Effing the Ineffable (2021) — Notes

Schwartz Reisman — March 17, 2021

I.Intro

A. History
   1. I needn't say, here, that recent progress in machine learning has revolutionized AI, and also raised challenges and issues.
   2. This seminar series has been a great venue in which to discuss the potentials and challenges raised by these systems.

B. Data-focused
   1. Though I haven't been to all the seminars, I believe it fair to say that the technical analyses of these systems have largely been carried on in terms of the data, representations, and algorithms we use to process them.
   2. I.e., as depicted in this image (◆ 02)
   3. That doesn't mean that the *concerns* people have brought to this seminar are entirely "data internal"
      a. The data & information are assumed to *represent some world or task domain* (◆ 03)
      b. When we label the data, for training purposes, we base the labels on our understanding of what it is (in the world) that we think the data we are looking at represents (◆ 04)
         i. We think this is a representation of a panda, this is a representation of pig, etc.
         ii. And that this is not a representation of a gibbon (whereas this is), not a representation of airplane, and so on.
      c. Similarly, understand bias, fairness, equity, etc., in terms of the people, society, medical practices, etc., that the data are about (◆ 05)
      d. Evaluation criteria and success rates are assessed in terms of measures on internal states and results, they are implicitly interpreted in terms of what is true of, or works in, the domains they represent.
      e. Insensitive Alexa responses, sexism and racism, etc. are implicitly referenced against our background understand of the world or worlds that the data represents.
   4. Nevertheless, the analyses we have heard here are framed in terms of the data itself (e.g., against distributions of identified categories of words) (◆ 06)

C. World
   1. What I am concerned with is with the relation between the networks, the data and representations encoded in the networks, the algorithms that run over them, etc., and the worlds that they represent. (◆ 07)
   2. Now many of the worlds that we train our networks on, and use them for, are *themselves representational*.
      a. X-ray, images of house numbers, maps, etc. (◆ 08)
      b. Even more are *words*—Twitter feeds (◆ 09)
      c. Or word-like, such as data sets that AI runs over are lexical in origin: data bases with explicit categories (age, sex, gender, nationality, blood type AB, price, in stock or not, etc.) (◆ 10)

3. We will get to those later
4. I want to start concretely
   a. With the ordinary physical world (◆ 11)
   b. I.e., want to understand what is going in AI systems interacting with, and directly connected to, the world
   c. Not mediated by our understanding of our understanding of the data …
5. I.e., start with cases of perception (◆ 12)
6. What is the relation between “input data” (e.g., pixel arrays on cameras, microphone samples for audio, etc.), the patterns and configurations of data that result from its processing in the machine, and the world that that data represents?

D. 〈 That was intro (◆ 13) 〉

II.Classical View

A. Start with the world (we will talk about its representation in a bit)

B. Classical view

1. In the first decades of AI—the 1960s and 1970s, back when I was a student in the MIT AI Lab—it was assumed that the world was relatively clear-cut
   a. An assumption that the world consists of **objects** of various types or kinds, manifesting various **properties**, and standing in various **relations**
      i. Also **sets** and collections (◆—group of houses)
      ii. Also **states of affairs** (◆—the house being on top of the hill)
      iii. Also **abstract objects**, such as numbers, those properties and relationships themselves, in the abstract, etc.
   b. I will call this the **classical view.**
      i. Basis of formal logic
      ii. Also: models of reasoning, set theory, formal semantics, etc.
2. The classical view underlay all the initial efforts in ai
   a. But still, this was assumed sufficient for AI
      i. (◆) KR (conceptual dependency, semantic nets, etc.
      ii. (◆) SHRDLU
   b. Robotics: made the world to fit the model! (◆ Shakey)
   c. More than that: it was assumed that the classical model was **correct**
      i. That is what the world consists of: objects, properties, etc.
      ii. So SHRDLU and SHAKEY were just simple versions.

C. Challenges

1. The classical view didn’t really work out
2. Here is a more realistic image of an almost-empty room (cabin) (◆ 21 — cabin)
3. History
   a. This was the time (mid-1970s) that digital cameras and digitizing imagers were just starting to be built (very first digital camera was released)
      i. (◆) Steve Sasson, Kodak, 8 lb, 0.01 megapixels[1]
   b. When people first looked at the data coming out of these imagers, they were

[1] https://www.diyphotography.net/worlds-first-digital-camera-introduced-man-invented/

flabbergasted by how *messy* the data looked

D. What is going on here?

1. Now as we all know, you can train a DL to “recognize” this picture

a. E.g., draw a box around the people, for example

b. The auto-focus algorithms on your camera—or the bounding boxes around people and cars and such in the videos showing what driverless cars are “seeing”

2. But how do we understand this?

III.World

A. There are two possibilities

B. Naïve

1. On the first option—which I will call the **naïve** option—the classical view is assumed still to be correct

a. There *is* a person there—or, rather, two people—and two windows (though you can see a third one in the mirror), and a table or bench or something (which is it?).

b. Any “messiness” in the incoming data merely reflects *uncertainty about the data-world relation*

c. We encode that with *probabilities*—probabilities that some way of classifying it is “correct”—or the one we want, or something like that.

2. This is the view that underlies our characterization of the problems with adversarial examples

a. (◆) Panda/gibbon

b. (◆) Macaw/bookcase

c. In such cases it is probably correct

3. This view also underlies the characterization of *certainty* of results, or *confidence in*

a. The world, we might say, is assumed to be “completely certain”

i. Or rather, since it doesn’t make sense to talk about the world’s being certain (certainty is only something you can say about knowledge of [or data about] the world), the world is *completely determinate*

b. Now in one sense the world *is* completely determinate (pace quantum mechanics): it *absolutely is what it is*

i. It is "complete in its being,” or however one wants to put such things.

4. But—and this is going to be important—what that is assumed to mean, in this context, is that the world is *100% determinate at the level at which it contains discrete objects, exemplifying determinate properties.*

C. Constructivism

1. But there is another interpretation—a second view about what is going on.

2. It addresses a variety of challenges that have been raised to the classical view

3. Properties

a. Determinacy/definitions

i. Take the property of being a *chair*

ii. As many people have pointed out, it is not clear that there is single thing that characterizes all chairs as chair

- iii. (Lots of pictures)
- iv. Wittgenstein: family resemblances
- b. Affordances
  - i. J J Gibson: a chair is what *affords sitting*.
    - a. Keep this in mind; because what is sitting? [⟸ mention later]
- c. Context
  - i. Context: log chair
- d. Boundaries of applicability
  - i. Large bean-bag chair
- e. Boundaries between and among properties themselves (not sharp-edged)
  - i. Egoistic, egotistic, pride, self-confidence, braggadocio, boastful, cocky, uppity, snooty, high-minded, pompous, …
- 4. Relationships—even worse!
  - a. “Next to” (“small,” “complicated,” etc.)
  - b. (ϕ thinks of them as *vague*—but I don’t think that is the right category)
- 5. Objects, too
  - a. Also contextual (example?)
  - b. Also not sharp-edged
    - i. Cf. the table in that workroom
      - a. Does it have a back, which is formed by the wall? Or does it *not* have a back, and instead is resting on the wall?
      - b. Who cares?
      - c. TDepends on what is at stake … move the table to the middle of the room?
      - d. “It is attached to the wall”
      - e. “Can you unattach it?”
      - f. “No, because the wall *forms* part of the table—forms the back”
  - c. So too with mereology:
    - i. Cf. relations between seat and arm; where does one start, other end?
  - d. Identity
    - i. Washington’s axe
    - ii. Ship of Theseus
  - e. Objectification
    - i. Clouds (story about camping and you look out of the tent)
    - ii. Fog (“Weather ‘it’ ”: foggy, raining, etc.)
  - f. Features
    - i. “It’s Mommy!” vs. “It’s Mommy-ing again!”
- 6. Constructivism
  - **a.** Considerations of this sort have led to what is sometimes called **constructivism**
    - *i.* Properties don’t exist or hold independent of us; Rather: *relative to our interests*
    - ii. *Objects* relative to our interests, too.

D. These considerations suggest a different view of what our networks are doing (or need to do)

- *1.* Rather than determine *what objects and properties there are out there*¨

2. Instead the task is to *find the world intelligible* in terms of relevant abstractions and idealizations
    a. In some sense, to impose some order on the booming buzzing confusion.
    b. But that is not right; there is order, or variety, or anyway stupefying richness, out there.
    c. It is a question of organizing it, finding patterns, and interpreting the world in terms of those patterns.
    d. Perhaps by doing cluster analysis on the perceptual array of real-number-valued high-dimensional vectors
3. But—and this is critical thing—finding those patterns, clustering that data, does not mean losing the richness.

