

## Registering the World — Skeleton

### I. Setup

#### A. Intro

1. It is widely recognized in metaphysics and epistemology that there is no way to explain our understanding of the world from outside of that understanding.
  - a. McDowell famously put it ... no “sideways on”
  - b. Or Nagel: no view from nowhere
  - c. Conant
    - i. An ‘impression that we are able to step outside the space of what is thinkable, judgeable, or possible.’
    - ii. “no explaining something that belongs to the order of knowledge from outside that order” (2020, 776)
  - d. To “think it is possible to step outside our understanding of the world —imagine life outside the limits of thought or language” [Browning]
2. That doesn’t mean that all metaphysics is cognitive metaphysics [ref].
  - a. It is not just the mind’s involvement in our account of how minds understand the world
  - b. Rather: no way to extract ourselves from our embedded understanding of the world
    - i. From our understanding that we exist, that we understand, that there is a world that we are embedded in that we understand, etc.
  - c. It applies not only to metaphysical reflection, but presumably to everything we understand, ...

#### B. AI — and allergy

1. Even endorsing McDowell’s injunction, I believe there are epistemic, ontological, and even metaphysical insights to be extracted from artificial intelligence (AI)—the project, now about 50 years old, to construct, as I will call them, **synthetic minds**.
2. Now mentioning AI as a potential source of philosophical insight, in my experience, is liable to trigger allergies.
  - a. For some philosophers, who treat AI as essentially an engineering endeavour, the idea that “mere technology” could have deep implications on truth, being, the nature of mind, etc. suggests a complete misunderstanding of what philosophical depth would actually involve.
3. < maybe: insert “problem” here >
4. The problems with such reactions are many.
  - a. One is idea that AI is “just technology”—as if it was clear in advance what ‘technology’ is, and what sort of understanding
    - i. The aim of AI, after all, is not simply to build devices with certain physical characteristics, but devices with epistemic and intentional properties.
    - ii. It may be that the way AI is taught, these days (e.g., in computer science departments) views it as a technological project, in a reductionist view of technology. I have some sympathy for this view—but so much the worse, I believe for CS.

- iii. Though writers use pens and paper, that doesn't reduce literature (or philosophy, for that matter) to a "merely technological" project.
  - iv. As I document in a forthcoming book, the understandings on which computational engineering are based include issues of semantic interpretation, normative evaluation, etc., which are not dealt with as such by the reigning theories of computation.
    - a. < fn: not 'semantics' in the sense that CS uses that term ... >
  - b. But independent of AI's particular aims, there are understandings that have developed in computational practice that are of genuine philosophical interest
    - i. E.g., dealing with complexity (easily  $10^{13}$  in time, same in space)
    - ii. E.g., dealing with implementation (realization, reduction, type and token supervenience, etc.)
  - c. Open up a  $\Delta$  between what we *say* and what we *think*, by giving us another window into the structures with which we think.
- C. < An example — Indexicality
- a. Standard philosophical view: certain kind of context dependence
    - i. I, here, now, etc.
  - b. Sometimes complicated a bit
    - i. Cf. Kaplan's *content* and *character* (■ = look up!)
    - ii. Cf. Millikan vs others on the mental indexical (■)
    - iii. Perry & Israel on *incremental* content (■)
    - iv. Maybe even two-dimensional semantics (■)
  - c. In computational systems—even classical ones (i.e., not contemporary deep learning) contextual dependence is often much more complicated than that
    - i. Some things go with one, some things with the other, etc.
    - ii. So to a computer architect, the proposals for two or three dimensions of context dependence seem likely to be woefully inadequate to deal with the sorts of complexity that human cognition will require
  - d. Ties into the adequacy of the theoretical construct of type vs. token (class vs. instance)
    - i. One can use that two-level characterization as orthogonal to the ones under discussion. But that violates something I call reflective integrity: ... [explain?]
- D. Vocabulary
1. More generally, the situation exemplifies something I have struggled with for the 50 years that I have been involved in AI.
  2. On the one hand, there are few philosophers who have a deep understanding and appreciation for what is going on, in AI.
  3. And perhaps needless to say, there are few AI engineers or practitioners trained in philosophy—competent to talk about the distinction between efficient and final causes, about what Kant is getting at in the B Deduction, on whether the norms on thought (how to reason correctly) and the norms of objectivity (what is the case) are one and the same, or two and different—or even independent.
- E. Now I am not trained in philosophy, either.
1. Talk about our inability to extract ourselves from the state of knowledge!

