pertaining to the place of ourselves in the world. If we take ourselves as a functional toy model of AGI, then we are dealing with at least two classes of metatheory: metatheories associated with the bulk of models we are using to identify the conditions required for the realization of the cognitive-practical abilities necessary for anything that can be treated as an index of general intelligence; and metatheoretical assumptions related to the conditions of observation under which these necessary capacities are distinguished and described. It is the latter that need to be subjected to a thoroughgoing critique of transcendental structures of experience in order for the former to be adequately objective. Absent a systematic attempt to render explicit our subjective experiential assumptions with regard to the cognitive acts we take to be necessary, 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, by-products of our local and contingently situated transcendental perspective.

An outside view of ourselves as a toy model AGI that allows us to conceptually come to grips with the problematics of these two classes of metatheoretical assumptions is exactly what the philosophy of German Idealism encapsulates. German Idealism is a theoretical system built at the intersection of the philosophy of action, philosophy of mind, and philosophy of knowledge. In other words, it is broadly concerned with the conditions of possibility of theoretical and practical cognitions, what they are and how they can be realized. Discussing AGI in the context of German Idealism may appear a retrogressive move to some, but this could not be any further from the truth. When it comes to philosophy of mind, German Idealism—particularly that of Kant and Hegel—poses the right kinds of questions.

For the most part, the problem with today's research on artificial general intelligence is that it is content with ephemeral smart answers to ill-posed questions. Tied up in its local achievements and complying with the demands of the market, the field of artificial intelligence has neither the time nor the ambition to think about what it means to pose the right kinds of questions regarding the nature of mind and the realization of cognitive acts. The idealisms of Kant and Hegel, on the other hand, outline the fundamental conceptual problems that are in fact still at the centre of debates in cognitive and theoretical computer sciences. Needless to say, if our aim is to understand the relevance of these problems for artificial general intelligence, then we will need to reframe them in terms of concepts that are much more in tune with contemporary cognitive science. And of course, throughout this process of synchronization certain aspects of Kant's and Hegel's programs will turn out to be—for various reasons—untenable. Nevertheless, our extended thought experiment remains an unabashedly Kantian-Hegelian reconstruction of ourselves as a toy model of general intelligence, starting with Kant's transcendental psychology and moving toward Hegel's formulation of language as the dasein of geist.

FORMALIZING A BIG TOY UNIVERSE

Before moving forward and concluding the discussion on toy models and what it means to view ourselves as a toy model AGI, it is important to provide a minimal formal description for the kind of big toy model we are considering here. I propose two similar candidates, the concept of Chu spaces, particularly as elaborated by Vaughan Pratt, and Virtual Machine Functionalism (VMF) as developed by Aaron Sloman.[96]

Roughly speaking, Chu space is a topological space in which computational dualities and interactions can be accurately expressed. Pratt's seminal paper 'Rational Mechanics and Natural Mathematics' attempts to capture Descartes's mind-body dualism not as a problematic metaphysical dualism but as a computational duality in the precise sense of 'duality' given in mathematics. The duality at stake here is the computational interaction—the interchange of roles—between the formal dimensions of thinking (Kant's

96 See V. Pratt, 'Rational Mechanics and Natural Mathematics', in *TAPSOFT '95: Theory and Practice of Software Development (Lecture Notes in Computer Science)*, vol. 915 (Heidelberg: Springer, 1995), 108–22; and A. Sloman, 'Architecture-Based Conceptions of Mind', in P. Gärdenfors et al. (eds.), *In the Scope of Logic, Methodology and Philosophy of Science*, vol. 316 (Heidelberg: Springer, 2002), 403–27.

a priori acts of cognition) and the substantive dimension of sensibility (which can be laid out in terms of the causal-mechanistic structure of the nervous system). How do these two categorically distinct spaces—of reasons and causes—interact? What other kinds of meta-interactions or computational dualities must be introduced into the duality of thinking and sensing, reasons and causes, for them to be coherently pictured as interacting without their qualitative distinctions being elided or ontologically flattened? We will examine the details of the logico-mathematical concept of duality—rather than metaphysical dualism—in chapter 6. But for now we can think of the duality of thinking and sensing, reasons and causes, as an *equivalence* relation between the two—as, for example, when one obtains a mirror image of something rather than a simple opposition or dualism (mind *vs.* body).