E. Some evidence
1. A critic might say
    a. “Of course there are objects and properties out there”
    b.
2. But there is a problem
    a. You have *already processed* that picture, using exactly the sorts of processes that we are trying to understand!
3. Leads to a question: what is the world like “prior to that processing?”
    a. Show Adam’s picture
    b. Analogy: Georgian Bay

IV.Registration

A. We need some vocabulary

B. Constructivism/Realism
1. For decades a debate has raged between
    a. What is called “**realism**”, which takes the structure of the world to be out there, independent of us; and
    b. **Constructivism**, which argues that, rather than being independent of us , objects and properties are human constructs which reflect our cultures and societies and interests and such.
2. I think that way of framing the debate is fatal—never going to be resolved.
3. Chair
    a. Suppose a network “sees” a chair
    b. Has it “seen a chair *out there*”, independent of us?
    c. In one sense the answer is *yes*
        i. There it that, out there—a patch of reality—which affords sitting
        ii. Even if I don’t see it, if I sit down, it will support me
    d. In one sense the answer is *no*
        i. There is no *chair out there*, qua chair
    e. As I have already suggested, and what DL shows, is that the chair, or the concept chair, or something like that, is an abstraction or “construction”—a coarse graining—of what is out there, where “what it is out there” is…
        i. The chair is not the coarse-graining, of course—or the abstraction, or the ide-

alization itself—but that patch of the world which has been coarse-grainedly categorized.

- ii. Moreover, the chair isn't "the world as coarse-grained"
- iii. The chair isn't has been coarse-grained. If there is a spike that is ignored by the coarse-graining, in its classification of the chair as a chair, and I sit down on it, I will sit down on the spike, and it won't help me to say "the spike isn't part of the chair, per se, because the abstraction according to which it has been classified as a chair ignores it."

- f. Rather, we want to say something like this:
  - i. The chair is that non-idealized patch of reality, in its fullness, which warrants the idealized or coarse-grainedly classification *as* a chair,
  - ii. I.e., it is that patch of reality (out there) which allows the system to take it (in virtue of our interests and projects and practices and such) to be a chair (out there!)
- g. But that is an awkward way to talk
  - i. Randy: "What kind of furniture did they give you for your new office?"
  - ii. Pat: "That which, in virtue of my cultural embedding, projects, and purposes, I take to be that which I classify using the English word 'chair'"
  - iii. Randy: "You need to get a life!"

- 4. Story … Zen master and the acolyte

- C. So here is how I think we need to talk.
  - 1. I will say that I **register** something (a chair, a storfm)
  - 2. (xx◆xx—Raymond Carver)
    - a. What we refer to, think about, are talking about
  - 3. The world: a surpassingly and ineffably rich plenum, which we register in ways that allow us to find it intelligible, with respect to our projects and purposes…
- D. Not a middle ground *between* realism and constructivism, but an approach that incorporates what is best about, both realism and constructivism.
- E. One more example
  - 1. When we register, we refer
    - a. To *the patch of the world (in-its-fullness) that we register*
    - *b.* Not to the *world as registered.*
  - 2. … Mondrian…
  - 3.

V.Effability

- A. Intro
  - 1. OK, almost ready to get to the subject of the talk
  - 2. Just two more preliminary points.
  - 3. Compositionality
    - a. I haven't yet what our conceptual categories—those abstractions and idealizations, those information-losing coarse-grained clusterings, with which words are associated—are *for*.
    - b. There are three obvious (and related) potential answers:

i. COMPOSITIONALITY: these properties, types, and concepts (and the words we use to express them) support a kind of recursive compositionality we know from grammar:

a. "The infinite use of finite means"—as von Humboldt[2] put it

b. So while they abstract away from a stupendous amount of detail "below" them, they nevertheless support staggering levels of complexity "above" them—in recursively-constituted composite forms.

c. So they allow finite communication.

d. We don't have USB-C plugs in our spines, with GHz transmitters and receivers, capable of sharing big swaths or swatches of our slow but massively complex brain states.

ii. ABSTRACTION: It turns out that there are relatively high-level generalizations—high-level regularities—that the world sustains. Even if we don't know exactly how situation A falls under concept α, and how situation B falls under concept β, in many cases it is nevertheless overwhelmingly likely that there will be (or will follow, or whatever) another situation G that falls under concept γ.

a. Even without knowing the configurational details of the leaves of a tree that is on fire, and can't predict the fine-structure of the swirls of smoke that result, we can, in general, say—and know—that "fire produces smoke."

b. Or maybe it doesn't "turn out" that there are such regularities: maybe the only regularities that we can register are those with these high-level structures.

— This is a serious suggestion…

*iii.*LEARNING: Not only are these high-level regularities, and their ability to be expressed, rely on the detail-shedding abstractions, it is only if one sheds all those details that the concepts can be *learned.*

a. That is one reason why intermediate layers cannot be too big: the appropriate generalizations only "hold"

4. On the other hand

a. It is clear that the complexity that can be represented by our networks is vastly greater than that that can be expressed

b. Our networks can represent millions or billions of real-valued vectors

i. GPT3 is 175 billion parameters

c. …

B. Effing

1. Intro

a. "Effable" means "able to be described in words"

b. From Latin *effābilis*, from *ef-fāri*, to utter, from *ex* (out) + *fāri* (to speak).

2. Non-conceptual content

a. Suppose, while driving in Paris, I suddenly go blind.

i. "Hey, I can't see any more," I tell the passenger beside me.

ii. "But I like driving. Just tell me what to do, and I'll control the vehicle."

[2] Wilhelm von Humboldt, *Introduction to General Linguistics*, 1836.

b. Another example
    i. Adrian, police: “Do you know how fast you were going?”, says the police.
    ii. “In one way yes; I knew how far over to lean; how to stay on the inside of the curve, etc.
    iii. “In another way, no—e.g., how many miles per hour”
c. More
    i. Evans: footstep in the night…
    ii. “Hit your second tennis serve at the speed paper shoots out of the big copier at work.”
d. This is called **non-conceptual content** in philosophy

Relations
Mix effable and uneffable “Do you see those deer about to leap out across the road?” someone could say; or “Be careful; the road is slick.” But just how slick—one can know, but one cannot say.
If one sees a soccer ball come flying out past a hedge and go across the road, and one abruptly slows down, knowing that where goes an errant ball, a child is sure to follow.

Now suppose, while looking out in good light, I register a chair
Perhaps 〈 … do I have a chair I want to focus on? The wooden log? … 〉
Let’s suppose, reasonably, that I take in a lot of the visual detail. And suppose that visual input is put through network algorithm to classify it—and ‘chair’ comes out best.
But the resultant state—even the state that was selected as “highest probability”—will likely have a huge amount of information in it above and beyond its winning at that particular contest.
Action will require those fine-grained details
When we register an object as an object, does that mean that we shave off all the details, package it up in brown paper box, so that it ends up in as one indistinguishable chair—or as a box tagged with labels from all the properties or types we deemed it to deserve?
No. DL shows the way.

If I am going to remember that chair—why would I need to throw away all that detail?
Morals that I want to convey
What I *understand* may be vastly richer that its articulation in words
Think of the affordance of the chair: that it “offers sitting upon”.
That may be linked, via complex webs of connection strength, to routines for sitting—routines that in turn are tied into my physical motor capacities, etc.
Or think of the word ‘laugh’
If I say ‘laugh,’ *and you are someone who has laughed*, chances are that that word (the token of the word) will cause the deployment in you of a richly complex vector with, again, implicit ties into emotional states, into muscular patterns, etc.
If *Alexa* says ‘laugh,’ there is no reason to suppose that its pattern of activations will have any such web of connectivity

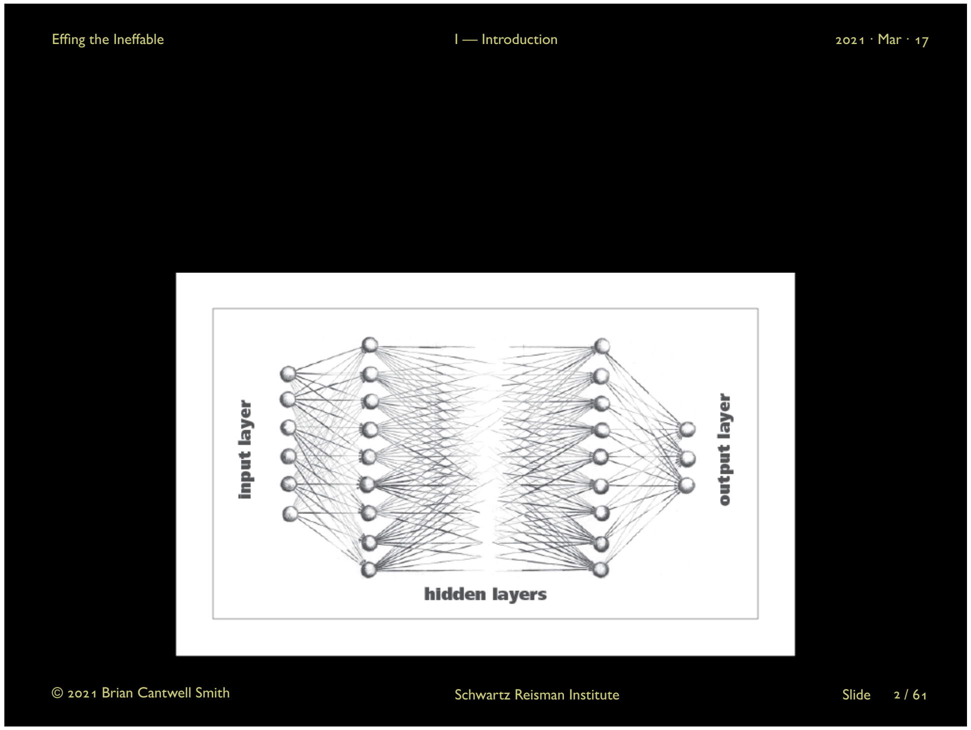
It might have webs of connectivity with other *words*—but not with the activity of laughing, or with emotional states that have engendered laughter, etc.
Cf. those horrifying Alexa responses that « … » showed two weeks ago, “Remind me to kill myself” ⟹ “I’ll put it on your schedule”
So conclusion #1
One thing that DL has shown us is that “expressed in words” doesn’t mean is captured merely in the word used and in its relations to other words. [■]