2. But what I want to do today is to reach out, from AI, and try to formulate some questions and issues that touch on philosophy, in the hopes of catalyzing a deeper engagement between the two.
3. Why I am so appreciative of participating in this workshop...

## II. Philosophical assumptions

- A. Now as I read it, a great many lines of metaphysical argument, within the philosophical literature, share some fundamental ontological assumptions:
- B. Assumption [1]
  1. That we understand the world in terms of objects, exemplifying properties, with identity conditions, ...
  2. There are great areas of debate:
    - a. Constitution
    - b. Reduction / supervenience ...
    - c. Nominalist, idealist, realist, etc.
      - i. And if realist, whether the type of existence they exemplify is “real existence,” independent of theory, or ...
      - ii. I.e., internal vs. external questions, in Carnap’s sense
  3. But on the basic framework of objects, properties, etc., there is widespread agreement
    - a. Fundamentally take the world—and language—to have a kind of compositional algebraic structure
      - i. Objects, property / relation / states of affairs—ontologically
      - ii. Terms, predicates, relation symbols, sentences—epistemically, in language and logic
    - b. Call it: OPRA—for objects, properties, relation, and states of affairs
      - i. Or equivalently, without a shred of etymological warrant, *operatic* (physicist have their charm, after all; we can’t let them be the only ones with colorful terminology).
      - ii. Whether these two algebraic, compositional structures are the same or  $\Delta$ ...
        - a. operatic-e (epistemic), vs. operatic-o (ontological)
      - iii. (Possibly the same; maybe that is the point of the B Deduction)
      - iv. Certainly related: Pred names prop; term names object; sentence true if object exemplifies prop; etc.
    - c. E.g., Thomasson: “Assuming that our language contains the general noun ‘fillow’ (and holding its meaning constant), ‘fillow’ applies just in case a fillow exists, and so the conditions under which ‘fillow’ applies are those conditions in which it is true to say ‘a fillow exists’.”
    - d. The standard story
    - e. So that is the first assumption: the operatic assumption
- C. Assumption [2]
  1. Another assumption that there is a single property designated by a predicate  $P$ , which property is instantiated by any object  $x$  of which it can truly be said that  $P(x)$ .
  2. I.e., that the ingredients in the algebraic, compositional structure are in some sense

*atomic*. That is not a mathematical or computational necessity.

D. Assumption [3]

1. Role of epistemology—or at least of reason
  - a. McDowell: “second nature” —“embedded into a lifeworld where meaningfulness and reasons are part of the furniture, and these are intelligible naturally in our enculturation” [Browning]
2. E.g., in Frege’s  $\Delta$  between sense and reference
3. In Thomasson’s “easy” vs “hard” ontology
  - a. Conceptual role
4. ... (■) I need to review these, what the cognition is for, how it relates to ontology, etc.

E. Assumption [4]

1. Perhaps most universally of all, that our knowledge of the world is best access through the **language** in terms of which we characterize the world
2. And that our best understanding of the world is as expressed in language.
3. I.e., “world as understood in thought and language,” as many people say
  - a. But language wears the trousers, in these analyses
  - b. It is presumed that thought follows language—roughly, that we think in language
    - i. cf. Fodor’s conception of a language as anything with a compositional syntax and semantics.
  - c. That may be...in some cases, for some purposes. But it may not be a complete understanding of the world—or of its metaphysical or ontological nature.

