N. Bulle (2022). Rationality as a meta-analytical capacity of the human mind: From the social sciences to Gödel. *Philosophy of the Social Sciences*. DOI :10.1177/00483931221115345. (27 pages).

**Rationality as a Meta-Analytical Capacity of the Human Mind:**

**From the Social Sciences to Gödel**

Received date: 18/02/2022

Contact address: [nathalie.bulle@sorbonne-universite.fr](mailto:nathalie.bulle@sorbonne-universite.fr)

Abstract

In contrast to dominant approaches to human reason involving essentially a logical and instrumental conception of rationality easily modeled by artificial intelligence mechanisms, I argue that the specific capacities of the human mind are meta-analytic in nature, understood as irreducible to the analytic or the logical, or else the computational. Firstly, the assumption of a meta-analytical level of rationality is derived from key insights developed in various branches of the social sciences. This meta-analytical level is then inferred from Gödel’s incompleteness theorem. On these bases, and with the help of psychological approaches to consciousness, I argue that human rationality may be characterized as the capacity to make meaningful use of signs.

Keywords

Rationality; Gödel; incompleteness theorems; mind and machines; consciousness; understanding; meaning.

**1. Rationality in the Social Sciences**

Rationality usually refers to a fundamental human ability that differentiates humanity from animality and, as such, expresses the unity of humanity regarding the powers of the mind. Nevertheless, understanding the foundations of this unity remains largely conjectural and, as many authors have argued, there is no agreement on the best theory of rationality (Wettersten 1995), which is confirmed by the title of a recent collective (Bronner & Di Iorio, 2018), evoking the “mystery of rationality”. The question is not to obtain a strict scientific definition, involving a particular interpretative framework, but it should not be an impossible task to identify the fundamental human ability to which the notion of rationality refers. The fact that no agreement has been found so far in this respect is undoubtedly linked to the situation, observed throughout the 20th century, of a “crisis” of psychology, torn between two opposing types of interpretive approaches, one causal and objective, the other teleological and subjective.[[1]](#footnote-1) This opposition of causal and teleological explanations runs through the whole of the social sciences, which consequently only agree on a narrow, logical and instrumental conception of the rational, referring to the capacity for using means to reach goals.

The instrumental conception of rationality easily fits with the naturalization of the human mind that places it in the continuity of physiological phenomena, so that a naturalist philosopher such as John Dewey thought that rationality was essentially “an affair of the relation of means and consequences” (Dewey, 1938, p. 9). Nevertheless, if the rational has to be limited to the instrumental, then people are fundamentally irrational: Apart from practical matters, they are not critical of the values, beliefs and ends to which they subscribe, so that, ultimately, these could just represent properties ascribed by simplification to complex physical systems (see Dennett, 1971).[[2]](#footnote-2) In spite of their ambition to go beyond the various forms of behaviorism in this respect, the cognitive sciences, that today dominate scientific viewpoints on the mind, have fostered a computational representation of cognition based on the models of artificial intelligence, and generally represent the human mind as an information processing system. In the same way, while the former great paradigms of the human sciences ‒ culturalism, functionalism, structuralism, Marxism, Freudism ‒ which placed the springs of social action elsewhere than in conscious motivations, have disappeared, but this disappearance does not give rise to a renewed, non-strictly instrumental, idea of rationality (Boudon, 2008). Instead, we are witnessing a sub-theorization of approaches, along descriptivist and positivist lines, which tacitly revive behaviorist assumptions. Current trends on these matters are even to rethink instrumental competences on the basis of various non rational mechanisms, described as “biased”, “embodied”, or “distributed”, whose virtues are supposed to be adaptive.[[3]](#footnote-3)

In contrast to dominant approaches to human reason involving essentially a logical and instrumental conception of rationality easily modeled by artificial intelligence mechanisms, I argue that the specific capacities of the human mind are meta-analytic in nature, understood as irreducible to the analytic or the logical or else, the computational.[[4]](#footnote-4) Firstly, the assumption of a meta-analytical level of rationality is derived from key insights developed in various branches of the social sciences. This meta-analytical level is then inferred from Gödel’s incompleteness theorem. On these bases, and with the help of psychological approaches to consciousness, I argue that human rationality may be characterized as the capacity to make meaningful use of signs.

