Inference, Language, and Situated Conceptualization

Alexander Porto

April 2025

The purpose of this writing is to exercise a philosophical lesson about the role of linguistic inference in cognitive science accounts of conceptualization. I argue that while it makes good sense to endorse the idea that our conceptual scheme comes about through a complex, multimodal, contextual, and simulative process of what Barsalou calls entrenched situated conceptualization, it nonetheless requires a linguistically-articulated inferential framework to get off the ground. I defend the thesis that the cognitive process of entrenched situated conceptualization requires a transposition into inference-first terms. I do this by showing that we can engage in contextual simulated cognition if and only if we have a normative conceptual repertoire, a common ground; and that we can establish common ground if and only if our cognitive processes that afford representational capacities are seen as a way of making second-order maps of an environment.

Keywords: inference, representation, entrenched situated conceptualization, cognitive linguistics, philosophy of cognitive science

Introduction

0. The purpose of this writing is to exercise a philosophical lesson about the role of linguistic inference in cognitive science accounts of conceptualization. I will argue that while it makes good sense to endorse the idea that our conceptual scheme comes about through a complex, multimodal, contextual, and simulative process of what Barsalou calls *entrenched situated conceptualization*, it nonetheless requires a linguistically-articulated inferential framework to get off the ground. The explicit form[1] of my argument is as follows:

- **Thesis:** The cognitive process of entrenched situated conceptualization requires a transposition into inference-first terms.

[1] I follow Seiberth (2025; *forthcoming*) in structuring my argument in this way.

- **Premise 2:** We can engage in contextual simulated cognition if and only if we have a normative conceptual repertoire, a *common ground*.
- **Premise 1:** We can establish common ground if and only if our cognitive processes that afford representational capacities are seen as a way of *making second-order maps of an environment.*

I will start by clarifying the terms used throughout this writing. This will include broad uses of the terms *semanticity* and *pragmatics*, but will also involve more demanding explanations of a set of terminology used in cognitive science (especially regarding language); I will focus on three main accounts: (1) Barsalou's *entrenched situated conceptualization*, (2) Clark's *staging theory*, and Ferretti's *navigational-narrative cognition*. I will then make a bridge between these accounts and the philosophical lineage of Wilfrid Sellars (especially his use of the term *picturing*). Along the way, I will provide examples from experimental cross-cultural research about concept use to support my overall claims.

Setting the Stage

1. Any account of what we are up to when we use language has at least two obligations. On the one hand, we are obligated to account for the *semanticity* of language. The fact that the multimodal systems of language that we use have repeatable forms which hold meaning is a remarkable achievement of human cognition. On the other hand, we are obligated to account for the *pragmatics* of language: how we use semanticity to expressively organize an environment with others. That is, our linguistic conventions allow us to communicate about our environment in a stable manner: when I say, "I want the red circle", each functional term in the statement (as well as the statement as a whole) contains semantically persistent information that allows others

to *do* things with it (for instance, you can hand me the object). This is the familiar Vygotskyan lesson that humans are able to tie a knot in a handkerchief *for a reason*; the pragmatic corollary is that we can *share that reason* with others and *tell them* when they get things wrong (Sellars, 1949, 2003; Wittgenstein, 2001, 2003). Our *communicative systems*, therefore, rely at the very least on semanticity and pragmatics in order to get off the ground.

Our communicative systems are also pattern-governed and normatively regulated. In order for a person to communicate, it is necessary that they do so in linguistic communities. We lay down grammatical rules, introduce and evaluate moral language to express and uphold shared standards, and establish conventions for getting about the world in ways that allow us to justify ourselves and predict the behavior of others. But our communicative systems likewise have statistical regularities beneath the surface. It is well-established that human language exhibits systematicity (form-meaning pairings across sets of signals) and that systematicity eventually gives way to *compositionality* (form-meaning pairings that become relatively stable and obligatory among a linguistic community) (Carr et al., 2017; Dingemanse et al., 2015; Winter, Woodin, and Perlman, *forthcoming*). In an important sense, those regularities support the development of shared reasoning through, for instance, our capacity to use indexicals to make causal inferences (Brandom, 2015; McDowell, 1994; 2013).

2. The regularities of language afford a straightforward view of what it is to be a member of a linguistic community. Our cognitive capacities for what Barsalou calls *entrenched situated conceptualization*—the architecture that brings together perception, thought, and language to reliably represent and predict contextually-embedded environmental information—cooperate with our social, normative disposition to produce representations that are intelligible to others (Barsalou, 2009). Barsalou's idea is that the repeatability of salient embedded concepts is an

achievement of our capacity to *simulate* perceptual information; and the argument that representation is situated allows Barsalou to say that our conceptualizations are multimodal; after all, we have to somehow make sense of the ways that the same kind of environmental information occurs in different contexts. For instance, the term "coffee" is obviously associated with distinctly multimodal sensorimotor situations: preparing it, drinking it, spilling it, socializing with it, etc.; each of these different situated embeddings of the term require different situated embedded cognitive processes in order to correctly represent the function of the term. For Barsalou, this is possible through being able to simulate subsets of integrated information from past experiences as *modal re-enactments*: distributed, multimodal chains of events (e.g., tasting, burning, smelling, seeing, talking) in associative neurons that are contingent on frequency, context, and recency.

From Barsalou's point of view, entrenched situated conceptualizations are what make successful language use possible. The idea is that prediction is the bedrock of linguistic behavior. This is a simple point: if Jones would like to say something to Jane, then Jones must be able to predict how Jane will receive his words; this involves simulating how she will respond to his utterance. But it also involves simulating the other background context relevant to whatever it is Jones would like to say. Suppose Jones is going to ask Jane if she would like to drink a martini together. On Barsalou's account, this involves Jones modally re-enacting—and inferring that Jane is also able to re-enact—some of what is involved in drinking a martini (importantly, in a particular embedded context). So, linguistic behavior—and being a member of a linguistic community—is to use entrenched situated conceptualizations in order to predict that same capacity in others. In our example, I focused on Jones saying something, but the same point

could be made for any other of Jones's and Jane's shared linguistic behavior (gesturing, for instance).

