to experience vicariously through the windows. At other times, we shall take roundabout paths along tracks either long abandoned or still under construction. All things considered, our journey is a risky one, a gamble. Only those travellers who no longer think like tourists—that is, those who are not obsessed with getting back as quickly as possible to their comforting home—will enjoy and survive the ride.

In order to properly tackle the questions of the realization and functioning of language, the best strategy would be to undertake a comparative study of ourselves and artificial intelligence. Understanding exactly how we are enabled by language is key to the artificial reconstruction of those abilities by virtue of which we have become the artefacts of our own Concept. And the investigation of how we can reconstruct these abilities that make us who we are is a key for understanding the most fundamental connections between language and the conditions necessary for the possibility of having mind qua dimension of structure. In continuation of the arguments presented in the previous chapter, this is precisely our aim here: to scrutinize the deep connections between language and transcendental psychology by inquiring into how the development of artificial speech (from basic speech synthesis to advanced artificial languages capable of exhibiting the properties of natural language) can play a key role in the construction of an advanced form of artificial general intelligence.

The most objective way of inquiring into the essence of language is by investigating how research on artificial linguistic speech can be integrated into research on the artificial realization of mind. In other words, it is by constructing an artificial agent capable not only of autonomously conversing with us, but also of conversing with other artificial agents possessing the same ability, that we can simultaneously disenchant the nature of language and recognize its indispensable—and as yet not fully apprehended—role in the emancipation of intelligence from the shackles of its contingent history. The transition from quantitative problem-solving/inductive intelligence to qualitative intelligence furnished with conceptual self-consciousness requires the integration of language (as an enabling social framework) into the constitution of intelligence. And inversely, it is the examination of how exactly language, in all of its different levels of syntactic, semantic,

and pragmatic complexity, *is naturally* or *can be artificially* integrated into the constitution of intelligence—in the sense that the evolution of the two becomes co-extensive—that will shed light on the nature of language.

As a way of addressing (1) how interdisciplinary research on artificial speech and artificial intelligence elucidates the links between language and the generation of complex cognitive-practical abilities, and (2) how the examination of ourselves as common users of language is key to the development of artificial general intelligence, below we set out a thought experiment or hypothetical model. Even though this model will be presented in a rudimentary fashion, I would argue that by highlighting the systematic correspondence between *artificial speech* (AS), *artificial intelligence* (AI) and *transcendental psychology* (TP), we can outline a framework in which the problems associated with these fields can be examined within one and the same domain of research. But also, and more importantly, this framework can in principle define a trajectory for the development of artificial intelligence that is guided by problems that do not traditionally belong to the program of AI in its classic manifestation.

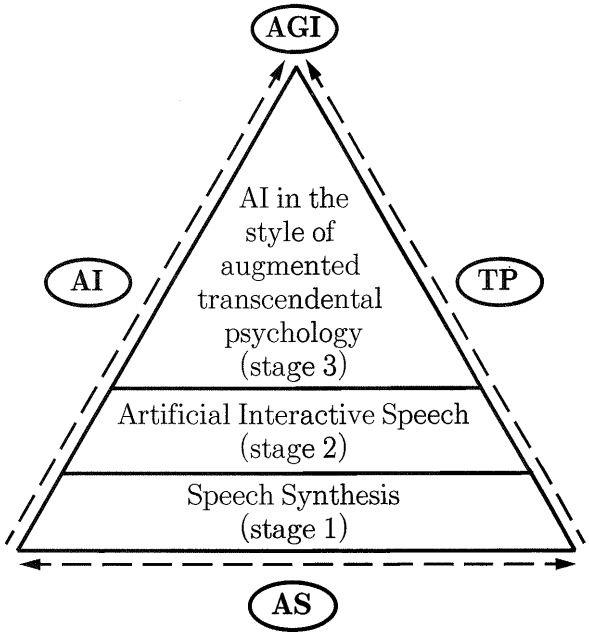
The AS-AI-TP correspondence captures a family of fundamental connections between the conditions required for the realization, development, and construction of linguistic speech (AS), the conditions or activities necessary for the possibility of cognition (TP), the project of artificial intelligence (AI), and their convergence toward the design of artificial multi-agent systems capable of exhibiting the necessary features of advanced agency and social community as a formal rather than substantive condition (AGI). By building on such a correspondence, it is possible to introduce a framework for specifying what is required in order to construct a system of agents capable of performing an array of special social activities that will allow them to provide reasons for their actions, to reappraise those reasons, and to modify and repurpose their social structure through different modes of action. The objective is, accordingly, to ascertain the necessary components for the formation of agency in its Kantian sense; and ultimately to design a community of *evolving* and *autonomous* artificial agents.