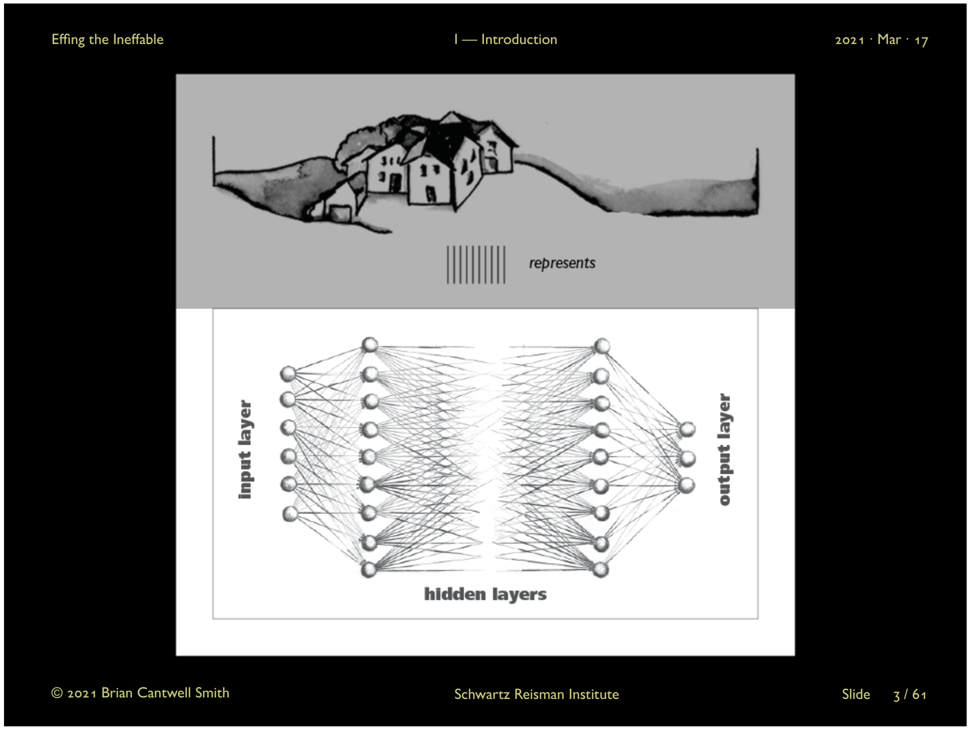
But what *is it* for something to be describable in words?
 Miscellaneous
Settling a lump into a variegated landscape (…)

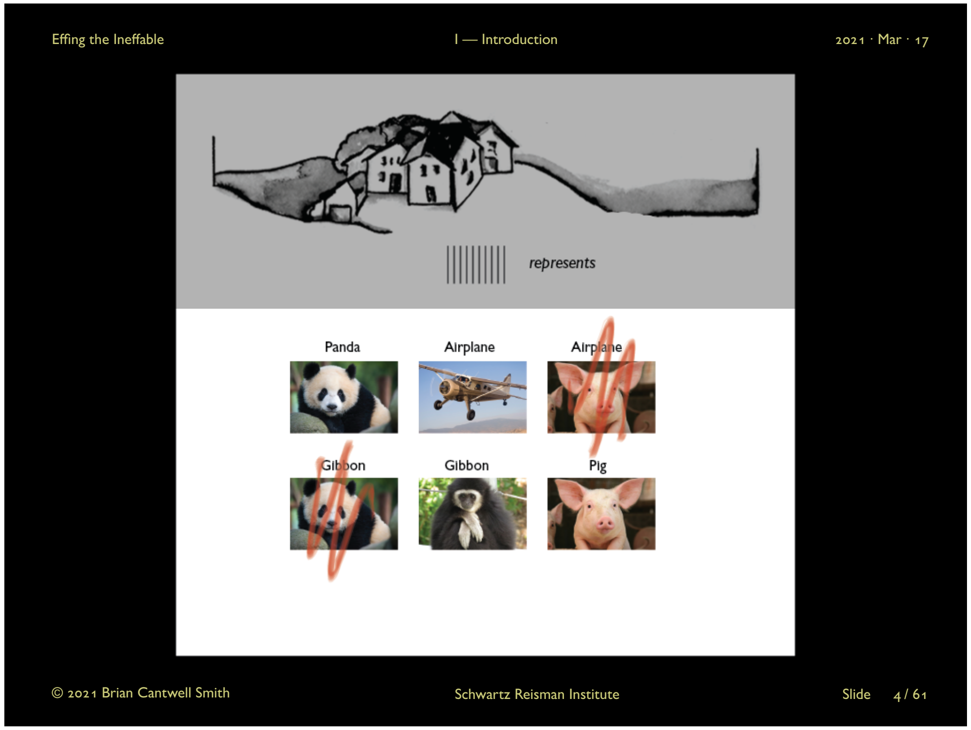
Effing the Ineffable

What AI Teaches Us About What Can and Cannot be Said

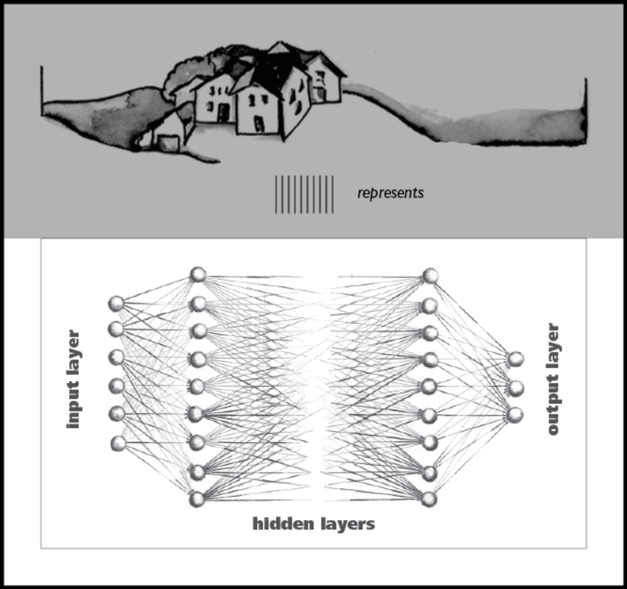
Brian Cantwell Smith
University of Toronto





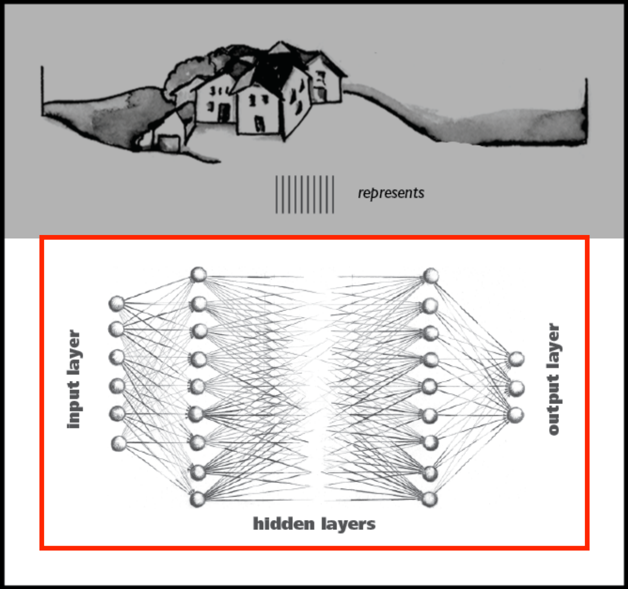


Effing the Ineffable I — Introduction 2021 · Mar · 17



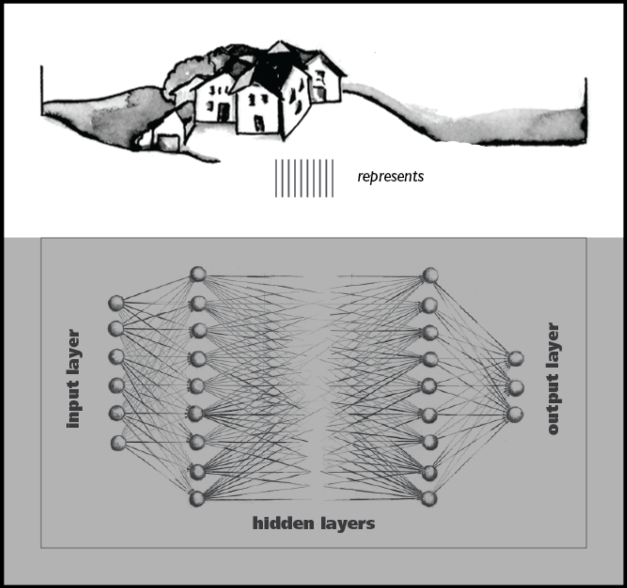
© 2021 Brian Cantwell Smith Schwartz Reisman Institute Slide 5 / 61