3. Robert Brandom calls a version of this the *modal Kant-Sellars thesis*.[2] It goes like this: by becoming able to use everyday empirical vocabulary (i.e., describing observables; e.g., saying "it's sunny outside"), one already has everything they need in order to make inferential statements about theoretically-postulated unobservables (so-called *alethic modal vocabulary*; being able to say "if it were not cloudy, then it would be sunny outside" or "if certain conditions are met in the future, cold fusion will have been achieved in laboratory contexts"). The radical aspect of Brandom's Kant-Sellars thesis is that, following Sellars, we are obligated to see that Jones's simulation of Jane's responsive linguistic behavior (and her reciprocal simulation of Jones's simulation) is *not* a bit of everyday empirical vocabulary but instead is theoretically-postulated. For instance, modal re-enactments require a capacity for abstract representation; it requires one to project themselves into a different context. The consequence is that when cognitive science postulates inner experiences (or introspections), it must not take for granted that the very idea of, say, happiness is itself a linguistically mediated situated re-enactment.[3] If one does is capable of imagining happiness, then it follows that thinking about happiness is an entrenched situated conceptualization. The point is that we would not have any entrenched

[2] Brandom's use of *modal* differs from Barsalou's. Brandom is referring to *alethic modal vocabulary*: describing possible states of affairs through inductive inferential reasoning (i.e., being able to say how things are, absent of any observable empirical information).

[3] This story is told compellingly by Sellars in his *Myth of Jones* (EPM; 2003).

situated conceptualization without a world-historical situation. *That* we call a particular, situated cognitive process *happiness* does nothing to secure the existence of the inner experience. All that it does is show that we have named a particular process and have learned how to functionally use the name in an expressively intelligible way. It is therefore unknowable outside of the functional naming practices of a linguistic community, what Sellars calls *a space of reasons*. For reasons that would take us too far away, I leave a commitment to this thought as-is.

4. This is all fine and well for a social account of language. After all, it is straightforward to see linguistic behavior as instances of entrenched situated conceptualizations: we have to be able to predict, via simulation, the responsive dispositions of our interlocuters on the level of embedded cognitive processes. And part of that prediction entails a commitment to believing that our interlocuters are making the same (or at least adjacent) multimodal sensorimotor simulations as we do when we communicate with them. However, this does not yet account for the fact that the information active in our modal re-enactments is *about* anything but shared linguistic behavior itself. We have not yet got the world in view. This opens up an important part of our story: we need to be able to say something about what it is for our communicatively predictive simulations to be able to pick out material particulars—bits of the physical world—and make functional terms for them. In other words, we need to account for the ways in which our language has external constraints (McDowell, 2013; O'Shea, 2007; Seibt, 1990; Sellars, 1992).

5. *Iconicity* in language is a way into such accounting. Because languages exhibit iconicity—cross-modal resemblances between form and meaning—we are afforded a view about how material particulars factor into our entrenched situated conceptualizations. This is especially in light of the fact that iconic representations are cross-linguistic, multimodal, and part of the development and evolution of language (Irvine, 2016; Poggi, 2008; Winter, Woodin, Perlman,

forthcoming). For instance, evidence for prosody, gesture, eye gaze, and other multimodal communicative aspects of linguistic behavior have supported the idea that, whatever it is that we are doing with language, we are always somehow representing the world (Perlman and Cain, 2014; Ferretti, 2022; Fibigerova and Guidetti, 2022; Irvine, 2016; Murgiano, Motamedi, and Vigliocco, 2021; Schlenker, 2018).

6. In parallel with Barsalou, Clark (2016) argues for a distinction between *depictive* and *descriptive* uses of language cognition. Though it is left implicit in his argument, Clark advocates the idea that depictive linguistic behavior (i.e., *setting a communicative stage*) involves unmediated representational access to an external world, whereas descriptive linguistic behavior occurs at a higher level of abstraction. For example, gesturing to a point in an environment is depictive whereas using writing to describe a location is descriptive. The implication of Clark's point is that depictive linguistic behavior is prior (in evolution and human development) to descriptive linguistic behavior as a result of its supposed direct correspondence to a physical environment. Clark (2016, p. 324–325, emphasis added) says:

> With descriptions, people rely mainly on their knowledge of the vocabulary, syntax, and semantics of, say, English or Japanese and their ability to categorize. With depictions, they rely mainly on their visual, auditory, tactile, and proprioceptive knowledge of physical scenes and on their **ability to use one scene in imagining another**. Describing depends on knowledge of a language or code. Depicting does not. (…) What, then, are depictions? In the theory proposed here, called the *staging theory*, they are **physical scenes that people stage for others to use in imagining the scenes depicted**.

Despite philosophically problematic assumptions that I will return to below, we already see clear affinities between Clark's *staging theory* and Barsalou's *entrenched situated conceptualization*, as well as clear ways that iconicity factors into simulated conceptual representations. It is also easy to imagine ways in which physical scenes are eligible for depiction before being eligible for description. The throughline is that certain salient features of the environment are available for use and re-use (with statistical regularity) in linguistic representation, hence communication. This also shows a clear phylogenic link between iconicity, systematicity, and compositionality: while iconicity is asymmetrically distinct from systematicity (i.e., languages can be systematic without iconicity, but iconicity is always at least partially systematic), it obviously plays a role in our capacity to abstract away from an environmental context of material particulars; likewise, as abstracted systematic bits of language become more statistically regular and normatively regulative toward compositionality (i.e., evaluatively available to a linguistic community), this clearly shows *one* way to understanding how language becomes compositionally stable over time, on an ontogenic and phylogenic point of view.[4]

7. Now, there are adjacent movements in philosophy of mind and language that partially resonate with the ideas of the relation between mind, language, and the world that Clark and Barsalou offer.[5] However, cognitive accounts of this relation have important lessons to learn from their

[4] At this point, it should be clear that I endorse the Vygotskyan idea that ontogeny recapitulates phylogeny. I leave this as a now-explicit commitment and do not elaborate it further.

