temporal view with a model of time-conscious experience that is neither necessarily directional nor dynamic, which neither pictures time as a flux nor takes for granted our experience as advancing from the past toward the future. What are the ramifications of a nondirectional/nondynamic model of time-consciousness for our existing theoretical and practical models within not only the physical sciences but also the social sciences? And more importantly, what can be gained, theoretically and practically, by adopting alternative models of time-consciousness, and specifically an atemporal model?

The third challenge is in fact the continuation of rational agency's struggle for self-conception and self-transformation, for a release from any residue of essentialist attachment to a particular transcendental structure of experience. In adopting a view from nowhen[185]—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 what Hegel calls Absolute Spirit: moving from particular, contingent consciousness to genuine self-consciousness. The view from the space of reason qua Concept (*Begriff*) is not only a view from nowhere—the impersonality of reason—but also a view from *nowhen*—the pure formlessness of time that is expressed by discursive rationality as *a project that takes time.* As we shall see in the following chapters, this time-generality will take the form of Plato's Ideas: Knowledge, Truth, Beauty and above all, the Good. History is an ongoing process of totalization under the aegis of the Concept; and the form of the Concept (not its content) is atemporal, if not timeless. Directionally temporalized history represents the ultimate limitation imposed by intuitions, or the fundamentalist attachment to a locally constituted experience over the form of the Concept and thus that of history. Once we unconditionally cast off this forced limitation step-by-step,

185 'The campaign for a view from nowhen is a campaign for self-improvement, then, and not a misguided attempt to do the impossible, to become something that we can never be. It promises only to enhance *our* understanding of ourselves and our world, and not to make us gods.' Price, *Time's Arrow and Archimedes' Point*, 267.

dismantle this cursed raft plank-by-plank, history transforms into a medium for the expression of time as the formlessness that conditions any possible form. We have tried to understand and make sense of the world in which we acquire a perspective on ourselves, the meaning of mind, and intelligence, by analysing it internally. But maybe the best solution is to go entirely outside of this world and analyse it from a viewpoint that is both possible and fecund with further possibilities. And yet for this transition to the *outside view* from nowhere and nowhen to be concrete, we must grasp it as a circuitous path of arduous task upon task. This adventurous yet demanding *umwege* is the course of the critique of transcendental structures, spanning different methods and frameworks of theory and practice, from science to technology, art, general pedagogy, politics, and so on. It would be pitifully naïve to think that we can liberate intelligence by means of technoscience alone without changing what and how we think about who we are, and correlatively what we think of, in comparison to ourselves, as intelligent. Changing the latter is a question of attaining a veritable self-consciousness of who we are and what we ought to do. And this change comes at a price which is the complete alienation of the human in itself: that we will never settle, we will never mistake anything for our home, for we have come to the understanding that the very vector of alienation—exodus between possible worlds—is actually our home and the source of what is good and satisfying. It would be equally credulous to believe that, in the course of such an enterprise, we will be able to maintain the conception of the human rooted in how we *experience* ourselves here and now. As the *umwegen* to the outside view of ourselves revises the very experience of who we think we are, we become that which no longer experiences itself in terms of what we experience ourselves as in this very moment. For an intelligence with a larger field of experience than ours, what the human means or what the appellation 'we' stands for no longer abides by the terms of our epochal particular field of experience. We become only an antecedent condition to what is *necessarily* us—that is, the *form* of the us, which is neither a transitory content nor an ephemeral particular object. Ultimately, genuine self-consciousness turns out to be the view of ourselves from nowhen. And the world conceived concretely from a view that is of

nowhere and nowhen is a world that at once is replete with possibilities and is possible, one whose possibility is no less actual than the actuality of the inhabitable world we currently inhabit. The actuality of our world is merely an abstractly determined absolute (a sedimented totality that feigns completion of the Absolute), but the actuality of possible worlds conceived from nowhen and nowhere is the Absolute as the concretely determined, never given in what appears to us in time but procured through the cunning plot of history to explore the meanings of time. From a Kantian perspective, in taking up the third of the challenges outlined above, 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 the thinking of the former and the being of the latter.[186]

186 Jay Rosenberg introduces Kant's mutuality thesis as follows: 'The same activities of synthesis which constitute the represented world as an intelligible objective unity constitute the representing self as an apperceptive subjective unity.' Rosenberg, *The Thinking Self*, 6.