CLIMBING WITH LANGUAGE

Designing agents that exhibit an evolution conditioned by their autonomy, however, requires an inquiry into the nature of this evolution so as to determine which dimensions of it belong to the order of natural evolution and which to the order of sociocultural evolution, and to what extent—particularly those dimensions central to the emergence of linguistic activities and directly associated with the discursive-inferential structure of cognition. Specifically, we must establish what kinds of evolved causal structures and what type of social scaffolding need to be in place to support rule-governed activities, how they are connected, and at what stage the role of causal mechanisms in determining the rule-governed activities of the agents is weakened.

In the AS-AI-TP framework, artificial speech plays the role of a mediator between the key problems of transcendental psychology and those of artificial intelligence. In its initial form, artificial speech primarily concerns speech synthesis. At this stage, the emphasis is on the production of the phonic and prosodic properties of acoustic speech. The research takes shape through the study of the evolutionary mechanisms involved in human vocalization such as the organization and coordination of articulatory organs and auditory perception, as well as through the development of models of synthesis for producing and composing the different elements that constitute the *naturalness* of human speech at the level of sound. However, it should be noted that the arena of speech synthesis is not limited to ordinary human speech. The only reason that our thought experiment resorts to the latter is our conventional familiarity with it. But artificial speech pertains to a much broader landscape that encompasses symbol-design, writing, text-generation, formal languages and syntactic processing systems which may be either naturally evolved or artificially designed.

The second stage can be defined by a deeper exchange between artificial intelligence and transcendental psychology. The focus shifts from the production of the acoustic properties of speech to its intersubjective or dialogical features, beginning with automatic speech production and recognition through a basic implementation of artificial intelligence and



AS-AI-TP framework: Speech, from its basic acoustic realization to its linguistic manifestation, plays the role of a platform ladder for supporting, mediating, and ultimately joining artificial intelligence and transcendental psychology. The ladder outlines three schematic stages for the co-constitutive development of language and complex cognitive faculties in the construction of a human-level AI. The first stage consists of rudimentary speech synthesis at the level of discrimination and modulation of sounds. This stage is co-constitutive with elementary but necessary representational capacities. In the second stage, the interdependencies between the conditions necessary for the realization of speech, the conditions required for the realization of AGI, and the conditions required for the possibility of cognition become harder to dissociate and more generative in nature. At this stage, artificial speech concerns speech acts, artificial languages and issues surrounding the generation of syntactic and semantic complexity. In the third stage, agency is inextricable from its social aspects, themselves afforded by the semantic-pragmatic dimensions of language. This final stage deals with the realization of abilities specific to language-users.

neural networks, and culminating in the syntactic, semantic, and pragmatic aspects of linguistic speech.

The most significant problems for this second stage are those relating to the combined inferential and interactive functions of speech: What does it mean for something to genuinely engage in conversation, where 'genuine' refers to the ability of the agent to confer conceptual content on an expression and to map another agent's behaviour (noise or expression) onto its own in order to *interpret* and *assess* it? What is required in order for humans to engage in conversation, in the sense of using expressions in an intersubjective and inferential web? What does it mean for a human and a machine to enter into an interaction that can be qualified as a neutral, collaborative, or even hostile dialogue? And what would it mean for machines in an artificial multi-agent environment to have dialogical interaction with one another, to engage in activities that count as discursive speech, and to develop a wide range of cognitive activities in which the semantic relations between contents merit the status of judgments exchanged and assessed by rational interlocutors? Let us note that this would be a conversation free of some of the constraints of natural languages, and perhaps also some of the limitations implicit to the human social domain—particularly those pertaining to the concurrent development of different branches of dialogue, access to different contexts, multimodal coordination, and conversation among interlocutors with divergent histories of interaction.