[5] Obviously among others. I use their names to represent the distinct school of thought in the cognitive accounts of language that I have described so far.

philosophical relatives. Sellars, for instance, argues for the idea that representation is an exercise of a *picturing system*. For reasons too distant for this writing, Sellars takes *picturing* to be part and parcel of the same philosophical function of mind and language; the point is that thinking and languaging are, for Sellars, two sides of the same coin.[6] So, when I say *picturing*, it should be kept in mind that this is a two-sided, parallel act of cognitive representation as well as language use.

Picturing accounts argue that our communicative systems are isomorphic with reality: the structure of a particular situation becomes modeled through linguistic expression. That is, in order for language to express reality, there must be something in common between our communicative systems and the material of which those systems are about. However, in an important sense described below, that isomorphism is always *analogous,* i.e., second-order. For example, statements like "this cube is red" or "it is raining in Pittsburgh" obviously do not make the external situations that depict true or false; instead, such statements are models that can be verified or rejected by consensus in a communicative system. Obviously this accords with the models proposed by Barsalou and Clark. Taken together, then, our language systems form a sort of map with which we navigate our environment. The items in our environment stand in causal

[6] Sellars shows that we should model thought on verbal behavior: while we are trained to use language, part of what we are doing is learning to think silently; then, our speech acts are a kind of thinking-out-loud. Imagine if language evolved without accounting for concepts "inner states" or "private thoughts". cf. Sellars, *EPM*, 2003, XII. §§48–50, XIII. §§51–52, and XIV. §§53–55.

relations to the ways we picture them through linguistic behavior: if the environment were otherwise, so too would our linguistic behavior be otherwise.

8. A similar argument is outlined by Ferretti (2022), who proposes a cognitive macrosystem for framing language (in this case, narrative language) as a pragmatic picturing achievement of human evolution. Ferretti demonstrates that in order for persons to use abstract concepts—to use language to detach from the present moment—they must exercise a complex, multimodal matrix that involves spatio-temporal cognition via self-projection (i.e., putting oneself in simulated contexts), scenario construction (i.e., representing perceptual information as narrative events; seeing the world as eligible to act in), and social cognition (viz., perspective-taking; emotional involvement with others). By Ferretti's lights, the way we can cash out the coherence this macrosystem is by connecting it to navigational cognition:

> The metaphor of storytelling as a form of spatial navigation is highly intuitive. When we tell someone about something, in constructing the story we are guided by the goal and the path to follow to achieve that goal. As in spatial navigation, maintaining the line of discourse is crucial to maintain the right trajectory towards the story end. Thus, the ability to keep a course in the right direction is a central skill for reaching a given destination in both space navigation and storytelling. In narrative terms, this ability fits with the construction of global coherence (p. 99).

Using clinical case examples that show how impairments in spatial domains entail impairments in narrative domains (e.g., difficulty in using landmarks on a territory entails difficulty interpreting important plot-points in narrative; traumatic brain injury in the prefrontal cortex as associated with impairments in planning and organizing motor activity outside of localized instances as well as difficulty with organizing globally coherent stories and reports), Ferretti

makes a strong case for the idea that there is an isomorphic cognitive relationship between telling a story and following a path. Ferretti refers to Trabasso et al.'s *Causal Network Model* (CNM) as a suitable explanatory model for grounding the idea that thought, language, and spatio-temporal cognition are part of the same cognitive system.

Inferring the Stage

9. But all of this talk of isomorphic picturing between the various sensorimotor, multimodal aspects of cognitive representation and the world is obligated to incorporate an important turn that Sellars urges us to take: we must build our accounts of mind and language without presuppositions of givenness. There is a problematic knowledge-inheritance from classical empiricism à la Hume, Berkeley, and Locke which advocates the idea that epistemology is possible if and only if it is derived from sense experience. That is, classical empiricism says that what we know hangs only on what we perceive; in this way, it is assumed that we have unmediated access to perceptual information that is given in experience. For instance, properties of perception (*red*), inner experience (*happiness*), and categories (*lionhood*) are taken to be available to our perceptual capacity, itself a tabula rasa, independent of any conceptual, linguistic mediation.

Versions of this thought are explicitly present across other philosophical schools (e.g., phenomenalism, phenomenology, and logical empiricism); and it is implicitly present across the cognitive sciences, most clearly in Bayesian theory of mind (i.e., iterated predictive processing of an environment via unmediated access to interaction with others), innatist linguistics (e.g., universal grammar) and empiricist accounts of perception. (Bottemanne, Longuet, and Gauld, 2022; Carlson & Zelazo, 2008; Perner, 1993; Chomsky, 1991; Stacy, Gong, Parab, Zhao, Jiang,

Gao, 2023). More, givenness is also assumed in the accounts of Barsalou, Clark, and Ferretti that I outlined above. Whether situated simulations, depictive theories of communication, or narrative-navigational cognitive architecture, the idea that we have direct epistemological access—that we can somehow make knowledge out of what Rorty calls "raw feels" and what Sellars calls noninferential "sense data"—is taken for granted (Rorty 1979/2018). Sellars attacks this idea under the banner of *The Myth of the Given* (Brandom, 2015; McDowell, 2013; O'Shea, 2007; Seibt, 1990; Sellars, 1949; 1954; 1992; 2003).