Effing the Ineffable I — Introduction 2021 · Mar · 17



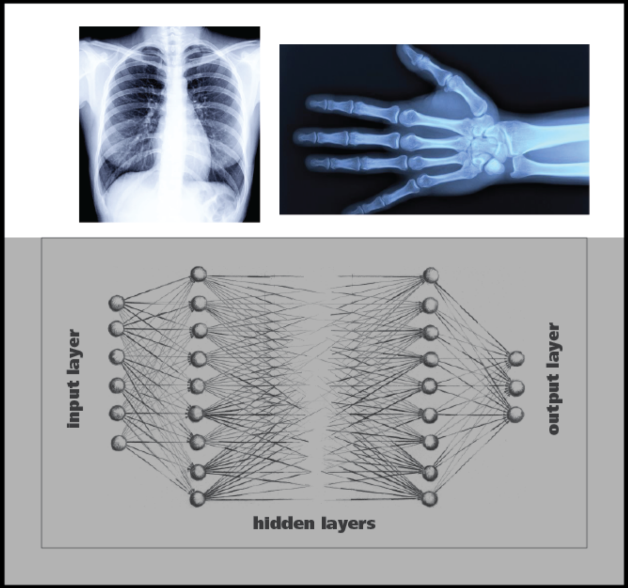
© 2021 Brian Cantwell Smith Schwartz Reisman Institute Slide 6 / 61

Effing the Ineffable I — Introduction 2021 · Mar · 17

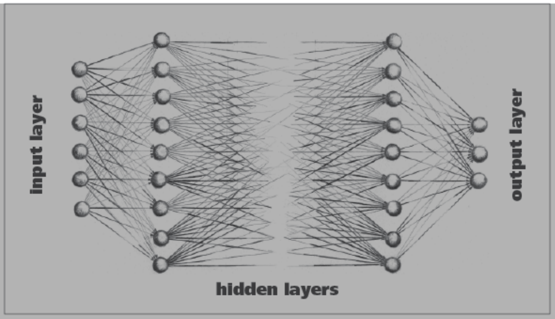


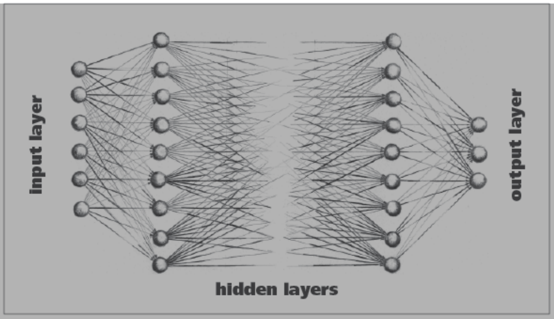
© 2021 Brian Cantwell Smith Schwartz Reisman Institute Slide 7 / 61

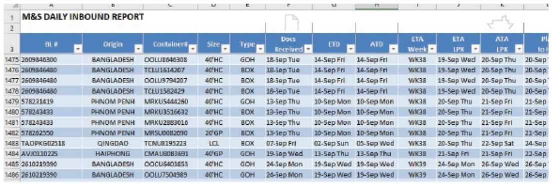
Effing the Ineffable I — Introduction 2021 · Mar · 17



© 2021 Brian Cantwell Smith Schwartz Reisman Institute Slide 8 / 61

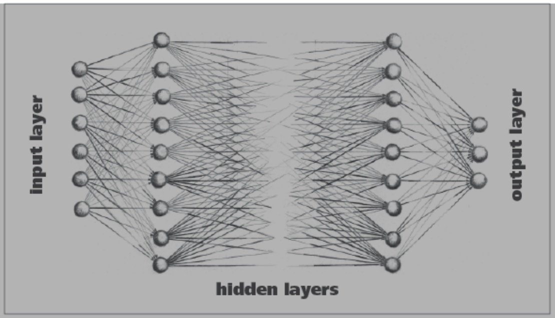


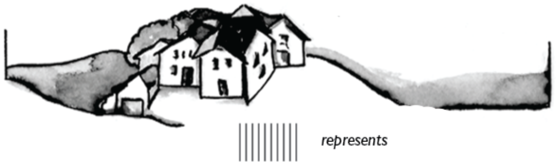


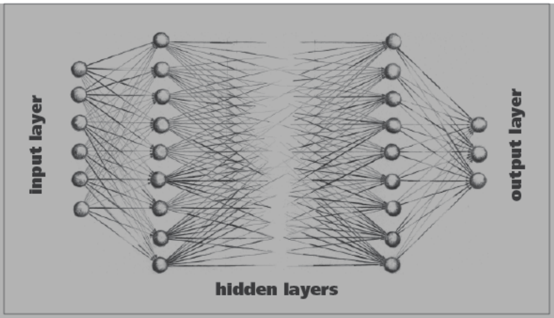


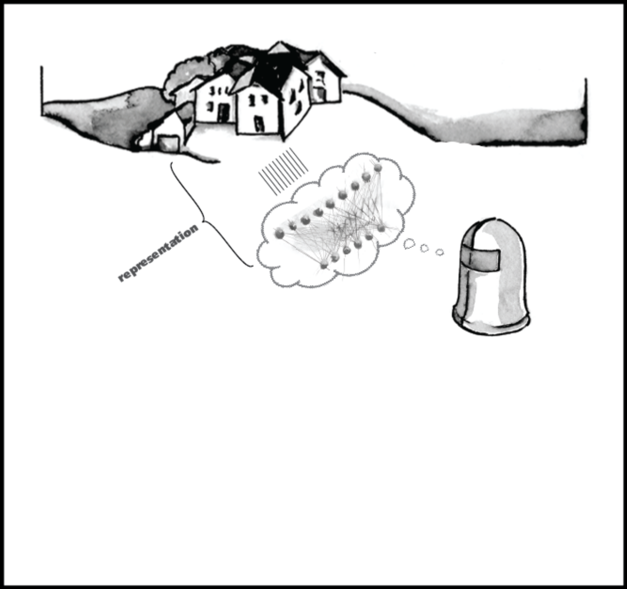
M&S DAILY INBOUND REPORT

BL #	Origin	Container#	Size	Type	Docs Received	ETD	ATD	ETA Week	ETA LPK	ATA LPK	Pl to
2609846300	BANGLADESH	OOLU8646308	40'HC	GOH	18-Sep Tue	14-Sep Fri	14-Sep Fri	WK38	19-Sep Wed	20-Sep Thu	20-Sep
2609846480	BANGLADESH	TCLU1614207	40'HC	BOX	18-Sep Tue	14-Sep Fri	14-Sep Fri	WK38	19-Sep Wed	20-Sep Thu	20-Sep
2609846480	BANGLADESH	OOLU9794207	40'HC	BOX	18-Sep Tue	14-Sep Fri	14-Sep Fri	WK38	19-Sep Wed	20-Sep Thu	20-Sep
2609846480	BANGLADESH	TCLU1582429	40'HC	BOX	18-Sep Tue	14-Sep Fri	14-Sep Fri	WK38	19-Sep Wed	20-Sep Thu	20-Sep
578231419	PHNOM PENH	MRKU5444260	40'HC	GOH	13-Sep Thu	10-Sep Mon	10-Sep Mon	WK38	20-Sep Thu	21-Sep Fri	21-Sep
578243433	PHNOM PENH	MRKU3516632	40'HC	BOX	13-Sep Thu	10-Sep Mon	10-Sep Mon	WK38	20-Sep Thu	21-Sep Fri	21-Sep
578243433	PHNOM PENH	MRKU2883016	40'HC	BOX	13-Sep Thu	10-Sep Mon	10-Sep Mon	WK38	20-Sep Thu	21-Sep Fri	21-Sep
578262550	PHNOM PENH	MRSU0082690	20'GP	BOX	13-Sep Thu	10-Sep Mon	10-Sep Mon	WK38	20-Sep Thu	21-Sep Fri	21-Sep
TAOPKG02518	QINGDAO	TCNU8195223	LCL	BOX	07-Sep Fri	02-Sep Sun	05-Sep Wed	WK38	20-Sep Thu	22-Sep Sat	24-Sep
AVJ0110225	HAIPHONG	CMAU8083691	40'GP	GOH	19-Sep Wed	13-Sep Thu	13-Sep Thu	WK38	21-Sep Fri	21-Sep Fri	22-Sep
2610219390	BANGLADESH	OOCU6403853	40'HC	GOH	24-Sep Mon	19-Sep Wed	19-Sep Wed	WK39	24-Sep Mon	26-Sep Wed	26-Sep
2610219390	BANGLADESH	OOLU7504989	40'HC	GOH	24-Sep Mon	19-Sep Wed	19-Sep Wed	WK39	24-Sep Mon	26-Sep Wed	26-Sep









