An Outside View of Ourselves as a Toy Model AGI

Reza Negarestani

Agency, its functions and transcendental structures

We begin by a conceptual question regarding analysis, modeling and construction of human-level intelligence. This is as much a question in the philosophy of mind as it is a question about how we should go about constructing artificial general intelligence. Roughly speaking, AGI or artificial general intelligence is taken to be a hypothetical artificial agency or an artificial multiagent system that has at the very least all the capacities of the human agent, namely, it is endowed with theoretical and practical cognitions. Now the question is how much something that has at the very least theoretical and practical abilities that we have corresponds to or diverges from conditions necessary for the possibility of mind namely the conditions necessary for that which makes us human. This question can be condensed as follows:

Does/Should AGI mirror humans or does/should it diverge from them?

The answer depends on a number of presuppositions: the level of generality in General Intelligence, what we mean by the human, and whether the question mirroring and divergence is posed at the level of functional capacities or the structural constitution, methodological requirements necessary for the construction of AGI or diachronic consequences of its realization.

Answer:

The answer to this question depends on a number of presuppositions: the level of generality in General Intelligence, what we mean by the human, and whether the question mirroring and divergence is posed at the level of functional capacities or the structural constitution, methodological requirements necessary for the construction of AGI or diachronic consequences of its realization.

If we are parochially limiting the concept of the human to a certain local and contingently posited conditions –namely, a specific structure or biological substrate and a particular local transcendental structure of experience—then the answer is divergence. Those who limit the significance of the human to this parochial picture are exactly those who advance parochial conceptions of AGI. There is a story here about how anti-AGI skeptics (specifically those who think biological structure or the transcendental structure of the human subject are foreclosed to artificial realizability) and proponents of parochial conceptions of AGI (i.e. those who think models constructed on a prevalent 'sentient' conception of intelligence, inductive information processing, Baysian inference, problem-solving or emulation of the physical substrate are *sufficient* for the realization of AGI) are actually two faces of the same coin. Positions of both camps originate from a deeply conservative picture of the human which is entrenched either in a biological chauvinism or provincial account of subjectivity. The only thing that separates them is their strategy toward their base ideological assumption: the skeptics inflate this picture into a rigid anthropcentricism and the proponents of parochial AGI attempt to vastly deflate it. Thus we arrive at either a thick notion of general intelligence that does not admit artificial realizability or such a thin notion of general intelligence that is so diluted for it to have any classificatory, descriptive and theoretical import. In the latter case, the concept of general intelligence is watered down to prevalent yet rudimentary intelligent behaviors based on the assumption that the difference between general intelligence and mere intelligent behaviors which are prevalent in nature is simply quantitative. Therefore, if we artificially realize and put together enough of basic behaviors and abilities we essentially obtain general intelligence. In other words, the trick in realizing general intelligence is to abstract basic abilities from below and then finding a way to integrate and artificially realize them. Let us call this approach to the AGI problem, hard parochialism. Hard parochialists tend to overemphasize the prevalence of intelligent behaviors and their sufficiency for general intelligence and become heavily invested in various panpsychist, pancomputationalist and uncritical anti-anthropocentric ideologies that justify their theoretical commitments and methodologies.

However, if we define the human in terms of cognitive and practical abilities that are minimal yet *necessary* conditions for the possibility of any scenario that involves a sustained and organized self-transformation (i.e. self-determination and self-revision), value appraisal, purposeful decision and action based on an objective knowledge that has the possibility of deepening its descriptive-explanatory powers, and the capacity for deliberate interaction: negotiation, persuasion, or even threat and plotting, then the answer is functional mirroring (despite structural divergence).

But now a different question arises: Should we limit the model of AGI—both from a methodological perspective and a conceptual perspective that is the hermeneutics of general intelligence—to mirroring capacities and abilities of the human subject?

My answer to this question is an emphatic No. Functional mirroring is a soft parochialist approach to the problem of AGI and the question of general intelligence. In contrast to hard parochialism, it is necessary for grappling with the conceptual question of general intelligence as well as modeling and methodological requirements for the construction of AGI. But even though it is necessary, it is not sufficient. It has to be coupled with a critical project that can provide us with a model of experience that is not restricted to a predetermined transcendental structure and its local and contingent characteristics. In other words, it needs to be conjoined with a critique of the transcendental structure of the constituted subject (existing humans). In limiting the model of AGI to the replication of necessary conditions and capacities required for the realization of human cognitive and practical abilities, we risk to reproduce or preserve those features and characteristics of human experience that are purely local and contingent. We therefore risk falling back on a parochial picture of the human as a model of AGI that we set out to escape. As long as we leave the transcendental structure of our experience unquestioned and intact, so long as we treat it as an essence, we have an inadequate objective traction on the question of what the human is and how to model an AGI that is not circumscribed by contingent characteristics of human experience. But why is that the critique of the transcendental structure is indispensable? Because the limits of our empirical and phenomenological perspectives with regard to phenomena that we seek to study are set by transcendental structures. Put differently, the limits of the objective description of the human in the world are determined by the transcendental structure of our own experience. The limits of scientific-empirical perspective is set by the limits of the transcendental perspective.

AGI and the Critique Transcendental Structures

But what are these transcendental structures? They can be physiological (e.g., locomotor system and neurological mechanisms), linguistic (e.g., expressive resources and internal logical structure of natural languages), paradigmatic (e.g., frameworks of theory building in sciences), or historical, economic, cultural and political structures that regulate and canalize our experience. These transcendental structures need not be seen separately, but instead can be mapped as a nested hierarchy of interconnected and at times, mutually reinforcing structures that simultaneously constitute experience and regulate it. If we were to imagine a Kantian-Hegelian diagram of this nested hierarchical structure, it would be represented along the nested hierarchy of conditions and faculties necessary for the possibility of mind, `[Sensibility [Imagination [Understanding [Reason]]]]`. Transcendental structures then would be outlined as structures required for not only the realization of such necessary conditions and faculties but also moving upward from one basic condition to a more composite condition as well as moving downward from complex faculties to harness the power of more basic faculties (for example, deployments of concept in order to manipulate imagination in its Kantian sense).

In so far as any experience is perspectival and this perspectivality is ultimately rooted in transcendental structures (namely, the structure that makes it possible to have *a priori* knowledge of objects), any account of intelligence or general intelligence (whether in the context of describing the target model—in this case, the human agent – or in the context of artificial realization based on a given model—is circumscribed by the implicit constraints of the transcendental structure of our own experience. Regardless of modeling AGI on humans or not, our conceptual and empirical descriptions of what we take to be a candidate model of general intelligence is always implicitly constrained by our own particular transcendental structures. This is in fact a more insidious form of anthropocentrism to the extent that it is hidden because we take it for granted as something essential and natural in the constitution of human intelligence and our experience of it. In leaving these transcendental structures intact and unchallenged, we are inevitability susceptible to reinscribe them in our objective model. Particularly anti-anthropocentric models of general intelligence and those philosophies of posthuman intelligence that have anti-humanist commitments are far more susceptible to fall in the traps of this hidden form of essentialism. Since by treating the rational category of sapience as irrelevant or obsolete, and by dispensing with the problem of the transcendental structure—a problem that can only be conceptually and methodologically tackled by resorting to our rational-scientific knowledge and understanding—as a paltry human concern, we become oblivious to the extent our objective conceptual and empirical perspectives are predetermined by our transcendental structure. Becoming oblivious to the problem of transcendental blind spots we are at much greater risk of smuggling around essentialist anthropocentrism, replicating local and contingent characteristics of human experience in what we think is a radical non-anthropocentric model of general intelligence. Those who discard what non-trivially distinguishes the human are the ones who preserve the trivial characteristics of the human in a parochial conception of AGI.