**2. Rationality and Explanation of Social Action**

2.1 The Object’s Rationality and the Subject’s Rationality

Rationality is commonly torn between a subjective and a normative point of view. On the one hand, to rationalize is to normalize, and as such, reason consecrates the established order, that of constituted knowledge (Dufrenne 1979, p. 20), so that rationality implies an external yardstick ‒ i.e., the principle of coherence, criteria of deductive or inductive logic, etc. ‒ against which to judge the conformity to reason of a definite object, whatever it may be, reasoning, action, belief, etc. From the point of view of the object, that which is rational therefore has an objective and substantive character. However, when referring to constituted knowledge that is the outcome of a construction, a work of reason, rationality involves considering reason also from a subjective point of view, as a faculty, that of constituent reason. As opposed to constituted reason, constituent reason represents something more fundamental, a law-making power that translates objectively into established laws. The opposition between these two points of view on rationality: that of the object, of constituted reason, and that of the subject, of constituent reason, has nourished much of the discussion about the nature of rationality in the literature. In this respect, when these two levels of the rational are distinguished, the subject point of view tends to be seen as a weakened, unfinished or even erroneous version of the object one. By distinguishing a substantive rationality “appropriate to the achievement of given goals within the limits imposed by given conditions and constraints,” and a procedural rationality taking into account human cognitive limitations, and referring to “the process that generates” behavior, Herbert Simon (1976, 1978), while adopting the subjective point of view, clearly fits into this framework. He defends that a theory of rationality is not a theory of the best solutions but a theory of the efficiency of computational procedures for finding good solutions, that is, a theory of methods. This one involves selective heuristics based on traits associated in memory with possible actions, and end-means analyses. The notion of “satisficing” allowing a search for a solution to be stopped as soon as a satisfactory one is reached (Simon, 1957) or that of “bounded rationality” (Simon 1982), are also defined in relation to the idea of an optimum unattainable by the individual subject. Assumptions underpinning rational capacity, when they confront its products with a normative ideal, are particularly cumbersome. For example, the model of rational choice in chess (Newel & Simon, 1972; Simon, 1978), implies a specific combinatorial structure, which cannot account for the ordinary challenges of human rational activity. Moreover, only the logical or instrumental dimension of the rational is taken into consideration: Principles, tastes, ultimate values are, according to Simon, impossible to justify rationally, so that human behavior develops on an arbitrary basis: “reason is wholly instrumental. It cannot tell us where to go; at best it can tell us how to get there” (Simon, 1983, p. 7).

This tendency to hold the rational capacities of individuals as weakened versions of normative models follows another way of development since the pioneering works of Daniel Kahneman and Amos Tversky (1974) in cognitive psychology. In these works, the predominantly false inferences made by test subjects in matters of logic or probability calculation (i.e., the modus ponens rule, the modus tollens rule, Bayes’ theorem of probability revision, etc.) are attributed to cognitive “biases” inherent to the functioning of the mind (i. e., for instance, to cognitive illusions similar to visual illusions). Kahneman and Tversky’s interpretive framework has the merit of involving a search for alternatives to the models of neoclassical economics that directly bring into play normative rules of choice, and constitutes a fruitful basis for reflection and hypotheses on ordinary thought. However, it originates in the same computational or syntactic types of models involving constituted standards, and leads to conceiving of the subjective processes based on their deviation from these models (criticism along these lines may be found, for instance, in Boudon 1989, Cohen 1981, Gigerenzer 1991, Gintis 2007 and Smedslund 1989).

In their influential article on the rationality of magic, Ian Jarvie and Joseph Agassi (1967) oppose a weak conception of the rational, applied to any goal-directed action, which underpins a basic human capacity, and a strong conception, applied to beliefs, provided that they meet certain adopted norms. Rational action on the basis of rational beliefs then corresponds to a strong sense of rationality. On these grounds, magic rituals are rational in a weak sense, unlike scientific actions which, by relying on normative criteria, involve a strong sense of rationality. The two philosophers point to the error of accounting for the failures of human rational capacity on the basis of normative criteria, when it is the tools of thought available to individuals that explain the particularities of their thinking. Therefore, the interpretation of the anthropologist John Beattie, according to which magic rituals have a primarily symbolic or expressive function, appears as an example of the conflation of action’s rationality (as a fundamental capacity) with beliefs’ rationality (involving normative criteria), or else, of the subject level of the rational with the object one. Such conflation involves interpreting the human actions in question as escaping the criteria of rationality or as embodying an altered and inferior version of them. But, according to Jarvie and Agassi, the question is not so much about knowing, and this is what preoccupied the symbolists, why otherwise rational individuals practice magic rites, but about how, while entertaining an all-encompassing magical world-view, people could have become critical.