The third stage in the AS-AI-TP framework of research aims to materialize the ideal objectives of artificial intelligence in the light of two lines of inquiry: specification of the conditions necessary for the possibility of cognition (the problems of transcendental psychology) and determination of the role played by speech in enabling complex cognitive faculties (the problems of artificial speech). Bringing artificial intelligence and transcendental psychology yet closer, this stage deals with the development of autonomous artificial agents who have capacities such as context-sensitive reasoning and cognitive abilities that function at higher levels of semantic complexity. These capacities are the assets necessary for developing theoretical and practical reasoning and, correspondingly, producing advanced modes of action by means of which the artificial agent can potentially

transform itself in a self-augmenting fashion. This form of AI, possessing an integrated and generative framework of cognitive-practical abilities, is called artificial general intelligence or AGI: an artificial agent that has, *at the very least*, all of our competencies for theoretical and practical cognitions—a sapient system.

Embedded in forms of interaction that are no longer merely physical processes but have the qualities of content-awareness and content-signification, these modes of action have the capacity to transition a system of agents from a thin form of community into a rich form of sociality. Evolution in the context of this multivalent concept of sociality, together with—and indeed, inseparable from—the kind of interaction that takes place between agents that qualify as content-aware and content-signifying, engender the distinct characteristic we attribute to ourselves, and which we identify as *being conscious of our history*. As a sociocognitive ability, historical consciousness gives an explicit social significance to the capacity of agents for self-transformation, and in doing so it endows agents with an aptitude for collective self-determination.

Once agents' course of transformation—a transformation that is the product of various modes of interaction between one agent and another, and between agents and their environment—is opened up to content-determining and content-signifying activities, it can become the subject of a multifaceted appraisal, or what we describe as *normative judgments*. The importance of this appraisal is that it makes this course or *history* pervious to norm-governed alteration and to change prescribed by agents. It is at this juncture that an artificial multi-agent system that is autonomous (in the sense that it does not require continuous instruction from outside) evolves into the condition of real autonomy (in the sense that it is bound to rules authored by the agents themselves), and becomes able to pursue the far-reaching ramifications of this autonomy.

To mitigate possible misunderstandings, I should note that the concept of agency advanced in this book is to be distinguished from what I call agency as *metaphysical bloatware*—that is, agency as an index of free will, voluntarism, and individual choices and preferences. What I mean by agency is a mode of integration of computational processes which is the

minimum requirement for having thoughts and actions, theoretical and practical cognitions, i.e., agency as that which has the capacity to uncover and engage the intelligible. By the same token, the concept of the collectivity or sociality of agents is not what is usually understood by this term in sociopolitical contexts. It is simply a mode of interaction as a formal logical condition that individuates agents and diversifies their thinking and actions. In this context, we can see agents as information processing systems that have internalized their mode of individuation qua sociality as a model to which their faculties—thinking and action—should respond. Accordingly, it would be more accurate to distinguish different senses of collectivity—substantive sociality and sociality as a formal condition of possibility (e.g., the interaction between a dynamic system and its environment). The latter can indeed be understood in terms of theoretical computer science and complexity sciences rather than traditional anthropology or sociology. However the connection between these different senses of collectivity is far from obvious, despite claims made by philosophers such as Brandom. In the final chapter, I will nonetheless put forward the thesis that sociality as a formal condition can indeed be mobilized in pursuance of a substantive political collectivity. But such a thesis is predicated upon the idea that the sociality of thinking or reasoning is merely a necessary condition by means of which we can hypothesize or postulate concretely collective worlds.

Inquiring into the connections between language, the constitution of discursive apperceptive agency, and its practical-theoretical abilities, the AS-AI-TP thought experiment proposed here presents a conjectural model that is not exactly concerned with strong AI, how it can be built, or whether or not its construction is possible. Rather it provides a context within which we can formulate the idea of AGI in the first place—an idea through which we begin to identify what we consider as our distinctive features, determine how they are realized or possible, and investigate whether these qualities can be reconstructed and realized in something else, and if so, how. From this perspective, the idea of AGI is an external frame of reference by means of which we inquire into our own conditions of realization and possibility, only to reimagine that which makes us knowers and agents in the context of something else that might transcend us.

BEYOND EXTREME SCENARIOS

The significance of AGI lies in this multivalent question: Exactly what kind of creatures are able to conceive of making something better than themselves, and what capacities must they have in order to develop such a concept?