5. This I, or We or It, the Thing, Which Speaks (Objectivity and Thought)

ASCENT TO THE INFANTILE

Our toy model automaton is now equipped with a spatially and temporally perspectival awareness. In other words, it has developed the capacity to handle space and time, the Objects (*objekt*) required for the rudimentary organization of an encountered item (an appearance) in the world—that is, an object (*gegenstand*) as distinguished from other items. However, at this point, both *objekt* and *gegenstand* are mere analogical correlates of our objects of thought (*Objekte*) and categorically determined sensible objects (*Gegenstände*). They are neither objective (factual or inter-subjective) nor subjective (in the full-blooded sense of the subject as one who is in a position to make veridical claims or critical judgements, rather than a thin notion of subject as that which, de facto—under the rudimentary transition laws of imagination—is able to discern uncritically empirical associations in the order of appearances). The achievement of subjective and objective thoughts requires that the automaton advance from rudimentary capacities (abilities_1) to advanced abilities (abilities_2). To enable it to do so, we must equip our automaton with a new structure—not a structure that belongs to the automaton itself, but one it is plugged into or bound up with, namely the structure of a community: a multi-agent system such as a framework in which multiple information processing systems are constrained by their dynamic or concurrent interaction with one another (i.e., every system is the environment of the other systems). We must therefore introduce two modifications to our picture. The first modification is simply necessary whereas the second, although in essence necessary, could be introduced in forms other than that depicted in our toy universe:

- The automaton is now a part of a multi-agent system comprised of automata with a differential responsiveness to the items in the world.

In its most basic configuration, the multi-agent system is designed to enable interaction between automata/agents as a way of increasing the probability of goal-attainment. This multi-agent system is then introduced into a specific environment wherein agents have to interact not only with one another, but also with the features of this environment. Inter-agent interaction is, accordingly, coupled with the dynamic inputs and constraints of the resulting ecology.

- The automaton is now furnished with built-in electromechanical devices whose coordination results in the production of quasi-continuous sounds. These quasi-continuous sounds are the primary means of communication between automata. In an alternative toy universe, this feature could be implemented in different ways that might not necessarily involve sound. However, insofar as this is a component of *our* particular toy universe, we have to abide by its characteristics and constraints.

Chapter 2 ended with the automaton endowed with the capacity of meta-awareness, or the causal analogue of inner sense or inner perception—the capacity of the mind to be affected inwardly and passively by its own thought-episodes, to report representations of items and occurrences in the world to the mind as temporally organized rudimentary *re*-presentations which are the mind's own presentations. But the automaton's inner sense was merely de facto. In other words, the automaton wasn't aware of these occurrences *as* its own, nor did it have thoughts in any meaningful sense. Indeed, even if this inner sense was not causal, even if the automaton did actually have thought-episodes at this level, the passive capacity of inner sense could not (pace Descartes) supply the automaton with the synthetic awareness that these inner thought-episodes or meta-awarenesses are *its own*. As we cursorily surveyed in chapter 2, such meta-awareness requires the active power of apperception rather than just the passive faculty of inner sense. Only when the passivity of its inner sense is brought under the apperceptive I as an active logical form (rather than, as in Descartes, a substance) can the automaton have anything resembling thoughts in even the minimal sense.

The apperceptive I is synchronically attached to all instances of representations (I think X, I think Y, I think Z) and diachronically extends over all thoughts (I think $[X + Y + Z]$). But if we are to build this synthetic apperceptive I as a necessary abstract and logical form, we have no choice but to finally depart from Kant's account of the apperceptive self—which Hegel reproaches for being an *empty* transcendental subjectivity—and to instead adopt a resolutely Hegelian approach: the apperceptive self is only a cognitive self in so far as it is part of geist. An individual is only an individual to the extent that it is individuated by social recognition, which is the form of self-consciousness. This logical self is at once one and many—and, as if to prove this, it can only be constructed, as we shall see, by way of a confrontation with another I^*.[187]

To recapitulate, our aim is to elevate the automaton from being an agent that de facto possesses the meta-awarenesses of inner sense to one that is inferentially—synchronically and diachronically—aware of having meta-awarenesses; from representations that are only causally and passively 'its own' to an I—a logical form—that actively accompanies and integrates such representations *as* its presentations. In short, the goal is to bring the automaton to a state where it not only has experiences and thought-episodes, but where such experiences and thoughts are de jure and by entitlement its own. The modifications we have made to the automaton are precisely the kinds of adjustments that will enable the automaton to make this transition.

