# Instantiating Sapience: An Inquiry into Reason

Andrea Giuseppe Ragno[1]

[1] Goldsmiths University, 8 Lewisham Way, London SE146NW, UK
andreagiuseppe@outlook.it

**Abstract.** In what follows I offer a brief methodological perspective on the project about the implementation of sapient conceptual skills in technical devices, i.e. part of what has been defined as Artificial General Intelligence (AGI). In the first section, I mention the revival of Kant and Hegel in the neo-rationalist philosophies followed by my critique of Popper's philosophy in order to outline the main features of logical inferentialism through the work of Brandom and Trafford. In the second section, I provide a quick genealogy of the theory of mind that I find most useful for this project, i.e. functionalism. Then, I specify this theoretical approach with the work of Putnam and Sellars. In the third and last section, Pearl's causation ladder and Brandom's conceptual hierarchy emerge from the framework elaborated so far and is made more sound. Finally, I give an account of the engineering approach to functionalism sketched by Negarestani as it is the one which might accomplish the objective pointed out in the premises of neo-rationalism.

**Keywords:** Brandom, Functionalism, Inferentialism, Reason, Sellars.

## 1 Introduction

Since Ancient Greek philosophy, reason has always had a special status. Within the history of philosophy, reason has been described in two ways: on the one hand epistemically and on the other instrumentally. The instrumental approach systematically accomplishes the targets and the intents set out by a specific relation between thinking and its surroundings — the enterprise of science, for instance. Instead, the epistemic approach focuses more on mapping each belief, model, system to their respective targets and on the efficiency of these cognitive and semantic maps. Explaining the relations between reason, or thought, and being, or nature, has been considered one of the tasks of reason.

Contemporary rationalist theories have emerged within the functionalist framework inaugurated by American pragmatist Wilfrid Sellars and have tried to clarify those relations without relapsing into a dogmatic understanding of the relation aforesaid. For examples, Reza Negarestani[1] emphasizes the autonomy of reason without returning to a Cartesian rationalism). Although, the demarcation between reason and nature — what Wilfrid Sellars defines respectively as space of reasons and the realm of causes — is fundamental to answer certain issues, this should not make forget that

[1] Following Hegel and Brandom, Negarestani sees reason as a protocol, a project which aims for self-transformation [23].

rationality and its capacities supervenes on causal patterns recurring in nature. Participating in the ‘game of giving and asking for reasons’ (GOGAR) is a key criterion to distinguish between rational — or, sapient — entities and the rest of the entities which inhabit the world [1]. In fact, the pragmatic independence of certain mental properties and states is made more robust by the computational turn, which sees computational neuroscience and artificial intelligence as contributing factors. Following the developments in computational theory and in formal logic started from the beginning of the 20th Century, the pressures which are endured by various types of reason are rising. An example of this kind might be offered by the enterprise behind Artificial General Intelligence (AGI) — that is a kind of artificial intelligence capable of performing human tasks, or capable of experiencing consciousness and awareness.

## 2 Contemporary Conjectures

### 2.1 Kant’s and Hegel’s Rekindling

In their way of responding and thinking how to articulate the relationship between rule-governed rationality and pattern-governed behaviors, normative inferential skills and bio-chemical structures, reason and nature, contemporary philosophers have been rekindling two key figures of German Idealism, Kant and Hegel. Following Enlightenment, Kant provides a vigorous theory of experience, which is justified for being individual. In fact, particularized singular consciousness, i.e. ‘the transcendental unity of apperception’ according to Kant, is constituted by a series of faculties, through which perceptual, conceptual and agential capacities are considered as different. Hegel is known for reinterpreting reason collectively and historically. He offers an insightful idea in his theory of the *Geist*. Society — and more broadly culture — gives rise to antinomies, schisms, divisions. Thus, a positive conception of antinomy should be preserved.

