

3 — Failure

Nevertheless, GOFAI failed. Or anyway it is deemed to have failed.¹ The failures can be summarized as of four main types.²

FI Neurological: The brain does not work the way that GOFAI does. All of today's mainstream CPUs, and all classical AI systems, consist of relatively few serial loci of activity (one to a few dozen threads) running at extreme speeds (on the order of 10^9 operations per second), pursuing relatively deep levels of inference or procedural nesting. So far as we know, the functioning of the brain is almost the complete opposite. Brain operations are roughly 50 million times slower, deriving their power from massive parallelism.³

1. GOFAI was not abandoned. Research in its vein has continued; in chapter 7 I argue that it addresses issues crucial to forward progress. But GOFAI no longer has imaginative grip as “how intelligence works”; it is widely disparaged in cognitive science and neuroscience; the center of gravity of AI thinking has shifted to machine learning and other second-wave approaches; and, as I argue here, a number of its founding assumptions (C1-C4) are untenable as a general model of intelligence.

2. This typology is loosely based on Part II of Dreyfus's landmark *What Computers Can't Do*, 1972.

3. According to what is known as Feldman's “100-step rule” (Jerome Feldman and Dana Ballard, “Connectionist Models and their Properties,” *Cognitive Science* 6, no. 3, 1982, 206), the brain can compute at

Whether these low-level architectural differences are germane to the achievement of human-level intelligence, however, is not a straightforward question. To answer would require knowing what it takes to have intelligence like ours, what intelligence of a nonhuman kind might be like, what sorts of implementation strategies the brain uses, and so forth. Nevertheless, GOFAI's decline has led to renewed interest in "brain-style" computing, of which today's "neural networks," deep learning systems, and other machine learning architectures are instances.⁴

- F2 **Perceptual:** Most GOFAI theorists thought that perception—recognizing or "parsing" the world based on perceptual input from sensors—would be conceptually simpler than simulating or creating "real intelligence." After all, in Descartes's phrase, "mere beasts" are good at it, in some cases notably better than people. Plus, the ontological assumption underlying GOFAI (C4; also F4, below), according to which the world consists of unambiguous objects exemplifying discrete properties, suggested that perception would merely involve determining what objects were "out there," what properties they manifested, what types or categories they belong to, and so on.

It did not work out that way. The difficulty of

...
 most 100 serial steps per second—sobering, given the brain's ability to perform many cognitive tasks in under a second. .

4. Neurally inspired architectures have a long history, going back to McCulloch and o. And not all massively parallel architectures were neurally inspired—for example, some early semantic nets. But it is unarguable that the imaginative center of attention in AI has shifted from serial architectures to massively parallel networks.



Figure 2

Otto Lowe, 2018

real-world perception was one of several profound humiliations to which we have been led by our experience with first-wave AI.

When people first attached digital cameras to computers, they were flabbergasted. This was the early 1970s, after all. The idea of using electronics instead of film had just been bruited (in 1961 in regard to satellites; the first digital camera was not released to the market until 1975, toward the end of GOFAI's reign). Why were people taken aback? Because what came through the sensors was a mess. It turns out that the idea that the world consists of a simple arrangement of straightforward objects is the delivery, to our consciousness, of the results

of an exquisitely sensitive, finely tuned perceptual apparatus running on a 100-billion neuron device with 100 trillion interconnections honed over 500 million years of evolution. Figure 2 may seem like a straightforward picture of an artist⁵ in a workshop; figure 3 is a rendering this artist once made to convey to human (visual) observers what he believes the world is actually like—that is, it is an image that, *after* human perceptual processing, gives us a rough sense of what the world looks like *prior* to human processing.



Adam Lowe, CR29, "Studio with phone," 1993⁶

Figure 3

5. Adam Lowe, of Factum Arte (<http://www.factum-arte.com>).

6. Collection of Adrian Cussins. I used this image as chapter 10 of *On the Origin of Objects*.

Descartes was no fool. His standards on rationality were much higher than commonly appreciated. But history suggests he was insufficiently impressed with the achievement of looking out and seeing a tree.

- F3 **Epistemological:** Thinking and intelligence, on the GOFAI model, consisted of rational, articulated steps, on its founding model of logical inference. As Dreyfus insisted, and many cognitive theorists have emphasized,⁷ in many situations a more accurate characterization of intelligence is one of skillful coping or navigation—of being “thrown” into individual and social projects in which we are embedded and enmeshed. Even thinking, these writers argue, rather than consisting in a series of readily articulable steps, emerges from an unconscious background—a horizon of ineffable knowledge and sense-making.