Pratt captures this interaction using Chu spaces—understood as topological spaces satisfying the criteria of interaction between a set and an antiset together with their corresponding functions and antifunctions, which stand for the bodily and the mental, causes and norms, the neural materialism of the cerebrum and the logical idealism of the psyche. In the simplest terms, Chu space can be defined as a triple $\langle P_o, \vDash_P, P_a \rangle$ over a set K, where P_o is a set of objects or events, P_a is a set of attributes or states, and $\vDash_P$ is a satisfying relation which is the subset of the interaction game $P_o \otimes P_a$, or the matrix $P_o \times P_a$ whose entries are drawn from K. In the most elementary form $K = 2 = \{0,1\}$, P_o stands for the set of causal events or sensings, P_a for the logical space of the mental or inferential states, and the satisfying relation $\vDash_P$ represents the interaction between the categorically distinct set of the causal events or structure and its dual, the antiset of the cognitive (rule-governed) functions.

Here, states and events refer to the fundamental dual concepts of computation. The states of a computing system or an abstract automaton (e.g., a Turing machine), bear information and change time, i.e., the total amount of time elapsed since the beginning of the process. Events in a computing system, on the other hand, are instantaneous: they change information and bear time. Succinctly speaking, states are information-stamped while events are time-stamped. In this respect, 'we may think of the state as bearing information representing the "knowledge" of the automaton when in that state, and the event as modifying that information'.[97] Employing the fundamental duality of states and events, then, any behaviour—whether monotonic or non-monotonic—of a process can be seen as an unfolding with regard to time and information. It can be plotted as a graph where the x-axis represents time and the y-axis represents information, with the states as horizontal line segments and the events as vertical segments.[98] The relevance of this schema to the big toy model of AGI is that it allows us to model not only the interactions between sensings and thinkings—physical events and noetic states—but also the *distinct* causal interactions between physical events on the one hand, and noetic states on the other, as different classes of computation.

The Chu space view offers advantages which are necessary and crucial in the construction of a big toy model of AGI:

- It supplies a precise formal framework in which the distinctions between sensings and thinkings but also the causal interactions between noetic states (thinking_1, thinking_2, ...) and physical events (sensing_1, sensing_2, ...) are preserved rather than being elided.

- In modelling the interactions between states-events, states-states, and events-events as different classes of computations, the Chu space view eliminates the vaguely generic notion of thinking as global information-processing. Different physical and mental behaviours are modelled as distinct forms of computation. Their unfolding behaviour is explained in terms of the duality of information and time, as series of states and

97 V. Pratt, 'The Duality of Time and Information', in W.R. Cleaveland (ed.), *CONCUR '92: Third International Conference on Concurrency Theory* (Dordrecht: Springer, 1992), 237.

98 In this diagram, which expresses the duality of information and time, schedules (i.e., sets of events distributed in time), and automata (i.e., sets of states distributed in information space), time appears to be flowing downwards in the schedules and upwards in the automata.

events in which the *logical* space of reason can be said to be swimming upstream against time and time moving upstream against logic.

- In treating complex behaviours in terms of interactions between distinct classes of processes and interactions as different classes of computations, the Chu space view eliminates the risk of the naïve or monolithic view of processes and behaviours. Behaviours are characterised strictly in terms of specific levels or types of interactions between, for example, an agent and its environment, or this agent's physical and noetic states.

- Finally, the Chu space view permits us to model physical and mental behaviours in terms of true concurrent interactions—i.e., synchronous as well as more generic asynchronic actions of processes on one another. Computational concurrency allows us to capture complex phenomena which are otherwise unavailable or hidden in conventional computational models: such as conflicts, asynchronous interaction, temporal precedence, supervenience, formal autonomy of noetic states, and the distinction between causing and enabling events.

To further clarify how Chu spaces can model the interactions between sensings and thinkings, we can now turn to Pratt's remarks on the mechanics of interactions between physical-causal events and noetic states:

> Events of the body interact with states of the mind. This interaction has two dual forms. A physical event a in the body A impresses its occurrence on a mental state x of the mind X, written $a \dashv x$. Dually, in state x the mind *infers* the prior occurrence of event a, written $x \models a$. States may be understood as corresponding more or less to the possible worlds of a Kripke structure, and events to propositions that may or may not hold in different worlds of that structure. With regard to orientation, impression is causal and its direction is that of time. Inference is *logical*, and logic swims upstream against time. Prolog's backward-chaining strategy dualizes this by viewing logic as primary and time as swimming

upstream against logic, but this amounts to the same thing. The basic idea is that time and logic flow in opposite directions.