10. Sellars shows how the Myth of the Given is a mistake. Take, for instance, the idea of perceiving something green. If we assume that we have direct access to a universal property *green*, available to us through raw perceptual activity, then we must offer an account of how it is possible to acquire those universal, conceptually-articulated concepts. In other words, accounts of cognition that treat the mind as a tabula rasa, whereby extralinguistic entities (e.g., *green* material particulars) impress upon us as sensations that form our knowledge of those entities, must also explain how we come to *know* those extralinguistic entities. Any attempt to do so from the tabula rasa perspective, however, collapses in vicious regress: it is *green* because it *is green* and we know that because that is what *green* objects look like, which we only know through other *green* objects, etc. So, for example, when Clark says that depictions "rely mainly on their visual, auditory, tactile, and proprioceptive knowledge of physical scenes" and that "[d]escribing depends on knowledge of a language or code. Depiction does not", we are left to wonder just what depiction *does* depend on. After all, isn't knowledge of physical scenes contingent, according to Barsalou, on my capacity for entrenched situated conceptualization? If depiction is based on that knowledge, but is also independent of a "language or code", how do I know what features of the environment are salient for the conceptualization to be intelligible? Because

things look a certain way? But how do I know that my conceptualization is accurate? And if I don't know, then how do I know that your conceptualization is accurate? And so on.

11. Similarly, when Barsalou (2009, p. 1282) argues that "simulators for abstract concepts generally capture complex multimodal simulations of temporally extended events, with simulation of introspections being central" and that "simulators develop to represent categories of internal experience just as they develop to represent categories of external experience", there is no explanation of where such categorial knowledge comes from. It therefore becomes imperative to see just how it is that persons are able to selectively identify distinct categorial knowledge, despite the fact that such categorial knowledge itself depends upon the capacity for one to exercise entrenched situated conceptualizations in order to make it intelligible. And simply saying persons have direct access to internal and external experiences cannot get off the ground, just as saying that we know what green is because we know what green *is* cannot either. Even though the Myth of the Given is most subtle in Barsalou's account, it remains unclear how we are supposed to build an epistemology out of simulating four basic types of information (people and objects, actions, introspections, and settings) over repeated category-specific exposure if the achievement of situated conceptualization hangs on whether or not one already knows what to simulate.

12. The lesson from Sellars is that those situations are themselves conceptually articulated *through language*. We cannot argue for situated conceptual representation while also argue for access to an unmediated perceptual environment with extralinguistic universal properties. It cannot be argued, for instance, that the conceptual simulation of the perceptual event *seeing that the square is red* refers to an abstract entity like redness. The perceptual event itself has no *red*

things, as it were, inside of it; the conceptual simulation of this event is linguistically articulated in a space of reasons. Sellars (2003) says:

> The essential point is that in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says (EPM §36).

The way out of the Myth of the Given is by seeing that we can only know which aspects of a situation are salient to cognition through a distinctly anti-foundationalist epistemology (Seibt, 1990). The basic idea is that, rather than try to identify an extralinguistic foundation to knowledge or conceptual representation, we ought to see our capacity to linguistically articulate (and narratively navigate, as Ferretti would have it) the world as trained patterns of responses, passed down through social traditions. On this normative picture, experiential knowledge is not autonomous to a linguistic community. It is instead an inference—a justified reason—that follows distinct patterns based on the ways that persons of a linguistic community have converged on, have been trained on, and continue to train others to respond to material particulars that in themselves do not *contain* the extralinguistic properties that epistemologies of givenness require.

13. The advantage to Sellars's alternative is that it takes Clark's depictive account of language, Barsalou's entrenched situated conceptualization, and the iconic features of language systems and shows a way to say, for instance, what we are *doing* when we make depictions, how situations become entrenched, and why language evolution involved form-meaning correspondences without falling into incomprehensible regress. Take, for instance, Barsalou's point that simulative prediction is the bedrock of language. Suppose that Jones would like to

indicate to Jane that he would like a blue object out of a set of green, blue, and yellow objects. But this time, suppose that Jane is a monolingual speaker of Tsimane' and does not have distinguishing color-predicates for those three objects; and suppose that Jones is a monolingual speaker of English. What are the necessary conditions for Jones to successfully predict that Jane will reciprocally simulate the color that Jones is trying to indicate in order to form a joint-attentional framework on the correct object? How can that situation become a mutually entrenched situated conceptualization?

14. The answer is in Sellars's twofold argument of anti-foundational nominalism and functional role semantics. The basic idea is that there must be conceptualization before there is conceptual perception; this conceptualization hinges on material particulars (and later, abstract concepts) having designated functional roles in a language system. Somehow, Jones and Jane must come to establish a mutually-intelligible functional role for distinguishing (at least) blue objects from green and yellow objects. Sellars's idea is that Jane would *not* have a distinguishing cognitive system that accounts for such functional role differences in the three color terms until, of course, she acquired one in a conceptually-articulated language-game. Seibt (1990) echoes Sellars:

> Not only is experience not the foundation of the formation of concepts; conversely experience presupposes concept formation. The language of observation, as part of the reticular functional system of concepts, is acquired all at once and only in connection with all other parts of that system. For example, mastering the use of color predicates means not only reacting under standard conditions with semantically correct language episodes; one must also be able to recognize when standard conditions obtain. Knowing that an object is not green if it is red under standard conditions, but that it can be green if it appears blue under

> non-standard conditions; that one can make a green object look blue; that a colored object is also extended: all this and more belongs to the meaning of color predicates.