Acquired from this theoretical background, reason distinguishes different modalities of thought and practice and it recognizes them as part of a common framework. Precisely, reason is able to organize the connection between identity and difference, finite and infinite. It follows that reason not only represents the world but, as an agent, it also progressively and effectively shapes it. Accordingly, the problem with Kant rises when he absolutizes the limits of human knowledge by introducing the notion of *noumenon*, which remains embroiled in causal determination. This removes the systematicity Hegel tries to pursue and undermines the possibility of knowledge from reason’s epistemological assumptions. Also, both Kant’s and Hegel’s conclusions might lead to a naive idealist and thus anti-realist standpoint, which ontologizes the demarcation between mind and nature, instead of rendering it pragmatical and inferential.

### 2.2 Instrumentalism, Essentialism, and Its Counterarguments

Along with the rekindling of Kant and Hegel, the rationalist tradition has received further underpinning by the practical results obtained by science [2]. The scientific enterprise can historically be considered as antidogmatic thanks to its skills of producing novel theories, hypotheses, and counterfactual arguments. In his short essay on human knowledge, Karl Popper defines science by explaining two major theoretical flaws which have been acquired during the course of history by the scientific endeavour. The most recent one is instrumentalism. Popper does not completely dismiss instrumentalism and, in fact, he tries to retain a feature of it. For this reason, his conception of science holds an experimental instrumentalism, which conjoins a critical discussion. Both are deployed to control scientific theories, opinions, and conjectures, which are distinguished from pseudo-scientific theories for their aptitudes of being falsified — rather than being supported — by any kind of evidence [2]. Thus, natural sciences — and all applied sciences — do not belong to the realm of *epistēmē* or *technē*, rather to the one of *doxa*, i.e. opinions and conjectures sensible to mistake and trial [2]. What makes humans capable of being critical of their own creations is, according to Popper, rational criticism and this specific human skill allows mind and knowledge to transcend themselves and grow. The second view criticized by Popper is the one of essentialism: as in the case of instrumentalism, he finds an obscurantist feature which characterizes essentialism. This obscurantist aspect is the belief that science has the ability to describe the *ultimate* essence of nature. It follows that essentialism is confident that science is capable of assigning an undoubtful truth to its own theories [2]. Popper argues that the essentialist point of view does not let thought generate novel theories, issues, and conjectures.

## 2.3 Inferentialism

As in instrumentalism, Popper retains one feature of essentialism: "the scientist aims at finding a true theory or description of the world, which shall also be an explanation of the observable facts" [2]. From here, Popper departs to describe the third view, which represents his standpoint. In short, the third view believes that while all efforts in science are pointed towards a real description and explanation of the world, scientists are aware that the results will never be definite — although, as Popper argues, theories can be proved wrong with a high degree of certitude. Following Kant, he argues that observable properties are falsifiable as data perceived through sense is always elaborated with respect to a theoretical framework. Non-*dispositional* terms have a relationship with experience, but the latter cannot verify the former. This leads to an arbitrary decision on whether a scientific theory can be falsified or not by non-*dispositional* assertations. In other words, testing if an observable property is a true falsifier of a conjecture becomes a trivial matter. Apart from being incoherent with his falsification theory, Popper's idea undermines his conception of truth based on correspondence theory. The correspondence theory of truth is usually linked to a metaphysical realism, which Popper defends. One among many oppositions of the correspondence theory [3], inferentialism includes crucial aspects of all classic objections of that theory of truth, i.e. coherentism, pluralism, pragmatism, and verificationism. Widely understood, inferentialism is a result of rationalistic expressivism and rationalistic pragmatism [1]; in fact, contemporary inferentialism is mostly elaborated by Robert Brandom [1][4-5][21]. By shifting from representationalism to inferentialism (hence, privileging inference over reference and denotation), Brandom draws inspiration from Frege, Dummett, and Sellars, explaining meaning by way of the inferential connections between linguistic expressions [6]. The pragmatic approach of languages is a reformulation of Sellars' doctrine 'meaning-as-use'. Therefore, inferentialism understands semantic statements as assertions which predicate an inferential relation between a term and an object or a property of such object: speaking does not assign any property to reality, rather it becomes a matter of establishing properties and roles.