7. For example, Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (New York: W. H. Freeman and Company, 1976); Lucy Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions* (Cambridge: Cambridge University Press, 2007); Humberto Maturana and Francisco Varela, *Autopoiesis and Cognition: The Realization of the Living* (Dordrecht: Reidel, 1980); Terry Winograd and Fernando Flores, *Understanding Computers and Cognition: A New Foundation for Design* (Norwood, MA: Ablex Publishing, 1986); Eleanor Rosch, Francisco Varela, and Evan Thompson, *The Embodied Mind* (Cambridge, MA: MIT Press, 1991); Evan Thompson and Francisco Varela, “Radical Embodiment: Neural Dynamics and Consciousness,” *Trends in Cognitive Sciences* 5, no. 10 (2001): 418–425; and Daniel Hutto, “Knowing What? Radical Versus Conservative Enactivism,” *Phenomenology and the Cognitive Sciences* 4, no. 4 (2005): 389–405.

F4 **Ontological:** The misunderstanding about perception, and perhaps about thinking as well, betray a much deeper failure of first-wave AI: its assumption, already mentioned (C4), that the world comes chopped up into neat, ontologically discrete objects. In fact one of the primary theses of this book is that the misconception of the world in first-wave AI is the “smoking gun” that explains its ultimate inadequacy. In chapter 6, I will argue that the successes of second-wave AI can be interpreted as evidence for the inadequacy of formal ontology both as a basis for intelligence and as a model of what the world is actually like.⁸

Of these four types, the ontological issues cut the deepest. They not only underlie the perceptual and epistemological difficulties, but also need to be appreciated in order to understand how and why second-wave AI has made at least some progress toward interpreting the world, especially at the perceptual level.

Three points are especially relevant.

First is a general thesis with which few GOFAI designers would disagree, even if no GOFAI system could itself be said to understand it. Though framed epistemically (as having to do with intentionality), the point rests on underlying ontological facts. It is universally accepted that the world can be described or conceptualized in multiple ways—as is often said, at various “levels of description.” Reality itself is assumed, at least for all practical purposes,

8. In chapter 7 I argue that preserving what was good about GOFAI requires extracting its insights from the presumption that they can be understood in terms of, and uniquely apply to, a world adequately characterized in terms of formal ontology.

to be surpassingly rich, about which any ontological “parse” provides only partial information. Representations, descriptions, models, and the like all interpret or picture or filter the world through abstractions or idealizations—conceptual “frames” that highlight or privilege some aspects of what is represented, minimize (or even distort) others, and ignore or abstract away from a potentially unbounded amount of in-the-world detail.⁹

This commonplace is easiest to see in the case of conceptually structured representations—including not only propositional structures of the sort exemplified in logic (predications, conjunctions, disjunctions, negations, implications, quantifications, etc.), but also other representational forms that characterize the world in terms of a pre-established stock of objects, types, properties, and so on, such as those used in computer-aided design (CAD) systems, architectural blueprints, databases, and the like.¹⁰

9. Some philosophical readers may object that *names*, at least, refer to objects “without loss”—that is, to objects *as such*, and therefore do not impose even a trace of a conceptual frame. More later; for now I would just note that this would be true only if objects were objects (i.e., had determinate identity conditions) independent of being taken *as* objects, a standard realist position with which I disagree. See *On the Origin of Objects*.