⬆ I. Introduction
II. Classical View
III. World
IV. Registration
V. Effability

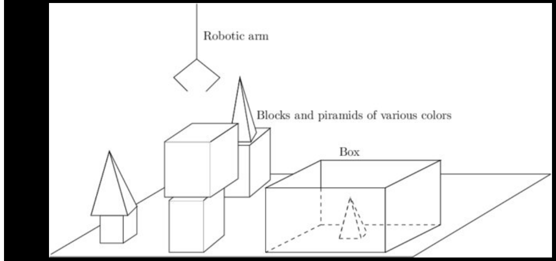
I. Introduction
➡ II. Classical View
III. World
IV. Registration
V. Effability

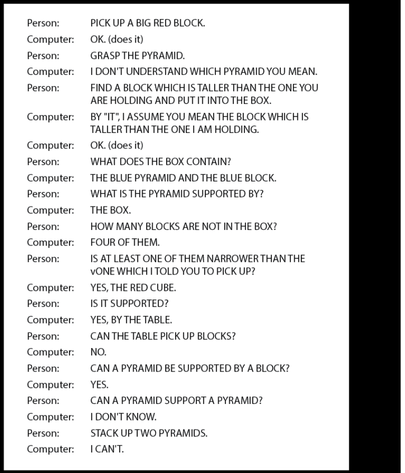
Classical View

1. The world consists of **objects**, **properties**, and **relations**
   a) Perhaps also **sets**, **states of affairs**, and **abstract objects** (like numbers)
   b) The entities
      i) Represented in data base entries
      ii) Assumed in GOFAI
      iii) The things we have words for in natural language (**maybe!**)

   } *Basis of formal logic*
2. A form of naive realism that GOFAI took to be the *fundamental structure of the world*
   a) Assumed in SHRDLU
   b) Assumed for SHAKEY—and the world adjusted to fit it!

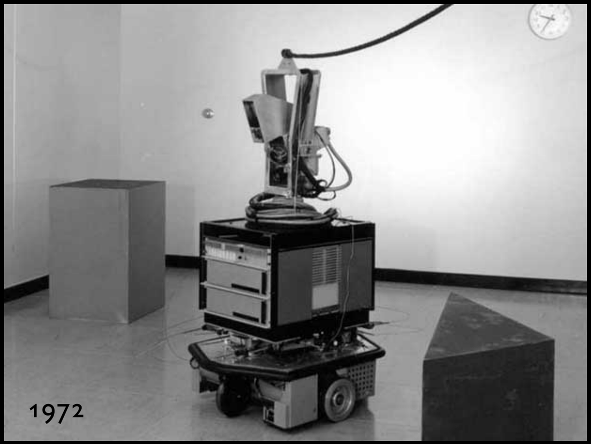
SHRDLU





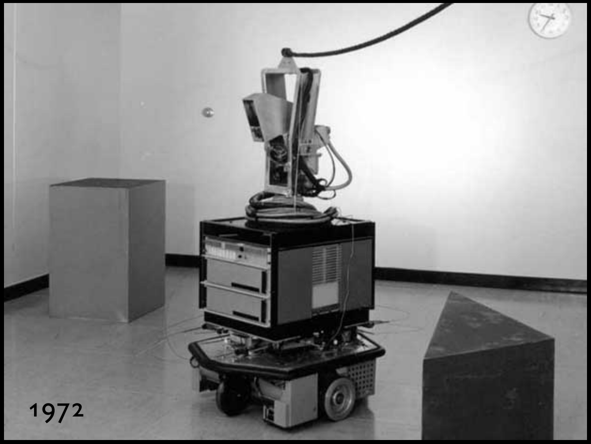
Person: PICK UP A BIG RED BLOCK.
Computer: OK. (does it)
Person: GRASP THE PYRAMID.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer: OK. (does it)
Person: WHAT DOES THE BOX CONTAIN?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
Person: WHAT IS THE PYRAMID SUPPORTED BY?
Computer: THE BOX.
Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
Computer: FOUR OF THEM.
Person: IS AT LEAST ONE OF THEM NARROWER THAN THE vONE WHICH I TOLD YOU TO PICK UP?
Computer: YES, THE RED CUBE.
Person: IS IT SUPPORTED?
Computer: YES, BY THE TABLE.
Person: CAN THE TABLE PICK UP BLOCKS?
Computer: NO.
Person: CAN A PYRAMID BE SUPPORTED BY A BLOCK?
Computer: YES.
Person: CAN A PYRAMID SUPPORT A PYRAMID?
Computer: I DON'T KNOW.
Person: STACK UP TWO PYRAMIDS.
Computer: I CAN'T.

SHAKEY — at the "other" SRI



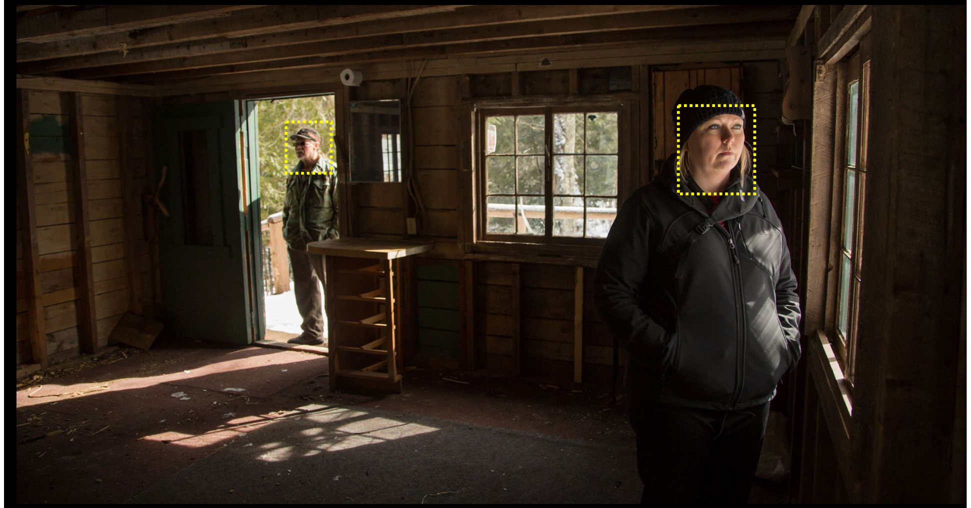
PDP10

SHAKEY (1972) — at the "other" SRI



SHAKEY (1972) — at the “other” SRI

First Digital Camera (1975) — 8 lb — 0.01 Megapixels (10K!)



I. Introduction
II. Classical View
➡ III. World
IV. Registration
V. Effability

Two ways to understand DL "recognizing" the cabin

1. **Naive realism**
   a) Assumption that the classical view (objects, properties, relations, etc.) is **correct**
      i) Messiness in the data reflects uncertainty about the situation
      ii) Measured in terms of *confidence* or *probabilities*
   b) This is the view that we use to analyze adversarial examples (panda, macaw, …)
   c) An assumption that the ontology of the world is completely determinate
      i) Not just at some fundamental or underlying level
      ii) But *at the level at which it contains discrete objects, exemplifying determinate properties*
2. **Pure constructivism**
   a) Objects and properties don't exist independently of us
   b) They are relative to our interests, perspectives, and projects
   c) Deals with a variety of challenges to the classical view

Two ways to understand DL "recognizing" the cabin

1. **Naive realism**
   a) Assumption that the classical view (objects, properties, relations, etc.) is **correct**
      i) Messiness in the data reflects uncertainty about the situation
      ii) Measured in terms of *confidence* or *probabilities*
   b) This is the view that we use to analyze adversarial examples (panda, macaw, …)
   c) An assumption that the ontology of the world is completely determinate
      i) Not just at some fundamental or underlying level
      ii) But *at the level at which it contains discrete objects, exemplifying determinate properties*
2. **Pure constructivism**
   a) Objects and properties don't exist independently of us
   b) They are relative to our interests, perspectives, and projects
   c) Deals with a variety of challenges to the classical view