Here, within the general inferentialist context, James Trafford goes beyond Popper's notion of truth. In his *Meaning in Dialogue* [6], his main aim is to show that meaning is a result of a dynamic interaction between agents, who act in a non-logically pre-determined space, as logical rules emerge through these interactions. While he is aware that Brandom's inferentialism eases agents from questioning their relationship with mind-independent entities (as inferentialism shifts the issue to the establishment of normative properties [6]), Trafford also claims that standard approaches to inferentialism — usually based on monotonic, static, rule-governed methods [6] — fail to meet his thesis; in fact, they dismiss the dialogical and interactive role of proofs, refutations, assertions, and denials. Trafford rejects Popper's belief: counterexamples for a theorem's validity cannot necessarily be found in the future [6] as this would mean dismissing any symmetry between proof and a conclusive refutation of a theorem. Popper's asymmetry is not accepted because it would allow a set of logical rules to casually determine interactions [6]. Trafford, instead, defends the dialogical interaction between proof and refutation free of normative constraints. The alternative offered is a constructive, defeasible, and pluralist approach to logics and inferentialism.

## 3 Contextualising Functionalism

### 3.1 The Origins of the Functionalist Approach

Functionalism is a theory about the constitution of mental states. According to functionalists, mental states are defined by what they do rather than by their internal constitution. Historically, functionalism emerged following the failure of behaviourism and the cognitive revolution in psychology. One of the prominent psychologists who describes stimulus-responsive (S-R) behaviourism's failure is George Mandler. In his paper *Origins of The Cognitive (R)evolution* [7], Mandler stresses that the gradual shift from a behaviourist psychology to a structuralist, cognitive, and functionalist one occurred between 1955 and 1966. His main opposition to behaviourism is that classic S- R theory and radical behaviourism fails to explain issues around human thinking and its agency and memory because of the undifferentiation between humans and nonhuman animals [7]. In fact, even Burrhus F. Skinner's functionalist behaviourism — which was aware that to illustrate a being as a functional system thereby entails characterizing it not in terms of basic regularities (defined as reliable differential responsive dispositions by Brandom [21]) reducible to patterns of S-R behaviour— lacks an analysis on specifically human functions [7].

Functionalism as a research strategy has another point of divergency with classic behaviourism, i.e. the multiple realizability thesis, which is one of the most important arguments for functionalism. By describing mental states in non-mental language — that is, logico-mathematical language — functionalism meets one of the requirements of behaviourism [8]. However, functionalism defines mental states in non-mental terms by quantifying over realizations of mental states [8]. Instead, behaviourism keeps its focus on environmental factors and only quantifies over external conditioning. Functionalism holds the ideas that functional states can be realized in different ways; vice versa, one physical state — e.g. pain — can realize several functional states in different devices [8]. Apart from clarifying the relation between mental states and their explanations in terms of logical and conceptual properties, functionalism also elucidates on how to distinguish sensible data from semantic inferences with regard to mental states' functional roles and relations.

### 3.2 Computational Functionalism

Alongside its recent history, functionalism has been understood in different ways within a philosophical outlook. The two main sources of it are the empirical computational theory of mind, chiefly explained by Hilary Putnam, and the functionalist theory of meaning sketched by Ludwig Wittgenstein and elaborated by Sellars. Machine state functionalism [9] corresponds to an early elaboration of a computationalist and functionalist theory of mind. Indeed, Putnam's theory is among the firsts computational theories of mind (CTMs) — today known as classic computational theory of mind (CCTM). According to Putnam's theory, mental states operate similarly to a Turing machines and thus they share the same properties, meaning that a set of instructions function as an algorithmic rule-set in relation to certain system parameters to determine a certain mental or machine state. It is important to remark that the resemblance between the two models has never been understood as an integral blending because abstract Turing machines are deterministic, and they have sequential processes and an infinite non-addressable memory. Whilst, minds are more stochastic, they can execute parallel computations by generating different types of outcomes from inputs — while inputs and outputs in a Turing Machine are always symbolic — and they have a finite addressable memory.