Since our automaton is now part of a multi-agent system, for the sake of efficiently tracking its course of development we can designate it with a

187 'Are you one self or many selves? [...] I immediately developed the second mechanism of proving my individuality: opposition to someone else's idea.' (V. Savchenko, *Self-discovery* [New York: MacMillan, 1980], ix, 13). From the seventies to the eighties, Soviet science fiction was a hybrid of German Idealism, cybernetics, and artificial general intelligence as a geistig multi-agent system. The most prominent exponents of this current are Vladimir Ivanovich Savchenko (*Self-discovery*) and Mikhail Tikhonovich Emtsev (*World Soul* [New York: MacMillan, 1978]). For a nonfictional take on the philosophical, scientific, and political commitments of this period see F. Mikhailov, *The Riddle of the Self* (Moscow: Progress Publishers, 1976).

proper name: Kanzi.[188] As a part of our multi-agent modification, we will also introduce two other automata and assign them proper names: Sue and Matata. Sue and Matata differ from Kanzi in that they are sapient automata that have mastered the use of language. In a nutshell, they are fully fledged concept-using adult guardians of Kanzi. In this configuration, Sue and Matata recognize and nurture Kanzi as what Rosenberg terms a CHILD (Concept Having Intelligence of Low Degree).[189] As a CHILD, Kanzi also has a universe of its own, an INFANTILE-World (INtelligible Familiar Appearance Naively Taken In Lieu of the External World). Let us denote Kanzi and its adult guardians respectively as $\mathbb{K}$, $\mathbb{S}$, and $\mathbb{M}$.

We now have the necessary resources to follow Rosenberg's Kant-Sellars inspired model of the transition from inner sense meta-awarenesses to thought-episodes—or from an automaton with pure perspectival awareness to a child automaton. $\mathbb{K}$ is endowed with behavioural dispositions. It has memories, interests, and anticipatory models which reflect the repeated impressional invariants of how it has seen$_1$ and encountered the world so far. $\mathbb{K}$ is a creature of habit, i.e., of invariants derived from its rudimentary organized sense-impressions or perspectival encounters with the world. Its representations and reliable responses to the world follow transition principles conditioned by how it has encountered the world so far. In this sense, its transitions from one awareness to another (e.g., from 'I_o be$_1$' be$_1$ to 'I_o be$_1$' be$_2$) are—analogically posited—*ur*-material inferences which are defeasible and formally incomplete. A case of a context-sensitive, defeasible, and formally incomplete material inference would be: *If the match was lit then it must have been rubbed against a frictional surface, but it could also be the case that it was lit because it was hit by an electric spark, or, if the match were in a vacuum, it would light but the stick would not burn*, etc. Like material inferences that open up a non-monotonic space of entailments, these protocol-like transitions also have their non-monotonic entailments: If A in circumstance C_i then B, if A in C_j then D, or if B then A in C_i and not C_j, etc.

188 See S. Savage-Rumbaugh and R. Lewin, *Kanzi: The Ape at the Brink of the Human Mind* (New York: Wiley, 1994).

189 Rosenberg, *The Thinking Self*, 135.

However, the range and complexity of $\mathbb{K}$'s transition protocols are limited—not just because their contexts are always restricted to circumstances that are of immediate sensory-behavioural *interest* to it, but above all because $\mathbb{K}$ has no conceptual resources. In particular, $\mathbb{K}$ does not have anything resembling modal concepts such as nomological relations and counterfactual dependencies, which are both context and resource-sensitive (e.g., '$A \nrightarrow B \nrightarrow A, C \nrightarrow B$', or '$A, d, d \nrightarrow B \nrightarrow A, d \nrightarrow B$', where $\nrightarrow$ is a counterfactual consequent relation).[190] Coming back to the match example, the context and resource-sensitivity of the counterfactual consequent relation between the non-lit match and the lit-match can be expressed as:

$$\Gamma, \textit{MH}, \textit{MS}, \textit{Sulphur}_{\textit{MH}}, \textit{Oxidizing agent}_{\textit{MH}}, \textit{Strike MH against a frictional surface} \nrightarrow_{\{\{wet\}\}} \textit{Lit MH} \otimes \textit{Burning MS}$$

where *MH* is the match head, *MS* the matchstick, the singleton $\{\{wet\}\}$ a control set which constrains the soundness of the sequent formula and the tensor product, and the compounding operator $\otimes$ signifies the combination of a lit match head and the burning matchstick *provided that the context* Γ *does not contain* $\{\{wet\}\}$. The context Γ is essentially a descriptive context in the sense that it includes the description of what a match is: empirical judgments about the properties and nomological relations of what can count or behave as a match.

---

190 This is of course the suspension of the structural rules in classical logic, monotonicity and idempotency of entailment, as reflected in Jean-Yves Girard's Linear Logic (LL) (see below). Formulas are treated as resources that cannot be used or reused under every condition. The LL operator that expresses this consumption of resources is linear implication: $A \multimap B$ (reads 'A lollipop B'), consuming A yields B. Once A (e.g. a resource, a belief or a piece of knowledge) is used, it cannot be unconditionally reused in the computation. The reusability of a resource is symbolized by the operator ! (reads 'of course'). !α means the *ability* to do α repeatedly. An object of type α is stored in such a way that it can be repeatedly accessed in a computation.