2.2 The Analytical and the Meta-Analytical Levels of Rationality

The conceptions of rationality mentioned above are centered on individuals’ instrumental capacities. Nevertheless, the instrumental conception does not account for the whole potential field of the rational. We know that Max Weber defines four major types of social action according to their orientation: instrumental, axiological (value-oriented), affectual and traditional (Weber 1921/1968, pp. 24ff). These categories do not represent complete types of activity – Weber specifies that social action rarely corresponds to only one of these types ‒ but to fundamental dimensions of it, of which the first two participate in rationality. For Weber, rationality implies that action is “meaningfully” oriented, that is, oriented by a “self-conscious” meaning. It does not apply to traditional and affectual types of action when they only involve infra-conscious mechanical processes (such as automatic reactions to habitual stimuli and uncontrolled reactions to exceptional stimuli). Weber thus distinguishes the first two types of social action as two main dimensions of rational action, according to whether the meaning that directs it refers to an extrinsic result or whether it is intrinsic to the action itself. Extrinsic ends are themselves defined in conformity with intrinsic meanings, so that the internal relationship of actions with the subjects’ ultimate values or life meanings drives the development of their instrumental goals: “in being pursued, these become translated into purposes, and thereby transformed into teleological-rational action” (Weber, 1905-1906/1974, pp. 26-27). On these bases, to be rationally understood, action has to be placed “in an intelligible and more inclusive context of meaning” so that it can be grasped as driven by the individuals’ intended meaning (Weber, 1921/1968, pp. 8-9). This context of meaning involves a dimension of the rational that is not reducible to the instrumental and that will be identified here as meta-analytical.

Jarvie and Agassi (1980) further developed their conception of goal-directed action, as a variant of Karl Popper’s, to include the choice of views, frameworks, norms, etc. According to Popper, individuals act in a way that is more or less adapted to the situation *as they see it*, so that actions are not “caused” by psychological states (as in Carl Hempel 1965 for instance) but involve the subjects’ understanding. Popper (1994, p. 168) especially opposes psychological explanations of human action that aim to replace Newton’s laws of motion by laws of human psychology. The subjective understanding involves an interaction between the three worlds he defines,[[5]](#footnote-5) the physical world outside, the individuals’ mental world and the human constructs at their disposal as tools of thought, language, conceptual systems, theories, etc. These abstract objects of the world 3 exist only if they are “grasped” by them (Popper 1978, p. 164), so that Popper’s conception of rationality involves a capacity of understanding irreducible to the analytical or instrumental capacity. To his basic rationality principle, Popper adds a stronger condition understood as a personal attitude of rationality: readiness to correct one’s beliefs. In so doing, he demarcates rationality in a strong sense from madness, because madness or fixed views may satisfy subjectively-defined situational logics, but not critical rationality (Popper 1994).

The assumptions made above in various ways concerning individuals’ capacity to manage their ends, knowledge and beliefs, suggest that neither the analytical nor the instrumental exhausts all rationality. Gilles-Gaston Granger (1985) distinguishes in this respect two levels of rationality that may be compared with the distinction between the analytical and the meta-analytical: the tactical level and the strategic level. At the first level, rationality underlies the observance of a link from principle to consequence and is reduced to logic in the narrowest sense (as induced by propositional calculation). However, the understanding of situations where the logical level of rationality is exercised involves a superior or strategic level of rationality according to which “it is no longer a question of observing the link between principle and consequence, but of the global determination of the field in which will play - or not play - or not play” – these instrumental or logical links. “Rationality is then more topical than logical, reflexive and meta-critical with respect to the logically rational” (Granger 1985, p. 361). At this level, rationality refers to the very definition of the objects, and the broad forms of organization of the knowledge in play, so that it is of the semantic order, as opposed to the syntactic order of logic. To account for the strategic horizon of rationality, Granger evokes the history of the problem of algebraic irrationals. It is because it appeared impossible to attribute an integer to both the side of a square and its diagonal, that a new field of mathematical objects, rigorously constituted and logically exploitable, was created: the Eudoxian relations of magnitudes, whose theory Euclid reported.[[6]](#footnote-6) In any case, a topical enlargement is necessary to solve problems of logical tactics.