Surprisingly, Putnam himself became a radical adversary [10] of his theory and, more generally, of every kind computational functionalism years later. Through Gödel's theorems of incompleteness, he develops four arguments to confute computational functionalism. All his theses are refuted by Jeff Buechner [11], who, by preserving the importance of Putnam's theory in cognitive sciences and inductive reasoning, provides a history of the applications of Gödel's theorems and the different degrees of relevance of computationalism within functionalism. In fact, in *Representation and Reality* [10], Putnam argues that inductive reasoning cannot be formalized, and he proves this with the incompleteness theorems — which have now gained an epistemic value [11]. It follows that there cannot be any computational description of the human mind or of general intelligence as both the computational and the non-computational inductive mind cannot prove mathematics' consistency. All methods of inquiry are thus susceptible to Gödel's theorems. Yet, Buechner questions why humans, who, according for Putnam, have no computational descriptions, can succeed in computational tasks [11]. On the one hand, he believes no convincing reasons to dismiss computationalism are offered by Putnam. On the other hand, what it is interesting to highlight from Putnam's *R&R* is the view that a general intelligence does not consist — only — in an inductive account of reasoning, but — also — in an abductive and deductive version. Inductive reasoning can only be dismissed if it was proposed as the only mode of mind's functioning. Rather, if the three forms of reasoning are all included in the debate around general intelligence's automatization, the inductive type of thinking can still play a crucial role.

### 3.3 Pragmatic Functionalism

Putnam takes part in solving the riddle of induction by extending his critique to a description of mind — i.e. inductive reasoning cannot be formalized by a human mind. According to Buechner, he fails encountering Kripke-Wittgenstein rule-following paradox [12]. Far from following Kripke's interpretation of the paradox, John McDowell [14] sees Wittgenstein's dismissal of his paradox [13] not as naïve, rather it is coherent in respect to his view, which conflates understandings with interpretations. Following a rule derives from the inculcation of the rule into a cultural practice. One learns how to use 'plus' because it was instilled into the practice of adding. This usage-based theory of language appearing in Wittgenstein's philosophy influences Sellars' version of pragmatic functionalism. Sellars' researches on *thought* represented a crucial shifting in analytic philosophy[2]. In contrast to the *scientific* image of humans, his *manifest* image of humans — which consists in the quasitranscendental property of reasoning instantiable in other systems than neurobiological ones — rekindles the Kantian project and other problematics regarding reason's infrastructure. In *Empiricism and the Philosophy of Mind* [15], Sellars defends a kantian theory of mind by refuting any Cartesian view of mind. His approach moves towards a naturalism, which does not neglect the normative dimension of mind. Accordingly, Sellars believes that psychological concepts should not be considered less relevant than theoretical concepts in the sciences: mental concepts have the same importance of any scientific concept within naturalism.

The implication of Sellars' top-down approach [16] is vital: mental states are not expressed in specifically human terms, rather they are articulated within a non-human framework as they correspond to functional tokens. Computational descriptions become possible because conceptual behaviours are now understood as transmitters of tokens which maintain inferential relations with other functional states. As long as a computational system becomes capable of conveying semantic content by instantiating interactive, dialogical processes between itself and other computational agents, it can enter Sellars' interactive and inferential space of reasons. Inferences are fundamental for the GOGAR as they establish normative properties between assessments.

[2] Under the guise of Kant's philosophy, Sellars' theories allowed to surpass the Humean deadlock in which analytic philosophy was stuck [1].

Within this framework, reason perceptively responds to an input, and elaborates an action in virtue of a protocol which aims to organize the cognitive and behavioural resources of a system in order to achieve a precise task. Across the space of reasons, terms can change their status from observational to theoretical, or vice versa, and multiple states can be implemented, realized, and instantiated in different technological devices in different ways. Furthermore, epistemic and ethical attitudes can be found accompanying reason. The responsibility in committing to a claim and the normative conceptuality which surround reasons allow minds — or, computational systems — to construct knowledge. It is thus necessary to introduce theoretical underpinnings to explain and track the divergence between the *scientific* image and the *manifest* image in order to describe how cognitive properties which yield sapience can be instantiated in machines.