If we add additional premises (qua resources), subtract existing premises, or change the control set (and hence the context), the hypotheses of the formula may change:

$$\Gamma, MH, MS, Sulphur_{MH}, Oxidizing\ agent_{MH}, Strike\ MH\ against\ a\ frictional\ surface, Vacuum \overset{}{\rightarrowtail}_{\{\{wet\}\}} Lit\ MH \otimes Not\ Burning\ MS$$

In the latter formula, remove the premise *Oxidizing agent*$_{MH}$ and the consequent will change to *Not Lit MH* ⊗ *Not Burning MS*. The introduction of this kind of defeasible nonformal inference is precisely what makes it possible to incorporate the unanticipated or contingent into the sphere of reason, in the form of a consequent that is not fixed, but is revisable given any change in its antecedents or context.[191] Now add a premise with a different context (e.g., *short matchstick*$_{\Lambda}$ of the descriptive context Λ pertaining to the size of the match, i.e., a match with a *different shape*). The addition of this premise does not change the consequent. We can, therefore, say that a good material inference (in this case, about lighting a match) exhibits and is defined by a range of counterfactual *robustness* (i.e., a dry matchstick whose head is made of a combination of sulphur and an oxidizing agent

191 It is often objected that reason is too rigid or fixed to leave any room for the dynamics of risk and contingency, and that rationality survives by means of a fundamental risk-aversion. This, however, is true only if reason is caricatured as consisting merely of classical logic or some arcane traditional version of epistemic logic. Other than the fact that both theoretical and practical reason incorporate the contingent and the unanticipated into their structure in order to afford new understandings and actions, the armamentarium of rationality is replete with modes of inference that are counterfactual, defeasible, non-monotonic, paracomplete, or paraconsistent. Such features allow the identification, assessment, and action of different types of risk or of the unanticipated without indiscriminately grouping them into an ineffable and exorbitantly ontological hyperchaos. In the absence of epistemic rationality, contingency and risk are always susceptible to radical ontologization in such a way that they become inevitably foreclosed to investigation. Lacking epistemic adequacy for the discrimination of different types and levels of contingency, the ontologization of risk turns into a blind faith in the radical powers of contingency in effect no different from a religious faith in an omnipotent god.

would still burn if the size of the matchstick were different and if it were ignited in the earth's atmosphere; but if it were lit in a vacuum, only the matchhead would burn). And here the counterfactual robustness depends on the descriptive context Γ (in relation to the control set $\{\{S\}\}$—which can only be obtained through modal vocabularies required for *explaining* what 'counterfactually robust' properties and lawlike relations x ought to have in order for it to be described as a match).[192]

The control set $\{\{S\}\}$ can be generally defined as a finite set of finite multisets of context formulas $\{\Gamma_1, \Gamma_2, \ldots \Gamma_n\}$ such that for all $1 < i < n$, $\Gamma \subset \Delta_{\otimes}$, where Δ is a set of precontexts. Correspondingly, the context Γ can be defined in terms of Δ and under the condition that if $\Pi_1, \ldots, \Pi_n$ are precontexts then $(\Pi_1, \ldots, \Pi_n)$ is a precontext. The context Γ is an ordered pair $\langle \Pi, f \rangle$ where f is a function assigning a control set to each node of the context tree Γ and not its branch nodes.[193]

Lacking alethic modalities of possibility and necessity, subjunctive and counterfactual conditionals which codify causal relations,[194] the transition protocols of the CHILD Kanzi's impressions of the world are much like the philosophical world of a fanatical empiricist who thinks he has stumbled upon some base empirical vocabulary with which he can describe and explain the furniture of the world, not knowing that the transition from

192 'It is only because the expressions in terms of which we describe objects, even such basic expressions as words for the perceptible characteristics of molar objects locate these objects in a *space of implications*, that they describe at all, rather than merely label. The descriptive and the explanatory resources of language advance hand in hand; and to abandon the search for explanation is to abandon the attempt to improve language, period.' W. Sellars, 'Counterfactuals, Dispositions, and the Causal Modalities', in *Minnesota Studies in the Philosophy of Science* vol.2 (Minneapolis: University of Minnesota Press, 1957), 306.

193 For more details on the formalism of context-sensitivity, see M. D'Agostino, 'How To Go Non-Monotonic Through Context-Sensitiveness', *Logic and Philosophy of Science* 8:1 (2015), 3–27.

194 For a disquisition on counterfactuals and causal relations, see D. Lewis, *Counterfactuals* (London: Wiley-Blackwell, 2001).

explanandum to explanans is always entangled within a modal web of relations and implications, and is inconceivable outside of this web.