**III.AI**

A. What does AI have to do with this?

1. I have already said that I do not believe—and will not argue—that taking the project of AI seriously < fn; maybe not the reality > means rejecting the **view from somewhere**
2. But does require seriously questioning, if not rejecting, all four assumptions
  - a. About OPRA
  - b. About atomicity
  - c. About reason
  - d. About language

B. At the beginning—in what Haugeland called GOFAI—AI shared the OPRAtic and atomic assumptions

1. But it did not work
2. Systems were brittle, and non-sensical, “sterile”—leading to a general sense that they didn’t know what they were talking about

C. So-called second wave AI changed the underlying architecture in major ways

1. Massive amounts of real-valued, multi-dimensional data
2. As I put it in a recent book:
  - a. GOFAI
    - i. Deep (many-step) Inference
    - ii. Serial process, using
    - iii. Modest amounts of information

- iv. Involving a relatively small number of
      - v. Strongly correlated variables
    - b. Machine Learning
      - i. Shallow (few-step) inference
      - ii. Massively parallel process, using
      - iii. Massive amounts of information
      - iv. Involving a very large number of
      - v. Weakly correlated variables
  - 3. Today, focus on the last 3 of these: the sheer complexity of the machine states
    - a. (■)... take numbers from GPT-3 ...
    - b. It takes only *one* 16-bit number to identify which word of English I am using
    - c. And just one more to identify what place within what grammatical construction any given word is.
    - d. Leaves 29,998 others...
    - e. [cite Wittgenstein's private language: he didn't expect it to be like *this*]
  - 4. Often *characterized* probabilistically: as probabilities that some realistic, atomic, OPRAtic characterization's being true.
    - a. That is retrospective evaluation—our external view about the systems, not a view they had
  - 5. What is telling is that one might think that “collapsing” the real numbers (the “fuzz”?) into an OPRAtic distillation would lead it to more certain and therefore more reliable results
  - 6. But the opposite has proved to be true
    - a. Cite “end-to-end” training, etc.
  - 7. It starts to undermine the idea that the way these systems “understand” is in terms of the four assumptions
  - 8. Rather, what OPRA beings to look like is
    - a. Not the way the system is perceiving, *or reasoning, or acting*, towards the world
    - b. But rather: an *abstraction* of those relations, allowing them to be *framed in language*
      - i. (the original discrete, compositional representation system)
    - c. (Lots of caveats; much of the complexity in GPT-3 is because the information is information about how words are embedded in structures of other words in online texts, rather than information directly about the world. I think this is a huge limitation to contemporary AI; I will get back to it in a moment. But there is no reason to suppose that the world is any less complex than the descriptions we make of it ...)
- D. Ineffability
1. If, instead of taking what is in front of it to be an object exemplifying a property, or two objects standing in a relation, or something like that, the system instead ... well, what are we to say?
    - a. “What does it think is in front of it?” you ask?
    - b. “It thinks *this*”, I say, printing out a list of 30,000 16-bit real numbers—i.e., a vector in a 30,000 dimensional space. And if you ask “what are those dimensions?”,

I can only recurse and ...

2. Nevertheless
  - a. In one sense, my printing out this list of numbers makes the understanding *public*—but in a useless sense.
3. One possible response
  - a. Well, *really* it is “understanding” the world in terms of OPRA,
  - b. These are just contextual parameters, probabilities, certainties, etc.
  - c. How do we know that?
  - d. The “end to end” training suggests that claiming that these things are approximating effable linguistic strings begins to sound like an attempt to save classical OPRAtic ontology from what is actually going on?
  - e. More plausible, given that this is how these systems are “*thinking*”, as it were, to wonder whether the real situation may not be the inverse (?)
  - f. And think about arguments about non-conceptual content (in Evans’ sense, not McDowell’s)
  - g. ...I have not knock-down argument *against* such a view, but I am not sure there is an argument *in favour* of it, either.