In order to embrace the meta-analytical level of rationality, Raymond Boudon (2003, 2007) refers to the notion of cognitive rationality, which applies to both descriptive and prescriptive issues, knowledge and action. Cognitive rationality in Boudon basically postulates that social actors are rational insofar as their actions are based on reasons (or “good reasons”), so that these are meaningful to them. Good reasons involve all forms of knowledge and beliefs that social actors are supposed to hold voluntarily, and that can be reconstructed by the analyst in order to be explicable, generalizable and thus intersubjectively understandable (Di Nuoscio, 2018). Boudon contrasts this broad conception of rationality with that of Rational Choice Theory (RCT), which represents only a specific and reductive version of it. Derived from empiricist and utilitarian approaches, and particularly adapted to mathematical modelling, RCT attributes to individuals consequentialist motivations, self-interest goals and decisional processes based on cost-benefit calculations. RCT can explain neither normative beliefs, which offer a solution to the voting paradox for example,[[7]](#footnote-7) nor non-instrumental cognitive beliefs, which underpin the construction of the object of knowledge and action evoked above. The Boudonian approach to this meta-analytical level of rationality involves the development of sense in terms of meaningful reasons: Boudon considers a system of arguments{S} -> P explaining a phenomenon P and assumes that cognitive rationality recognizes {S} as a valid explanation of P if all the components of {S} are acceptable, compatible and if no alternative {S}’ available is preferable to it. Rationality is of an axiological nature when the system of arguments in play contains at least one normative or appreciative proposition (see Boudon 2007, p. 63).

2.3 Levels of Rationality: A Synthesis

The approaches to rationality mentioned previously highlight the existence of different levels or dimensions of rationality that can be articulated. Two first levels distinguish the object’s rationality and the subject’s rationality. The former implies the application of a norm to a representation (of an action, belief, reasoning, etc.) while the latter identifies a fundamental capacity of the human mind. The object and subject levels can themselves be distributed on two other levels, which are respectively identified as analytical and meta-analytical. The analytical level refers to relationships between objects of thought, as for instance means and ends in action’s conduct. It is of the order of logic and computation, or else syntax, whereas the meta-analytical level refers to the apprehension by the mind of these objects of thought and puts into play the very conditions of their subjective meaning. It is, to take up the linguistic analogy, of a semantic order.

The strength of the analytical level of the rational is concentrated on the object dimension, which is normative, because the analysis develops in a field where the meaning is fixed so that it lends itself well to standardization. On the contrary, the strength of the meta-analytical level of the rational is concentrated on the subject dimension because it involves a power of understanding and creativity, underpinning a possible evolution of meaning. This is why computational conceptions of the mind tend to maintain an analytical or instrumental interpretation of rationality, because as soon as one does not limit oneself to this interpretation, subjective capacities of understanding become essential and, as I will argue below, make rationality appear as a fundamental capacity of a semantic nature.

Ultimately, the four levels distinguished can be represented in a double entry table crossing the distinction of the object and subject levels with the distinction of the analytical and meta-analytical levels of rationality (see Table 1). The examples given previously may illustrate various conceptions that lie at the intersection of the levels in play, but it must be kept in mind that these conceptions may not allow a total disjunction of these levels.

Table 1. The Levels of Rationality: Synthetic Table

|  |  |  |
| --- | --- | --- |
|  | **META-ANALYTICAL**  **LEVEL** | **ANALYTICAL**  **LEVEL** |
| **OBJECT**  **LEVEL** | *Topical, Meta-critical,*  Granger  *Openness to criticism,*  Jarvie & Agassi | *Logical,*  Granger  *Rationality of beliefs,*  Jarvie & Agassi |
|  | *Substantive rationality,*  Simon |
| **SUBJECT**  **LEVEL** | *Rationality in value, axiological,* Weber  *Cognitive rationality (non-consequentialist),* Boudon  *Understanding, readiness to correct one’s belief,*  Popper | *Rationality in finality, instrumental,* Weber  *Goal-directed action,*  Jarvie & Agassi  *Action appropriate to the situation (as the subject sees it),*  Popper  *Bounded, procedural rationality*  Simon |
|  |  |