In Brandom's words,

> Just how—they would want to know—did what seemed most urgently in need of philosophical explanation and defense suddenly become transformed so as to be unproblematically available to explain other puzzling phenomena? Surely such a major transformation of explanandum into explanans could not be the result merely of a change of fashion, the onset of amnesia, or the accumulation of fatigue? But if not, what secret did we find out, what new understanding did we achieve, to justify this change of philosophical attitude and practice?[195]

Lacking modal vocabularies, the extremist empiricist $\mathbb{K}$'s entrenched statistical regularities are limited to transition protocols (*ur-* or proto-inferences) such as

> 'Each time $\mathbb{K}$ sees_1 a protruding mass of fuzzy grey (G_1) *contacting* a heap of black (B_c), it hears_1 a shrieking noise (N).': $\mathbb{K}(G_1, B_c, \rightarrow N)$.

And since these transitions have their own corresponding basic *precluding* or *inhibitory* transitions, $\mathbb{K}$ also has conditioned preclusions or transition obstructions:

> 'The mass of fuzzy grey retreated *behind* the heap of black (B_b). The mass of fuzzy grey did not cease to exist nor is it in *front* of the heap of black (B_f).': $\mathbb{K}(G_1, B_b, \nrightarrow \textit{Not}\ G_1, B_f)$.

However, $\mathbb{K}$ does not have available the transitions or obstructions modally encoded or embedded in counterfactual situations such as the following:

---

195 R. Brandom, *Between Saying and Doing* (Oxford: Oxford University Press, 2008), 93.

'**If** the mass of fuzzy grey **were not** coming into contact with the heap of black and another mass of fuzzy grey (G_2) **were** appearing, there **would be** a shrieking noise.': $\mathbb{K}(G_1, G_2, Not\ B_c \nrightarrow N)$.

Nor does $\mathbb{K}$ have anything like an account of causation since, once again, it has no grip on the modal relations that codify causal ones. It has 'Contact with the heap of black → Shrieking noise from the direction of the mass of fuzzy grey', but not 'Contact with the heap of black $\overset{caused}{\nrightarrow}$ mass of fuzzy grey to make shrieking noise'.

Aside from having a small universe of transitions and precluded transitions between its dispositional awarenesses, $\mathbb{K}$ also naively takes its meta-awarenesses (or meta-representings) as evidence of a corresponding state of the world or of an item. Put simply, $\mathbb{K}$'s impressions of the world are naïve because it takes conditioned transitions between its awarenesses to stand in a *one-to-one* relation with transitions in the states of the world. It is thus predisposed to take everything at face value, to presumptuously infer from its world-awarenesses

I_o be_1, b_2, b_3 (or, I_o was, is, or will be)

their corresponding constituting awarenesses qua meta-awareness/metarepresentings

'I_o be_1 / b_2 / b_3' be_1, be_2, be_3 (or, 'I_o was/is/will be' was, is, or will be)

—and vice versa, to infer from each constituting awareness a corresponding world-awareness.

Accordingly, as well as being a naive empiricist, Kanzi also happens to be a naive idealist, taking its familiar constituting awarenesses for what is actually going on in the external world. Therefore, $\mathbb{K}$'s INFANTILE-World is susceptible to Cartesian scepticism to the extent that, if everything concerning the external world can be inferred from apparently immediate constituting awarenesses (or occurrent thought-episodes)—i.e., what is going on in Kanzi's head—then $\mathbb{K}$ might also infer that there are no items

of the external world, and indeed no external world at all. In fact, $\mathbb{K}$ with its INFANTILE-World is properly speaking a child of Descartes's universe of mind where not only is the mind mistakenly regarded as a *tabula rasa*, but the immediacy of meta-awarenesses and their one-to-one correspondence with world-awarenesses leave much room for scepticism about the existence of an external world.

In this scenario, $\mathbb{K}$'s 'I_o be_1 / b_2 / b_3' be_1, be_2, be_3 are—from our perspective—the analogical counterparts of Kanzi's own occurrent thought-episodes, that is (* I think I_o be_1, b_2, b_3 *). In other words, $\mathbb{K}$'s de facto 'I_o be_1' be_1 is analogically $\mathbb{K}$'s own mentioned or quoted I-thought, i.e., $\mathbb{K}$(* I think I_o be_1 *). Similarly, 'I_o be_1' be_2 is analogically $\mathbb{K}$(* I think I_o be_2 *), as in $\mathbb{K}$(* There is a mass of fuzzy grey approaching *), and so on. These instances of $\mathbb{K}$(*...*) are nothing but the occurrences of a nonsubstantive apperceptive I that synchronically accompanies all of its distinct thought-episodes: I think *X*, I think *Y*, I think *Z*.