#### IV. Registration

##### A. Intro

1. Now before I am drowned out by howls about antirealism, idealism, or mysticism, I want to introduce a way of speaking
  - a. Not a philosophical position, so much as a way of talking about philosophical positions
2. With that in hand, will then address
  - a. Realism
  - b. Truth
  - c. Objectivity
  - d. And a variety of other objections to the proposal I am making
3. Inevitably, the way of speaking must of course still be in a public language
4. It is also intended to be relatively basic
  - a. My aim for it is to separate, more than usual:
    - i. What we are talking about, or perceiving, or otherwise intentionally directed towards;
    - ii. What we say or mean about it—i.e., about the patch of the world towards which we are intentionally directed.
  - b. But beyond that, I want it to be applicable to a wide variety of forms of intentional directedness—perception, assertion, belief, even imagining

##### B. In particular: I will say that if a system or agent **registers** the world, or a situation or patch of the world, if that system or agent is intentionally directed towards it, and in virtue of that intentional directedness *takes it to be a certain way*

1. If the system is operating with an *opratic* scheme, takes it to be object of a certain type exemplifying certain properties, or a state of affairs, or some such
  - a. I.e., in terms of the objects, properties, situations, states of affairs, features, etc.,

- that the system takes (the world, that is) to consist in.
2. Thus, as I write these sentences, I register a cup of coffee, register a *lake on the horizon*, register a *thorny academic dispute I just heard about as petty and unfortunate*, etc.
  3. When I say that to register ‘something’ is to be intentionally directed towards it, I mean ‘something’ very generally—as something like “a patch of the world.”
    - a. ‘Something’ is a singular referring term—but it is a term in the metalanguage I am using. I do not imply by that use that the system under discussion is taking that which it registers to be an object. The fact that my language treats it as such should be taken as a metatheoretic fact about my *characterization* of the registrational act—carrying as little ontological significance about the act itself, and that towards which it is directed, as possible.
    - b. Unlike the phrase “X is intentionally directed towards,” or “X refers to something,” in both of which cases the term ‘something’ would be taken to be in a transparent context in the metalanguage, if I say that “X registers something,” I meant to imply that they are intentionally directed towards it *as being a certain way*.
      - i. I.e., registration, as I am using it, is ineliminably *as*; it involves aspectual shape. If the system is operating within operatic frame, then it is likely to be directed towards objects exemplifying properties, standing in relations, etc.
      - ii. < fn: What do I want to say about representation? Maybe: that I don’t necessarily want to imply anything about the subpersonal mechanisms that underlie it—and thus to assume that to say that people registers situations in the world is automatically to endorse a representational theory of mind. >
  4. I take the term to be liberal as to the form of intentional directedness—and thus to be neutral as to any distinction between or among *sense, perception, thought, judgment, description, speaking*, etc. Thus, it is uses a singular term in a linguistic context, an agent or system may be said to *refer to* an object; if instead they see the object, that is not typically taken to be a case of reference.
  5. The term is also intended to be neutral on the division of responsibility between person and world for the resulting ontological “take”—i.e., neutral as among
    - a. Brute realism (I successfully register a table *as* a table because it *is* a table)
    - b. Constructivism (I successfully register it as a table because of the contingent and historical forces constituting the social community of which I am a member, or even due to the particular exigencies of my own individual history)
    - c. Pragmatism (I register it that way because it is practical and pragmatic to do so given the normative aims governing the projects in which I am engaged, etc.
    - d. ... o say nothing of idealism, solipsism, and many other epistemic, ontological, and metaphysical proposals.
  6. These neutralities are by design. The point is to develop an apparatus with which we can talk about registration in order to have conversations, in its terms, about realism, constructivism, idealism, pragmatism, normative standards, etc.—and about the intentional capacities of animals and machines, as well as of people.
  7. Normativity
    - a. Unlike ‘conceive’ or ‘cognize’ (and in this respect more like ‘see’ and ‘perceive’), I take ‘register’ to be a *success* verb. That is, if, in ordinary circumstances, and un-