Can a body meet a body? Only indirectly. All direct interaction in our account of Cartesian dualism is between mind and body. Any hypothesized interaction of two events is an inference from respective interactions between each of those events and all possible states of the mind. Dually, any claimed interaction of two states is inferred from their respective interactions with all possible events of the body. The general nature of these inferences depends on the set K of values that events can impress on states. The simplest nontrivial case is $K = 2 = \{0,1\}$, permitting the simple recording of respectively nonoccurrence or occurrence of a given event in a given state. In this case body-body and mind-mind interactions are computed via a process called residuation. Specifically, event a *necessarily precedes* event b when every state x witnessing the occurrence of b also witnesses a. This inferred relationship is calculated formally by left residuation, which we describe in detail later. The dual calculation, right residuation, *permits* a transition from state x to state y when every event a impressing itself on x does so also on y. That is, any transition is permitted just so long as it forgets no event. These simple-minded criteria are the appropriate ones for the small set $K = 2$.[99]

The 2-valued interaction between the causal and the rational/normative is an elementary Chu space. For this interaction to be nontrivial, the value of K must be greater than 2 (i.e., containing fuzzy values), which in turn means more complex rules of transition between them are obtained and the entries of the matrix of interaction grow. Another interesting property of a Chu space is that it can accommodate additional dualities and interaction spaces, permitting complex Chu transforms or mappings between different Chu spaces.[100] In contrast to Descartes's metaphysical dualism and Spinoza's

99 Pratt, *Rational Mechanics and Natural Mathematics*, 110.

100 Formally, a Chu transform is a morphism between two Chu spaces $\langle P_o, \models_P, P_a \rangle$ and $\langle Q_o, \models_Q, Q_a \rangle$. This morphism is a pair of functions (f_a, f_o) with $f_o : P_o \rightarrow Q_o$ and $f_a : Q_a \rightarrow P_a$

substance monism of the mind, Kant's picture of the mind—as elaborated within his transcendental psychology—can be expressed as a Chu space toy model, where the interaction between sensing and thinking, empirical computation and logical computation, permits a larger matrix of interaction and the accommodation of additional Chu spaces, thus obtaining more complex rules of transition and Chu transforms between what is causal and what is normative. Kant's schema of the three syntheses (synthesis of apprehension in the intuition, synthesis of reproduction in the imagination, and synthesis of recognition in the concept) is precisely an interaction matrix with complex rules of transition and Chu transports between sensing and thinking, the causal and the formal (i.e., logico-linguistic).

Lastly, the interaction between the set of events and the set of states in a Chu space can also be formulated in terms of schedules and automata, therefore turning Chu spaces into an ideal framework for the study of true computational concurrency, i.e., interactions between asynchronous processes of different types. The fact that Chu spaces can adequately model true concurrency—rather than merely sequential computation—makes them strong candidates for capturing the richness of the big toy model of general intelligence where the interactions between diverse processes as sets of behaviours require complex forms of coordination and scheduling.[101]

The intuitive idea behind Sloman's VMF is comparable to the interaction matrix formally presented by Chu spaces:

---

such that for any $x \in P_0$ and $x \in Q_0$ the following condition must be satisfied: $f_0(x) \vDash_Q y$ iff $x \vDash_p f_a(y)$. In more layman-friendly terms, the Chu transform can be interpreted under certain caveats as an adjointness condition between two distinct spaces. Think of an arrow forwards from sense-given materials to a priori acts of cognition and an arrow backward from a priori acts of cognition to sense-given materials. The Chu transform consists of these two arrows satisfying the dynamic condition of adjoint or mutual realization of the space of causes and the space of reasons, with all the intermediary back-and-forth movements or mappings required for the interaction between the two.

101 See for example, V. Gupta, *Chu Spaces: A Model of Concurrency* (1994), <http://i.stanford.edu/pub/cstr/reports/cs/tr/94/1521/CS-TR-94-1521.pdf>.