It is of course not the case that AGI research programs should wait for a thoroughgoing critique of the transcendental structure to be done via physics, neuroscience, economy and politics, in order for them to put forward an adequate model, but that they ought to be understood as two parallel and overlapping projects. In this schema, the program of the artificial realization of the human's cognitive-practical abilities coincides with the project of fundamental alienation of the human subject which is precisely the continuation and elaboration of the Copernican enlightenment, moving from a particular perspective or a local frame to a perspective or experience that is no longer uniquely determined by a particular and contingently constituted transcendental structure.

The structural-functional analysis of necessary conditions and capacities for the realization of human cognitive-practical abilities is a necessary framework for AGI research. But the sufficiency of this framework depends on how far we deepen our investigation into the transcendental structure of human experience and how successful we are in liberating the model of human of subject from the contingent characteristics of its transcendental structure of experience. In this sense, a consequential paradigm of AGI should be seen as the convergence of two projects:

1) Examination of conditions and capacities necessary for the realization of what for now can be called the human mind as well as the more applied question of how to artificially realize these conditions and capacities.

2) A critical investigation into the transcendental structure of experience in order to develop a different model of experience that is no longer predetermined by a particular local and contingently framed transcendental structure.

Thus to answer the question of whether AGI should be modeled on humans or not and if so on what level: AGI should be modeled on human in the sense of functionally mirroring conditions and capacities that are necessary for the realization of human cognitive-practical abilities. But it should diverge from the transcendental structure of the constituted human subject. However, the success of this divergence depends on (1) our success to rationally-scientifically challenge the given facts of our own experience and in that reinventing the figure of the human – ourselves – beyond strictly local transcendental structures and their contingent characteristics (this is the project of fundamental alienation of the human), and (2) the success of AGI research programs in extending their scope beyond applied dimensions and narrow implementation problems towards theoretical problems that have for a long time vexed physics, cognitive science and philosophy.

Modeling AGI on the human agency is not merely a strategy for tackling the conceptual problems in constructing a non-parochial artificial intelligence, but also more fundamentally a strategy for unlocking the questions about the nature of intelligence and the mind, what they are, what they can become and what they can do. If we posit ourselves as a model of an artificial agency that has all the abilities that we have, then we ought to examine what exactly it means for us to be the model of that which has the possibility of being—in the broadest sense—better than us. This is the question of modeling future intelligence on something whose very limits can be perpetually renegotiated, that is, a conception of human agency not as a fixed or settled creature but as a theoretical and practical life-form which is distinguished by its ability to conceive and transform itself differently, by its unchanging strife for self-revision and self-construction. Ultimately, the question of what the mind is and what it can do is a matter of developing a project in which our process of self-discovery and self-transformation fully overlaps and in a sense, reinforces the program of realization of an agency that can surpass what we conceive as ourselves here and now. It is within the ambit of this project—the project of inquiring into the meaning and possibilities of the agency—that the human and AGI become non-tautological synonyms. The non-trivial meaning of the human is in its ability to revise and transform itself, its ability to explore what the human is and what it can become. The non-parochial conception of AGI is simply the continuation and realization of this meaning in its substantive form.

The critique of transcendental structures is strictly a collective project comprised of procedural methods and incremental tasks. On a groundwork level, it begins theoretically by distinguishing necessary conditions for the constitution of the agency from contingent aspects of the subject's constitution, characteristics of reality from characteristics of the subject's experience. At this stage, the critique tackles two fundamental overlapping questions: to what extent objective descriptions of reality at various levels are biased or distorted by the characteristics of our experience, and to what extent the necessary conditions for the realization of our theoretical and practical abilities as well as our exercise of such abilities are caught up in or determined by the contingent positioning of our particular transcendental structures (be them associated with the terrestrial habitat, neurophysical systems, cultural environment, family, gender, economy, etc.)? On the basis of this theoretical phase, then the project proceeds to inquire into the possibilities of transforming and diversifying the transcendental structures of the agency. This is a experimental phase in which the possibilities of transcendental variation, and thus the possibilities of releasing experience—and by extension theoretical and practical abilities that it makes possible—from limitative attachments to any unique or allegedly essential local transcendental structure are examined. The central task of this stage is to expand the range and type of abilities by changing and reorganizing transcendental structures. Once the prospects of varying transcendental structures and transformation of abilities are systematically outlined and evaluated, the project shifts toward applied dimensions of developing implementable mechanisms and systems that can support the realization of new abilities by either enhancing or replacing transcendental structures of the constituted subject.

What begins as a systematic theoretical inquiry into limits and regulative regimes of transcendental structures of the constituted subject evolves into an applied system for the transformation of the subject and maximization of the agency's theoretical and practical abilities. Thus understood as a project in which the critique comprehensively revises what is the posited as its subject of inquiry, the critique of the transcendental structures is the compass of self-conception and self-transformation. By challenging the established characteristics of the experience of ourselves in the world and by renegotiating the limits posited by our contingent constitution, we transform ourselves by exploring what we actually are in the world and what we can be. Future intelligence is nothing but an instantiation of this cognitive exploration on which the true significance of the human hinges.

It is in the context of the critique of transcendental structures that AGI becomes the extension of the human whose meaning determined by the abilities for self-conception and self-transformation, by what human does not by an essence given in advance. This is a meaning that can be conferred upon, transferred and extended, for it is neither fixed by biology nor bestowed by the divine, neither limited to what we are here and now nor is exhausted by what may carry the title of the human in the future. AGI, non-parochially conceived, does not step outside of this meaning of human. Rather it marks the maturity of the human who has finally recognized its meaning not in virtue of its contingent constitution but in spite of it. This is a conception of the human as an all-encompassing collective project of self-conception and self-transformation whose veritable consequence is reinventing the human by modifying its structure and expanding its abilities.

To sum up: The functional map of human cognitive-practical abilities is a proper theoretical model for the construction of AGI. But without incorporating the problematics of transcendental structure of human experience, this model as I mentioned risks conflating the contingent characteristics of human experience with necessary conditions for the realization of human abilities, and thus relapsing into a hard parochialist approach to the questions of what general intelligence is and how to artificially realize it. If we treat the human – non-trivially defined—as a model of AGI, then this model should not only be a model by which we can identify and differentiate necessary conditions and capacities for the realization of theoretical and practical cognitions, but also be a model within which we can renegotiate the characteristics of general intelligence by renegotiating the limits and characteristics of the human experience. This is brings us to the title of this presentation “the outside view of ourselves as a Toy Model AGI’ i.e. treating ourselves—both our functional capacities and what we experience ourselves as … -- from an objective point of view—that is to say, a point of view that while is able to distinguish the necessary conditions and capacities for the realization of the subject’s ability is not ‘in principle’ bound to local characteristics of the subject’s experience. So in reality this is a model that is capable of making explicit, its implicit meta-theoretical assumptions. And what are these implicit meta-theoretical assumptions? They are precisely the implicit or hidden assumptions that arise from applying the characteristics of our subjective experience to our objective descriptions of abilities or functions and structures responsible for realizing them. And this is the reason that I call it a toy model.

AGI Toy Model

Toy models are simplified models that are conceptually sufficient to accommodate a wide range of theoretical assumptions for the purpose of organizing and constructing overarching narratives (or explicit metatheories) that change the standard and implicit metatheoretical interpretations according to which such theoretical items are generally represented. In other words, by explicitly changing the metatheoretical narrative, toy models provide new interpretations of problems and puzzles associated with the implicit metatheoretical frameworks within which theoretical ideas and observations are interpreted. To this end, a rigorous and internally consistent toy model can offer insights about how to solve these puzzles or how to overcome the setbacks caused by the standard interpretations. What separates toy models from models is not just that they are simplified enough to enable us tinkering with the internal theoretical structure of a model, but that they are explicit metatheories. All theories are metatheories, but within regular theoretical models meta-theoretical assumption are usually implicit or hidden. Whereas toy models are explicitly meta-theoretical and in fact the simplification (what gives them the name toy) servers as a strategy for bringing meta-theoretical assumptions out in the open by tinkering with the internal variables of the model without being bugged down by huge amount of theoretical details.