- less otherwise remarked, it I say that a system or person X registers—i.e., takes there to be—a tree, then it is fair to assume that there was a tree there to be so taken, and that X did so take it, in the full semantically and normatively appropriate way.<sup>1</sup> < Say: modify this in due course ... >
- b. As should be evident from the above (including occasional use of the qualifier ‘successful’), I take registration to be normatively laden, in the philosophical sense of serving as the subject of such issues as truth, objectivity, worth, etc.
8. Notwithstanding the foregoing, however, by following the verb *register*’s direct object with ‘as ...’, the parlance facilitates at least a first step towards distinguishing how, as theoreticians (cognitive scientists, epistemologists, etc.):
    - a. We register situations or phenomena in the world, and
    - b. How we take them to be registered by the subjects or systems people or agents, of whom we may be speaking.
  9. That is, if ‘register’ is followed with ‘as’, I take the direct object to be a transparent reference, and the intensional character of the registrational act to be shouldered by the ‘as’ phrase. Thus I might say, of an infant, “She does not register her mother’s arrival as the re-appearance of a recognized individual [object], but as something more like a repetition of a feature placement (in a Strawsonian or Millikanesque sense) of the feature Mama.”<sup>2</sup>
    - a. Cf. Millikan’s “More Mama”
  10. If *not* used with an explicit ‘as’ construction, however—i.e., notwithstanding the above—I will assume that the direct object of ‘register’ to include the aspectual nature of the way in which that phenomenon or entity is registered by the (individual designated by) the sentence’s subject. In this way ‘register’ differs from at least common uses of such perceptual verbs as ‘see.’ Thus while some would claim that it is possible for the sentence “Randy saw the Northern Lights” to be true even if Randy did not recognize them *as* the Northern Lights, I consider it an implication of the sentence “Randy registered the Northern Lights” (without any following ‘as’ clause) that Randy did so take them.<sup>3</sup>
    - a. Talk about the debate I’ve been talking about with Güven (Fred vs. Dan?)

## V. Discussion

### A. Reference

1. An immediate problem arises.
2. If I say that X registers  $\alpha$  as  $\beta$ , how do I know that the delineated region of the world that I register as  $\alpha$  is the “same” as that which X registers as  $\beta$ ?
3. I don’t.

---

<sup>1</sup>This is a rather realist characterisation of “success”; it should be replaced as appropriate for other metaphysical views. The point is simply that “a registered b” should be true just in case something roughly of the form “There is b and a took it to be b” is true.

<sup>2</sup>It is only a first step, because of the evident but fraught issue of *how we register the subject’s registration* (e.g., in the ‘g’ part of the sentence ‘a registers b as g’)—including whether we can, and if so how much, and in what respects.

«For the infant case, ref Jun’s Duke dissertation.»

<sup>3</sup>i.e., the direct object position of the verb ‘register’ is thus not assumed to be referentially transparent.



4. But two comments
  5. [1] If one believes that registration schemes are independent of the practices of systems and agents, then there is no problem.
    - a. Of course if one believes that objects exist and have identity conditions independent of one's registration scheme (even if properties and relations are relative to scheme), there is still no problem
    - b. First, it is widely recognized that there must be *some* overlap in the ontological commitments or perspectives of two intentional systems in order for one to be intelligible to the other.
      - i. Luna, Selene, Artemis, Thoth, or other; made out of cheese, or that a rabbit lives there.
      - ii. Cross-registration scheme talk is commonplace...
    - c. But also the question eases a bit if the requirements on referential overlap are not as strict.
  6. [2] But there is another option
    - a. If I say that x registers a a chair, détente, a question, a centaur
      - i. It does not strain credulity to say that it registers *something* as
    - b. What is that "something"?
    - c. A patch of reality, we might say.
      - i. The thing is, there is no way (in public language) to refer to that which it registers except via its registration of it.
      - ii. So: I will say that **that which it registers as a chair**
  7. That is: it responds appropriately towards that which I call a chair.
    - a. It may *call* it a chair
    - b. But more importantly—and more importantly with respect to a pragmatist understanding of the norms governing its actions, ...
- B. ... transition...
- C. This introduction of a gap between
1. The registration scheme of the person or system describing the situation; and
  2. The registration scheme of the subject being described
  3. ... proves advantageous in talking not only about AI but also cultural difference
- D. For example:
1. History (constructivism)
    - a. Debates rage (perhaps not in this audience) between social constructionists and realists as to the provenance of items referred to in cultural and historically specific ways. Debates no more sophisticated than the following:
      - i. Protons and electrons have existed only since Thompson and Rutherford, vs.
      - ii. That's nonsense! Protons and electrons have existed since the Big Bang.
    - b. In registrational language, it is straightforward to say "There have been protons and electrons since the Big Bang, but it has only been since Thompson and Rutherford that they have *been registered as protons and electrons*.;
    - c. Similarly: electrons were first *registered as such* in ...;
      - i. Millikan; 1909
      - ii. Thompson; 1897