Malik-Moraleda et al., (2023) validate this idea by showing that monolingual speakers of Tsimane' do not reliably distinguish between blue and green colors, whereas bilingual Spanish-Tsimane' speakers do. The point is that if and only if one has explicit nominal, linguistic functional roles active in experience does one have the ability to recognitively know meaningful differences in hue. This suggests that persons' capacity to perceive color (in an epistemological sense of knowing that one is perceiving a color) is indeed linguistically-mediated; visual perception, then, hangs on whether or not entrenched situated conceptualizations are themselves conceptually structured, or, in the case of bilingual speakers of Spanish-Tsimane', become re-structured through second language acquisition (Conway, Malik-Moraleda, and Gibson, 2022; Gibson et al., 2017; Malik-Moraleda et al., 2023).

15. To take stock: we now have a picturing account of mind and language that integrates Barsalou's entrenched situated conceptualization, Clark's depictive account of language, and Ferretti narrative-navigational cognitive macrosystem that does not rely on assumptions of givenness. We have shown how in order to simulate and predict, in order to use depictive language, and in order to structure perceptual experience in narrative-navigational form, it is necessary that shared linguistic conceptualization is already in play. We have also indicated toward a way of seeing how it is that visual perception becomes more granular once it is re-mapped into a new conceptual system. What is left to show is how we can make sense of the requirement that our language corresponds to the world as cognitively complex picturing system.

Mapping the Stage

16. Pragmatic inferential cognitive capacities are a promising starting point. In a study investigating communicative differences across language systems, Ryskin et al. (2023) tested how monolinguals speakers of English and Tsimane' responded to noun-adjective tokens (e.g., *a large cup*; *a small banana*) in a referential communication task and a visual world eye-tracking task. In the referential task, participants were paired with a conversation partner and were tasked with describing one of four objects to their partner in order to probe the production of referential expressions. In the eye-tracking task, participants heard audio recordings of short imperative statements (e.g., *show me the big cup*; *show me the small banana*) while looking at a digital display of four objects and were asked to point to the object. The eye-tracking task was based on past findings that suggest that hearing referring expressions supports real-time interactive inference about a speaker's intentions. Of relevance here, Ryskin et al. found that "both English and Tsimane' participants were more likely to make anticipator looks at the target (and fewer to the competitor) when it was in a contrast set than when it was not" (2023, p. 1255). Taken together, the authors conclude that, regardless of degree of how referential language is conventionalized, interactive conceptualization requires real-time inferences of perceptual context (visuo-spatial information) and reasoning about the knowledge and communicative goals of others. Ryskin et al. conclude that sophisticated inferential prediction plausibly relies on language systems but that predictions about the intentions of others emerge earlier in development and across cultures.

If that is the case, then we can put forward the idea that pragmatic social inference—reasoning about and forming joint-attentional common ground with others—supports the development of linguistic conceptualization of a situated, simulated environment. Our

conceptualizations and situated representations, then, can be framed in terms of dynamic and interactive negotiations of categorial knowledge (e.g., first, joint-attentional organizations of perceptually available information in the environment, and later, abstract theoretical language about postulated unobservables). The point is that the cognitive processes required for making inferences comes logically prior to those required for conceptualizing perception into bits of knowledge. By contrast to latent classical empiricist thought in cognitive science, we instead have a justified position to believe that we do not first acquire perceptual knowledge and then come to make inferences using it; it is instead that we cultivate the capacity for inferential reasoning (i.e., having a sense of how things stand in a shared representational scheme) and organize perception in a way that accords.

17. If we take Barsalou's account without assumptions of givenness and apply it to English-speaking Jones and Tsimane'-speaking Jane, it is clear that what precedes Jones successfully having Jane give him the blue object is a pragmatic, inferential framework wherein Jones and Jane can mutually simulate situated conceptualized ways of getting on the same page. There cannot be the simulation of salient features of an environment without first coming to know—developing an inferential interactional position—which features of an environment are relevant to and eligible for simulation. So, we do not lapse into assuming that the world is endowed with an extralinguistic categorial structure; but we also do not lapse into assuming that every conceptual categorial representational scheme we develop is relatively constructed, without the world serving as an external constraint. For while it is obviously the case the human visual field perceives different hues of physical light, we would not want to thereby state that physical light is itself colored according to our visual field.

We are left, then, with an account of cognition that does justice to the fact that we can use depictive language via entrenched situated conceptualization—after all, we *can* point at material particulars in the world. But we also have an account that does justice to the fact that we do not have direct, unmediated access to conceptual contents. Call this account: *inferentially-entrenched situated conceptualization*.

18. I argue that our capacity to engage in inferentially-entrenched situated conceptualization affords us a cognitive system that makes second-order maps of our environment through the production of semantic artifacts. For, while we cannot say that we come to know what *green* is through learning the extralinguistic categorial structure of the world, we *can* say that, all things equal, persons *do* reliably respond to material particulars based upon the semantic artifacts that members of a linguistic community have converged upon. And this convergence of semantic artifacts is what makes representational capacities second-order: our language obviously does not stand in a one-to-one relation to the world (e.g., the term DOG looks nothing like the animal the term plays a role for).

The use of map-making as analogous to our representational capacities is a productive example: if we wanted to make a first-order isomorphic map of Pittsburgh, it would be necessary to make it in full, three dimensional scale; in such a situation, we would have nothing other than a direct replica of Pittsburgh, perhaps made of different materials (but then, it wouldn't truly be a first-order isomorphism). Surely that is not what I suggest is going on. Instead, when we make a second-order isomorphic map of Pittsburgh, we know that, for example, the first-order Penn Avenue is roughly 14km long but that our second-order picture of it is 140cm long. We neither expect nor require that our second-order Penn Avenue to be 14km long in order to know how to use our map to get from, say, Garfield to Wilkinsburg. Then, quite naturally, we can develop a

common ground with others for understanding how to use our second-order map—a semantic artifact—in order to *put it to use*.[7] In the end, I am suggesting that our inferentially-entrenched situated conceptualizations are nothing other than producing and learning how to re-produce maps with others. That is, I am arguing that what it is to be representational is nothing other than what it is to be linguistic: to put to use semantic artifacts in a shared space of reasons. In accord with Ferretti, we can think of this second-order isomorphic mapping as the operationalization of a complex, multimodal, sensorimotor, and inferential cognitive macrosystem.