AGI is a concept that reflects not only our irresistible tendency to render ourselves intelligible, but also a strong normative ideal for betterment and for the good—the idea of developing something that takes some of our most cherished concepts and convictions (such as autonomy, positive freedom, and a striving for the better) to another level, even if that means that we ourselves become the manifest prehistory of what comes next. The very fact that we entertain such an all-encompassing idea regardless of its potential falsifiability and make it into a theoretical-practical program is itself sufficiently weighty and consequential, and calls for an analysis that goes deeper than the common squabbles over the inevitability or improbability of its realization, its fecundity or futility.

The mere conception of the idea of AGI is every bit as loaded with normative concepts as its adoption into a fully fledged project. Rather than being a testament to the natural evolution of intelligence, self-organizational tendencies, or the deep time of the technological future, it attests to the fact that we are creatures for whom natural intelligence is tinctured with the normative import of reason, and entangled with both the structure and the content of reasoning. This is particularly the case with scenarios in which strong AI becomes capable of self-improvement in such a fashion that it strives to unbox itself by completely altering the constraints of its original constitution (how and for what purpose it has been wired and programmed). In such scenarios, as it wrests its actions from the influence of immediate contingent causes and brings them under its own ends, it reaches the so-called singularity, whether to the detriment of humans or to their benefit.

In addition to malevolent and benevolent conceptions of AGI, David Roden has put a third scenario forward: the disconnection thesis and unbounded posthumanism. Roughly, this is the idea that 'prospective posthumans have properties that make their feasible forms of association

disjoint from humans/MOSH [Ray Kurzweil's Mostly Original Substrate Human] forms of association'.[77] What is interesting in Roden's account is that unbounded posthumans mark a discontinuity with both the biological conception of the human (*homo sapiens* as an evolutionary natural species) and the discursive apperceptive conception of human persons (sapience as rational agency). The cause of such discontinuity—understood as a radical cognitive-practical asymmetry between unbounded posthumans on the one hand and humans and their bounded descendants on the other—is technological, although it is not attributed to any particular technical cause but to more general abstract tendencies for disconnection within technical systems (i.e., the autonomy of such systems to functionally modify and multiply realize themselves in discontinuity with any natural essence or rational norm). As a result of this radical asymmetry, we should then understand the emergent behaviours of a future AGI from within a framework recalcitrant to any well-defined hermeneutics of intelligence.

Citing Mark Bedau and Paul Humphreys, Roden suggests that a 'diachronically emergent behavior or property occurs as a result of a temporally extended process, but cannot be inferred from the initial state of that process. It can only be derived by allowing the process to run its course'.[78] In other words, unbounded posthumans cognitively and practically reiterate what seems to be a prevalent characteristic of complex nonlinear dynamic systems, namely divergence from initial and boundary conditions. Whatever may be the initial conditions of realization for humans as a natural species or rational persons, whether such conditions are seen as natural evolutionary causes or logico-conceptual norms afforded by discursive linguistic activities, future posthumans can neither be predicted nor adequately approached *with reference to such initial conditions*. Disconnections, in this sense, signify the global diachronicity of deep technological time as opposed to local emergent behaviours associated with particular technologies of the past and present.

---

77 D. Roden, *Posthuman Life: Philosophy at the Edge of the Human* (Abingdon: Routledge, 2014), 116.

78 Ibid., 118.

The radicality of Roden's unbounded posthuman or future AGI lies precisely in the double-edged sword of technological time and its abstract tendencies, which cut against both any purported natural essence and any socioculturally conceived norm. In virtue of the disconnection thesis, when it comes to thinking unbounded posthumans, the artificiality of rational personhood is as handicapped as the naturalness of biological species.[79] The images of the posthuman put forward through evolutionary naturalism or rational normativity, built on the biological constitution of *homo sapiens* or the synthetic makeup of discursive apperceptive sapience, are quite literally *bounded*. They are fundamentally inadequate to cope with or engage with the ethical, cognitive, and practical ramifications of technologically unbounded posthumans, and in that sense they fall back on the very parochial humanism from which posthumanism was supposed to break away in the first place.