iii. Rutherford; 1911, 1917

## 2. Cultures

- a. Similarly, a conservative zealot might register the assault on the US capital on Jan 6, 2021 as a heroic act of patriotism, where the registration involved in the phrase “assault on the US capital” is my own, and “heroic act of patriotism” theirs, while nevertheless implying, plus or minus (more on this below) that we were intentionally directed at the same patch of reality.

## E. Issues

1. The basic idea: (an imperfect practice to allow) reference preservation across  $\Delta$ s in registrational schemes (or “registers”)
2. If I say “X registers  $\alpha$  as  $\beta$ ,” what assurance do I have that the object or entity or patch of reality towards which X is directed—the patch that it takes to be  $\beta$ —coincides with the patch towards which I am oriented in using  $\alpha$ . None, per se. Suppose I say ‘he registers the chair of this committee to be a fool.’ My use of “the chair of this committee” may be ambiguous as to whether I use it attributively,  $\langle$ check Donellan? No. Not ‘marry a Norwegian’ either $\rangle$  to refer to the person who plays this role, whoever they are—a reference that spans times and changes of chair, or the person who is currently chair.
3. What do I want to say about this? Sort it out pragmatically?
4. But what if

## VI. Considerations

### A. Realism

1. Is this way of speaking per se antirealist?
2. Have I fallen off a cliff into the “lunacy of idealism” (Eklund?)

### B. Truthk

1. What makes a registration of something *true*
2. Or more generally, what undergirds whatever norms govern registration?
3. When philosophy talks about the normative criteria, it is straightforward to focus on truth
  - a. Reference (what is referred to) is thought to be settled—i.e., as specified by causal theories of reference
4. In AI, though, there is a prior and more profound question:
  - a. What makes the actions of an AI system *be about the world at all*?
    - i. Kant’s problem of objective purport: how thoughts can be *about* the world, at all?
    - ii. What gives them *intentional directedness* (to use a term of Brentano)?
  - b. I.e., a Kantian form of skepticism, not merely a Cartesian one.
  - c. This is not a simple question at all.
5. It undergirds Searle’s concerns about weak vs strong AI
  - a. (Not what contemporary AI *thinks* the distinction between “weak vs. strong AI” is about—i.e., particular vs. general purpose [cite, even quote, Descartes])
    - i. Another casualty of the discursive gap between  $\phi$  and AI.
6. This is what I wrote a short book about a couple of years ago

- a. Existential commitment to and involvement in the world
- b. Cf. Haugeland: “give a damn”
7. Find the world intelligible, in terms of the registration
8. An inversion of the traditional picture
  - a. The local statement is true if, overall, the pictures is of the world.
  - b. More detail: if the overall picture has sufficiently “grip” on the world that a piece of it can be held accountable to a patch of the world onto which it fits...
9. Is this pragmatism?
10. < seems weak; doesn't take advantage of the surpassing detail ... >

Notes

- C. Talk about registration scheme vs. conceptual scheme
- D. Talk about “conceptual analysis”  $\Leftarrow$  requires a registration scheme
- E. Emphasize: what it registers is not *vague*.
  1. What is vague is our *calling it chair*.