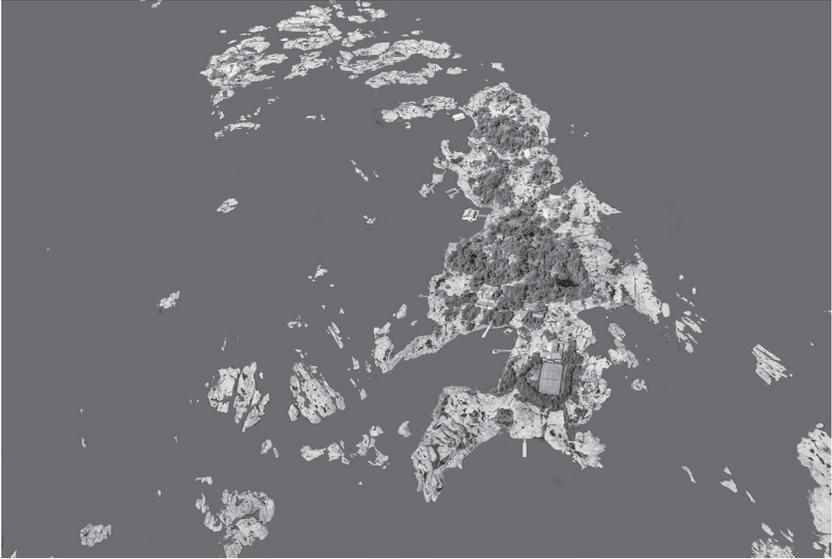
10. Evans ties this notion of conceptuality to what he calls a “Generality Condition” (Gareth Evans, *Varieties of Reference*, Oxford: Oxford University Press, 1982, 100–105): a requirement that for a system to be able to entertain a “thought” or representation that *a* is *f*, it must also be able to entertain the thought that *b* is *f*, *c* is *f*, etc., for any terms *b*, *c*, etc. that represent other objects, and also that *a* is *g*, *a* is *h*, etc., for any predicates *g*, *h*, etc. representing other properties (modulo appropriate type restrictions). This is the sense of conceptuality that underwrites the characterization of the nonconceptual given in the sidebar on the next page; it differs from the one McDowell critiques in Lecture III of *Mind and World* (Cambridge, MA: Harvard University Press, 1996).

Nonconceptual Content

“Nonconceptual content” is a philosophical term used to explain the fact, evident upon reflection, that we are capable of thoughts and judgments that, although grounded in the ways of the world, and capable of being right or wrong, are not framed (in our minds) in terms of anything like a discrete set of articulable concepts. Classic examples include the speed at which one rides a bicycle, the position in one’s egocentric personal space where items are located, shades of color, and so on. Adrian Cussins, an early theorist of the notion, when stopped for speeding on his motorcycle, famously answered a policeman’s question about whether he knew how fast he was going with the statement “In some ways yes; in some ways no” (personal communication). Evans (MIT seminar 1978) reported on a footstep behind him when working in his study; in one sense he knew with exquisite precision where it was, yet would have been entirely unable to describe that location precisely in publicly shared concepts—feet, angle in degrees, and the like (see Gareth Evans, *Varieties of Reference*, 105). Similarly, it would be of limited help, if you were having trouble with your second tennis serve, to suggest that you hit the ball at the speed at which copies come out of your office copier.

But it also applies to most analog representations (those that represent continuous quantities in the represented domain—the domain of the problem to be solved, typically—with analogous real-valued quantities in the representational medium), since the property-to-property correspondence in such cases is typically assumed to be discrete at a higher-level of abstraction.¹¹

11. As analog computers demonstrate, characterizing a situation in terms of discrete properties and relations does not mean that the *values* of those properties need to be discrete. The use of the differential calculus in classical physics is fully conceptual, on this account, even if any given velocity or acceleration or mass can be measured with a real

Figure 4¹²

The second point is also traditionally framed epistemologically. As noted in F4, theorists in numerous fields have emphasized the fact that not all human understanding has a “conceptual” form at all. The point is most familiar from phenomenological theorizing, but similar views underwrite several contemporary currents in cognitive science, including enactivism, connectionism, and deep learning. Analogous intuitions have been explored in analytic philosophy in the notion of *nonconceptual content* (sidebar on the previous page). These approaches are united in viewing

... number. What matters is that the properties of velocity, mass, etc., are discrete. There is no such thing as something “halfway between a mass and charge,” though even that might be expressible in terms of a higher-order discrete categorization (John Haugeland, “Analog and Analog,” *Philosophical Topics* 12, no. 1, 1981, 213–225).

12. A masked version of figure 6, p. 34.



Figure 5

intelligence as emerging against an unconscious background—as I put it above, against an ineffable horizon of tacit knowledge and sense-making.

Together with the perceptual points made earlier, these views provide indirect support for an ontological world view I have sketched elsewhere,¹³ for which figures 4–6 provide a metaphor. Figure 4 (previous page) is a photograph of some islands in Ontario’s Georgian Bay. Already, one can see that the real-world topography fails to support the cut-and-dried ontological registration that GOFAI assumed. Figure 5¹⁴ “cleans the picture up” in a

13. *On the Origin of Objects*.