But since, as discussed in chapter 3, $\mathbb{K}$'s meta-awarenesses are only meta-awarenesses in so far as they are part of a web of *equivalence* relationships extended through time, the nonsubstantive I that synchronically accompanies each meta-awareness is formally identical or equivalent to the nonsubstantive I that diachronically accompanies *all* combinations of meta-awarenesses. This I that thinks *X*, thinks *Y*, thinks *Z* is the I that thinks [*X*+*Y*+*Z*]—and without whose *logical form* there would be no thought-episodes and no experience. In other words, the diachronic and apperceptive I (I think [*X*+*Y*+*Z*]) is the sufficient condition for the possibility of having thought-episodes or synchronic Is (I think *X*, I think *Y*, ...). The network of equivalence relationships between meta-awarenesses and their corresponding world-awarenesses is what Kant identifies as the *manifold* of given presentations. The diachronic I is the formal unity of consciousness in which thoughts of *X*, *Y*, and *Z* are combined (the I who thinks [*X*, *Y*, *Z*] = the I that thinks *X* + the I that thinks *Y* + the I that thinks *Z*).[196]

196 Think of a simple example: the person who ate a hot dog is the person who ate the bun, the sausage and the ketchup sauce. The *I*s which are synchronically attached to the acts of eating these ingredients are the diachronic apperceptive I. However,

But as we saw in the first chapter, this formal unity of consciousness (rather than consciousness *per se*) is precisely what is afforded by the movement of self-consciousness, the $I \rightleftharpoons I$, or more accurately $Id_{map\text{–}I} \rightleftharpoons Id_{map\text{–}I^*}$.

It is only in virtue of the recognitive space of language that there are apperceptive cognitions. And correspondingly, it is only through the movement of self-consciousness, which extends from one consciousness to another through the public semantic space of language, that there is a diachronic unity of consciousness and a synthetic unity of apperception. The I that thinks is the encapsulation of this formal movement in which ordinary consciousness is caught up. The apperceptive I is neither an empirical self nor a phenomenal self-model nor a field of awareness, but a necessary *formal* condition brought about by a recognitive movement through the space of language. Anything and anyone who fulfils this formal condition is not only a person but also the bearer of thoughts.

---

for us to posit such a connection between the I that ate the hot dog and the different instances of I attached to the gustatory acts of eating ingredients, we cannot resort—à la Hume—to empirical evidence (e.g., the person who ate the bun and was of such and such empirical characteristics is also the same person who ate the dog and ketchup by virtue of having the same characteristics). Nor can we conjecture à la Descartes that the instances of the I that ate the ingredients are the same as the I that ate the hot dog in virtue of the substantive persistence of the I over time. Kant's critique shows that such a connection requires something more: a time-conscious judging subject or the analytic unity of apperception which brings about the possibility of combining objects of different acts of thinking or in this case eating into one single complex. However, this analytic unity is itself dependent on a synthetic unity of apperception since the synthesis or integration of different intuitive representations under one concept, or several concepts under one higher integrative concept already implies the priority of the synthetic unity of apperception: Only in so far as I can synthetically combine the manifold of sensible intuition and be conscious of the unity of this act of synthesis can I also analyze this manifold (dog, ketchup, bun) into different concepts and a more universal concept (the hot dog) in one and the same critical consciousness.

INFANT AGI

So far we have discussed what the world of Kanzi could be like had Kanzi reached the status of a child; but it hasn't yet attained this goal. We shall follow Rosenberg's path in order to turn this into reality, while at the same time making necessary modifications in tandem with the Hegelian revision of Kant's transcendental subjectivity. To achieve this objective, necessary modifications have been made: the introduction of $\mathbb{S}$ and $\mathbb{M}$ as language-using adult guardians.

For now, $\mathbb{K}$ has de facto rudimentary transitions between its constituting awarenesses. In other words, it has an analogical counterpart of the ur-inferences between the orders of before and after (e.g., what was encountered and what is now being encountered). In order for $\mathbb{K}$ to distinguish the orders of before and after and to incorporate them into a growing space of implications regarding one and the same world, it must integrate its meta-awarenesses under one formal synthetic unity, one and the same I. To satisfy this necessary condition, $\mathbb{K}$ must model its private meta-awarenesses on a public and deprivatized language. In other words, it must interact with $\mathbb{S}$ and $\mathbb{M}$ as two fully fledged linguistically competent agents.