## 4 An Engineering View of AI

### 4.1 From Correlation to Counterfactual Judgments

Judea Pearl is one the few machine learning (ML) researchers who have recently tried to go beyond the current paradigm ubiquitous in this field, i.e. deep learning. During his academic years, Pearl focused on probabilistic inference [17-20]. His aim is to challenge standard probability calculations — i.e. reasoning by association — embedded in deep learning. In fact, he is one of the pioneers of Bayesian neural networks (BNNs), which, according to him, allowed machines "to think probabilistically" [17]. He believes that current machine learning techniques pay too much attention to uncertainty, while they are not concerned with causal reasoning — that is, thinking cause and effects. The pure statistics applied in ANNs with the curve fitting method shows that their form of predicting, a correlationist approach to AI cannot be accounted as a form of intelligence. This general idea is what leads Pearl to describe the ascension towards the 'ladder of causation'. Thus, he reconstructs the history of probability and causality starting from Thomas Bayes and David Hume.

During the 18th Century, Bayes gave an analysis of 'inverse probability', which is the opposite of forward probability. Bayes was interested in finding the probabilities of two events: the hypothesis and the evidence. Given that one knows some of the prior conditions, Bayes calculated what the probability for a certain event to occur is. He eventually came out with a general solution, which is known as Bayes' rule: by estimating the conditional probability in one direction, it is possible to mathematically derive the conditional probability in the opposite direction., which is less reliable than the former [17]. Thus, Bayes' rule expresses how a degree of belief, accounted as a probability, should rationally be updated to justify for accessibility of associated evidence. In other words, Bayes' rule is needed to update beliefs in a certain hypothesis as evidence can never be fully certain. In this fashion, Pearl and other scholars imagined a hierarchical architecture for ANNs — i.e. BNNs — which could pass information on higher orders through conditional probabilities and likelihood ratios. BNNs are gifted with building blocks similar to causal networks [17] in order to imitate the rational process, which understand cause-effect relationships.

As it is easy to imagine, Bayesian networks highlight the crucial connection between causes and probabilities, which helps estimating the probability of a pattern of dependencies in the observable data, given a pattern of network in the building block [17]. Pearl's desiderata involves eliminating *do*-operators from Bayes' rule to reduce the estimation to a less complex calculus. In other words, reducing the 'do' quantity to a manipulation of the observable data [17]. *Do*-operators are introduced by Pearl [18] to resolve epistemic problems in probabilistic causation. They are expressed as *P(effect|do(cause))* in contrast to classic Bayes' rule expressed as: *P(effect|cause)* [17]. While the latter indicates a passive observation, the former shows an external intervention. Mere observational cases cannot establish any cause-effect property between two cases. This confirms the most famous phrase in statistics, i.e. *correlation does not imply causation*. However, conscious manipulation (the *do*-operators) may estimate causal effects from observational experiments. Pearl argues that Bayesian networks can only perform the *do*-calculus, which generates the reduction — i.e. replace an experiment with a mathematical model — he aims to achieve as they can preserve the meaning of a probabilistic expression [19]. Furthermore, completeness is a property of the *do*-calculus axiomatic system, meaning that that decision problems are controllable [19]. If decision problems are manageable, it follows that an algorithm can decide whether there is a solution to an expression or not in a finite and fast time, i.e. in polynomial time [19]. Thus, Pearl[3] shows how to automate statistical expressions strongly related to causation with mathematics.

However, the final element which makes Pearl climb the 'ladder of causation' is not the algorithmic version of the *do*-calculus. The quintessential element for causal reasoning is the counterfactuals as it helps producing more efficient and precise statements on several issues [19]. Imagining different outcomes becomes possible by adding or removing certain causes from an observable phenomenon. Therefore, 'what-if' form facilitates defining a cause as necessary, sufficient, or necessary-sufficient. According to Pearl, Hume was the first philosopher to include counterfactuals in his account of causality. Hume suggests two opposite definitions for causality. On the one hand, he defines causality as a mere observable pattern regularity which occurs in nature. On the other hand, he introduces the notion of counterfactual judgment and thus he establishes a cause-effect relationship. Pearl retains the latter definition. Counterfactual judgments consist in imagining alternative worlds where a different action would have produced another outcome. These structural models are the short-cut which allows to simulate and compress infinite imagined worlds, to compute the closest to our reality and, which ultimately allows causal reasoning.