14. Figure 5 is by design simplistic; the filters used to create it are not especially fine-grained. One could construct an image in which they encoded vastly more detail—as, for example, is current practice in digital GIS systems. But the conceptual point would remain:

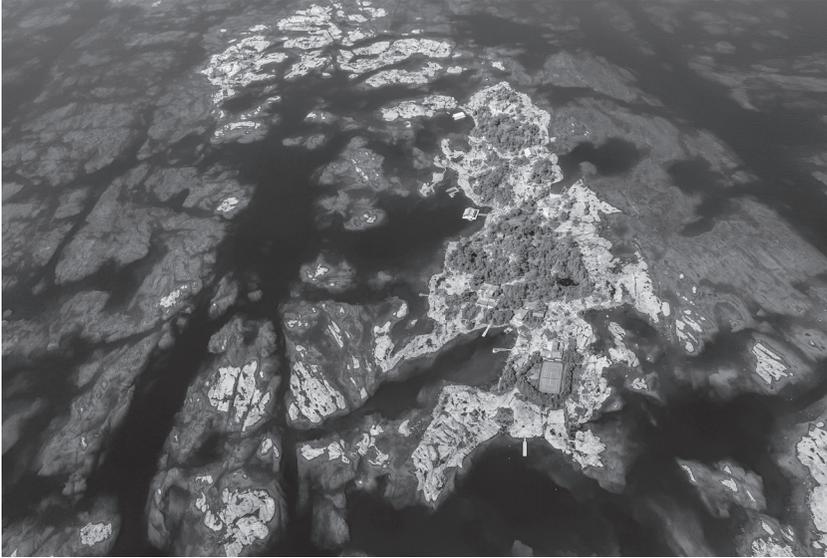
way reminiscent of GOFAI knowledge representation—making the islands, though still relatively detailed, “clear and distinct,” and also internally homogenous, in the way in which conceptual models (such as data bases) all ultimately assume. While the question of “how many islands are there” may have a determinate answer in figure 5, the same is not true of the world depicted in figure 4. Distinctness flees, as realism increases. In the world itself, the question lacks a determinate answer.

More telling yet, though, is figure 6 (next page)—the same photographic image as figure 4, except now revealing the submarine topography. Compared to the world’s messiness, the image is still simple: gravity is a single dimension of salience, the water line is relatively sharp, the image is gray scale, and so on. Nevertheless, if the islands in the image are taken as analogs for properties, then the images suggest what in fact is true: that as soon as one presses for detail, distinctions multiply without limit.

* * *

... any (first-wave) knowledge representation structure is formulated in terms of a well-specified particular level of granularity; the representation “individuates” the world at a specific level of grain—no finer. In figure 5, as in all conceptual representations, there is a definite fact of the matter about how many islands there are, where exactly the boundaries are specified as being, etc. One would not learn more about the islands by studying the representation in any finer detail than necessary to reveal those facts. (This is what Dretske nonstandardly means by “digital” representation, as opposed to “analog”; see his *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press 1981, 135–141.)

Someone might argue that the same is true of (even analog) photographs, because of the ultimate granularity of film. What is crucial, though, is *the same is not true of the reality that these images represent*. The world itself is arbitrarily detailed.



Paul Bennett Photography

Figure 6

Though our concepts and the properties they represent may seem discrete, in other words, the failures of GOF AI and the merit of the critiques mounted against it suggest that the idea of a “clear and distinct” world is an artifact of how it is represented. The issue for AI is that, in order to function in the world, AI systems need to be able to deal with reality *as it actually is*, not with the way that we think it is—not with the way that our thoughts or language represent it as being. And our growing experience with constructing synthetic systems and deploying them in the world gives us every reason to suppose that “beneath the level of the concepts”—beneath the level of the objects and properties that the conceptual representations represent—the world itself is permeated by arbitrarily much more thickly integrative connective detail. It is not just that

our concepts sometimes have vague or unclear boundaries; it is that these facts tell on a world that itself is not itself clear cut. (Is a boisterous child the same as or different from a rambunctious child—or a boisterous CEO? If we are climbing the tallest peaks in Canada, and there is another local maximum 100 meters away from the summit we have just reached, do we need to go over there as well? Where does one “fog” end and another start? *Reality will not tell us*. If we want “clear and distinct” answers, we need to employ conceptual schemes that impose them.)

