding

- Given a **two-dimensional image of a pattern of blocks** (left), the General Query Network artificial intelligence **can infer their three-dimensional arrangement in space (right).**
- The system relies on some of the same fundamental insights that underlie the **neuroscience theory known as predictive coding.**

predictive coding hypothesis

- The **prediction errors** that can't be explained away get passed up through connections to higher levels (as "feedforward" signals, rather than feedback), where **they're considered newsworthy**, something for the system to pay attention to and deal with accordingly.
- "The game is now about **adjusting the internal models**, the brain dynamics, so as to **suppress prediction error**," said Karl Friston of University College London, a renowned neuroscientist and one of the pioneers of the predictive coding hypothesis.

'explain away' versus adjusting the model to reduce the error – the latter is straightforward

Just perception or all of cognition

- Over the past decade, cognitive scientists, philosophers and psychologists have taken up predictive coding as a compelling idea, **especially for describing how perception works**, but also as a more ambitious, all-encompassing theory about what the entire brain is doing.
- Experimental tools have only recently made it possible to start directly testing specific mechanisms of the hypothesis, and some papers published in the past two years have provided striking evidence for the theory.
- Even so, it remains controversial, as is perhaps best evidenced by a recent debate over whether some landmark results were replicable.

Coffee, Cream and Dogs

- "I take coffee with cream and ____."
  - It seems only natural to fill in the blank with "sugar."
- That's the instinct cognitive scientists Marta Kutas and Steven Hillyard of the University of California, San Diego, were banking on in 1980 when they performed a series of experiments in which they **presented the sentence to people, one word at a time on a screen, and recorded their brain activity.**
- Only, **instead of ending with "sugar," when the last word popped into place, the sentence read: "I take coffee with cream and *dog*."**

greater brain response when the study's subjects came across the unexpected word "dog,"

- The researchers observed a greater brain response when the study's subjects came across the unexpected word "dog," characterized by a specific pattern of **electrical activity, known as the "N400 effect,"** that peaked approximately 400 milliseconds after the word was revealed.
- But how to **interpret it** remained unclear.
  - Was the brain reacting because the word's meaning was **nonsensical** in the context of the sentence?
  - Or might it have been reacting because the word was simply **unanticipated**, violating whatever predictions the brain had made about what to expect?

N400 effect - https://en.wikipedia.org/wiki/N400_(neuroscience)

- The N400 is a negative component, relative to reference electrodes placed on the mastoid processes (the bony ridge behind the ear), and relative to a 100 ms pre-stimulus baseline. Its amplitude can range from -**5 to 5 microvolts**
- The **N400** is a component of time-locked EEG signals known as event-related potentials (ERP). It is a negative-going deflection that peaks around 400 milliseconds post-stimulus onset, although it can extend from 250-500 ms, and is typically maximal over centro-parietal electrode sites. The N400 is part of the normal brain response to words and other meaningful (or potentially meaningful) stimuli, including visual and auditory words, sign language signs, pictures, faces, environmental sounds, and smells

- “
- Different brain regions … trade in
- different kinds of prediction.
- ”

In support of violating predictions hypothesis

- In 2005, Kutas and her team conducted **another study** that pointed to the latter possibility. **** unanticipated ****
- People were again asked to read a sentence one word at a time on a screen:
  - "**The day was breezy so the boy went outside to fly ____.**"
  - Because "a kite" seemed the most likely way to finish the sentence, the subjects expected to see "a" next, a word that had no intrinsic meaning but did predict the word to follow.
  - When the participants saw **"an"** instead (as in "an airplane"), they experienced an N400 effect, seemingly because the brain had to process a mismatch between its expectation and reality.
  - The **effect was apparently unrelated to the meaning of the word or any difficulty in processing the presented stimulus itself.**

Or might it have been reacting because the word was simply unanticipated, violating whatever predictions the brain had made about what to expect? --- the latter possibility

several labs have been unable to replicate the result.

- The 2005 finding seemed like a great fit for the predictive coding framework.
- But this past April, a paper published in *eLife* reported that several labs have been unable to replicate the result.
- Now, other researchers have started to respond, some claiming that subtleties in the replication methods still favor the prediction-based interpretation.

Experiments like Kutas' are subject to many interpretations

- This back-and-forth reflects much of the debate that's surrounded predictive coding.
- **Experiments like Kutas' are subject to many interpretations**.
- They **can be explained by models other than predictive coding**, and they fall short of definitive proof of the hypothesis because they don't delve into the actual mechanisms at play.
- While the idea that the brain is constantly making inferences (and comparing them to reality) is fairly well-established at this point, proponents of predictive coding have been seeking ways to prove that their particular version of the story is the right one — and that it **extends to all of cognition**.

Bayesian Brains and Efficient Computing

- The foundational insight that the brain perpetually makes and evaluates its own predictions about ongoing experiences wasn't always taken for granted.
- The view of neuroscience that dominated the 20th century characterized the brain's function as that of a feature detector:
  - It registers **the presence of a stimulus, processes it, and then sends signals to produce a behavioral response**.
  - Activity in **specific cells reflects the presence or absence of stimuli in the physical world.**
  - Some neurons in the **visual cortex, for instance, respond to the edges** of objects in view; others fire to indicate the objects' orientation, coloring or shading.

detector neurons for lines stop firing even though the line hasn't disappeared

- But the process turned out to be far less straightforward than it seemed.
- Further tests found that as the brain perceives, say, a longer and longer line, the **detector neurons for lines stop firing even though the line hasn't disappeared.**
- And the fact that **so much information seemed to be communicated through mysterious top-down feedback connections suggested that something else was going on**.

brain is an inference machine

- “
- The idea is that if the **brain is an inference machine**, an organ of statistics, then when it goes wrong, it’ll make the same sorts of mistakes a statistician will make.
- ”
- Karl Friston, University College London

“Bayesian brain”

- That’s where the **“Bayesian brain”** comes into play, a general framework with roots dating back to the 1860s that flips the traditional model on its head.
- The theory proposes that the brain makes probabilistic inferences about the world based on an internal model, essentially calculating a **“best guess” about how to interpret** what it’s perceiving (in line with the rules of Bayesian statistics, which quantifies the probability of an event based on relevant information gleaned from prior experiences).

Perception as “controlled hallucination.”

- Rather than waiting for sensory information to drive cognition, the brain is always actively constructing hypotheses about how the world works and using them to explain **experiences and fill in missing data**.
- That’s why, according to some experts, we might think of perception **as “controlled hallucination.”**

Bayesian brain also explains why visual illusions work

- In that vein, the Bayesian brain also explains why visual illusions work:
  - Two dots blinking in rapid alternation on a screen, for example, look like a **single dot moving back and forth**, so our brains unconsciously start to treat them like a single object.
  - Understanding how objects move is a higher-level type of knowledge, but it fundamentally influences how we perceive.
  - **The brain is simply filling in gaps in information — in this case, about motion — to paint a picture that's not entirely accurate.**

Maybe a subtle point but I am NOT convinced that 'filling in' is the right phrase – there is no need to fill in but to be able to predict – the concept of 'filling in' is a way to explain the perception but shouldn't be a guiding construct on what is actually happening – reminded of jared comment about 'virtual' world – just have to be able to construct it but you don't actually have to construct it to process it