**3. The Meta-Analytical Level of Rationality and Gödel’s Theorem**

3.1 The Limitations of Formalism: Gödel’s Theorem

The canons of rationality classically involve the normative and analytical levels distinguished above (top right box of Table 1), the rational par excellence being represented by the rules of formal logic. Alternatively, as evoked previously, the subjective level is often assumed to underpin altered forms of rationality, inherent to the cognitive and computational limits of the human subjects, as well as to their sensitivity to factors that escape reason or thwart its own exercise. The meta-analytical level, on its part, is not recognized in strictly instrumental-oriented interpretations of human action. The exemplarity of formal logic in the matter of rationality has nevertheless been singularly challenged by Kurt Gödel’s incompleteness theorem published in 1931. According to this theorem, in any formal system (i.e., consisting of axioms and rules of inference permitting to derive theorems) sufficient to demonstrate the basic theorems of arithmetic (i.e., containing the natural numbers and the operations of addition and multiplication), there exist undecidable statements, that is, propositions that can neither be affirmed nor denied on the basis of the axioms of the system.

It is interesting here to recall the main elements of the challenge thus posed to analytical rationality. Let F be a formal system and let Cl be a class of elements of the system (propositional forms, predicates, class of functions, etc.) In the case of Gödel (1931)’s theorem, the Cl class is one of propositional forms. Gödel (1931) takes the case of formulae with one free variable.[[8]](#footnote-8) Propositions may be formed by applying these formulae to the integers n. The metatheoretical property P “to be provable” may then be considered. This property P allows us to associate to each formula of the class Cl the set of integers that verify the property P. We can order the formulae *f* of the class Cl, on any basis, by numerical indices *i*: *f1*, *f2*, *f3*, ..., *fi*, ... The integers thus identify formulae (by their rank) on the one hand, and the variables involved in each formula, on the other hand. Let designate the formula derived by replacing the free variable in the formula *fi* with the natural number *n*.

The diagonal argument (which goes back to Cantor, and allows to intuitively acknowledge the existence of a set that cannot be found in the formalized system) can then be put forward. The sequence of propositional formulae *fi* determines an ordered sequence S of the sets Ei of integers *n* such that the formula verifies the property P “to be provable”: E1, E2, E3 ..., Ei, ... Let us consider each of these set Ei, pick the integers *n* which are not part of the set En they serve to rank, and define the set E of these integers *n.*

[1]

This set E is composed of integers referring to non-derivable propositions formed on the basis of propositional forms of Cl as defined . To E corresponds a propositional form of the class Cl with one free variable, it is therefore listed in the suite S. Let *k* be the integer defining its rank, and *fk* the defined formula, we thus have E=Ek.

The proposition is then undecidable in the formal system: neither nor not- are provable. For if were provable, *k* would belong to Ek, so that the condition [1] would be verified, which contradicts the prior assumption, and conversely, if were non provable, *k* would not belong to Ek, so that, according to the condition [1], would be provable, which also contradicts the prior assumption.

Gödel’s undecidable proposition is in close relationship, as Gödel himself notes, with the Liar paradox (for the undecidable proposition “says” that *k* belongs to Ek, implying, according to [1], that it is not provable). Gödel adds that the proposition is true because it asserts its own unprovability and it is truly unprovable since it is undecidable: “The proposition undecidable in the system PM [Principia Mathematica] is thus decided by metamathematical arguments” (Gödel 1931/1964 p. 9).

A corollary of Gödel’s first theorem is Gödel’s second theorem, according to which a consistent theoretical system powerful enough to contain a representation of arithmetic is (syntactically) incomplete: It cannot demonstrate its own consistency. Therefore, something that is not related to the formal links defined by the system always exceeds the system. This is reputed to reveal the incapacity of the formal links involved to entirely account for the meaning of the mathematical objects they refer to. From then on, Gödel’s theorems underlie a questioning of any conception of rationality restricted to the analytical dimension.