> Virtual Machine Functionalism (VMF) attempts to account for the nature and causal powers of mental mechanisms and the states and processes they produce, by showing how the powers, states and processes depend on and can be explained by complex running virtual machines that are made up of interacting concurrently active (but not necessarily synchronized) chunks of virtual machinery which not only interact with one another and with their physical substrates (which may be partly shared, and also frequently modified by garbage collection, metabolism, or whatever) but can also concurrently interact with and refer to various things in the immediate and remote environment (via sensory/motor channels, and possible future technologies also). I.e. virtual machinery can include mechanisms that create and manipulate semantic content, not only syntactic structures or bit patterns as digital virtual machines do.[102]

Rather than a functionalist approach to the general architecture of the mind framed in terms of the supervenience of states and properties on well-defined input-output transitions, VMF is about the layering of virtual machines that form a complex network of causally and logically interacting processes operating on different time scales or schedules. In other words, the architecture of the mind is presented as a complex of virtual machines that can be implemented in any sufficient physical machine or system. But because these virtual machines are not physical as such, they can also form matrices of interaction and mereological (part-whole) relationships that cannot obtain in physical machines. The notion of virtuality, in this sense, does not suggest that virtual machines are not real, but that they are primarily machines that are not describable in physical terms. Virtuality refers to emulated machines capable of modelling qualitative properties and characteristics of consciousness or mind, as distinguished from those of the physical systems that support and embody them.

---

102 A. Sloman, *Virtual Machine Functionalism* (2013), <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vm-functionalism.html>.

Sloman's VMF presents a functional diagram or architectural schema of mind built around supervenience and emergent behaviours. Yet in contrast to the more traditional theories, the concepts of supervenience and emergence are principally defined in terms of virtual machines (VMs) and their causal interactions, rather than the causal interactions between physical systems or machines (PMs). As such, VMF avoids the pitfalls of the traditional theories of supervenience and emergence, which have been justly criticized by the likes of Jaegwon Kim.[103] An intuitive example of VM-supervenience would be the interaction between a series of running applications on a computer: an operating system or a platform VM, a word processor, and a number of plug-ins such as equation and graphic editors which can run independently or within the word processing software. A script can be written or a new piece of software can be introduced that monitors the resource-consumption of these VM-interactions and, if necessary under such-and-such parameters, distributes the processing resources among them, or continuously changes the temperature by regulating the physical components of the computer such as the fan speed, etc.

Similarly to the Chu space model, VMF discriminates between different types of causal interactions: those that take place between VMs, those that take place between a VM's events and processes and a PM's events and processes, and finally, those causal interactions that take place between PMs. This schema allows the modelling of VM-interactions as a nested or generative hierarchy—rather than a traditional control hierarchy—of concurrent and interacting virtual machines distributed along different scales or levels of granularity, shifting from a fine-grained granularity at the level of physical supporting systems to the coarse-grained granularity of virtual machines linking their supporting systems. The multiscale view enables VMF to treat characteristics of the mind and the constraints or conditions necessary for their realization on different descriptive levels, thereby avoiding the risk of the flat picture of functions discussed in

103 See J. Kim, *Essays in the Metaphysics of Mind* (Oxford: Oxford University Press, 2010), particularly chapters 3, 11, and 13.

the previous chapter. The descriptive hierarchy of VMF has a number of advantages over traditional functionalist approaches:

- It calls for an integrative engineering-philosophical approach to mind. Methods and models must be thought and implemented *across scales*. The pluralism entailed by VMF's multilevel architecture requires an objective ranking of models and methods, in the sense that for each descriptive layer there are specific sets of models and methods which must be prioritized. For example, rather than taking Bayesian learning methods to be the principal methods for modelling general intelligence, they can be effectively employed as the most appropriate methods for modelling the process of low-level information such as visual and audio signals. Salient local variations of an object such as a chair can be singled out and learned via Bayesian networks. Using threshold mechanisms, the output of the process can be sorted into a set of discrete propositions representing the candidate features from which a perceptual invariance of the object can be extracted.