Being capable of extrapolating patterns within certain degrees of uncertainty is not enough for Pearl. He is aware that Bayesian networks are not capable of establishing causal properties [19]. Instead of focusing on opaque systems — e.g. CNNs — and mere detection of patterns within a data set, Pearl is still trying to instantiate consciousness, awareness, and intentional states in technical devices. These metaphysical elements constitute the notion of agency. Therefore, next steps in AI should aim to provide a software model for its agency. This hard engineering problem is pursuable

[3] With the help of other scholars and his students.

for two reasons. Following Sellars, mental properties are virtually instantiable in technical device. Following Pearl, algorithms for causal and counterfactual issues already exist [20]. Accordingly, two further questions follow: what is the hierarchy behind cognitive skills identifiable in sapient agents? Which theoretical framework is necessary to instantiate mental properties in technology? An overview and a critique of Brandom's conceptual hierarchy; an illustration of a novel functionalist theory, i.e. the engineering approach to functionalism.

### 4.2 Brandom's Conceptual Hierarchy

Pearl seems suggesting an engineering approach to computational functionalism. Before elaborating this idea, it is necessary to highlight a certain hierarchy among non-sapient and sapient cognitive skills. This hierarchy removes any kind of naivety from theories aimed to level non-sapient and sapient skills, which — as it has described so far — only omits fundamental characters of what is understood as consciousness, agency, and reason. The peculiarity of functionalism is that it does not treat consciousness, agency, and reason as something exclusively human meaning that implementing sapient cognitive skills in technology is possible — which is scientifically (and partly) confirmed by Pearl. The conceptual hierarchy developed by Brandom is a ladder to rational inferentialism, which, by virtue of the theoretical assumptions illustrated so far, is more robust and less vague than Pearl's ladder to causation and is coherent with Sellars' functionalism [15-16].

In *How Analytic Philosophy Has Failed Cognitive Science* [21], Brandom describes how analytic philosophy failed to investigate and strengthen Frege's theories[4] present in his *Begriffsschrift* which gave birth to analytic philosophy itself. Following this failure, analytic philosophy has not convincingly clarified issues around the principles around mind, reason, and thought. Thus, analytic philosophy has also failed cognitive sciences [21] which, in turn, have made several important progresses in explaining how to empirically and experimentally perform cognition. Philosophy, instead, must move in another direction according to Brandom. Following Sellars, Brandom argues that philosophy should ask normative questions around conceptual, sapient, discursive practices of sapient entities. To pursue such enterprise, philosophers should analyse the structural cognitive hierarchy within concept-use practices [21]. Repurposing this ladder towards the achievement of causal inferential modalities of AGI means refuting several accounts of experimentalism and non-pragmatic heuristics. The scientific enterprise is a crucial epistemic practice in the neo-rationalist movement and Brandom's ladder confirms the developments made in the first chapter, where, following Kant and Hegel's contemporary revitalizations, debate within philosophy of science were introduced.

Brandom's hierarchy involves four basic steps: labelling-classification — the level of bare reliable differential responsive dispositions; describing — the level of rational classification; formal inference — the level of logical classification; complex predica-

[4] I.e. the mathematical characterization of semantics and conceptual content, and so of the structure of sapience [1].