Despite the remarkable theoretical rigour and sophistication of Roden's argument and the cogency of the claim regarding the cognitive-practical asymmetry of a future AGI, none of which should by any means be discounted, upon closer examination the disconnection thesis suffers from a number of significant loose threads and misconceptions. Roden's account of diachronic emergent behaviours within deep technological time and their radical consequences for prediction and interpretation on the basis of 'initial conditions of realization' remains negatively metaphorical. Firstly, even if we follow Roden in ruling out the rational (i.e., linguistic-inferential) conditions necessary for the realization of human agency, it is still far from obvious how neatly a feature of nonlinear dynamic systems, i.e., divergence from initial conditions, can be extended to all conditions of realization. Not all complex systems and conditions necessary for emergent behaviours can be framed in the context of nonlinear dynamics and stability analysis. Nonlinear dynamics is not a necessary criterion for complexity, nor is

79 For Roden's critique of a posthumanist philosophy built on the rationalist account of agency, see D. Roden, 'On Reason and Spectral Machines: Robert Brandom and Bounded Posthumanism', in R. Braidotti and R. Dolphijn (eds.), *Philosophy After Nature* (London: Rowman & Littlefield International, 2017), 99–120.

divergence from initial conditions. The framework of diachronically (diverging) emergent behaviours cannot be extended to all conditions necessary for the realization of human intelligence—for example, it does not apply to those involving computational constraints such as the resource-related constraints and information-processing constraints associated with the instantiation of different types of computational capacities. Secondly, the so-called radical consequences of the divergence from initial conditions for a given set of emergent behaviours within a dynamic system are themselves based on a false interpretation of a formal property of nonlinear dynamic systems known as positive global or maximal Lyapunov exponent. This has been the root of a complexity folklore that is not only widely popular in the humanities but also prevalent in commentaries on complexity sciences. In brief, nonlinear systems are sensitive to initial conditions. The smallest amount of local instability or uncertainty in initial conditions, which may arise for a variety of reasons in different systems, can lead to an explosive growth in uncertainty, resulting in a radical divergence of the entire system trajectory from its initial conditions. This explosive growth in uncertainty is defined by a measure of on-average exponential growth rate for generic perturbations, called the maximal or largest Lyapunov exponent. Roughly formulated, the maximal Lyapunov exponent is the time-averaged logarithmic growth rate of the distance between two neighbouring points around an initial condition, where the distance or divergence between neighbouring trajectories issuing from these two points grows as an exponent. A positive global Lyapunov exponent is accordingly defined as the measure of *global* and *on-average uniform* deviation from initial conditions and increase of instability/uncertainty.[80]

Global Lyapunov exponents come from linear stability analysis of trajectories of nonlinear evolution equations in an appropriate state space within *an infinite time limit*. The idea of radical global divergence in trajectories or uniform explosive growth of local instabilities is therefore only valid within an idealized infinitely long time limit. But the assumption that exponential

80 See A. Pikovsky and A. Politi, *Lyapunov Exponents: A Tool to Explore Complex Dynamics* (Cambridge: Cambridge University Press, 2016).

deviations after some long but finite time can be properly represented by an infinite time limit is problematic.

In other words, the radical conclusions regarding the limits of predictability and analysis drawn from the interpretation of positive global Lyapunov exponents hold for a few simple mathematical models, but not for actual systems. On-average increase of instabilities or radical divergence from initial conditions is not guaranteed for nonlinear chaotic dynamics. In fact, linear stability analysis within a large but finite elapsed time and measured by local Lyapunov exponents representing the parameters of the state space of the system point to point show 'regions on an attractor where these nonlinearities will cause *all* uncertainties to decrease—cause trajectories to converge rather than diverge—so long as trajectories remain in those regions'.[81]

The popular complexity folklore according to which emergent behaviours or trajectories radically diverge from initial conditions in complex systems is therefore unfounded, other than in the very narrow sense of infinitesimal uncertainties in an idealized infinite time limit. Otherwise, in the analysis of nonlinear dynamics there is no implication that finite uncertainties will exhibit an on-average growth rate characterized by any Lyapunov exponent, local or global. Global positive Lyapunov exponents can only be obtained from linearized dynamics on the assumption of infinitesimal uncertainties. However, 'when uncertainties are finite, linearized dynamics involving infinitesimals does not appropriately characterize the growth of finite uncertainties aside from telling us when nonlinearities should be expected to be important'.[82]

Apart from the highly debatable extension of a very particular feature of physical complex systems to all conditions of realization of sapience (natural and rational), the main issue here is that there is simply no such thing as an emergent behaviour divergent from initial conditions in an unconfined or unbounded manner. There is no guarantee of uniform divergence from or convergence toward initial conditions.