- Again similarly to the Chu space model, VMF's multiscale view does not admit the vaguely general and unhelpful concept of information processing. The question of different types of information and methods of information processing is central to VMF. By reapplying Bayesian methods to this candidate set of local variations, a rudimentary *perspectival* image-model qua singular representation of an item can be constructed, in a process similar to that of Husserlian adumbrations (*Abschattungen*). The set of local variations or the rudimentary image-model obtained by the Bayesian method by themselves cannot construct an object, since object-construction in its proper sense involves semantic-conceptual ingredients. While a chair qua image-model can be rudimentarily constructed or recognized using such methods, the judgement pertaining to the *concept* of the chair—which is not perspectival—is an entirely different issue (e.g., this is a chair with its legs missing, this is a chair and not a small table, or this is a chair therefore...).

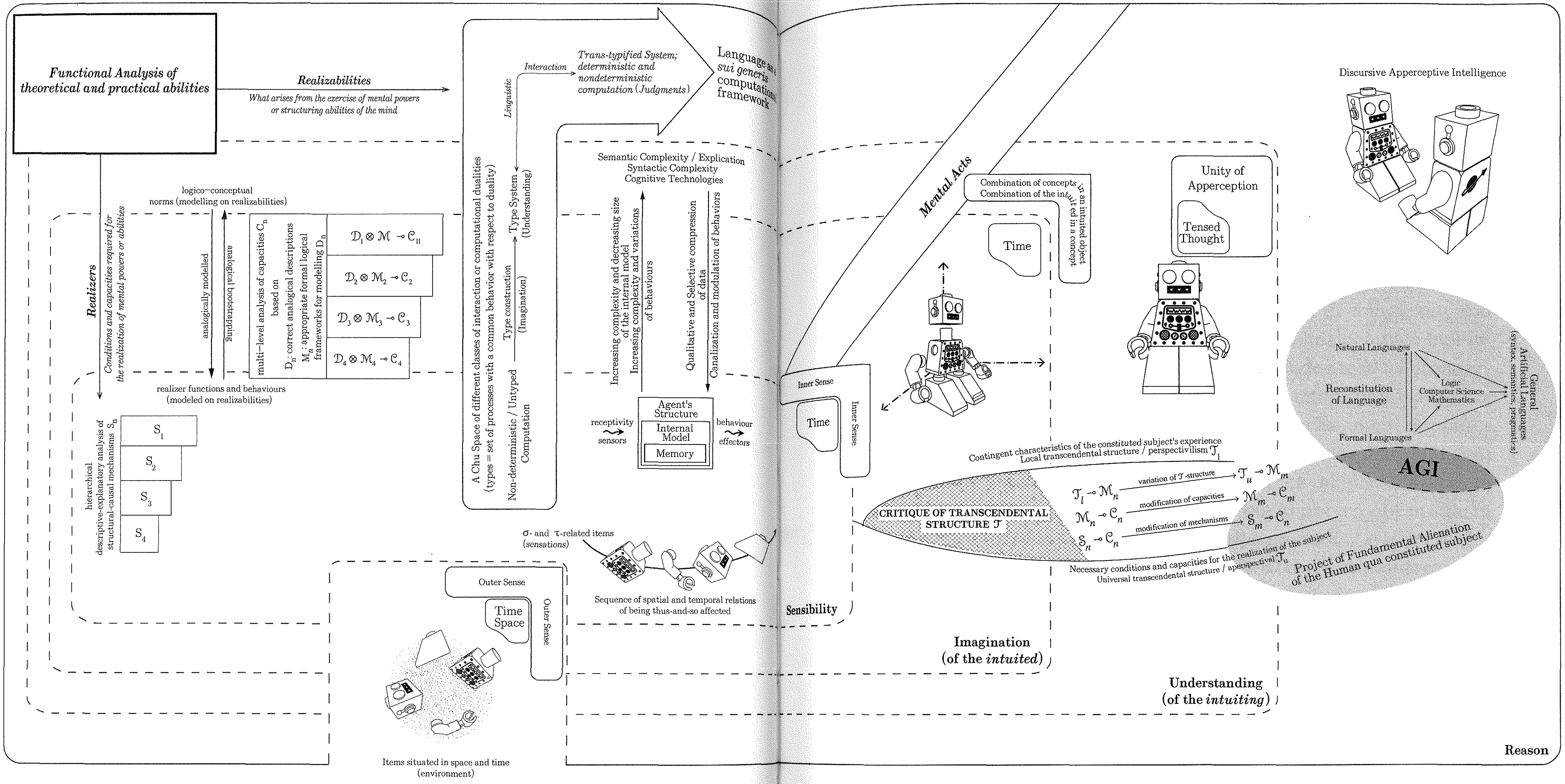
- The differentiation of scales and descriptive levels effectively connects VMF with the question of ontologies. In information science, the concept of ontology refers to a system for the formal naming and definition of the types, properties, roles, and interrelations of entities/particulars in a specific domain of discourse. An upper-level or mid-level ontology supports broad semantic interoperability between a large amount of ontologies accessible under it. In other words, it is a framework of complex categorization through which data across an expansive range of different domains can be exchanged, tracked, and computed. Within this framework, each level should be described by vocabularies which are formally and semantically adequate to capture the types, properties, and roles of particulars inhabiting that level. VMF thus emphasises the mesoscopic layer, refusing either purely top-down or purely bottom-up ontologies. The architecture of mind cannot be solely described in information-theoretic terms or those of any other supposed base-vocabulary. The use of representational, functional, and normative vocabularies borrowed from the transcendental tradition of philosophy of mind are not just informative here, they are descriptively indispensable and methodologically compulsory. It would not be controversial to claim that a multiscale view such as VMF is where computer science converges with transcendental philosophy of mind in the vein of German Idealism.