tion —the level of invariance under substitution. Conceptual, rational classification is an attribute of sentient entities and it encompasses an abstract clustering of something particular into a general term, which expresses a similarity emerging among particularities. Applying a concept to something means describing it, which is informative as it implies that something follows the description. The semantic import arises when something is inferred from a description [21]. Instead, labelling an object has no informative content and thus a label does not have any conceptual consequence. In fact, the last distinction Brandom stresses is between simple and complex predicates. While a simple predicate occurs as a component in a sentence, a complex one include any multi-place simple predicate. According to Brandom, through Frege's logical notation it is possible to form complex predicates, and logically codify the inferences that articulate its descriptive content. In fact, Frege's notation allows to invariantly substitute singular terms with variables and quantify these variables in order to form complex predicates. Thus, the final step of Brandom's hierarchy, i.e. the *analytical* concept formation [21] described with Frege's notation, accounts for concepts formed at the highest cognitive stage, which represent the most expressive concepts of the whole ladder.

Making sense of the world, organizing it, and changing it by optimizing one's own models are sapient capacities, which occur through inferential semantics, rational pragmatics, and systematic theoretical trajectories. The challenge of AGI should be slanted towards the development of these inferential skills. Although Brandom's ladder is clearly useful for the project behind AGI, an engineering approach to functionalism is still needed to account for the conditions which determine the way sentient and sapient skills could be implemented in technical systems.

### 4.3 An Engineering Approach to Functionalism

Among all the functionalist distinctions, the engineering approach is one that asks about how we might be able to replicate and simulate mind and its particular states in other, inorganic media. This question is not only about how we replicate the human, but also about how we can engage in processes of synthetic production by virtue of which different media can serve as a prosthetic outsourcing of sentient and sapient abilities, e.g. new kinds of perceptual experience, new forms of cognitive heuristics, systems that have non-textual, non-discursive or vocal languages but that communicate in other ways, etc. It can be argued that modelling, prediction, pattern recognition, memory, language processing and probability are cognitive procedures that can be instantiated in media with new shapes and methods[5]. However, imagining artificial entities as epistemic agents is not fiction. Imagining them as conceptual agents might still be. This means that even though technical devices influences human's way of reasoning, act, think, they cannot perform the high-order conceptual behaviours yet. Engineering functional properties means avoiding any trivial levelling of sapience,

[5] Far from endorsing any panpsychist view of mind, — both Brandom and Pearl agree that it is necessary to state that state-of-art of AI is stuck at the level of non-conceptual labelling, i.e. pattern detection through correlational statistical means.

reason to nonconceptual activities. And, an engineering approach to functionalism never starts with the presupposition of a vague superiority of bio-chemical human structures over other synthetic systems by virtue of the well-known multi-realizability argument.

Michael Weisberg [22] is instrumental in highlighting how the models used by engineers do not include those assumptions. This shows how local applications of models have always global unpredictable consequences in the process of engineering. For Negarestani [23], the pragmatic and epistemic method which engineers exploit is a way to see the project behind constructing intelligence as a problem of scaling different models, which interact as the explanandum and explanans of them. Thus, every theorist can access the epistemic metatheoretical assumptions as Weisberg pragmatically argues that the core of scientific evaluation and judgments consist in the analysis of the relations between models and reality. These relations are complex and plural, rather than merely ineffable and inaccessible. Interestingly, one of the consequences of the engineering approach is not a reduction of reason to a mere experimental heuristic, but a coherent conceptual image of reality globally valid and in constant development. Models in engineering are continually weighing their assumptions, tuning their theories on different scales, calibrating their scope between a global image and a local phenomenon as both universal and narrow perspective are always needed. An AGI, a project about the instantiation of conceptual capacities typical of sapient entities in technical devices, should thus move in this direction, i.e. towards an updated dialectics between different degree of scopes by understanding the peculiarity of each level, the multi-level relations, the proper method of explanation, description, and prediction of the either global or local stage of representation, and the inter-play of the many dynamic, defeasible, non-monotonic inferential logics.