At this point, however, for the prelinguistic $\mathbb{K}$, its fellow automata $\mathbb{S}$ and $\mathbb{M}$ are not speaking subjects, they are items of the world that not only satisfy its dispositions, but also reward and cultivate its new developments, its moves against its dispositional routines. As such $\mathbb{S}$ and $\mathbb{M}$ are not just any items, but especially salient ones. Their linguistic utterances (from $\mathbb{K}$'s perspective, their engaging and exciting noises) are also of the utmost interest, *as if* they conveyed something important.

In this scenario we can talk about linguistic interaction between the prelinguistic $\mathbb{K}$ and its linguistic guardians as communication, but with the proviso that, were $\mathbb{K}$ also a language-user, we could *not* use the term 'communication', for linguistic interaction is precisely *not* communication.

Within this communicative regime, the noises qua utterances that $\mathbb{S}$ and $\mathbb{M}$ make are for Kanzi—analogically speaking—instances of saying something, conveying something important that may be false or true. But at this time, $\mathbb{K}$ takes everything at face value; it believes everything it is told,

so to speak. For $\mathbb{K}$ to understand or, more accurately, to hear_2 these noises, it would have to be aware of them as representations (of importance) and not just simply as noises, since these noises are precisely—at least at this stage—in contiguity with its dispositional interests. More accurately, $\mathbb{K}$ hears such noises as representations because they are in contiguity with—or rather, because they fit—its dispositional or conditioned world-representings.

Accordingly, for $\mathbb{K}$ to hear_2 the communicated noises *as* representations would be for it to make sense of or to understand the *functional role* they play in its world-representings and the rudimentary conditioned transitions between them.[197] This functional role is the meaning (semantic value) of these noises or utterances which are mapped onto $\mathbb{K}$'s meta-awarenesses or awarenesses of awarenesses of items in the world. To put it simply, for the conditioned Kanzi, Sue and Matata's exciting noises communicate something of importance to the extent that they are in continuity with the dispositional transitions between its meta-awarenesses of interesting items in the world. In other words, $\mathbb{K}$'s world is inductively biased, first and foremost, by the items of interest that satisfy its behavioural regularities and dispositions. To the extent that $\mathbb{K}$'s transition protocols are inductively biased in such a manner, $\mathbb{K}$ also tends to recognize and respond to any communicative noise or representation that can be incorporated into its inductively biased world. And by virtue of this space of shared recognition between Kanzi and its adult guardians, all of $\mathbb{K}$'s awarenesses, meta-awarenesses, and their corresponding transitions (what leads to or precludes what) are now also implicated in the expanding web of $\mathbb{S}$ and $\mathbb{M}$'s representations—that is, the sayings and doings of its constantly appraising language-using guardians.

Imagine again the example drawn from *2001: A Space Odyssey*: $\mathbb{K}$ sees_1 a mass of fuzzy grey moving from the right of the heap of black at time t_1 to the left of the heap of black at time t_3. At time t_2, the mass of fuzzy grey $\text{disappears behind}_1$ the heap of black.

In our multi-agent scenario, Sue and Matata are to the left and right of the heap of black (i.e., the monolith). $\mathbb{S}$ and $\mathbb{M}$ tell $\mathbb{K}$ that the fuzzy item (the monkey) is moving from the right of the monolith to its left and, some

197 See W. Sellars, 'Meaning as Functional Classification', in *In the Space of Reasons*, 81–100.

time between these two occurrences, the monkey disappears behind the monolith. To $\mathbb{K}$, $\mathbb{S}$ and $\mathbb{M}$'s reports look like this:

At t_1, the fuzzy item moved from the right of the heap of black to its left: Item t_1.

At t_2, the fuzzy item moved behind the heap of black: Item t_2.

At t_3, the fuzzy item is to the left of the heap of black: Item t_3.

As a creature of dispositional regularities and interests, $\mathbb{K}$ is able to recognize these reports and, additionally, to map them to its own de facto meta-awarenesses. Consequently, $\mathbb{K}$ now acquires labelled meta-awarenesses. They are labelled because they have been received by $\mathbb{S}$ and $\mathbb{M}$ as contrastive reports mappable to $\mathbb{K}$'s meta-awarenesses or, more accurately, as reports that play a functional role in its awareness$_1$ of the world. Accordingly, in addition to its as yet unlabelled meta-awarenesses, the prelinguistic $\mathbb{K}$ has a family of labelled meta-awarenesses:

**the fuzzy item moved from the right of the heap of black to its left* * $_{\mathbb{S},\mathbb{M}}$ is: **Item t_1* * $_{\mathbb{S},\mathbb{M}}$ *be.*

* *the fuzzy item moved behind the heap of black* * $_{\mathbb{S},\mathbb{M}}$ is: **Item t_2* * $_{\mathbb{S},\mathbb{M}}$ *be.*

* *the fuzzy item is to the left of the heap of black* * $_{\mathbb{S},\mathbb{M}}$ is: **Item t_3* * $_{\mathbb{S},\mathbb{M}}$ *be.*

These reports and their matching labelled meta-awarenesses correspondingly activate the transition or obstruction protocols between $\mathbb{K}$'s meta-awarenesses:

* *the fuzzy item moved behind the heap of black* * is: $\xrightarrow{\mathbb{S},\mathbb{M}}$

* *the fuzzy item did not cease to exist* * *and*

* *the fuzzy item is not in front of the heap of black* *.