3.2 Gödel’s Theorem and the Opposition between Mind and Machines

Gödel’s incompleteness theorems’ challenge to the reducibility of the rational to the analytical dimension was largely interpreted as requiring the arbitration of computational models of the mind. This echoes Alan Turing’s (1950) question: “Can machines think?”, translated into an imitation game in the form of the question “can machines imitate the human mind?”. Simple imitation corresponds to the “weak” version of AI, which nevertheless underlies the possibility of a closer convergence, implying, with strong AI, the imitation of the very processes of thinking, and its corollary, the fundamentally computational character of the human mind.

Gödel remarks that Turing machines, which represent a general model of the functioning of mechanical calculation devices (such as computers), offer their “most satisfactory” form to the incompleteness theorems by “reducing the concept of finite procedure to that of a machine with a finite number of parts”[[9]](#footnote-9) (Gödel, 1951, pp. 304-305). Since these machines embody the computational model of the mind, while giving an operational form to the limitation theorems, the question of whether, on the basis of these theorems, they allow the refutation of the computational model of the mind arose. For his part, Gödel believes that a consequence of the incompleteness theorems is that “it is not possible to mechanise mathematical reasoning, i. e. it will never be possible to replace the mathematician by a machine” (Gödel, 193?/1995, p. 164) so that the human mind “infinitely surpasses the powers of any finite machine” (Gödel 1951, p. 310).

The challenge posed by the theorems of formalisms limitation to the comparison of the human mind with a computational machine, that is, a device for manipulating uninterpreted (formal) symbols, calls into question the dominant approaches to human thought developed by cognitive science in the second part of the 20th century. John Lucas (1961), by arguing in a philosophy journal that Gödel’s theorem implies a refutation of the computational interpretation of the mind, stimulated the first wave of debate, revived more recently by the arguments of the mathematician physicist Roger Penrose (1989, 1994). Lucas (1961, p. 112) states that, according to Gödel’s theorems, there are formulae in mathematical systems which cannot be proved-in-the-system, but “which we can see to be true”. In other words, the theorems of formalisms limitation demonstrate the existence of an effective power of the mind irreducible to analytical mechanisms alone, and implying the role of consciousness in understanding.

A basic counter-argument, which can be found in various forms in the literature on the subject, relies on the possibility of extending the analytical capabilities of machines by various logical-mathematical or other methods in order to imitate the meta-analytical insight in question.[[10]](#footnote-10) In his defense of weak AI, John Searle (1997) argues, for example, that the anti-mechanist argument derived from the limitation theorems would be valid if the human mind were to be simulated by programs using only sound methods of mathematical proof, but that this may be overcome. Nevertheless, Penrose (1994, pp. 81-82) prevents such objections invoking the potentially simulable character of human insight concerning the decidability of Gödel’s undecidable proposition by defending that such a program is always drawn by a programmer outside the system, because to conceive it, it is necessary to simulate the interpretation, which is of another nature. Moreover, it would be necessary to design the program at a higher level from the start, which leads to the reiteration of the same problem at this level and thus leads to an infinite regress, which is unsuitable for substituting itself to human insight.[[11]](#footnote-11)

Other arguments tend to reject the relevance of mind’s comparisons with computers on the basis of the incompleteness theorems, invoking for example the idealization of the human mind’s analytical capabilities (computation, consistency etc.) necessary for a mathematical insight into decidability (Putnam, 1995; Shapiro, 1998), or the possibility to relax the consistency requirements of formal systems so that the incompleteness argument no longer holds (Buechner 2010) or else, the limitations of minds similar to those of machines in matters of inability to compute a self-referring statement (Slezak, 1982).

A broad objection is that, even if the human mind has specificities irreducible to mechanical processes of a computational kind, this cannot be deduced from the limitation theorems.[[12]](#footnote-12) This objection may be justified if it concerns the empirical use of Gödel’s theorem to compare the potentialities of minds and machines on the basis of simulation issues, and it must be extended to the criticisms of the Gödel-Lucas-Penrose argument that have been developed mainly on this ground. Nevertheless, I argue in the following that Gödel’s theorem should be used as such for what it reveals of the analytical level of rationality’s limitation.