Framed as a virtual machine toy model, the only necessary aspect of the mind is its virtual *architecture*, which must satisfy the different levels of integration of information and computational criteria required for semantic and causal interactions between its components. The physical system supporting the VM architecture and the exact number or nature of the virtual machines or information processing systems can all be toyed with. Constraining the nature of the mind to a necessary general architecture that can accommodate new pieces of virtual machinery, Sloman's VMF allows us to treat the idea of artificial general intelligence in analogy to, but also with the exact same scope as, the development of a prelinguistic infant who is capable of growing into a language-using adult.

Let us recapitulate and conclude this section before embarking on our thought experiment. The treatment of ourselves as a toy model AGI should be seen as an attempt—incomplete at best and fundamentally crude at worst—to distinguish what is necessary for the realization of general intelligence (in organic species or inorganic systems) from what is contingent, and to investigate the extent to which the description of what we take to be necessary may indeed be distorted by what is in reality contingent. That is to say, the view of ourselves as a toy model AGI should allow us to eject the final residues of essentialism from the concept of the human, and thereby create an opportunity to examine the possibility of a nonparochial conception of artificial general intelligence. This is the core of our extended thought experiment: to entertain an outside view of ourselves as a prototype AGI—or more precisely, an artificial multi-agent system. If we were able to adopt this 'outside view' then what would our world look like from the perspective of artificial intelligence design? What kinds of basic capacities must these agents have in order to support complex schemas of self-conception and self-revision? And what would be the ramifications of these elementary abilities for their world, particularly when they become integrated and fully mobilized? Finally, what would be the test for determining whether this entry-level AI—i.e., ourselves here and now—in fact qualifies as a general intelligence?

However, what makes the universe of this entry-level AI world different from ours is that it is tailored and tapered to accentuate the *need to integrate accounts of cognition in linguistic interaction* and to highlight the role of speech, language, and interaction as special computational frameworks necessary for the construction of human-level AI. The method of inquiry presented here is essentially an integration of what might be called a constrained dynamic 'neural materialist' approach (Jean-Pierre Changeux, Stanislas Dehaene, et al.) and an interactive 'semantic inferentialist' approach (Wilfrid Sellars, Robert Brandom, Jonathan Ginzburg, and others) to cognition. Also, this world is assembled from myriad theoretical components that are potentially falsifiable, but which, once assembled, serve as a metatheory that provides us with insights into certain significant but non-self-evident features of a possible AGI.

Our toy universe will be constructed in two stages. In the first stage, in the next chapter we shall introduce those necessary capacities or faculties required for the realization of a discursive apperceptive intelligence that originate from causal-structural aspects of agency. In the second stage (chapters 4 and 5), we shall investigate the transformation of our rudimentary agent equipped with these structurally-causally originated capacities into a fully fledged discursive apperceptive agent whose basic capacities are not only caught up in language but which also, by virtue of being a language-user, is in possession of a generative framework of theoretical and practical abilities. For reasons that will soon become clear, in the first stage, chapter 3, we shift the focus to Kant, returning to Hegel in chapter 5 when the real protagonist of our toy universe becomes not a complex system of agents—in contrast to a rudimentary singular agent—but language itself. These two stages of construction, accordingly, set out the Kantian-Hegelian outlines of a programme of artificial general intelligence in which technical and social realizations of sapience coincide within a philosophical program for investigating the meaning of agency and bringing about its realization—a theme we shall explore in the final chapter.