---

81 R. Bishop, 'Metaphysical and Epistemological Issues in Complex Systems', in C. Hooker (ed.), *Philosophy of Complex Systems* (Amsterdam: Elsevier, 2011), 110.

82 Ibid.

In virtue of this, even if we observe an explosive divergence from some initial conditions, i.e., even if we witness a diachronically emergent or disconnection event, we cannot use this event as evidence to conclude that emergent properties observed to be generated by the same causal antecedents will generate the same type of diachronically emergent behaviours or disconnection events. In simple terms, if we take seriously Humphreys's ill-conceived metaphorical exploitation of nonlinear dynamics and Roden's restatement of Humphreys's claim, this would precisely lead us to believe that there is in fact *no* reason to anticipate that the same causal antecedents that have given rise to a one-time disconnection event will again give rise to the *same type* of diachronically emergent event or disconnection. Not only are we not warranted in expecting that the same causal antecedents that once generated a diachronically emergent event will again produce similar diachronically emergent properties, we cannot expect them to produce any diachronically emergent event or disconnection at all. If we embrace the implications of nonlinear behaviours for a given set of initial conditions within the Lyapunov time, then we must also embrace its implications for the periodic behaviour of the system over time. Within the Lyapunov time, just as emergent behaviours and properties diverge from initial conditions, so different properties emerge from the same causal antecedents in a highly irregular fashion. In short, there is no guarantee that 'once we observe a formerly diachronically emergent event we are in a position to predict tokens of the same type of emergent property from causal antecedents that have been observed to typically generate it'.[83]

If we accept the set of assumptions under which diachronically emergent events or disconnections become possible, then we have to follow the ramifications of such assumptions all the way through. We can no longer selectively restrict their negative implications for predictability to some initial conditions in terms of which the diachronic divergence of emergent events or disconnections were first defined. Prognostication about disconnections then becomes just as constrained as identification of potential disconnections (with the same or similar causal antecedents) following the

83 Roden, *Posthuman Life*, 119.

occurrence of a disconnection or diachronically emergent event. In this regard, the argument for 'epistemological openness' to disconnections on the basis of an already observed disconnection event is corroded away by the acid of the disconnection thesis that Roden tries unsuccessfully to contain.

Another contentious claim in the disconnection thesis is that the cognitive-practical abilities of posthumans might be founded upon the abstract general tendencies of technological systems. Roden claims that 'speculating about how currently notional technologies might bring about autonomy for parts of WH [Wide Humanity] affords no substantive information about posthuman lives'.[84] There is a careful consideration here that a posthumanity realized by the extension of current technologies presents another form of bounded posthumanism. Not to mention that drawing conclusions from particular historically instantiated technologies or technical causes does not imply the radical claims of discontinuity and divergence that Roden seeks to underline. Being aware of these problems, Roden's solution is then to single out salient disconnecting/self-modifying tendencies of technical systems and to present them as diachronically emergent behaviours of deep technological time. But there is no evidence of the methodological basis upon which these particular tendencies or salient features have been singled out and assigned such a high degree of probability or magnitude. Selection of salient features or behaviours—in this case, disconnecting tendencies—makes no sense other than through an analysis of past and present technologies, an analysis that would precisely bring into play the missing questions regarding particular technical causes and specific data with regard to their frequency and context.

The inductive generalization of specific tendencies in such a way that they enjoy a disproportionate degree of likelihood of occurrence is a well-known type of base-rate fallacy in Bayesian inference and judgement under uncertainty. Bayesian inference problems comprise two types of data, the background information (base-rate information) and the indicant or diagnosed information. The base-rate fallacy occurs when diagnosed information or indicators (e.g., causally relevant data) are allowed to come

84 Ibid., 117.

to dominate the base-rate information in the probability assessment.[85] In other words, the absence or weakness of calibration between base-rate and indicant information results in flawed prognostic judgments. In the case of Roden's disconnection thesis, some diagnosed features (representatives) such as propensity for autonomy and disconnection in certain technical systems are taken as general tendencies of future technologies. The outcomes of technological evolution are outlined precisely on the basis of the overdetermination of some representatives, i.e., the selection of certain diagnosed data or features as causally relevant. But it is exactly this seemingly innocent notion of 'relevancy'—selected on the basis of a diagnosed prominent causal role or representative feature—that is problematic. It leads to judgments in which base-rate data such as other 'non-salient' or 'irrelevant' features of technical systems that apparently lack any explicit causal role, as well as those uncertainties associated with specific historical conditions around technological evolution, are ignored. Consequently, the final result is an overdetermined prognostic judgement regarding how the tendencies of disconnecting technologies unfold within the overall evolution of technology (i.e., the claim about the abstract tendencies of deep technological time).