3.3 The Meaningful Horizon of Formal Systems

It is legitimate to draw genuine teachings from the incompleteness theorems with regard to the shortcomings of the analytical level of rationality they involve. It is interesting in this respect to refer to Jean Ladrière (1960)’s insights on the phenomena of formalisms’ limitation referring to the syntactic limitations, which concern the deductive possibilities of formal systems, as well as to the semantic limitations, which concern the relations between elements belonging to a formal system and elements outside that system. Especially, the logician Leon Albert Henkin demonstrates that these systems (which are not contradictory, and which contain the theory of numbers or a formal equivalent) are non-categorical, that is, they do not admit a unique type of model. Knowing that a model represents a domain of realization of the properties derived from the system, a non-categorical system admits as a model not only the domain of objects for the analysis of whose properties it was originally intended (for example the integers), but also other models that do not have any relationship and do not maintain any form of isomorphism with it. Henkin thus generalizes the notion of model by considering not only “normal” or “natural”, and in a sense, intuitive models, but also other “abnormal” models (see Ladrière 1960, pp. 292-298).

When we refer only to natural models, there is no equivalence between truth relating to these models and the derivability in formal systems, since some propositions appear undecidable. This is the case of propositions that are true only in natural models and are no longer true as soon as we consider non-natural interpretations. The extension of the notion of models to non-natural models allows us to obtain complete equivalence between “true” in a general sense (i.e., in non-natural models as well as in natural models) and “derivable” in the system. Gödel’s undecidable proposition is true only for the natural interpretations of the system, and not for all possible interpretations. The anomaly revealed by Gödel’s theorem reflects this impossible adequacy between the system and its natural interpretations, taking into account all its possible interpretations.

A natural model of a (non-categorical) formal system represents for us a non-formal theoretical domain that we can apprehend in an intuitive way, but that we cannot represent in a totally adequate way by the formal system itself. There is, Ladrière (1960, p. 306) explains, inside the formalism, the impact of a horizon from which the latter draws its latent meaning (latent because only an interpretation can bring it to light) but which irremediably exceeds it. Gödel’s theorem expresses this irreducible duality of the intuitive and the formal, which represents an essential overflow of mathematical meanings outside the framework of the formalizable.

The foregoing allows us to understand the price of the relative autonomy of formal systems, with regard to mind. As productions of mind detached from a subject, that is, as “objective thought contents,” they participate in Popper’s world 3 and are likely to support operations of knowledge, for instance through derivation of logical consequences, independently of any actual subjective thought processes (Popper 1972; 1978). But Henkin’s works evoked above through Ladrière’s synthesis involve an interesting interpretation of their intrinsic limitation, namely that without mental grasp they show, in relation to their possible domains of interpretation, a form of semantic insensitivity, or else, a deficit in semantic discrimination.

Therefore, by the example of one class of world 3 object, Gödel’s theorems allow to support the existence of a proper capacity of mind involving the issue of understanding, neither more, nor less. In this respect, discussions that pit mind against machine on the basis of Gödel’s theorem, using Turing’s imitation game arbitration, go the wrong way. Even if, as Penrose defends, an imitation of the causal power of mind in play is not possible, empirical proofs are likely to never be held sufficiently conclusive. But they are unnecessary.

Gödel defends in different texts that mathematics draws its meaning from a content that is ultimately grasped only by the mind. This is why, in order to prove the non-contradiction of classical mathematical systems (i.e., of the natural type), we need intuitive proofs, because something is lost in the transition from evidence to formalism, something of the form of a mathematical insight not derivable within the formalism. The conclusion Gödel derives on this issue concerns first of all mathematics: If the mind were equivalent to a finite machine, there would exist absolutely undecidable mathematical propositions (Gödel, 193?/1995, p. 164; 1951, pp. 309-310). Intuitive proofs involve abstract notions, that is to say, of a higher order than the concrete combinatorial relations between signs. We make use of insights of these abstract notions through mental constructs that spring not from the combinatorial properties of the signs representing them, but only from their meaning (Gödel, 1958/1990, 241).[[13]](#footnote-13) Such “use of abstract terms on the basis of their meaning” (Gödel, 1934/1986, p. 370) is, according to Gödel, the feature of human reason that radically distinguishes it from purely formal or mechanical processes. I come back to this in the following.