19. Though motivated by different reasons, the inferentially-entrenched situated conceptualization model I advocate has already been, at least partially, experimentally vindicated. In a pioneering series of studies, Galantucci (2005) shows that it is possible to investigate the emergence of language (which, in the case I am arguing here, is the emergence of conceptualization) in the absence of pre-established forms of communication like speech and writing. Galantucci had dyads of participants engage in computer-based problem-solving tasks that required communication without verbal contact or the ability to use conventional writing. Specifically, participants were required to send graphical messages to each other using a stylus on a digitizing pad; the digitizing pad was designed so that horizontal stylus movements were stable but vertical stylus movements were abstracted in real-time through continuous downward

[7] To go a step further, when we use abstract language (say, of theoretical science), we are still in the business of producing second-order maps; the important difference is only in degree of abstraction.

drift on the canvas. Galantucci found that communication systems did emerge relatively quickly across ten dyads, despite the constraints put on means of communication.

20. Since Galantucci, different versions of so-called artificial language learning designs have developed with increasing complexity (usually through constraint and abstraction) (Cuskley, 2019). A pivotal development in artificial language learning experiments which investigate the cultural evolution of language has been through iterated learning designs. Kirby et al. (2008), for example, introduce a way of leveraging diffusion chain studies (i.e., a participant observes a behavior and then is required to replicate the behavior in front of a second participant, who does the same before a third, etc.) that integrates artificial language learning tasks. Implementations of Kirby's iterated learning design—ranging from slide whistles, drawing and writing tasks, robotic communication, and other multiplayer interactive symbol games—have consistently shown that the development of communicative systems can successfully be studied in laboratory contexts (Delgado et al., 2018; Morin et al., 2022; Scott-Phillips & Kirby, 2010; Verhoef, 2012; Verhoef & Ravignani, 2021).

21. The throughline of these examples is that, in the absence of pre-established communicative norms, persons had to develop an inferential framework about material particulars in order to simulatively predict what their interlocuter was trying to convey. The Sellarsian point is that participants in Galantucci's experiment could not have, say, successfully carried out the instructions of their interlocuter without first making complex relational inferences between the interlocuter's intention, depicted signal, and salient aspects of an environment. Only then was the emerge of conventionalized categorial knowledge—second-order isomorphic cognitive maps—possible. And in Kirby et al.'s (2008; 2014) case, the language systems developed over

generations of learners became less complex over iterations, especially in comparison to language systems in closed-dyad interactions.

This makes sense on the inferentially-entrenched situated conceptualization model: as the range where one makes inferences about person-environment dynamics increases (e.g., needing to account for more persons or more features of an environment), communicative signals must become simple enough that cognitive load is adaptively minimized (e.g., as demonstrated by compositionality in language evolution). The affordance of minimizing cognitive load in inference systems is that simulatively predicting the contextually-embedded actions and situations (of oneself and others) becomes efficient for transmission to new members of a community. So, when we need to convey features of an environment to others, we better be able to do so in a way that maximizes inference and minimizes complexity.[8]

22. The inferentially-entrenched situated conceptualization model also fits well with inferential theories of learning (Kaufman, 2012; Seel, 2012). In inferential learning, distinctions are made between different kinds of reasoning: deductive (i.e., conclusions materially derived conditional propositions), inductive (i.e., conclusions inductively inferred from conditional probability statements), and abductive (i.e., conclusions identified to be the simplest and most plausible from

[8] This idea converges nicely with the second-order cybernetics conceptualization of eigenforms. While a discussion of this is besides the point here, I think that exploring eigenforms as a cognitive-mathematical model of language evolution would provide promising results. cf. Kauffman, 2017; 2003; von Foerster, 1985; 2003.

conditional probability statements).[9] More, a relevant upshot of inferential theories of learning is that understanding how thought and language are extendable (e.g., systematic and compositional) is hand-in-glove with the fact that thought and language are contextually-embedded. That is, once one has a grasp on inference-rules in a particular context—how to use a map in a shared context—it is possible to extend those rules into unknown or unobserved contexts.

23. The need for an inferential ground for representational capacities is also minimally present in Barsalou (2017) in a short argument against the idea that concepts have a stable core. Barsalou suggests that because humans share similar bodies, cognitive systems, and coordinated environments, it is reasonable to expect similar distributed simulative abilities (i.e., coming to simulate diverse multimodal information in similar ways) allow persons to identify common ground (Clark, 1996, as cited in Barsalou, 2017). Importantly, common ground provides extensional feedback which serves as a pragmatic (cognitive) situational constraint on concept processing and learning. To use Barsalou's example, when I say, "Here's my *pet fish*", the salient, multimodal, contextually-embedded features of the distinct, semantically dissimilar concepts *pet* and *fish* allow for efficient conceptual processing if and only if extensional feedback is present. If you were of a culture that did not have pet fish, it would be necessary to pragmatically establish a concept that would be mutually intelligible in order to build common

---

[9] Examples:

Deductive: All men are mortal. Socrates is a man. So: Socrates is mortal.

Inductive: *Pr(0.92)* that *x* is *y* in *C*. So: If *x* were in any *C*-circumstance, it would be *y*.

Abductive: In light of the fact that all known *x* are *y* in *C*-circumstances, it plausible to believe that *x* would continue to be *y* in any *C*-circumstance.

ground, which would perhaps lack some of the multimodal features that facilitate concept processing.