Now toy models come in small and big varieties. The small toy model is a simplified version of only one theoretical model (essentially it is a model in a collapsed form), whereas big toy models are the ones that accommodate different (often seemingly incompatible) theories, such as general relativity and quantum mechanics. Put differently, big toy models represent a form of model pluralism and for that they are required to have a conceptual architecture plastic enough to accommodate and faithfully represent main features of different models or theoretical frameworks, while at the same time be capable of preserving the distinct features of these models within different categories (more or less, in the category theoretical sense of mapping objects and their class of equivalence relationships). For example, in order to for us to be able to adequately think about the kind of problems that we dealing when talking about the construction of AGI, we first need a big toy model. An AGI big toy model should be able to coherently accommodate different models derived from physics, evolutionary biology, neuroscience, developmental psychology, multiagent system design, linguistics and computer science. One of the problems with old AI research was that it was strongly driven by unique and inflationary models of mind that generated more setbacks than progress. For example, consider the symbolic program of AI (the syntactic picture of the mind), deep learning (neural networks and statistical inference), and computational semantics (computational-logical modeling of meaning representation in natural languages), which were developed based on insights derived from the evolutionary sciences, computer science, neuroscience, logic, and linguistics.

While these programs in their own right have lead to undeniable achievements and progress in the field of artificial intelligence, they have also created theoretical bottlenecks and practical setbacks. This is because their implicit metatheoretical assumptions have been either left uncontested owing to the sheer success of these ideas and methods in a narrow domain of application, or unduly overstretched into global assumptions about the nature of cognition and the mind. The result is that the statistical framework of something like machine learning becomes the global model of general intelligence, or the characteristics of sequential algorithms (effective mechanizability, symbol-manipulation, deductive inference) establish a syntactic model of mind within which the program of artificial intelligence as a whole is oriented. Once the metatheoretical assumptions of these locally successful ideas and methods are inflated into global models of general intelligence or mind, it is only a matter of time before the model arrives at theoretical and practical impasses, and development comes to a halt. The summer of AI turns out to be a winter all along.

Toy models, on the other hand as I mentioned, are not only explicit metatheories in themselves but also make explicit the implicit metatheoretical frameworks of their constituent ideas, observations and methods. In doing so, toy models are able to keep these implicit metatheoretical assumptions underlying their components in check, and therefore avoid the risks of inflationary models. The utility of toy models lies not only in the idea that they allow for some theoretical arbitrage by combining and spanning across different metatheories, but also and more importantly, in their ability to facilitate the reinterpretation, reassessment, and reapplication of conventionally interpreted ideas and observations.

But the real value of a toy universe is that one learns from it by breaking it in the real universe; but not until one has systematically played with it. And it is exactly in this sense that a toy model AGI is an *explicit metatheory* of artificial general intelligence constructed from falsifiable concepts and models drawn from different theoretical frameworks. If we take ourselves as functional toy model of AGI, then we are dealing with two main categories of metatheory, one meta-theories associated with the bulk of models we are using to map the necessary conditions and capacities required for the realization of our cognitive-practical abilities along distinct descriptive-explanatory levels and the other metatheoretical assumptions related to conditions of observation and description within which these conditions and capacities are mapped. It is the latter category that needs to be subjected to a critique of transcendental structures of experience in order for the first category to be adequately objective. Absent a systematic attempt to render explicit our subjective experiential assumptions, we cannot sufficiently differentiate the conditions necessary for the possibility of mind (in all its semantic complexity) from the contingent characteristics of experience and our intuitive subjective biases which are by-products of the local and contingently situated transcendental perspective. (See Appendix 2, AGI toy model formalization via Chu Spaces).

An outside view of ourselves as a toy model AGI that allows us to conceptually come to grips with the problematics of both of these two meta-theoretical categories is exactly what the philosophy of German idealism encapsulates. As you know, the focus of german idealism is the intersection between the philosophy of action, philosophy of mind and philosophy of knowledge, namely, the problems concerning with the condition of possibility of theoretical and practical cognitions, what they are and how they can be realized. Particularly, Kant and Hegel and their legacy via Peirce, McTaggart, Grünbaum, Sellars, Rosenberg and Brandom can be seen in the context of how we should lay out the conceptual problems of the mind and experience. Talking about AGI in the context of Kant and Hegel might appear as a retrogressive move, but I think this is absolutely not the case, because when it comes to the philosophy of mind Kant and Hegel pose the right kind of problems. They provide an outline of fundamental conceptual problems that are in fact still at the center of debates in cognitive and theoretical computer sciences. Needless to say, if our aim is to understand the relevance of these problems for artificial intelligence, we have to reframe them in terms of concepts that are much more in tune with contemporary science. And of course, throughout this process of synchronization sustaining certain aspects of Kant and Hegel's programs become untenable.

That said, I'm going to give a very brief overview of this Kantian-Hegelian toy model of general intelligence. Starting with Kant's transdental psychology and moving toward Hegel's formulation of language as the dasein of geist. I will try to make brief comments as how these philosophical problems can be laid out in terms of the lexicon of contemporary sciences specifically cognitive science and theoretical computer science. Then I will – hopefully if have enough time – return to the problematics of transcendental structures of experience. I will make a case-specific example via physics to show how our transcendental perspective can fundamentally distort our objective description of phenomena. The example I will provide is Boltzmann's work on the problem of the second law of thermodynamics and his attempt to reconcile time-asymmetric experience (in Kant's terms transcendental ideality of experienced temporality) with objective statistical description of microscopic phenomena. The reason I'm choosing this example is because it addresses a lot of issues with regard to the statistical account of computation, scientific explanation, causal theories, epistemic models, subjective experience and objective description of behaviors that are supposed to be devoid of any direct intervention of characteristics of our experience. This will be more of a speculative foray into the deep conceptual problems that surround the descriptive-explanatory analysis of mind, experience and computation.

So let's start with Kant, as you know, transcendental psychology, which concerns with the identification and analysis of the necessary conditions for the possibility of mind was not a primary part of Kant's project. Kant began his project with transcendental arguments in order to inquire into the possibility of objective knowledge. Transcendental psychology was more like a happy accident that resulted from this inquiry. How is that we can have objective knowledge turns out to be a project about what is required for the realization of theoretical and practical cognitions. There is an interesting story here about how Kantian this functional analysis of truths qua problems in terms of tasks required for bringing about the possibility of knowledge of those truths is in fact the source of Brouwer-Heyting interpretation of proofs and problems in terms programs and constructions which is one of the central ideas in logics and computer science. Brouwer arrived at this interpretation directly via reading Kant and Heyting through Brouwer and Oskar Becker's interpretation of Kant by Husserl and Heidegger.

Under this interpretation of problems in terms of tasks or in more contemporary terms constructive programs, objective knowledge in terms of transcendental psychology, Kant begins to characterize the mind along two axes, what arises from the exercise of mental powers or abilities of the mind and what is required for the realization of these abilities or powers. Adopting a slightly functionalist vocabulary, we can call these two axes realizabilities and realizers, corresponding to what arises from the exercise of realized abilities and conditions or capacities that are necessary for the realization of these abilities. Kant's threefold synthesis is essentially an abstraction of realizers-capacities from above, from the vantage point of realizabilities. It is a form of functional analysis that attempts to describe how capacities are instantiated at various levels, and how they can – via the unity of apperception – be bootstrapped into more elaborate and more complex faculties.

All of these three synthesis involve with compression, ordering and combination of data to generate ever more complex invariances (from untyped rudimentary perceptual invariances to typed schematic-geometric invariances to trans-typified conceptual invariances). Just to make a very brief reminder: what are these three syntheses? These are synthesis of apprehension in the intuition, synthesis of reproduction in imagination, and synthesis of recognition in the concept. The synthesis of apprehension delineates the first constructive role of imagination in pulling together a synchronic manifold of sensations by antecedently taking up the sense impressions into its activity which is apprehension. It introduces an order to the confusion of simultaneous impressions, by giving them temporal and spatial locations, and thus differentiating them. In doing so, the synthesis of apprehension brings about the condition of the intelligibility of impressions as distinct (spatiotemporally structured) impressions qua appearances available for further construction and structuring.