Criticisms of the Classical View

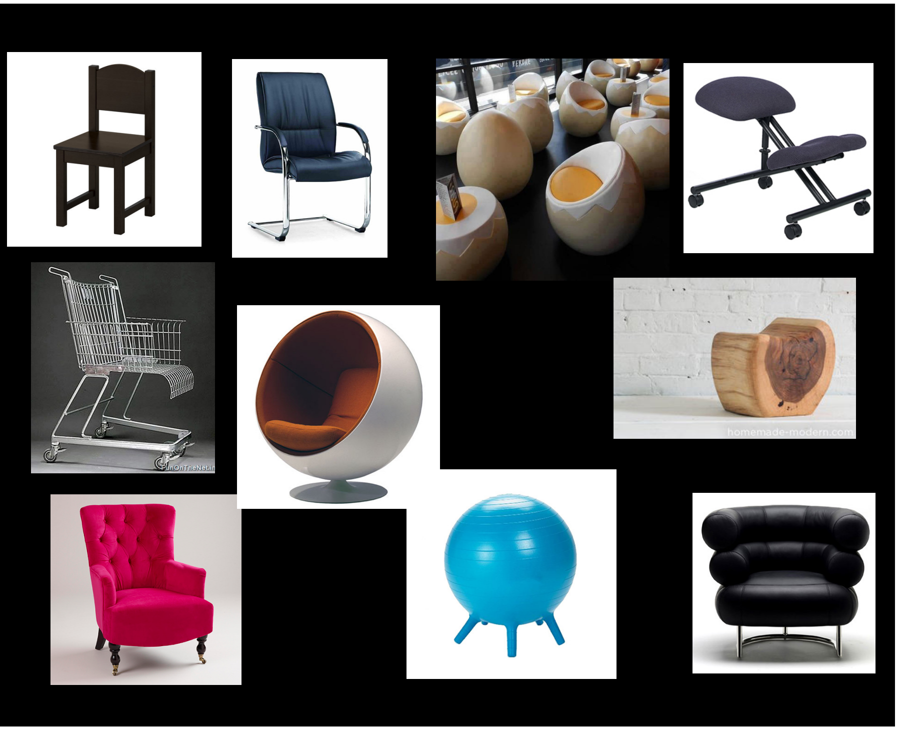
1. **Properties and types**
   a) Determinacy
   - No single property or feature, or group of them, defines a category or type
   - Instances of a property (type, class) at best share **family resemblances** (Wittgenstein)
   - E.g., chairs… (•)

   b) Affordances
   - Chairs are medium-scale objects that **afford sitting** (J. J. Gibson)

   c) Context
   - What properties objects exhibit—what type they are—may be affected by **context** (•)

   d) Boundaries
   - Chair vs. bed (•)

   e) Boundaries
   - *Egotist, egoist, proud, arrogant, self-confident, braggadocious, boastful, cocky, uppity, snooty, high-minded, pompous …*



Criticisms of the Classical View

1. **Properties and types**
   a) Determinacy
   - No single property or feature, or group of them, defines a category or type
   - Instances of a property (type, class) at best share **family resemblances** (Wittgenstein)
   - E.g., chairs… (•)

   b) Affordances
   - Chairs are medium-scale objects that **afford sitting** (J. J. Gibson) *<= will be relevant later*

   c) Context
   - What properties objects exhibit—what type they are—may be affected by **context** (•)

   d) Boundaries
   - Chair vs. bed (•)

   e) Boundaries
   - *Egotist, egoist, proud, arrogant, self-confident, braggadocious, boastful, cocky, uppity, snooty, high-minded, pompous …*

Criticisms of the Classical View (cont'd)

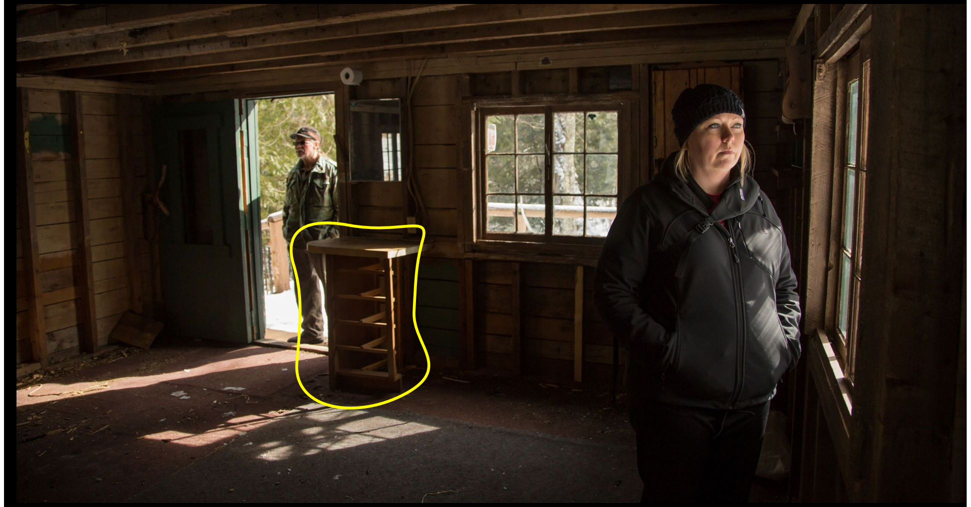
2. **Objects**
   a) All the issues about properties apply to objects as well

   b) E.g., the table in the cabin picture (•)
   - Does it have a back?
   - Or does it *not* have a back, and is resting on the wall?
   - Or does it have a back, which is *formed* by the wall?
   - (And who the hell cares? Depends on what one wants…e.g., to move it?)

   c) Mereology
   - Where does the arm end, and the leg begin? (•)

   d) Identity
   - Washington's axe
   - Ship of Theseus (•)

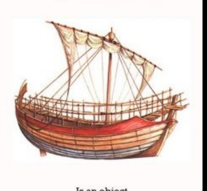
   e) Objectification
   - Clouds (•)
   - Fog
   - "It's mommy-ing again!"



Criticisms of the Classical View (cont'd)

2. **Objects**
   a) All the issues about properties apply to objects as well
   b) E.g., the table in the cabin picture (•)
      - Does it have a back?
      - Or does it *not* have a back, and is resting on the wall?
      - Or does it have a back, which is *formed* by the wall?
      - (And who the hell cares? Depends on what one wants…e.g., to move it?)
   c) Mereology
      - Where does the arm end, and the leg begin? (•)
   d) Identity
      - Washington's axe
      - Ship of Theseus (•)
   e) Objectification
      - Clouds (•)
      - Fog
      - "It's mommy-ing again!"

Ship of Theseus

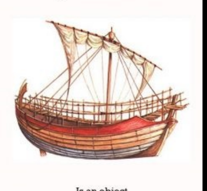


Is an object that has had all of its components replaced still the same object?

Criticisms of the Classical View (cont'd)

2. **Objects**
   a) All the issues about properties apply to objects as well
   b) E.g., the table in the cabin picture (•)
      - Does it have a back?
      - Or does it *not* have a back, and is resting on the wall?
      - Or does it have a back, which is *formed* by the wall?
      - (And who the hell cares? Depends on what one wants…e.g., to move it?)
   c) Mereology
      - Where does the arm end, and the leg begin? (•)
   d) Identity
      - Washington's axe
      - Ship of Theseus (•)
   e) Objectification
      - Clouds (•)
      - Fog
      - "It's mommy-ing again!"

Ship of Theseus



Is an object that has had all of its components replaced still the same object?

I. Introduction
II. Classical View
III. World
IV. Registration
V. Effability

Irreconcilable Debate

1. For decades a debate has raged between these two views
   a) **Realism**, which takes the structure of the world to be out there, independent of us; and
   b) **Constructivism**, which claims that, rather than being independent of us, objects and properties are human constructs which reflect our cultures and societies and interests and such
2. That way of framing the debate is **fatal** (imho) … it will never be resolved

Example: Chairs

1. Suppose our robot (or a network "sees" a chair)
2. Has it "seen a chair out there", independent of us?
3. In one sense the answer is **yes** ← *there's something right about realism*
   a) There it that, out there—a patch of reality—which affords sitting
   b) It is really out there, too. Even if I don't see it, if I sit down, it will support me
4. In one sense the answer is **no** ← *there's something right about constructivism*
   a) There is no chair out there, *qua chair*
   b) "Chair" is a culturally-specific category; someone else might take it to be something else
5. As I've argued—and as DL shows—the "chair qua chair" (or the concept chair, or something like that) involves (but is not itself!) an abstraction or "coarse graining" of what is out there
   a) The chair itself is not the coarse-graining—or the abstraction, or the idealization itself
   b) The chair itself is not "the world *as coarse-grained*"
   c) That is: the *chair* itself hasn't been coarse-grained

Example: Chairs (cont'd)

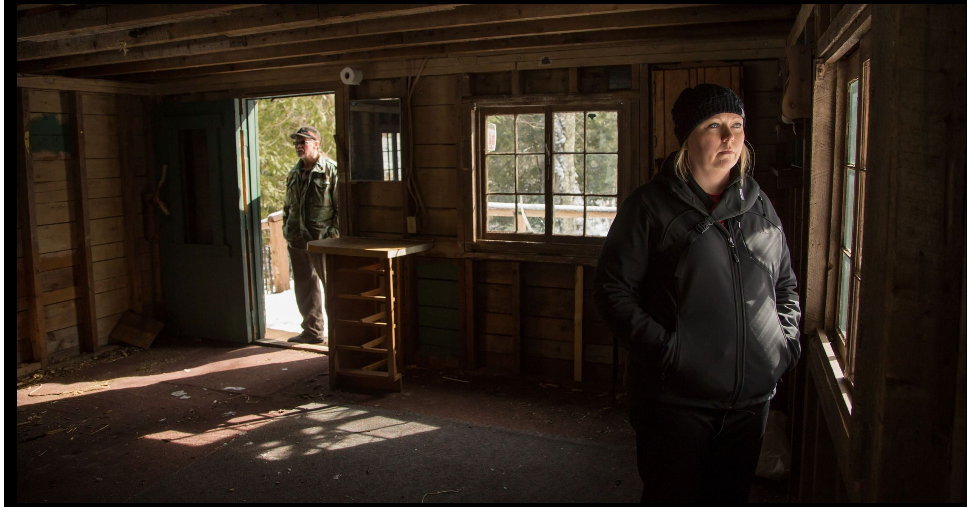
6. Rather, we want to say something like this
   a) The chair is that *non-idealized* (not coarse-grained) patch of reality, in its complete existential richness, which warrants a coarse-grained idealization (classification) *as a chair*—because, as a patch of reality (not as an idealization), it affords a non-idealized pattern of activity that in turn warrants being coarse-grainedly idealized (classified) as *sitting*
7. Needless to stay, that is an awkward way to talk…