## 5 Conclusion

Throughout this paper I have showed which theoretical position should be favourited within the debate about AGI. I have therefore provided a broader theoretical framework which has explained the developments of the argument I aim to support. In fact, in the first section I reviewed Kant's and Hegel's rekindling and juxtaposed it to the hopes of neo-rationalist philosophies. A sketched account of rational and pragmatical inferentialism has then emerged in contrast to Popper's scientific realism. The second section opened with a quick genealogy on functionalism. I specified this theory of with Sellars' theory of language, chiefly influenced by later Wittgenstein. However, I believe that this pragmatic form of functionalism should be associated with an interactional-computational form in order to explain mind's mechanisms and properties, which have been treated as processes and protocols. Treating them as such means that cognitive properties can be instantiated in machines. Finally, the last section provided a summary of Pearl's conception of AI. By refuting to level the complex differentiations within the process of conceptual reasoning, I presented Brandom's hierarchy in order to account the cognitive and conceptual capacities of sentient and sapient entities — the latter were defined in the last stages of Brandom's ladder by virtue of their discursive, apperceptive, and conceptual faculties. Brandom and Pearl seemed agreeing on the limitations of the current state-of-art of AI. Thus, I concluded by showing a new functionalist paradigm, i.e. the engineering approach to functionalism, in order to overcome and exploit flaws, vagueness, and doubts about contemporary AI and to advance novel theories about the implementation of conceptual skills in machines.

**References**

1. Brandom, R. B.: Articulating Reasons: An Introduction to Inferentialism. Harvard University Press, Cambridge (2000).
2. Popper, K. R.: Conjectures and Refutations: The Growth of Scientific Knowledge. 7th edn. Routledge, London (1963).
3. David, M.: The Correspondence Theory of Truth, https://plato.stanford.edu/entries/truth-correspondence [Accessed May 2019].
4. Brandom, R. B.: Between Saying and Doing: Towards an Analytic Pragmatism. Oxford University Press, Oxford (2008).
5. Brandom, R. B.: Making It Explicit: Reasoning, Representing, and Discursive Commitment. Harvard University Press, Cambridge (1998).
6. Trafford, J.: Meaning in Dialogue: An Interactive Approach to Logic and Reasoning. Springer Nature, Cham (2017).
7. Mandler, G.: Origins of The Cognitive (R)evolution. Journal of History of the Behavioral Sciences, 38(4), p. 339 – 353 (2002).
8. Block, N.: What Is Functionalism?. In: D. M. Borchert, ed. The Encyclopedia of Philosophy. MacMillan, New York (1996).
9. Putnam, H.: Minds and Machines. In: Mind, Language, and Reality, p. 362–385, 429-440. Cambridge University Press, Cambridge (1975).
10. Putnam, H.: Representation and Reality. MIT Press, Cambridge (1988).
11. Buechner, J.: Gödel, Putnam, and Functionalism. MIT Press, Cambridge (2008).
12. Kripke, S.: Wittgenstein on Rules and Private Language. Blackwell, Oxford (1982).
13. Wittgenstein, L.: Philosophical Investigations. 3rd edn. Basil Blackwell, Oxford (1953).
14. McDowell, J.: Wittgenstein on following a rule. Synthese, 58(3), p. 325–363 (1984).
15. Sellars, W.: Empiricism and the philosophy of mind. Harvard University Press, Cambridge (1997).
16. Brandom, R. B., Scharp, K., Sellars, W.: In the Space of Reasons: Selected Essays of Wilfrid Sellars. Harvard University Press, Cambridge (2007).
17. Mackenzie, D., Pearl, J.: The Book of Why: The New Science of Cause and Effect. London: Allen Lane, London (2018).
18. Pearl, J.: Causality: Models, Reasoning, and Inference. 2nd edn. Cambridge University Press, Cambridge (2009).
19. Pearl, J.: Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution. University of California Press, Berkley (2018).
20. Pearl, J.: The Algorithmization of Counterfactuals. Annals for Mathematics and Artificial Intelligence, 61(1), p. 29-39 (2011).
21. Brandom, R. B.: How Analytic Philosophy Has Failed Cognitive Science. CEUR Workshop, Genoa (2009).
22. Weisberg, M.: Simulation and Similarity: Using Models to Understand the World. Oxford University Press, Oxford (2013).
23. Negarestani, R.: Intelligence and Spirit. Urbanomic, Falmouth (2018).