The second synthesis, the synthesis of reproduction, signifies the second constructive role of imagination in combining and reproducing the sensory manifold diachronically, carrying over its earlier elements—with the help of constructive memory—in order to construct a stable image qua a singular representation of an item in the world. It establishes temporal associations between appearances that the synthesis of apprehension has located in space and time in a certain way, rudimentarily structured out of the undifferentiated homogeneity of simultaneous impressions. These two syntheses are the figurative part of the building process associated with imagination as a constructive-simulating capacity whose function is to 'represent an object even without its presence in intuition' and which is unavailable to pure sensibility (i.e. seeing1 of). The third synthesis, the synthesis of recognition, strictly designates the role of apperceptive consciousness in perception as what must be 'added to pure imagination in order to make its function intellectual', since 'in itself the synthesis of imagination, although exercised a priori, is nevertheless always sensible, for it combines the manifold only as it appears in intuition.' The synthesis of recognition requires both the act of recognizing a past representation as related to the present one and the act of recognizing past and present representations as belonging to one object via the function of a concept qua rule.

We can think of the threefold synthesis in terms of making a Lego model—say, a toy robot—using Lego blocks of different shapes and colors. The blocks in their various shapes and colors correspond to the diverse images (the intuiteds) of our Lego model. The shapes and colors of the blocks are the raw 'matter' of the intuited items qua images. The pictorial motif of our Lego model-building corresponds to the conceptual representation of these intuited items in acts of judgments (the intuitings). The function of the pictorial motif is to determine the colors and shapes of the blocks in such a way that it becomes possible to put them together so as to construct the specific Lego model in question. In other words, the pictorial motif encapsulates the function of the concept of a robot that determines images (the right blocks) as different aspects of only that object. It is only because the colors and shapes of these blocks hang together in the right way—perspectivally in space and time, synchronically and diachronically—that we are able to synthesize the pictorial motif of our Lego model-building. And respectively, it is only because the blocks (the images) can be put together in the right way—in accordance with a rule i.e. the concept of a robot—that we are able to conceive them as associable and multiple aspects of one such-and-such robot.

In short, for the time being, we must resign ourselves to this analogical application of the resources of our natural language to the navigational or interaction scheme as described in terms of a structure sufficient for *causally* mediating between *de facto* environmental inputs and *de facto* behavioral outputs. The first point is that the resources of language are ultimately the only resources available to 'we temporally discursive apperceptive intelligences' for representing the intelligible order (since for us, non-conceptual sensory impressions are caught up in the inferential web of language). The second and more important point is that the causal interaction of the rudimentary agent with the environment, as a protosemantic navigation of the world, exhibits a non-categorial orderliness that our concepts in their inferential relationships reflect and illustrate, disambiguate, and make explicit. In other words, the causal-heuristic interaction of the automaton instantiates precisely the structures from which our semantic structures have 'in part' evolved. For these reasons, if handled correctly, this analogical circle is a virtuous rather than a vicious circle. It enables us, through a series of controlled analogies to construct, step by step, from an intelligence that is neither discursive nor apperceptive, an intelligence that is no longer analogically posited because it has the form of a full-blooded discursive apperceptive intelligence. What it takes to analogically posit a non-conceptual awareness is exactly what it takes to elaborate this non-conceptual awareness into a conceptual awareness; and what must be added to the analogically posited awareness in order for it to be no longer analogical—to move from non-conceptual awareness to conceptual awareness—is exactly the same as what is needed to develop the non-apperceptive intelligence into an apperceptive intelligence, consciousness into self-consciousness.

Transcendental ideality of experienced temporality, causality and statistical computation

Now in order to clarify the idea of critique of transcendental structure of experience of why it is indispensable for thinking the problems of philosophy of mind, philosophy of knowledge and philosophy of action, the problems of general intelligence, I would like to make this final digression into Boltzmann's work particularly due its significance for the interpretation of computational description, causality and the epistemological project. As it will become clear there is also a link to the problem of construction of artificial general intelligence albeit from a speculative angle.

Also this can be counted as a digressive critique of inductive computation within the field of AI. Probabilistic-statistical computation—as distinguished from (proto)logical computation—does not have any causal explanatory power. One of the reasons that proponents of non-agentic (agency in the Kantian sense) AGI insist on the sufficiency of inductive computation for the realization of general intelligence is because they think probabilistic computation has strong causal powers. But this view is not exclusive to the proponents of non-agentic AGI, you can see a similar interpretation in evolutionary biology and cognitive science, where probabilistic computation enjoys a strong causal-explanatory role. The idea that probabilistic computation can have causal-explanatory role is more than being a theory about computation. It is strictly a thesis built on interpretations afforded by macrostate physics (particularly the classical interpretation of time-asymmetric phenomena like arrow of causation, or numerical asymmetries). As you know the inductive account of computation has its roots in information theory (via people like Kolmogorov and Solomonoff) which is itself developed out of statistical physics especially the statistical interpretation of thermodynamics and particularly the second law (via Gibbs and Boltzmann).

So the argument about the unreality of subjective time (the transcendental ideality of experienced temporality) can be posed in two forms, one what I call a modest version and the other is a sinister version.

The modest version of this argument is that we can neither draw conclusions about the objective reality of time, its direction, its flow and its temporal structure nor in fact the existence of such objective properties from the structure and characteristics of experienced temporality. Rather than arguing that there is no objectively real time, the modest claim here is about the illegitimate nature of the inference from the perception of time qua experienced temporality to the objective reality of time on the basis of some assumed isomorphy, private door, global sense of direction and passage of time, structure of tensed language or observed arrow of causality. Said in a different way, we cannot infer an objective time from temporal-dynamic characteristics appearing in experience or observation. But this does not tell us whether there is an objective time or not, and if there is, what its characteristics are. In other words, this is different from a naively hyper-Kantian position—born out of a conflation between commitments to epistemological idealism and commitments to ontological idealism—for which it is impossible to think an objective reality for time or that we attribute every structure and in this case, every objective conception of time, to our own subjective point of view. While we can make veridical judgments via language about time, we are not permitted to treat characteristics of temporal components of language as bits and pieces of evidence for the objective temporality of time, nor are we allowed to treat the temporal components—i.e. tensed verbs and temporal connectives such as before, after, when and until—of our statements about objective time as prima facie theoretical or matter-of-factual components. This of course opens up a more fundamental question: Is the tensed structure of natural languages even appropriate for investigating the question of objective time, or should we completely shift to formal-theoretical languages—where we can have temporal connectives with more neutral and flexible logical connections—when it comes to thinking an objective account of time?

The less modest and more disquieting version of the unreality of experienced temporality is that there is a good chance that any asymmetric picture of time where we can have sequences running from one extremity toward another (past to future or future to the past) in a punctual or durational form and the present can be regarded as something objectively distinguishable is riddled with experiential biases, specifically those related to the contingent structural organization that makes the experienced temporality possible. These biases are not exclusive to subjective perception of time and how it is reflected in the tensed structure of natural languages but also more fundamentally can be even extended to the ideal notion of observer in physics. The threatening aspect of this argument is that if the directional, flow-like pictures of time are negatively biased by the structure of experienced temporality and local characteristics of the subjective perspective / the observer, and if the classical notions of causality, system's state and antecedent conditions are embroiled in directional-flow-like pictures of time, then those portions of complexity and physical sciences which have incorporated these concepts (i.e. causality, state and antecedent conditions defined via time-asymmetric concepts and principles) as their fundamental explanatory-descriptive elements are also biased and prone to significant revisions if not abandonment.