Randy: *What kind of furniture did they give you for your new office?*
Pat: *That which, in virtue of my cultural embedding, projects, and purposes,*
I take to be that which I classify using the English word 'chair'
Randy: *Get a life!*

8. Zen master and acolyte …

Potential Criticism

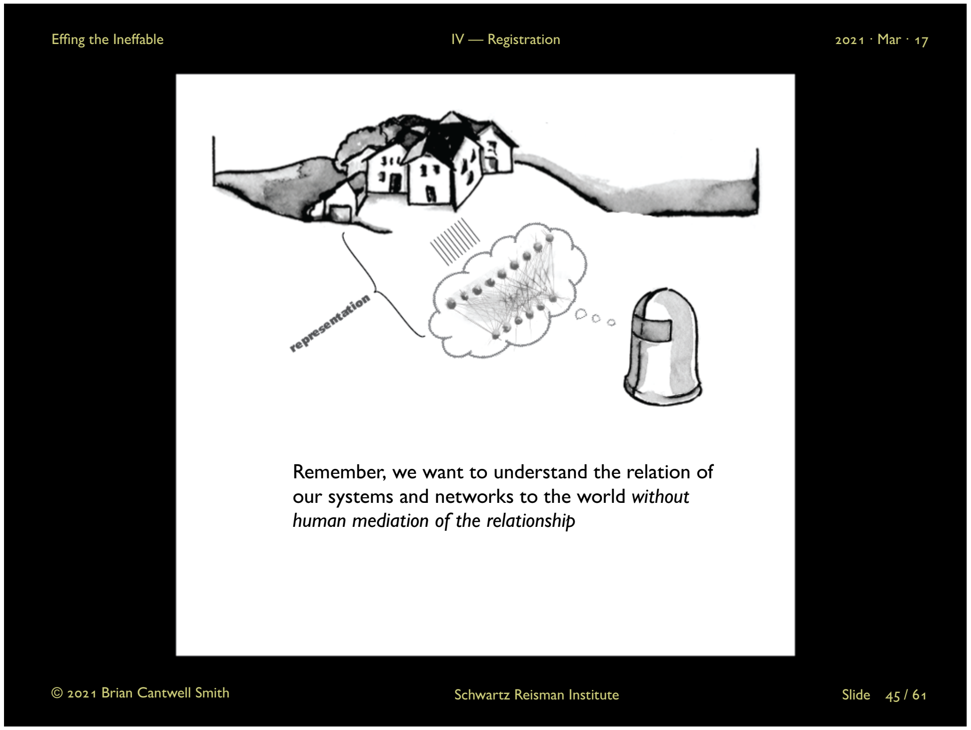
1. Someone (with realist sensibilities) might object
   a) "Of course there are objects and properties out there"
   b) "You can see them perfectly clearly in the room of the cabin"
2. But that argument makes a mistake …



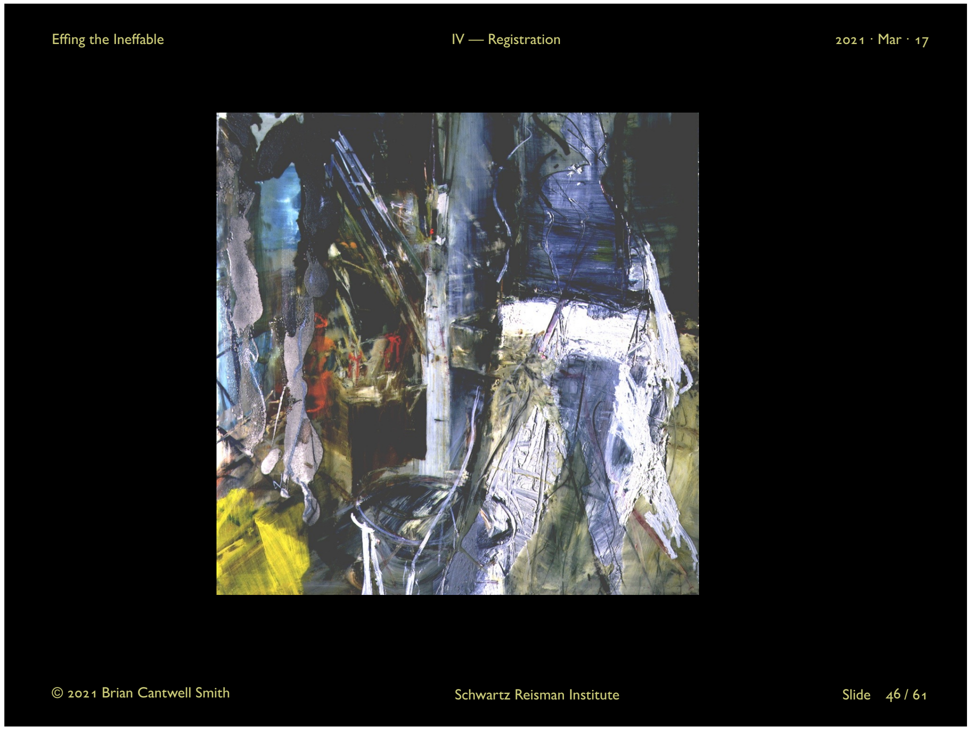
You just processed this image using a neuronal device comprising **100 billion elements** *with* **100 trillion interconnections** *honed for this explicit purpose over* **500 million years** *of evolution!*

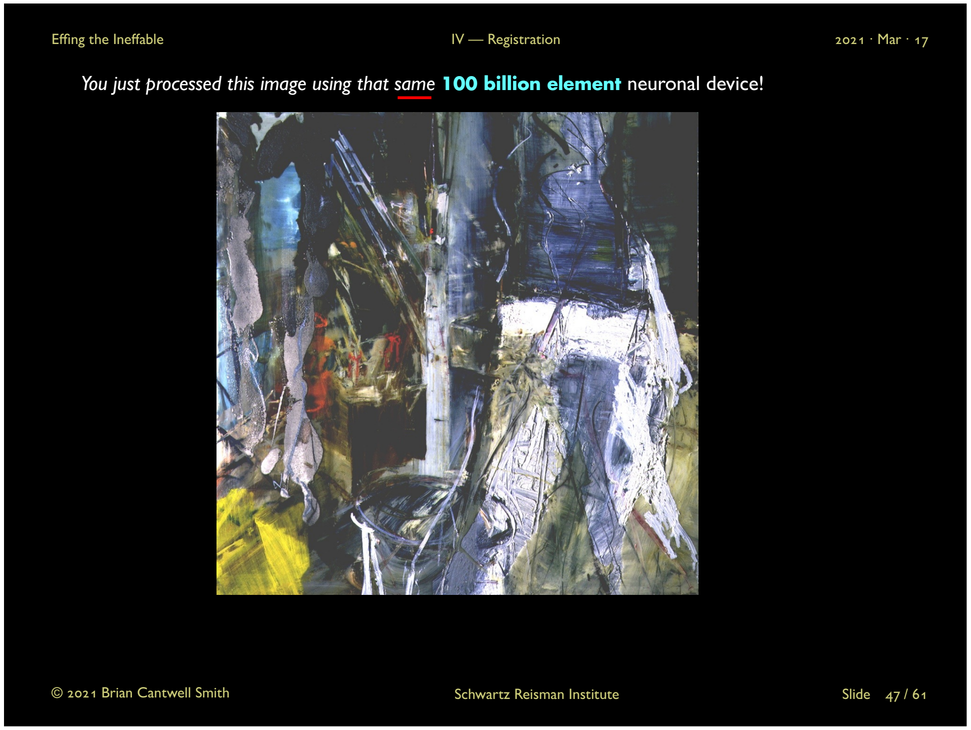
You just processed this image using a neuronal device comprising **100 billion elements** *with* **100 trillion interconnections** *honed for this explicit purpose over* **500 million years** *of evolution!*