What are the consequences of these insights for AI? What follows from recognizing that the nature of reality is as suggested in figure 6: a plenum of surpassingly rich differentiation, which intelligent creatures ontologically “parse” in ways that suit their projects?

Some of the technical implications will be explored in chapter 5. Overall, though, it means that AI needs to take on board one of the deepest intellectual realizations of the last 50 years, joining fields as diverse as social construction, quantum mechanics, and psychological and anthropological studies of cultural diversity: that taking the world to consist of discrete intelligible mesoscale objects is an *achievement* of intelligence, not a premise on top of which intelligence runs. AI needs to *explain* objects, properties, and relations, and the ability of creatures to find the world intelligible in terms of them; it cannot assume them.¹⁵

How we **register** the world, as I put it¹⁶—find it ontologically intelligible in such a way as to support our projects

15. Ontology requires naturalization too, to put this in philosophical terms.

16. See *On the Origin of Objects*, and “Rehabilitating Representation” (discussed in chapter 2, page 15, note 18).

and practices—is in my judgment the most important task to which intelligence is devoted. As I will describe in more detail in subsequent chapters, developing appropriate registrations does not involve merely “taking in what arrives at our senses,” but—no mean feat—developing a whole and integrated picture accountable to being in the world. It is not just a question of finding a registration scheme¹⁷ that “fits” the world in ways locally appropriate to the project at hand, but of relentless attunement to the fact that registration schemes necessarily impose non-innocent idealizations—inscribe boundaries, establish identities, privilege some regularities¹⁸ over others, ignore details, and in general impose idealizations and do an inevitable amount of violence to the sustaining underlying richness. This process of stewardship and accountability for registration, never imagined in the GOFAI project, is of the essence of intelligence.¹⁹

+ + +

One can (people do) say that GOFAI misconceived cognition, but as the foregoing discussion makes clear, I believe the deeper problem is that it misconceived the world. Why did it do so? One thing can be said, involving a subtlety that bedevils all discussion of AI systems’ engagement with the world.

Two “views” or registrations of the world are relevant to

17. This process is sometimes described as identifying an appropriate “level of description” or “level of abstraction” (sometimes even as “coarse-graining”), but the metaphor of “levels” can be misleading.

18. That is: the metaphysical warrant for that which we register as a regularity.

19. The failure to imagine and therefore to take up these issues is a major reason, I believe, for GOFAI’s inability to make progress on common sense. See the sidebar on the next page.

Common Sense

GOFAI's ontological presumptiveness, its blindness to the subtleties of registration, and its inadequate appreciation of the world's richness are the primary reason, in my judgment, for its dismal record on common sense. Notorious examples of early failure in this regard include systems suggesting boiling a kidney to cure an infection, and attributing reddish spots on a Chevrolet to a case of measles.^a

The initial response to shortfalls of this sort, most famously taken up in the CYC project,^b was two-pronged: first to codify, in a logical representation, all common knowledge to be found in an encyclopedia, and then, when that proved insufficient, to encode all common knowledge *not* in an encyclopedia. Those projects, and the wider “commonsense reasoning” program of first-wave AI, though still pursued in some quarters, have largely faded out of sight as models on which to base commonsense reasoners. They live on in such structures as Google’s “Knowledge Graph,” which is used to organize and provide short answers to queries in search engines. Tellingly, though, the results of such searches are snippets for interpretation by human intelligence, not the basis of the machine’s own intelligence.^c

a. Hamid Ekbia, *Artificial Dreams* (Cambridge: Cambridge University Press, 2008), 96–97.

b. See <http://www.cyc.com>, where this project lives on.

c. See my “The Owl and the Electric Encyclopaedia,” *Artificial Intelligence* 47 (1991): 251–288.

the design or analysis of an AI system, or indeed of any intentional system at all, including people. The first is the designer’s or theorist’s registration of the world in which they are building the system, and into which they will deploy it, or in which they are analyzing a system or creature—that is, the world, as registered by the theorist or