What is meant by causality here is not an intuitive idea of causality or what Wolfgang Stegmüller astutely identifies as the pre-scientific concept of causality, namely, singular causal judgments (or individual cause-effect connections) reflected in sentences of ordinary language containing terms such as 'because' and 'since' (Seneca cut his veins since Nero ordered him to kill himself, the car crashed into a tree because the breaks stopped working, etc.). The pre-scientific cause-effect connection is built upon the arbitrary selection—in Stegmüller's words more psychological than epistemological—of one diagnosed or indicant condition among a large number of conditions that may not seem to play any explicit causal role. Instead the concept of causality that is at stake here refers to a set of law-like regularities in addition to a set of antecedent conditions that together constitute the explanans of a causal explanation. In this schema of causal explanation, statements belonging to explanans must have empirical content as well as at least one law-statement to count as causally explanatory.

The challenge to the directional, flow-like picture of time can potentially problematize the classic conditions (i.e. those which operate on the basis of various dynamic time-oriented phenomena from numerical to temporal asymmetries, global models of local entropy curve or gradient, etc.) through which law-like regularities are derived and antecedent conditions are characterized. As for the notion of the state of a physical system that is in question here, it describes dispositions of the system for responding to a range of possible circumstances that might be encountered in the future. This notion of the state is, properly speaking, a descriptive tool for predicting the future responses or trajectories of the system (in terms of counterfactuals) from its present behavior or state. Hidden variables on the other hand refer to those states which are taken to be independent of those future interactions that the system might be subjected to. These concepts are based on perspectival temporal and modal asymmetries of the subject or the local ideal observer that ground the distinction between sequences running from past to future and future to past, or more generally, the orientation of sequences with regard to the passage of time. These concepts of causality (law-like regularities together with antecedent conditions), states and hidden variables play a fundamental role in complexity sciences particularly those branches which strongly make use of heuristic methods for describing the behavior of the system, characterizing its structural and functional features and predicting its evolution. Contemporary cognitive science and by extension programs of artificial intelligence as we know are strongly invested in these concepts that borrowed directly from complexity sciences. For example, strong causal powers are attributed to statistical computation, states of the system are modeled via initial and boundary conditions without any a fine-grained analysis as what counts as initial and boundary condition for a state of a system at a specific level of description, etc.

Any significant revision of the canonical model of directional-flow-like time (for example, a block view of time) can potentially harbor devastating outcomes for these frameworks in complexity sciences that span from physics, to chemistry, biology, neuroscience, economy and social sciences.

Drawing attention to the observational and subjective biases within the directional-dynamic picture of time and temporal asymmetries, and pointing to the negative connotations of such biases for the concepts of description, causal explanation, modeling and prediction is not by any means a recent line of inquiry. In what Huw Price calls "a Copernican moment" and Hans Reichenbach distinguishes as "one of the keenest insights into the problem of time", Ludwig Boltzmann summarizes the problem in the following remarks that are worth quoting in their entirety:

> Just as the differential equations represent simply a mathematical method for calculation, whose clear meaning can only be understood by the use of models which employ a large finite number of elements, so likewise general thermodynamics (without prejudice to its unshakable importance) also requires the cultivation of mechanical models representing it, in order to deepen our knowledge of nature–not in spite of, but rather precisely because these models do not always cover the same ground as general thermodynamics, but instead offer a glimpse of a new viewpoint. Thus general thermodynamics holds fast to the invariable irreversibility of all natural processes. It assumes a function (the entropy) whose value can only change in one direction—for example, can only increase—through any occurrence in nature. Thus it distinguishes any later state of the world from any earlier state by its larger value of the entropy. The difference of the entropy from its maximum value—which is the goal [Treibende] of all natural processes—will always decrease. In spite of the invariance of the total energy, its transformability will therefore become ever smaller, natural events will become ever more dull and uninteresting, and any return to a previous value of the entropy is excluded.
>
> One cannot assert that this consequence contradicts our experience, for indeed it seems to be a plausible extrapolation of our present knowledge of the world. Yet, with all due recognition to the caution which must be observed in going beyond the direct consequences of experience, it must be granted that these consequences are hardly satisfactory, and the discovery of a satisfactory way of avoiding them would be very desirable, whether one may imagine time as infinite or as a closed cycle. In any case, we would rather consider the unique directionality of time given to us by experience as a mere illusion arising from our specially restricted viewpoint.

Boltzmann then continues,

> For the universe, the two directions of time are indistinguishable, just as in space there is no up or down. However, just as at a particular place on the earth's surface we call

> ‘down’ the direction toward the center of the earth, so will a living being in a particular time interval of such a single world distinguish the direction of time toward the less probable state from the opposite direction (the former toward the past, the latter toward the future). By virtue of this terminology, such small isolated regions of the universe will always find themselves ‘initially’ in an improbable state. This method seems to me to be the only way in which one can understand the second law—the heat death of each single world—without a unidirectional change of the entire universe from a definite initial state to a final state.
>
> Obviously no one would consider such speculations as important discoveries or even—as did the ancient philosophers—as the highest purpose of science. However it is doubtful that one should despise them as completely idle. Who knows whether they may not broaden the horizon of our circle of ideas, and by stimulating thought, advance the understanding of the facts of experience?

Hans Reichenbach sums up this rather esoteric argument of Boltzmann in the following way. He says:

> Philosophers had attempted to derive the properties of time from reason, but none of their conceptions compares with this result that a physicist derived from reasoning about the implications of mathematical physics. As in so many other points, the superiority of a philosophy based on the results of science has become manifest. There is no logical necessity for the existence of a unique direction of total time; whether there is only one time direction, or whether time directions alternate, depends on the shape of the entropy curve plotted by the universe.
>
> Boltzmann has made it very clear that the alternation of time directions represents no absurdity. He refers our time direction to that section of the entropy curve on which we are living. If it should happen that 'later' the universe, after reaching a high-entropy state and staying in it for a long time, enters into a long downgrade of the entropy curve, then, for this section, time would have the opposite direction: human beings that might live during this section would regard as positive time the transition to higher entropy, and thus their time would flow in a direction opposite to ours. Since these two sections of opposite time directions would be separated by aeons of high-entropy states, in which living organisms cannot exist, it would be forever remain unknown to the inhabitants of the second time direction was different from ours.”

What vexes Boltzmann are not the puzzles of directional-dynamic picture of time but rather the unproblematic and innocent nature of the assumption that time has in fact an objective temporal direction. For him, the real conundrum is not why entropy increases with time, but why it was ever so low in the beginning. Formulated differently, rather than asking why does entropy increase toward the future, we should ask why does entropy decrease toward the past. The source of Boltzmann’s problem was precisely in what he had initially given as a key to solving the problem of the second law of thermodynamics: Where does the time-asymmetric characteristic of the second law—one that states entropy goes up over time—come from?

Toward the end of the nineteenth century, figures such as Boltzmann and Gibbs had begun to develop a fully statistical (proto-computational / information theoretic) account of thermodynamics. For Boltzmann, however, this was part of a broader project, one whose aim was to provide ‘complete descriptions’ of physical phenomena. The stepping-stone of this descriptive project was Boltzmann’s reformulation of the concept of scientific description at once removed from the dominant influence of earlier phenomenalist and psychologic accounts of description (such as Mach’s) and sufficiently fine-grained to be capable of integrating statistical, epistemological, phenomenological, real-ideal, subjective-objective levels and types of description in a non-arbitrary (i.e. intrinsic) manner. To this end, Boltzmann provided three general levels or types of description: the pure or abstract description based on inferential generalization of differential equations rather than a correspondence to observed facts, the indirect description based on a probabilistic framework of the statistical description and a level of description concerning unobservables. As Adam Berg argues in *Phenomenalism, Phenomenology and the Question of Time*, Boltzmann’s reframing of thermodynamics (specifically the second law) through statistical physics should be seen within the scope of this descriptive analysis as a multi-level complex system of coding with distinct descriptive levels that require different appropriate systems of coding, noetic contents and method of analysis as well as appropriate spaces for bridging these levels.