What does the world look like, **before** *all that processing?*

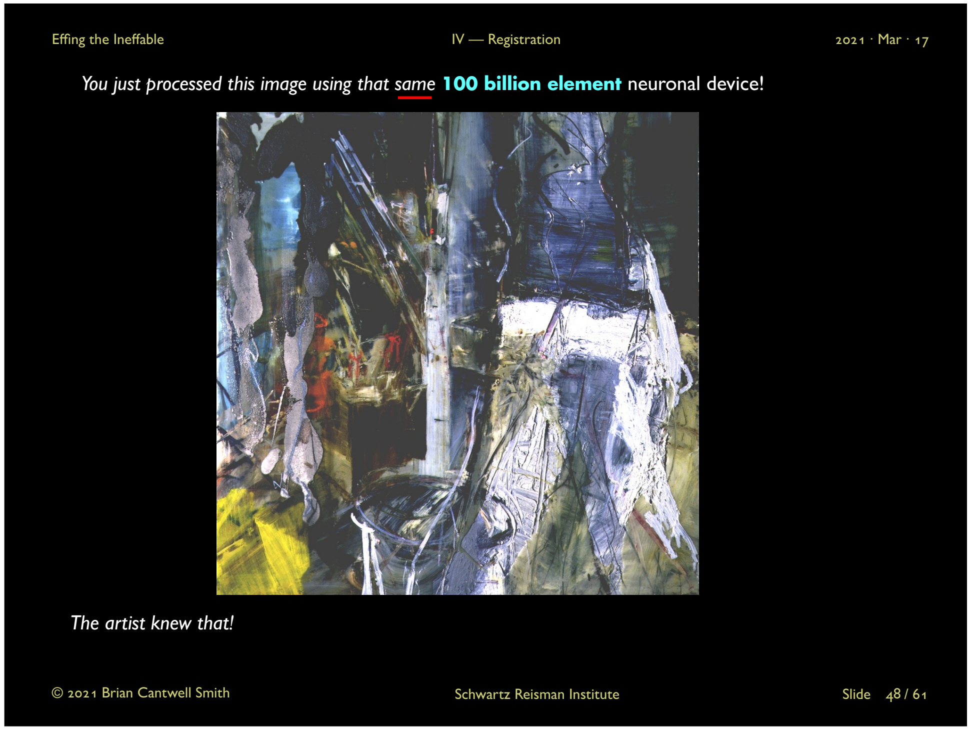


Remember, we want to understand the relation of our systems and networks to the world *without human mediation of the relationship*





You just processed this image using that same **100 billion element** neuronal device!



You just processed this image using that same **100 billion element** neuronal device!

The artist knew that!

You just processed this image using that same **100 billion element** neuronal device!

The artist knew that!

If "perception" computes function f(scene), this is his conception of f^{-1}(scene)

An Analogy

An island in Georgian Bay

Our concepts and words …

An Analogy

An island in Georgian Bay

The fine-grained detail "below the words"

Registration

1. I will say that a system (person or machine) **registers** the world—registers the surpassingly rich plenum that it encounters—by *finding it intelligible:*

Both {

   a) **Classifying** it in terms of coarse-grained, non-absolute patterns and abstractions, consistent with our interests,

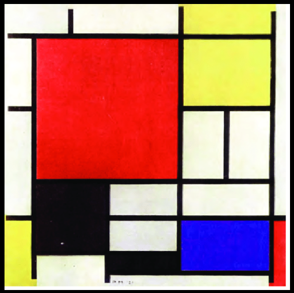
   b) While at the same time **not losing any of the underlying richness**

2. Examples

   a) Register a *chair* …

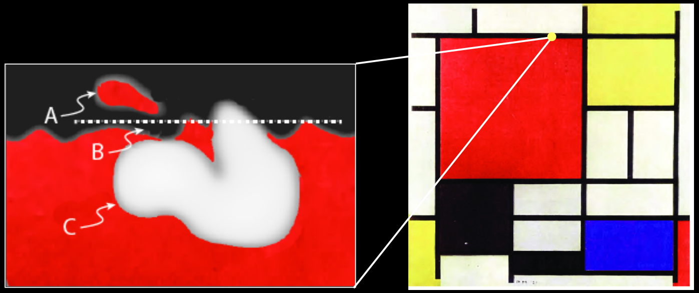
   b) Register the *length of the wall* between the sofa and the door …

   c) Register a *Mondrian painting* …



Registration (cont'd)

3. Crucially, what we register is
   a) Not the world **as registered**
   b) Rather, **that which we register**
4. Examples
   a) Register a *red square on a Mondrian painting* (•)
   b) What we register includes A, does not include B—and who knows about C?



I. Introduction
II. Classical View
III. World
IV. Registration
➡ V. Effability

Why do we "Ef"?

1. **Compositionality**
   a) The high-level (coarse-grained) **concepts and categories** (chair, person, sitting, etc.) support a kind of "algebraic" recursive compositionality that we know from grammar.
      - The "infinite use of finite means" (von Humboldt)
   b) Allows for language—a very *low-bandwidth* finite communication
      - We don't have USB-C connectors and GHz transmitters and receivers that would be necessary to share big swaths or swatches of our slow but massive network states
2. **Abstraction**
   a) It turns out that the world sustains relatively stable high-level generalizations and regularities that can only be framed by **discarding much of the massively complex finely structured details** of the world "underneath" them
   b) Or maybe it doesn't "turn out" this way—maybe these are the only high-level regularities that we can register! ⟵ *think about this!*
3. **Learning**
   a) Only if one sheds the underlying detail can these regularities—and the words we use for them—be **learned**

The Ineffable

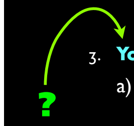
Nevertheless, a great deal of our understanding relies on the **sub-conceptual** details that anchor our registrations in the world— ineffable fine details that we cannot express in words

1. Rush hour in New York City
   a) Driver: "Oops! I've just been struck blind.
      But it's OK. I'll keep driving. Just tell me what to do."
   b) Passenger: "Eeek!"
2. Footstep in your study late at night
   a) You know *exactly* where the sound came from
   b) But you don't know in terms of feet & inches, degrees, or any other *concepts*
3. Taste: The differences between the tastes of nutmeg, cinnamon, and allspice
4. Tennis
   a) It doesn't help to say: "Hit your second serve at the speed that pieces of paper come out of that big office copier down the hall"
5. Motorcycle
   a) Police: "Do you know how fast you were going?"
   b) Rider: "In one way *yes*; in one way *no*"

These sub-conceptual details can be encoded in complex values of high-dimensional vectors

Morals

1. **Ineffable**
   a) The world itself is incomparable rich—vastly more than can be "captured" in words, or that can be finitely represented at all…
   b) We understand the world in ways that, though undoubtedly less rich than the world itself, are still vastly more complex than can be articulated ("effed") in words
2. **Effable**
   a) In order to track and understand high-level regularities, in order to categorize and classify, in order to learn, and in order to talk, we register the world in terms of much higher-level (coarse-grained) concepts and categories



3. **You might think**
   a) That this would result in a two-level system
      i) At the low level—for perception and action—we rely on the subconceptual richness
      ii) At the high level—for rational inference, language, and communication—we take refuge in effable concepts and categories, which work in ways roughly similar to the classical view (underlying logic, GOFAI, traditional linguistics, etc.)
   b) Attempts to develop "explainable AI" may rely on something like this picture

Morals (cont'd)

4. **But it is not so!**
   a) As we have seen, even when we register something in terms of a concept, we don't let go of the fine-grained (subconceptual) detail
   b) Hilary "Let's buy that sofa"
      Jordan "No. Just think about it. It won't fit between the piano and the door."
5. **Even more interestingly**
   a) Although, in order to categorize, we have to be **able** to let go of the particular details of the patch of the world we are registering, we do not have to let go of (ineffable) details that apply across the *category as a whole*
   b) Go back to chairs
   c) *All* chairs need to afford sitting
   d) That means that registering something as a chair may
      i) Require abstracting away from the details of various different *kinds* of chair
      ii) It may still include a (relatively general, but still ineffable) representation of the *act of sitting*—something we all do, but don't know how to give a conceptual account of

Morals (cont'd)

6. **Meaning**
   a) What does the word 'chair' mean?
   b) Classically, the word 'chair' is taken to mean "exemplifying the property of being a chair"
      i) Similarly, 'red' means that what it is predicated of is red
      ii) But those accounts are neither very satisfying nor insightful …
   c) Better: a theory of meaning should explain the **cognitive significance** of using or hearing a word
   d) As we have seen, using the word 'chair' may rely on, and evoke, an ineffable non-conceptual understanding of sitting
   e) That in turn suggests something interesting:

      Words themselves (may) have ineffable meanings

   f) Something that poets and translators have known for millennia …

Conclusions

1. AI and machine learning have implications for our understanding of words and the world
2. The meanings of words (may) involve complex ineffable understandings of the stupefyingly rich worlds we use them to register
3. Just because an AI system can **produce** a word doesn't imply that it can **mean** it
   a) E.g., it may not be possible for a system that *does not and cannot sit* to meaningfully utter the word 'chair'
4. These results have (as yet unexplored) implications for
   a) Explainable AI
   b) Theories of language and semantics
   c) Criteria for trusting AI systems (especially if they cannot mean the words they use)

hmmm …