AGI roadmap: charting the territory of the AS-AI-TP thought experiment at the intersections of the functional analysis of mind, the critique of transcendental structures, and the development of structuring abilities.

3. This I, or We or It, the Thing, Which Speaks (Forms of Intuition)

THE TOY UNIVERSE OF AN EMBODIED AUTOMATON

Suppose a hypothetical embodied automaton possessing the following properties and features:

FEATURE 1: It has been programmed to instantiate a number of diffuse and recurring goals centred on the maintenance and preservation of the system qua agent. The goals are at different levels of complexity, and exert various degrees of pressure on the system, analogous to biological needs. Furthermore, such pressures can be understood as necessary physical constraints imposed upon the agent as an information processing system, its computational capacities, its situated behavioural responses and even its range of possible behaviours. This means that our agent is essentially a limited being possessing neither infinite information processing time nor infinite resources. We would not have been able to even characterize this agent as an information processing system, were it not restricted by the physical and computational cost constraints existent in its sensory-behavioural matrix. In a nutshell, just as we cannot speak of computation without constraints of run time and computational cost, so we cannot speak of an agent as an information processing system without the physical constraints inherent to how it interacts with the environment by way of goal-specific behaviours. We might as well talk about supernatural beings. As we will soon see, the sense in which these goals are comparable to our own is complicated, but it is from here that we must begin.

FEATURE 2: It has been wired to engage in activities that increase the probability of goal-fulfilment. The wiring of the automaton can be thought of in terms of interacting levels of structure that can produce transient multiple variations in the internal states of the system.

The wiring or neural structure comprises primary low-level information processing modules, additional higher-level modules for intermediating primary modules, and workspaces through which processed information can be selected and made available to a higher-level global workspace that can be accessed by executive functions and goal-oriented behaviours.[104] The wiring should be sufficient to causally mediate between complex environmental inputs and the complex behavioural outputs of the agent. In short, the agent can be defined as a teleological system with *sufficient wiring* (structure) to be capable of reliable differential and adaptive responsiveness with respect to its environment. It is, however, crucial to grasp these goal-oriented activities as 'proto-intentions' whose objects are *causally*—not conceptually—entangled with the structure of the agent's goals. An example of such proto-intentional activities would be a predator chasing its prey (the object of the hunt). Without implying that the predator is aware of the content of its experience or capable of attributing the experience to itself, we can still speak of the predator's *awareness of* the prey in the sense of information from different information processing modules in the wiring structure of the predator being *globally* made available to a *workspace*. This workspace can be accessed by the attentional and evaluative systems that supervise the execution of the predator's goal-oriented activities. For example, as the predator orients itself toward the prey, the information provided by other (unconscious) processes at the level of wiring structures/processing modules is

104 The term 'global workspace' was first coined by Bernard Baars, to refer to 'a central information exchange that allows many different specialized processors to interact. Processors that gain access to the global workspace can broadcast a message to the entire system. [...] The word "global" in this context simply refers to information that is usable across many different subsystems of a larger system. It is the need to provide global information to potentially any subsystem that makes conscious experience different from many specialized local processors in the nervous system.' B.J. Baars, *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press, 1988), 43.

temporarily mobilized and becomes globally (or consciously) available to various processes required for the execution of a proto-intentional goal-oriented activity.[105] In this schema, the global workspace can be thought of as a distributed network with incoming and outgoing links from and to the underlying information-processing modules.

FEATURE 3: With regard to its capacity for reliable differential responsiveness (i.e., the sufficient structure required for causal mediation between environmental inputs and behavioural outputs), the agent is equipped with different specialized sensors and different modules for integrating sensory data both within a specific sensory modality (multiple data associated with one sensor) and across different modalities (data associated with different sensors).