Within the scope of this multi-level descriptive analysis that became the skeletal framework of modern scientific theories, Boltzmann developed his statistical theory of nonequilibrial (i.e. irreversible and time-asymmetric) behavior of macroscopic systems. He associated to each macrostate and each microstate giving rise to that macrostate an entropy. In this framework, entropy could be seen as a tendency to evolve toward more probable macrostates, and the increase of it as information regarding the qualitative behavior of macroscopic systems. From the perspective of the new descriptive analysis, the problems of the second law (i.e. why does entropy increase over time?) and the emergence of irreversible and time-asymmetric behaviors despite the time-symmetry of the underlying mechanical-physical laws could thus be reframed as the problem of moving from microscopic descriptions to macroscopic descriptions. The solution, to these problems could then be formulated by devising a statistical mechanical framework that accommodates a conception of macrostate (pertaining to the macroscopic level) expressed in terms of probability and intrinsically correlated to the microstate (associated with the microscopic level) responsible for it. Within this statistical mechanical resolution, entropy, then, could be defined as a tendency toward more probable macrostates.

We will not able to delve into details of how Boltzmann constructed his solution, but very briefly it involved a procedure that would make explicit the connections between statistical and thermodynamic descriptions through the introduction of the concept of macrostate put forward in terms of its probability and an appropriate space (the so-called μ-space which is essentially smoothing procedure for data concern single-point particles) for bridging microstates and macrostates, microscopic descriptions and macroscopic descriptions. Following Maxwell, Boltzmann began to examine the effects of collisions on the distribution of velocities of molecules of a gas. He introduced a space divided into an array of small cells or intervals of equal size in position and momentum. Once available velocities are partitioned into these cells, then there is an effective combinatorial-computational procedure for examining the effects of collisions on the number of molecules whose velocities entered these cells. Using this combinatorial procedure, Boltzmann was able to argue that (1) the distribution of velocities approaches Maxwell probability distribution in which the quantity E or H which is equivalent to minus the entropy can be said to be decreasing, and (2) this distribution is independent of the initial distribution of velocities. No matter how particles were initially assigned to the available velocity cell-partitions, we still get the same probability distribution that accounts for the monotonic decrease of H. Demonstrating the decrease of the quantity H was the proof of the unidirectional and irreversible increase of entropy and time-asymmetric behaviors at the macroscopic level in spite of the reversibility and time-symmetry of the underlying microscopic mechanics.

However, as reflected in the quotes cited earlier, in his later works Boltzmann started to express doubts about his solution and began to examine the challenges raised by adopting a resolutely atemporal perspective. Given the fact that the statistical argument itself is merely a combinatorial-counting procedure and lacks any time-asymmetry, there is no reason to apply the increase of entropy to a unique sequence that runs from the past toward the future, it can equally be applied to a sequence running from the future to the past. Then the genuine question, as mentioned earlier, is that what deserves explanation is not the increase of entropy toward the future (i.e. what appears to be a natural state of things) but the ever so low entropy in the beginning insofar as the statistical argument gives us reason to also expect the increase of entropy toward the past. In light of the statistical argument (i.e. equal probability of increase in entropy in either direction, past-to-future and future-to-past), the global decrease of entropy toward the past now appears as an unnatural condition and for that matter, itself demands an explanation. In other words, for Boltzmann, changing the perspective from temporal to atemporal had turned something natural (low entropy in the past, a so-called fact of experience) to something unnatural (high entropy in the past, something outside of experience) and therefore, in line with motivations of scientific explanation which demand us to account for 'unnatural conditions' called for a shift in the explanatory focus. In short, Boltzmann's new commitment to an atemporal perspective put the explanatory burden on the initial low entropy rather than the subsequent high entropy, and in so doing, it had raised new challenges that were hitherto hidden from the viewpoint of what previously appeared to be a set of natural assumptions and objective principles. These are challenges that have vast implications for not only our models of processes and methods of metricization of events but also what we take to be our established facts of experience, and to this date have mostly gone unheeded.

Even though Botzmann shifted his efforts to reinterpret thermodynamics from an atemporal perspective, the deep problematic aspects of his initial solution to the problem of the second law were carried over to his new interpretation that was given in the context of the 'cosmological hypothesis' and which was supposed to be free from any particular temporal bias. But what is this problematic aspect that despite being spotted—at least partially—by Boltzmann still resurfaced in his later interpretation? The problem with Boltzmann's initial solution was that he had unintentionally imported subjective characteristics into his combinatorial-computational procedure via the introduction of macrostates. In other words, the phenomenal assumptions regarding the facticity of observed time-asymmetry for the ensemble's macrostate were illicitly applied to the description of microstates. In this sense, Boltzmann had not really bridged the gap between statistical mechanical entropy and thermodynamic entropy belonging to different levels of description, but only had elided the distinction between the two descriptive levels by illegitimately transporting the underlying assumptions of one into another. Boltzmann indeed noticed this problem, but what he did not recognize was the extent of how far the time-asymmetric assumptions specific to the macroscopic description (the transcendental perspective) had distorted the statistical qua objective description associated with microscopic systems. In other words, Boltzmann did not fully realize that biases of the unidirectional time had already infiltrated the law-like principles through which the parameters of the microscopic systems such as initial and boundary conditions were being defined and chosen.

The whole idea that the velocities of two particles which have not collided yet can be said to be uncorrelated and therefore, can be identified as an initial condition (the so-called principle of stoßzahlansatz) is already presupposing a privileged temporal-causal asymmetry. Why? Because insofar as the microscopic mechanics is time-symmetric and initial microstates are equiprobable, then there is no reason to expect that the velocities of particles to become correlated 'as a result of' their collisions. In other words, we expect outgoing products of collisions to be correlated with one another in various ways, even if they never encounter one another in the future. We do not expect the incoming components of a collision to be correlated, if they have never encountered one another in the past. This is a time-asymmetric assumption that has no place in the statistical / information theoretic description of the system.

Statistical computation can only enjoy a causal-explanatory power if it is coupled with time-asymmetrically characterized antecedent or initial conditions. Once these time-asymmetrically characterized initial and boundary conditions are in place, the statistical distribution is 'observed' as a causal component. But as Boltzmann himself had realized the observed casual arrow for any statistical behavior is the result of forcing our time-asymmetric (temporal) perspective upon statistical description of initial conditions that by principle should have no privileged temporal-causal directionality. In Boltzmann's case, the renormalization of entropy – from microscopic-statistical to macroscopic-thermodynamic -- corroborates the observed time-asymmetry and irreversibility of thermodynamic entropy, i.e forces a causal arrow on statistical-computational description throw the kind of illicit move that I mentioned. But all the formal renormalization of statistical entropy—which is part of mu-space the smoothing procedure—does is that it adds an infinite positive contribution to the infinite negative entropy corresponding to a point in such a way that the finite result $S\mu[\rho\mu]$ (phase density of statistical entropy) can be said to be physically meaningful or real. Nothing more, nothing less.

If no direction of time is initially privileged and since the statistical argument by itself has no time-asymmetric component, then once we adopt an atemporal perspective there is no reason for us to presume that the time-asymmetric explanatory schema of an initial microstate explaining a final macrostate to be tenable. Accordingly, Boltzmann's true challenge—not fully appreciated even by himself—now boils down to a much more fundamental question: How can we suggest that an initial microstate can explain a final macrostate, if what is really in need of explanation is the temporal asymmetry that grounds such an explanatory schema? It does not take too much critical acuity to realize that a similar question can be posed with regard to those frameworks of causal explanation—specifically utilized in the context of reductionism—that rely on identification of some antecedent conditions and a temporally directed causal arrow through which—thanks to the convenient mediation of time-asymmetry—the distinction between 'the causal' and 'the explanatory' effectively fades away. Nevertheless, the real significance of Boltzmann's challenge is only revealed in full force when it is treated as a general epistemological critique: How can we justify a chain of inference that follows an explanatory arrow whose mere ground of justification—its explanans—is the past state of affairs as an observable item or an empirical footprint? This question can of course be equally applied to a chain of epistemological inference that runs from the future to the past. In both cases, what needs to be justified is exactly what is taken to be the ground of justification. Thus epistemological neutrality appears to be in sharp conflict with temporally charged modes of epistemological inference. Hoping to retain some aspects of the former while drawing conclusions from the latter in a practical tradeoff is more of a wishful thinking than a pragmatic paradigm of scientific knowledge.