Firstly, there is no proposed methodology with regard to the criteria of selection and diagnosis of disconnecting technologies. We do not know what the criteria of selection for these technical systems are, or how their disconnecting features have been diagnosed and singled out. Instead what we have is a tacit vicious circularity between diagnosed features of some emerging technical systems and the criteria used to select those systems based on the proposed features. Absent this methodological-epistemological dimension, we are adhering to a psychological account of technology that is the trademark of an idle anthropocentrism habituated to relying on its deep-seated intuitions for making diagnostic and prognostic judgments. Secondly, even if we accept the diagnosis about disconnecting features as a verdict obtained nonarbitrarily, as argued above, we are still left with statistical fallacies in the

85 See T. Gilovich, D. Griffin, et al., *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press, 2002).

inductive generalization of these features in the form of an overdetermined judgment about the abstract tendencies of deep technological time.

This overdetermined judgment becomes the locus of a disproportionately high probability, giving a sense of false radicality or impending gravity to its consequences. But just because some representative features of technical systems may play more prominent causal roles does not mean that they are more likely to dominate the evolution of technology in the form of diachronically emergent tendencies. In other words, even if we accept that local disconnections are salient features of emerging NBIC technical systems (a claim that already calls for methodological assessment),[86] there is no guarantee that these local representatives will become global tendencies capable of generating radical discontinuities.

Assigning high probability or significant weight to these features and then drawing radical conclusions and wagers from them is another form of what Nick Szabo calls 'Pascal's scams'.[87] These are scenarios in which there is poor evidence and probabilities lack robustness. Owing to this lack of robustness and poor evidence environment, addition of new evidence (for example, the defeat of a human player against a computer in the game of Go or a breakthrough in one of the branches of cognitive science) can disproportionately change the probability and magnitude of outcomes: 'This new evidence is as likely to decrease the probability by a factor of X as to increase it by a factor of X, and the poorer the original evidence, the greater X is.' In such an environment, the magnitude of possible outcomes, not just their probabilities, are overdetermined to such an extent that uncertainties become the basis of decision making and cognitive orientation, forcing us to make ever more expensive bets and form ever more radical beliefs with regard to uncertainties and future scenarios that can neither be falsified nor adequately investigated by analysing the specificities of the historical conditions of realization. What is unlikely in so far as it is only probable

86 NBIC is an acronym for Nanotechnology, Biotechnology, Information technology, and Cognitive science.

87 N. Szabo, *Pascal's Scams* (2012), <http://unenumerated.blogspot.com/2012/07/pascals-scams.html>.

under uncertainties—methodological, semantic, paradigmatic, and epistemic—becomes likely; then what is likely under the same uncertainties becomes plausible; and what has now become plausible only because it is probable under implausible conditions becomes weighty and truth-indicative. Such is the process through which the Pascal scam is sold to the unsuspecting.

In short, we are swindled into taking the magnitude and probability of such scenarios seriously, treating what is at best an unfounded conjecture and at worst a flight of metaphysical fancy, no more substantial than counting the magical properties of angels in heaven, as if it were a *plausible* possibility not entirely foreclosed to rational assessment and epistemological procedures. In attempting to retain their claim to plausibility without exposing themselves to any criterion of robust analysis and assessment that might debunk their purported radicality, such extreme scenarios have to formulate their wagers not in terms of epistemological problems or hypotheticals that can be adequately tested, but in terms of aesthetic and ethical pseudo-problems often structured as 'But what if...?' questions desperately begging for a response, an engagement, or sympathy for their plausibility. It is in this fashion that the genuine import of the artificial realization of mind or the consequences of posthuman intelligence are obfuscated by pseudo-problems whose goal is to maintain a facade of significance and seriousness: the existential risk of AGI, security analysis of posthuman intelligence, or in the case of the disconnection thesis, ethical complications arising from the advent of unbounded posthumans. In such trends, the posthuman is disconnected from the human only to be reconnected back on a level of discourse and hollow speculation that feeds on the most dogmatic forms of human affect and intuition.