Barsalou (2017, p. 22) says: "Counter-intuitively, perhaps, understanding a phrase or sentence about a non-present situation may often draw heavily on extensional feedback, whereas understanding what a phrase of sentence means in the current situation may sometimes rely more on concept composition." But I do not think that this is at all counter-intuitive: in order to have common ground representations based on entrenched situated conceptualization, one must have first mastered the inferential capacities necessary for processing complex, multimodal information. Only then is common ground possible.[10]

24. Finally, there is an evolutionary point to be made about representational language hinging on inferentially-entrenched situated conceptualization. Prior to early depictive symbolic representation (paradigmatically, parietal arts), humans had nothing more than vocal tracts to produce rapidly fading signals in order to communicate with one another. It is reasonable to imagine that the capacity for abstract conceptualization was minimal; after all, if we have no persistent, conventionalized linguistic register, what is able to be stored for later simulation must be limited (Burling, 2007; Hockett and Hockett, 1960). At some point, however, it became necessary for humans to establish common ground with one another. Said in terms of the inference-first account of language I am urging, it became important for persons to infer under which circumstances certain modes of expression were relevant and which were not. The consequence of this pragmatic, inferential need was to establish more stable modes of expression

[10] The Tsimane' example above demonstrates this.

that enabled the transmission of contextually-embedded information to others. Hence, the advent of cave paintings and other pictorial representations.[11] Increased complexity in the inferentially-mediated conceptual register served as an external constraint for those semi-stable, contextually-fixed signals to become modular, extendable, and efficiently replicable. Hence, the advent of writing systems. As our conceptual scheme continues to increase in complexity, we continue to establish more permanent, distributed, and accessible forms of transmitting contextually-embedded concepts to others (e.g., digital texts, email, photocopies, etc.). The point is that the evolution of our semantic artifacts can be clearly traced along a phylogeny once those semantic artifacts are put in terms of inferentially-mediated map-making practices.

Closing the Stage

25. The purpose of this writing has been to show that language and conceptual representation—two sides of one map-making coin—hinges on two fundamental, cross-modal, and contextually-embedded capacities: (1) the capacity to establish second-order semantic artifacts that are eligible for conventionalization; and (2) the capacity to engage in inferential reasoning in a space of reasons. I have shown this through reframing Barsalou's model of simulative conceptualization in inference-first terms; I have shown how second-order cognitive mapping supports the emergence of well-established features of language (iconicity, systematicity, and compositionality) and how those features are inferentially-mediated; and I have offered a philosophical treatment of ways in which contemporary cognitive science risks taking

[11] Given the scope of this proposal, I gloss over other arguments (cf. Pinker, Boyes, Miyagawa) about the evolutionary role of the parietal arts during the Paleolithic.

conceptualization as given. Along the way, I have provided evidence from ecologically-valid experimental research that supports the philosophical commitments underscoring this argument. The result is that language is to be seen as the exercise of an inferentially-entrenched situated conceptualization that occurs only in a shared space of reasons.

References

Barsalou, L. W. (2017). Cognitively plausible theories of concept composition. *Compositionality and concepts in linguistics and psychology*, 9-30.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. Philosophical transactions of The Royal Society B: biological sciences, 364(1521), 1281-1289. https://doi.org/10.1098/rstb.2008.0319

Bottemanne, H., Longuet, Y., & Gauld, C. (2022). L'esprit predictif : introduction à la théorie du cerveau bayésien [The predictive mind: An introduction to Bayesian Brain Theory]. *L'Encephale, 48*(4), 436–444. https://doi.org/10.1016/j.encep.2021.09.011

Brandom, R. (2015). *From empiricism to expressivism: Brandom reads Sellars*. Harvard University Press.

Burling, R. (2007). *The talking ape: How language evolved* (Vol. 5). Oxford University Press.

Carlson, S. M., & Zelazo, P. D. (2008). Symbolic Thought. In *Encyclopedia of Infant and Early Childhood Development* (pp. 288–297). Elsevier. https://doi.org/10.1016/B978-012370877-9.00158-4

Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The Cultural Evolution of Structured Languages in an Open-Ended, Continuous World. *Cognitive Science*, *41*(4), 892–923. https://doi.org/10.1111/cogs.12371

Chomsky, N. (1991). Universal grammar. *The New York Review of Books, 38.*

Clark, H. H. (2016). Depicting as a method of communication. *Psychological review, 123*(3), 324. https://psycnet.apa.org/doi/10.1037/rev0000026

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Conway, B., Malik-Moraleda, S., & Gibson, E. (2022). Hering's opponent-colors theory fails a key test in a non-Western culture. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44, No. 44). https://escholarship.org/uc/item/15j5h368

Cuskley, C. (2019). Alien forms for alien language: Investigating novel form spaces in cultural evolution. *Palgrave Communications*, *5*(1), 1–15. https://doi.org/10.1057/s41599-019-0299-5

Delgado, T., Ravignani, A., Verhoef, T., Thompson, B., Grossi, T., & Kirby, S. (2018). Cultural transmission of melodic and rhythmic universals: 4 experiments and a model. *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)*. The Evolution of Language. Proceedings of the 12th International Conference on the Evolution of Language (Evolang12). https://doi.org/10.12775/3991-1.019

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Ferretti, F. (2022). *Narrative persuasion. A cognitive perspective on language evolution* (Vol. 7). Springer Nature.