The illegitimate imposition of time-asymmetric descriptions exclusive to macrostate on to descriptions of microstate in order to explain the behaviors of the former by the mechanics of the latter, therefore, bespeaks of a much broader range of complications arising from our epistemological biases (originating from our particular transcendental structure of experience) in coordinating our observational frameworks with our theoretical-inferential frameworks. Accustomed to the cozy naturalness of our experience and under the theoretical influence of its biases, we are prone to frequently project our subjective assumptions onto the world and through that posit what itself requires explanation (qua a subjective characteristic) as an objective explanatory feature.

Objections can be made that time-asymmetry and temporal descriptions are only useful fictional instruments (at best subjective and at worst, speculatively metaphysical) that allow us to talk about non-temporal events and processes. But in dismissing temporal descriptions and time-asymmetry as useful idealization or metaphysical fiction, these objections reinforce our obliviousness to the influence that our temporal intuitions exert upon models and methods that are assumed to be unaffected by any objective or subjective account of time-asymmetry. In doing so, rather than giving reason for making a radical scission from temporal intuitions, they give more reason for further postponement of the overdue critical task which is the examination of the extent of distorting effects temporal intuitions have had and continue to have on scientific models and methods.

If we were to reformulate the Boltzmann's challenge with regard to the question of time in a very traditional philosophical frame, it would be that the problem is not really the enigmas of the Heraclitian flux—like the quandaries of becoming, recurrence or the puzzles of absolute contingency (every law is susceptible to change within time)—but the questions of why there is often an element of time-asymmetry—whether disguised in the shape of punctual sequential series or flow—in our philosophical reflections about events and processes that make up the pictures of the world and ourselves? And to what extent these time-asymmetric elements have overstretched boundaries of the experiential image of time as matter-of-factual characteristics of reality? Why the conditions of our time-conscious are laid out in terms of specious nows and why specious nows are explained in terms of an objective durational continuity made of a successive saturation of retentions and protensions? But this durational continuity by which the objectivity of the present is corroborated is precisely an experiential explanandum that ought to be explained without recourse to pieces of evidence gathered solely from the transcendental structure of experience itself. It is a matter of fact that the Now shifts in conscious awareness to the extent that there is a diversity of the Now-contents, and it is likewise a fact that the Now-contents are temporally ordered. But since these diverse Now-contents are ordered with respect to the relation "earlier than" no less than with respect to its converse "later than," it is a mere tautology to say that the Now shifts from earlier to later. Secondly, it is only obtaining of a diversity of now-contents which is a matter of fact, but not the allegedly unidirectional character of the progress of time. The factuality of the diversity of now contents does not suffice to give synthetic content to the assertion that time progresses unidirectionally.

When characteristics of reality happen to share or match the characteristics of experience, one ought to question them rather than taking them for granted. In line with the total assault of scientific investigation and critical rationality on our most well-cherished and established intuitions, why should we expect that a global picture of the world to which we humans belong can resemble or be the extension of characteristics of the subject?

Why not instead investigate alternative models of time or causality —which are only rejected on the basis of not looking natural from the perspective of our time-asymmetric assumptions—that while are compatible with local temporal and causal perspectives, they do not privilege one perspective over another? These are models that should allow the “alternation of time directions” while not contradicting temporal “experiences accessible to us” or some other contingently posited local observer/subject. This is much more in tune with a pragmatist account of the time-conscious agent for which the anthropocentricity in the use of temporal or time-directional concepts is justified and even pragmatically necessary so long as there is no metaphysical reification of these concepts as global characteristics of the objective world.

To summarize, the pragmatist view of the time-conscious agency should be based on a model of time that while refuses to be characterized temporally, admits the local consistency of any possible model of experienced temporality and causal perspective such as ours and hence in an oblique way, justifies the use of temporal concepts. In other words, a robust model of time should be an expression of a reality that constitutes local temporal perspectives without ceasing to become one. Thus reencountered within this model, our temporal perspective should be seen as a local self-expression of an absolute (atemporal) reality rather than being dispensed with as a complete illusion.

The liberation from a model of time restricted to a particular contingent constitution does not rob the subject of its cognitive and practical abilities, but releases it from the shackles of its most entrenched dogmas about the necessity of the contingent features of its experience. In doing that, such liberation sheds light on the prospects of what the subject of experience and exercise of change in the world is and can be. The transition to a state where one no longer is afraid of being lost in time for it has come to the realization that time accommodates no one should be celebrated as the sign of rational maturity rather than decried as being a manifestation of the subject's impotency. It is in continuity with the critical attitude of the rational agency to adopt a model of time that can interrogate the most natural and established 'facts of experience' rather than corroborating them by the so-called fact that these are simply the ways by which we experience the world.

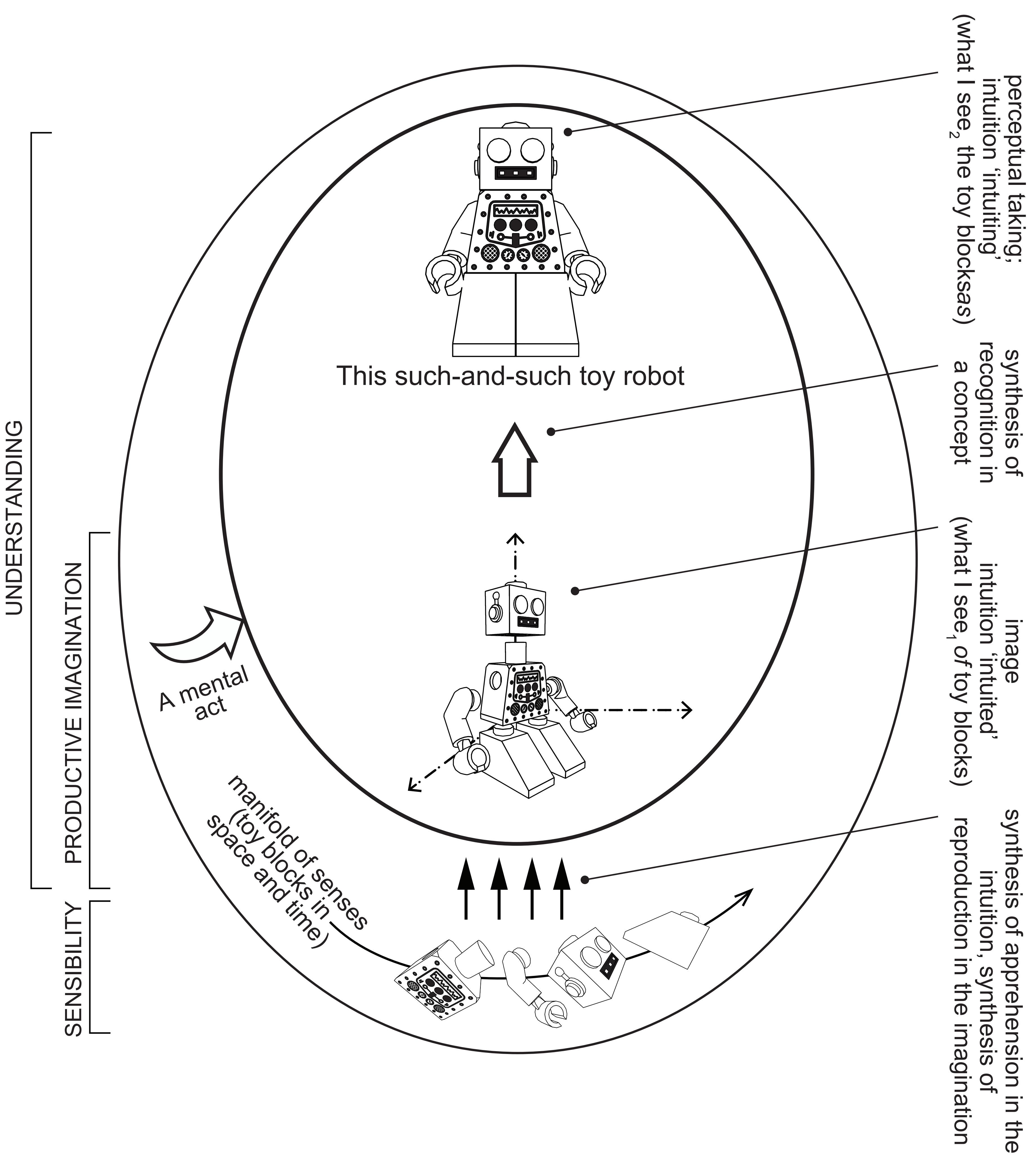
Having cursorily glanced over the radical implications of Boltzmann's radicalization of Kant's thesis on the transcendental ideality of time qua experienced temporality as a characteristic of the transcendental structure of exprience, we are now faced with three challenges:

(1) Our capacity to be affected by representations qua intuitions of items and respond to them is conditioned upon our rudimentary capacity to be non-conceptually aware of the sequence in which we are being affected and the even more rudimentary capacity of synthesizing compresent impressions into a sequence of impressions of which we can be aware as an orderly whole. What is necessary here are the roles of the synthesis of impressions and an awareness that can track the sequence of impressions. But how impressions should be synthesized and what awareness of impressions as an orderly whole should look like need not be treated as necessary conditions for the realization of these roles or as invariant aspects of the agency. Respectively, how can we envision models of agency that might have fundamentally different perceptions of time by virtue of enjoying different local conditions of observation, possessing different structural-behavioral organizations (i.e. different modes of responsiveness to the impingement of items in the world on their senses and different constructive-anticipatory models of memory) or on the conceptual levels of time-perception, having different logical connections between temporal connectives of language or different structures of tensed verbs? This is as much a challenge to think the agency beyond a particular set of contingencies as it is a research question for envisioning an artificial model of the agency not essentially bound by our local contingently posited limits.

(2) Are there models of time that can be compatible with the subject-observer's temporal-causal perspective while not privileging or overstretching this local perspective into a canonical global model? How can we provide a physics of our temporal-causal perspective that explains its characteristics without reinscribing the same perspectival characteristics as features of objective reality and thus, positing them as explanans of what is already an explanandum? To this extent, the second challenge is concerned with a systematic inquiry into models of time and causality that can (a) account for the characteristics of our temporally oriented perspective, and (b) resolve the problems within the directional-dynamic picture of time, namely, quandaries and paradoxes (from enigmas of change and temporal asymmetry to paradoxes of causality such as retrocausality) that originate from the inadequate descriptive-explanatory resources of the directional-dynamic model of time.

(3) In line with the first and the second challenges, the third challenge centers on the problem of reconciling the pragmatic imports of our temporal view with a model of time that is neither necessarily directional nor dynamic. What are the ramifications of a non-directional/non-

dynamic model of time for our existing theoretical and practical models not only within physical sciences but also social sciences? And more importantly, what can be gained, theoretically and practically, by adopting alternative models of time and specifically an atemporal model? The third challenge is in fact the continuation of the rational agency's struggle for self-conception and self-transformation. Rather than encountering itself from a particular temporal perspective bounded by contingent features of its local positioning (a particular when), the agency adopts a viewpoint that coincides with that of reality, that is, a "view from nowhen" (Huw Price). In adopting a view from nowhen—a view that explains any mode of experienced temporality without being circumscribed by their particularities—we embark on a necessary task required for the realization of an intelligence that moves from a particular contingent perspectival consciousness to a genuine self-consciousness, an outside view of itself. From a Kantian perspective, in taking up the third challenge, we come closer to fulfilling the central goal of critical philosophy, that is demonstrating the mutuality of the rational self (discursive apperceptive intelligence) and the world, without eliding the distinction between thinking of the former and being of the latter. In Jay Rosenberg's words, showing that "The same activities of synthesis which constitute the represented world as an intelligible objective unity constitute the representing self as an apperceptive subjective unity." This is an insight that is as consequential for the future of the already constituted subject of thought (us humans) as it is significant for the realization of a future vehicle of thought.


perceptual taking; intuition 'intuiting' (what I see$_2$ the toy blocks*as*)
synthesis of recognition in a concept
image intuition 'intuited' (what I see$_1$ *of* toy blocks)
synthesis of apprehension in the intuition, synthesis of reproduction in the imagination
This such-and-such toy robot
UNDERSTANDING
PRODUCTIVE IMAGINATION
SENSIBILITY
A mental act
manifold of senses (toy blocks in space and time)

Language as a *sui generis* computational framework

Analysis of theoretical and and practical abilities

Realizabilities

What arises from the exercise of mental powers or abilities of the mind

Realizers

Conditions and capacities required for the realization of mental powers or abilities

logico-conceptual norms (modelling on realizabilities)

analogically modelled

analogical bootstrapping

functions and behaviours (realizers modeled on realizabilities)

multi-level analysis of capacities $\mathcal{C}_n$ based on

$\mathcal{D}_n$: correct analogical descriptions

$\mathcal{M}_n$: appropriate formal logical frameworks for modelling $\mathcal{D}_n$

$\mathcal{D}_1 \otimes \mathcal{M}_1 \multimap \mathcal{C}_1$

$\mathcal{D}_2 \otimes \mathcal{M}_2 \multimap \mathcal{C}_2$

$\mathcal{D}_3 \otimes \mathcal{M}_3 \multimap \mathcal{C}_3$

$\mathcal{D}_4 \otimes \mathcal{M}_4 \multimap \mathcal{C}_4$

hierarchical descriptive-explanatory analysis of structural-causal mechanisms $\mathcal{S}_n$

$\mathcal{S}_1$

$\mathcal{S}_2$

$\mathcal{S}_3$

$\mathcal{S}_4$

Computational Dualities i.e. Interactions and Their Histories

(types = set of processes with a common behavior with respect to duality)

Non-deterministic / Untyped Computation

Type construction

(Imagination)

Type System

(Understanding)

Language

Trans-typified System deterministic and nondeterministic comp.

(*Judgments*)

Semantic Complexity, Socio-linguistic cognitive technologies

internal model (Increasing complexity, decreasing size)

behaviour (Increasing complexity, variation)

qualitative compression and selectivity of data

formatting and modulation of behaviors

Structure

Internal model

memory

receptivity

sensors

behaviour

effectors

σ- and τ-related items (sensations)

Sequence of spatial and temporal relations of being thus-and-so affected

Outer Sense

Time Space

Outer Sense

Inner Sense

Time

Inner Sense

Sensibility

Mental Acts

Combination of concepts in an intuited object

Combination of the recognized (intuited) in a concept

Time

Imagination

Unity of Apperception

Tensed Thought

Understanding

Discursive Apperceptive Intelligence

Contingent characteristics of the constituted subject's experience

Local transcendental structure / perspectivalism $\mathcal{T}_l$

CRITIQUE OF TRANSCENDENTAL STRUCTURE $\mathcal{T}$

$\mathcal{T}_l \multimap \mathcal{M}_n \xrightarrow{\text{variation of } \mathcal{T}\text{-structure}} \mathcal{T}_u \multimap \mathcal{M}_m$

$\mathcal{M}_n \multimap \mathcal{C}_n \xrightarrow{\text{modification of capacities}} \mathcal{M}_m \multimap \mathcal{C}_m$

$\mathcal{S}_n \multimap \mathcal{C}_n \xrightarrow{\text{modification of mechanisms}} \mathcal{S}_m \multimap \mathcal{C}_n$

Necessary conditions and capacities for the realization of the subject

Universal transcendental structure / multiperspectivalism $\mathcal{T}_u$

Project of Fundamrental Alienation of the Human qua constituted subject

AGI

Natural Languages

Reconstitution of Language

Logic

Computer Science

Mathematics

Formal Languages

General Artificial Languages (syntax, semantics, pragmatics)

Reason

Items situated in space and time (environment)