But they also create a family of mutually precluding perspectives:

t_1: **the fuzzy item moved from right to left**

t_2: **the fuzzy item moved behind the heap of black**

$*item\ t_1* \not\equiv *item\ t_1*$

As a result, $\mathbb{K}$'s pure perspectival world-representings are now in tension. In fact, with the introduction of the labelled meta-awarenesses, $\mathbb{K}$ has transitioned into an automaton that is no longer pure and one-dimensionally perspectival. Its world is proto-inferentially multiperspectival.

To clarify, currently, in addition to its de facto world-representings and the dispositional transitions or obstructions between them, Kanzi has labelled meta-awarenesses generated based on reports received from Sue and Matata. Even though these labelled meta-awarenesses are still consistent with its perspectival world-representings, their transitions and obstructions are not seamlessly consistent with its rudimentary perspectival transitions, i.e., with what Kanzi $sees_1$ of the world. The labelled meta-awarenesses preclude the corresponding element of Kanzi's perspectival transitions: e.g., in moving from right to left, the fuzzy item does not fall out of the world but is simply occluded behind the monolith, or 'the fuzzy item to the left of the heap of black at t_3' precludes 'the fuzzy item moved behind the heap of black at t_2'.

The introduction of labelled world-representings into Kanzi's world-representings creates a source of tension or multiperspectival disturbance with which, finally, the pure perspectival world of Kanzi comes to an end, since from now on Kanzi's seamless world (of awarenesses and awarenesses of awarenesses) is continually being decohered and recohered by its guardians' perspectives. These inferential reports at once obstruct or preclude some of Kanzi's perspectival transitions and, on the other hand, facilitate a new group of transitions (proto-inferences) between awarenesses or meta-awarenesses which either lay outside of its dispositional interests or were previously absent or implicit from its pure perspectival worldview.

To eliminate this tension and recohere its de facto decohered worldview, 𝕂 has no recourse other than to no longer take the reports by 𝕊 and 𝕄 at face value. In fact, it has to dispense with 𝕊 and 𝕄's reports, to stop inferring$_1$ world-representings from labelled meta-representations. In Rosenberg's terms, 𝕂 is now disposed to trust only the evidence of its senses. But in this elementarily recohered world, 𝕂 also comes to possess the ability to distinguish between seeming and being—even if, at this stage, for 𝕂 'being' is what perspectivally seems to it, and 'seeming' is how things seem to 𝕊 and 𝕄. Nevertheless, it is precisely this distinction between seeming and being that germinates veridical normative judgements later on. 𝕂's own *SEEMINGS* are the seeds of thoughts: *SEEMS* 𝕂 (∗ *there is a mass of fuzzy grey* ∗) is in reality 𝕂(*thinking there is a mass of fuzzy grey*), a thought-episode corresponding to a perceptual experience. In other words, 𝕂 now has the ability to remove the quotation marks (∗...∗) from its meta-awarenesses and to take ownership of its awarenesses of items, as its own rudimentary thoughts qua uncritical perceptual experiences.

However, 𝕂's SEEMINGS qua thoughts are not objective: they are not beliefs, and as such enjoy no epistemic status. Even though these inner thoughts are modelled on a public language (𝕂's communication with the language-using 𝕊 and 𝕄), they lack propositional attitudes. Only when they are linguistically asserted, i.e., committed to as beliefs (with all the requirements that such an endorsement or commitment entails) will they possess an epistemic status subject to assessment and revision. Even if 𝕂 could assert out loud SEEMS (*There is fuzzy item over there*) as in 'It seems there is a fuzzy item over there', that asserted seeming would only be an instance of 'experiencing-out-loud'.[198] As thought-episodes combined with sensations, seemings only attain epistemic status when the speaker can commit to the correctness of their content by way of saying or judging that 'There is a fuzzy item over there'. Of course, there needs to be an experiential readiness—a perceptual experience qua thought-episodes-cum-sensations—in order for 𝕂 to have perceptual judgments and to be capable of justifying the assertion

198 J. Rosenberg, *Thinking About Knowing* (Oxford: Oxford University Press, 2002), 87.