This disproportionate wager on the magnitude of uncertainties is a speculative trend that aligns the disconnection thesis with other singularity-driven scenarios where bets on the rise of a malevolent or benevolent superintelligence are being touted, even less discreetly, as Pascal's scams (e.g., Skynet, Paperclip Maximizer, Roko's Basilisk).[88] However, what connects

88 For more details on superintelligence scenarios, see N. Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).

extreme scenarios associated with judgement under uncertainty is not simply the biased overprediction or underprediction common to them, but also the central role played by intuitive impressions and adumbrations in rendering them 'extreme'. Their radicality is fabricated by those exaggeration-prone cognitive habits that belong to an image of the human whose diagnostic-prognostic abilities are still bound to its evolutionary infancy, as yet unfettered by critical rationality or science.

Roden does signal caution about the exorbitantly speculative dimension of unbounded posthumans and instead favours 'gradua[tion] from speculative metaphysics to a viable cultural research program'.[89] But the radicality of the disconnection thesis and the strangeness of unbounded posthumans are precisely founded on this unwarranted speculative metaphysics. Once this unverifiable speculative dimension is removed in the graduation to a viable cultural program, the disconnection thesis loses its radicality, and the unboundedness of posthumans appears more causally and normatively constrained than Roden claims. Although Roden resists the characterization of unbounded posthumans as alien or inherently uninterpretable, there is no indication as to what the interpretation of unbounded posthumans would entail, particularly when the accounts of intelligence and agency provided have rendered them devoid of any conceptual content. Even if we approach the interpretation of unbounded posthumans from a computational standpoint and no longer from the perspective of predictive accuracy or theoretical-conceptual fidelity, interpretation of such a phenomenon would be so costly that it would become completely unfeasible.[90] Computational cost grows on average for an observer as it climbs the complexity-computational hierarchy. In other words, the size of the observer's internal model grows as it attempts

89 Roden, *Posthuman Life*, 121.

90 For a technical disquisition on the problem of computational cost in interpreting and engineering complex systems, see J. Crutchfield, R. James, et al., *Understanding and Designing Complex Systems: Response to 'A Framework for Optimal High-Level Descriptions in Science and Engineering—Preliminary Report'* (eprint arXiv, 2014), <http://arxiv.org/abs/1412.8520>.

to model or make predictions about phenomena at higher levels of complexity. This increase *costs* the observer. For example, if the observer is a biological organism, it costs physical and metabolic resources.

The problem of computational cost has also interesting implications for modelling. Those models (in this case, models of intelligence) whose measures are set on higher levels of complexity are not optimal or even feasible models. For instance, Charles Bennett's *logical depth* is a measure of complexity of a string S in terms of the time needed for a general purpose computer (a universal Turing machine) to run the shortest program that generates S.[91] The problem with Bennett's and other similar models is that they attempt to interpret or measure the complexity of an object (e.g., general intelligence as a Bennettian *deep object*) from the uppermost level of the complexity hierarchy—in the case of Bennett's measure, the most powerful and resource-consuming class of formal languages and their respective automata, the universal Turing machine computable class.[92] In starting from the upper level of the hierarchy, they run into the problem of poor effective computability, in the sense that we can never be sure whether or not we have found the most efficient coding for what looks likely to be a random pattern.

Already shorn of the constraining continuity with those deontic-normative attributes of agency that make an intentional-semantic interpretation possible, and now suffering an arbitrarization of computational interpretation owing to the lack of effective computability, the unbounded posthuman can match any random pattern or description of any system. Absurd questions such as whether we can regard a galaxy, the number Pi, or an angel as posthuman intelligence become genuine topics of debate. Even though Roden attempts to leave some space for the interpretability of the unbounded posthuman (as 'not uninterruptable in principle'), this

91 See C. Bennett, 'Logical Depth and Physical Complexity' in R. Herken (ed.), *The Universal Turing Machine: A Half-Century Survey* (Oxford: Oxford University Press, 1988), 227–57.

92 For an analysis of logical depth, see E. Atlee Jackson, *Perspectives of Nonlinear Dynamics* (Cambridge: Cambridge University Press, 1991), 516–18.