Fibigerova, K., & Guidetti, M. (2022). Iconicity in gesture. *Iconicity in Cognition and across Semiotic Systems, 18*, 245. https://doi.org/10.1075/ill.18.12fib?locatt=mode:legacy

Galantucci, B. (2005). An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science*, *29*(5), 737–767. https://doi.org/10.1207/s15516709cog0000_34

Galantucci, B., Kroos, C., & Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies, 11*(1), 100-111. https://doi.org/10.1075/is.11.1.03gal

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences, 114*(40), 10785-10790. https://doi.org/10.1073/pnas.1619666114

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American, 203*(3), 88-97. https://www.jstor.org/stable/24940617 https://repository.rit.edu/cgi/viewcontent.cgi?article=1819&context=theses

Irvine, E. (2016). Method and evidence: gesture and iconicity in the evolution of language. *Mind & Language, 31(2)*, 221-247. https://doi.org/10.1111/mila.12102

Kauffman, L. H. (2017). Eigenform and reflexivity. *Constructivist Foundations, 12*(3), 246-252. http://constructivist.info/12/3/246

Kauffman, L. (2003). Eigenforms: Objects as tokens for eigenbehaviors. *Cybernetics & Human Knowing, 10*(3-4), 73-90. https://journals.isss.org/index.php/proceedings50th/article/view/811

Kaufman, K.A. (2012). Inferential Theory of Learning. In: Seel, N.M. (eds) *Encyclopedia of the Sciences of Learning*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6_1787

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the*

National Academy of Sciences, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Malik-Moraleda, S., Mahowald, K., Conway, B. R., & Gibson, E. (2023). Concepts Are Restructured During Language Contact: The Birth of Blue and Other Color Concepts in Tsimane'-Spanish Bilinguals. *Psychological Science, 34*(12), 1350-1362. https://doi.org/10.1177/09567976231199742

McDowell, J. H. (1994). *Mind and world*. Harvard University Press.

McDowell, J. H. (Ed.). (2013). *Having the world in view: Essays on Kant, Hegel, and Sellars* (1st paperback ed). Harvard Univ. Press.

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating language in the real-world: the role of multimodal iconicity and indexicality. *Journal of Cognition, 4*(1). https://doi.org/10.5334/joc.113.

O'Shea, J. R. (2007). *Wilfrid Sellars: Naturalism with a normative turn* (1. publ). Polity Press.

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture, 14(3),* 320-350. https://psycnet.apa.org/doi/10.1075/gest.14.3.03per

Perner, J. (1993). *Understanding the representational mind* (1. MIT Press paperback ed). MIT Press.

Poggi, I. (2008). Iconicity in different types of gestures. *Gesture*, 8(1), 45-61. http://dx.doi.org/10.1075/gest.8.1.05pog

Rorty, R. (1979/2018). *Philosophy and the mirror of nature*. First Princeton Classics edition.

Ryskin, R., Salinas, M., Piantadosi, S., & Gibson, E. (2023). Real-time inference in communication across cultures: Evidence from a nonindustrialized society. J*ournal of*

Experimental Psychology: General, 152(5), 1245–1263. https://doi.org/10.1037/xge0001345

Schlenker, P. (2018). Iconic pragmatics. *Natural Language & Linguistic Theory, 36*, 877-936. https://doi.org/10.1007/s11049-017-9392-x

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*(9), 411–417. https://doi.org/10.1016/j.tics.2010.06.006

Seel, N.M. (2012). Inferential Learning and Reasoning. In: Seel, N.M. (eds) *Encyclopedia of the Sciences of Learning. Springer, Boston*, MA. https://doi.org/10.1007/978-1-4419-1428-6_583

Seiberth, L.C., (2025; forthcoming). Hypergraphing the space of reason. To appear in (L. Koscis & Pete, Eds.) Nature and Norms in a Stereoscopic View: Sellars' Images and the Philosophy Between. Bloomsbury.

Seiberth, L.C. (2022). *Intentionality in Sellars: A transcendental account of finite knowledge*. Routledge, Taylor & Francis Group.

Seibt, J. (1990). *Properties as processes: A synoptic study of Wilfrid Sellars' Nominalism*. Ridgeview Publ.

Sellars, W. (1949). Language, rules and behavior. In Hook, S. (Ed.), *John Dewey: Philosopher of science and freedom* (pp. 289–315). Dial Press New York. https://sites.pitt.edu/~rbrandom/Courses/2023%20Sellars/Sellars%20texts/Language%20Rules%20and%20Behavior%201949.pdf

Sellars, W. (1954). Some Reflections on Language Games. *Philosophy of Science*, *21*(3), 204–228. https://doi.org/10.1086/287344

Sellars, W. (1992). *Science and metaphysics: Variations on Kantian themes*. Ridgeview.

Sellars, W. (2003). *Empiricism and the philosophy of mind* (4. print). Harvard University Press.

Stacy, S., Gong, S., Parab, A., Zhao, M., Jiang, K., & Gao, T. (2024). A Bayesian theory of mind approach to modeling cooperation and communication. *Wiley Interdisciplinary Reviews: Computational Statistics, 16*(1), e1631. https://doi.org/10.1002/wics.1631

Verhoef, T., Kirby, S. and de Boer, B. (2016), Iconicity and the Emergence of Combinatorial Structure in Language. Cogn Sci, 40: 1969-1994. https://doi.org/10.1111/cogs.12326

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*(4), 357–380. https://doi.org/10.1515/langcog-2012-0019

Verhoef, T., & Ravignani, A. (2021). Melodic Universals Emerge or Are Sustained Through Cultural Evolution. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.668300

von Foerster, H. (1985). Gegenstände: greifbare Symbole für (Eigen-)Verhalten. In: *Sicht und Einsicht. Wissenschaftstheorie Wissenschaft und Philosophie*, vol 21. Vieweg+Teubner Verlag, Wiesbaden. https://doi.org/10.1007/978-3-663-13942-3_11

von Foerster, H. (2003). Objects: tokens for (eigen-) behaviors. *Understanding understanding: Essays on cybernetics and cognition*, 261-271. https://doi.org/10.1007/0-387-21722-3_11

Wittgenstein, L. (2001). *Tractatus logico-philosophicus* (D. Pears & B. McGuinness, Trans.). Routledge classics.

Wittgenstein, L. (2003). *Philosophical investigations: The German text, with a revised English translation* (Anscombe, G.E.M., Trans.; 3rd ed). Blackwell Pub.