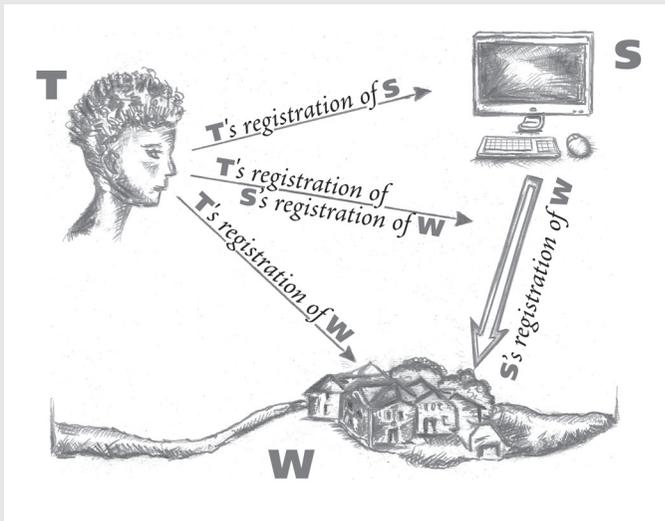
designer, that they expect that system to find intelligible, deal with, behave appropriately in. The second is the system's own registration of that world—the registration that will presumably lead it to behave and deal with the world in the way that it does. *There is no reason to suppose the two registrations will necessarily align.* As explored in greater detail in the sidebar on the following pages, the suggestion about why GOFAI may have misconceived the world is this: because first-wave AI and GOFAI took a technoscientific attitude to their subject matter of constructing an intelligent system—analyzing it in terms of its causal or mechanical constituents formally conceived²⁰—and because they did not deal with the phenomenon of registration at all, they may simply have assumed that intelligence could be constructed in a system that *itself* took a technoscientific attitude to its world.

It did not work out that way.

20. For example, analyzing it in terms of what Sellars would call the “scientific” image, as opposed to the manifest image likely used by the system being modeled.

Theorist vs. Subject Registration

An issue that comes up when theorizing intentional systems is the relation between (i) the theorist's registration of the world toward which the system in question is oriented, and (ii) the registration of the world in terms of which that system finds that world intelligible. Take T, S, and W to be the theorist, system under investigation, and world or task domain, respectively. If T is a simple realist, taking W simply to be "is as it is," then the issue for T is straightforward: T can characterize W, or at least those aspects that S deals with, "correctly," and assume that S does—or should do so—as well.



If registration is taken seriously, however—as we must in order to do justice not only to AI, but to the world and intentionality more generally—then at best it would be preemptively inscriptive, and at worst false, for T to presume that S's registration matches their own. T's registration may be adequate with respect to *nonsemantic* aspects of S's engagement—that is, adequate in terms of which to

understand S's causal interactions with W (except in as much as those causal interactions need to mesh with S's intentional registration, such as serving as satisfaction conditions for S's planned actions). A much-touted feature of dynamical systems accounts of cognitive behavior, for example, is their ability to theorize the agent-world relation in the form of the differential equations required by dynamical systems theory.

When it comes to intentional or semantic relations between S and W, however, the situation grows more complex—as depicted in the figure on the previous page. Three (related) issues are at stake: (i) how T registers W, (ii) how S registers W, and (iii) *how T registers S's registration of W*.

Although this is not the place to engage in a detailed technical analysis of how these registrations can be meshed or appropriately coordinated, a few remarks can be made. First, to the extent that T's and S's registrations align, the issue is unlikely to be problematic. The problem grows more challenging to the extent that they differ—akin to that of understanding a foreign culture or alien species. There is no reason to suppose that T will necessarily be able to “grasp,” at all, how it is that S registers W, no matter how committed T may be to doing so.

One might imagine that understanding human intelligence would be an instance of the former case, and therefore straightforward, but that is where the issue mentioned in the text comes to the fore. If T brings a scientific or technical theoretic attitude to bear on its theory of S, S's intentionality, how S perceives and thinks about the world, and so on, then T's registration of W is perforce likely to be *scientific*—treating W in terms of formal ontology: discrete objects, exemplifying clearly-defined properties and standing in unambiguous relations, and so on. Unless T is theorizing how S is a scientist, however, T's scientific registration of W and S's nonscientific registration are liable to part company.

These discussions illustrate one way to understand how first-wave AI and GOFAI may have “got the ontology wrong.” It is easy to see how, given that GOFAI was blind to registration overall, it would have made an implicit presumption that elementary intelligence could be fashioned out of the GOFAI researcher’s scientific or technical attitude toward W. Or put another way, GOFAI may have failed (ontologically) because it projected its own default *theoretic* attitude onto S.

I take it to be a consequence of the metaphysical position described in *On the Origin of Objects* that no deferential or accountable registration can be constructed from such a base.