This sensory integration extends the spatial and temporal coverage of the sensory information processing system and increases the robustness and reliability of sensory information. In addition, it reduces the

105 For example, see the work of Stanislas Dehaene on the role of the attentional system in mobilizing a global workspace: 'Top-down attentional amplification is the mechanism by which modular processes can be temporarily mobilized and made available to the global workspace, and therefore to consciousness. According to this theory, the same cerebral processes may, at different times, contribute to the content of consciousness or not. To enter consciousness, it is not sufficient for a process to have on-going activity; this activity must also be amplified and maintained over a sufficient duration for it to become accessible to multiple other processes. Without such "dynamic mobilization", a process may still contribute to cognitive performance, but only unconsciously. A consequence of this hypothesis is the absence of a sharp anatomical delineation of the workspace system. In time, the contours of the workspace fluctuate as different brain circuits are temporarily mobilized, then demobilized. It would therefore be incorrect to identify the workspace, and therefore consciousness, with a fixed set of brain areas. Rather, many brain areas contain workspace neurons with the appropriate long-distance and widespread connectivity, and at any given time only a fraction of these neurons constitute the mobilized workspace.' S. Dehaene and L. Naccache, 'Towards a Cognitive Neuroscience of Consciousness: Basic Evidence', *Cognition* 79 (2001), 1–37: 14.

ambiguity of sensory input, and thus increases salience. Specifically, integration across different sensory modalities expands the range of behaviours and yields a stronger effect on the behavioural output. The sensory integration is carried out at different levels, distinguished in terms of models of information processing (sequential or concurrent), temporal links between different streams of input (synchronous and asynchronous), and various frames of reference.

FEATURE 4: The automaton is furnished with a sufficiently complex and functionally flexible memory capable of encoding, retrieval, consolidation, discarding, and transfer of sensory impressions.

Rather than operating as an established storage space for the retention of past impressions, this memory is modelled as an adaptive dynamic process that plays a constructive—or more accurately, simulative—role in the behaviours of the agent. The retrospective retrieval of information (recalling) is correlated to its prospective constructive role in the agent's ongoing interaction with its environment. Every time a memory needs to be accessed, it is constructed. Recalling the memory of an original impression is tantamount to reconstructing it. But this process of reconstruction is guided by the situation at the time of access or demand for retrieval. In other words, everything that has happened since the time of original impression determines the result of the construction. Moreover, each constructed memory becomes a part of the situation and, accordingly, influences the construction of further memories. The memory system is therefore not a predefined fixed state, but is governed by the situatedness of the constructive process that creates simulations of the environment. This constructive process links the external representation of the environment, the interpreted model of the environment, and the predicted environment model produced as the result of an expected action with variables different than those of the action performed at the time of the original impression. The predicted environment is a simulated internal model of the environment constructed based on the current goal and the current interpretation of the relevant external occurrence. In this model, constructed memories may not match original impressions since they alter according to when, where, and with what the memory system is cued.

THE VIRTUOUS CIRCLE OF ANALOGY

In this thought experiment, we can now formulate a story to describe how our automaton navigates the world. Even though this story is elucidating, it is circumscribed to the extent that it chronicles the automaton's interactions with the environment, how it approaches the world, and the obstacles it encounters, only in relation to our full-blooded conceptual and linguistically structured beliefs about what the automaton is actually doing. Its limitation is that it models the representational component of this causally conditioned navigation (how the agent's nonconceptual impressions of the environment exhibit an orderly structure of their own) on our *theoretical reasoning*, just as it posits the behavioural component of the automaton's navigation of its universe (how the automaton functions and interacts with the environment to satisfy its goals) in analogy to our *practical reasoning*. In short, we are stuck with talking about 'what this automaton is up to' or 'what the automaton's nonconceptual experience looks like' in analogy to our own discursive (concept-using) apperceptive consciousness. Even at the rudimentary level of the nonconceptual sense impressions of the automaton, the analogical framework is inevitable. We can only talk about them in analogy to our inexorably 'concept-laden' introspection into our inner states.

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 behavioural outputs. But this in itself contributes to the fruitfulness of our story. For in *carefully* extending our conceptual linguistic resources to describe how this agent structures its awareness of the environment and nonconceptually navigates it, the story underlines two crucial points: The first point is that the resources of language are ultimately the only resources available to us (temporally discursive apperceptive intelligences)