**4. The Semantic Roots of Rationality**

4.1 The Semantic Mastery of the Mind

The statement of one of the pioneers of modern mathematical logic, Emil Post, enounced as early as 1924, according to which if we could be “completely conscious” of something, it could be mechanized, seems to undermine the specific, non-computational role assigned to human understanding in thought.[[14]](#footnote-14) But a characteristic feature of consciousness, which appears in Post’s observation, is the distinction between a “conscious” subject and an object whose complete comprehension would allow complete mechanization. The “consciousness of something” thus implies at least two levels, one of which underpins the act of apprehension of the other, justifying the notion of “consciousness of”, so that what remains when all the possible consciousness of a thing is objectified still involves the act of aiming, which is irreducible to its object.

The fact that consciousness is typically of, or about, something underpins the idea of intentionality, which is particularly developed by phenomenology. Meaning towards objects translates into directedness of consciousness. Our minds appear to be content-driven, or, semantics-driven, whereas Turing machines, Richard Tieszen (2006, p. 239) notes on this subject, do not understand meanings and thus are not goal-directed, they have no intentionality.

The driving role of meanings is an important aspect of Lev Vygotsky’s psychology which links in his last works the dynamics of human action to an internal dynamic of meanings (Bulle, 2021; Zavershneva, 2016, p. 139) and in so doing, considers an act of will as consciousness animated by meaning, or else, “a concept that has become an affect” (Vygotsky 1934/1986, pp. 9-10; Zavershneva, 2010, p. 66). The exact psychological nature of the inner mediation by meanings is unknown but its reality is beyond doubt, according to Vygotsky. Especially, to be a meaning involves links of generalization and such links allow logical and deliberate use (Vygotsky, 1934/1986, p. 171-172). Moral consciousness also assumes forms of abstraction underpinning intended meanings. Vygotsky (1933/2016) explains, for example, that children’s moral awareness develops in play with the invention of imaginary situations whose essential attribute, he explains with reference to Spinoza, is “a rule that has become a desire”. In so doing, children abstract themselves from the immediate demands of their environment and act not according to what they perceive, but according to the meaning they give to their action. Similar insights are present in David Ausubel’s “psychology of meaningful verbal learning” where concepts “acquire their meanings and are stored hierarchically (not linearly) in memory,” they are “related in particular semantic (not associative or syntactic) ways to particular ideas in a hierarchically organized cognitive structure with stable and explicit meanings” (Ausubel 2000, p. 139).

These psychological theories of consciousness are consistent with the premises of the social science works mentioned previously, which conceive of social action and human thought as driven through meanings developed at a meta-analytical level of rationality.

4.2 The Capacity to Make Meaningful Use of Signs

The power to manipulate ideas on the basis of the meaning they have for us, which assumes the power to apprehend them in consciousness as objects of thought, implies, as we have seen, an act of aiming that is situated in the internal semantic system of consciousness (this notion is borrowed from Vygotsky 1934/1986) at a higher level of abstraction or generalization than that of the developed ideas.

The human fundamental capacity underpinning the possibility to apprehend ideas as objects for the mind refers to a recursion capacity ‒ the ability to fit elements into one another in a hierarchical way. It is interesting to remark here that, of all the differences that have been evoked by scientists as marking the distinction between humans and animals, such as articulated language and the use of tools and symbols, it is the recursive capacity of thought that appears to be at the foundation of other human skills (Corballis, 2011). The recursive capacities have been developed adaptively by natural selection and especially make use of a large working memory allowing ideas to be embedded within ideas. First and foremost, they underpin mental time travel (the ability to insert events into consciousness) and theory of mind (the ability to put oneself in the place of other persons in order to understand them). Recursion distinguishes human psychology from that of animals, since we only find elementary recursion in some animals and, as I argue here, allows the use of abstract objects of thought on the basis of their meaning.

The relatively recent (on an evolutionary scale) human capacity for abstract thinking may ride on the older functioning of the sensory cortex. Especially, sensory perception is the input mode of the nervous system, the gateway to consciousness (Baars 1997, pp. 63-64). The mind uses the sensory impressions of conventional signs, or in the case of written language, of signs of signs, to control itself. In Vygotsky’s psychology, the general capacities of human thought are based on the possibility of the mind to manipulate at will the sensorial impressions which, when they refer to conventional signs, imply abstract meanings. These “symbolic stimuli” (Vygotsky 1930/1999, p. 36) present us with an internal world, which is that of the meaning attached to the signs. This inner world of abstract ideas and relations underlies models of the outer world that we can manipulate through the use of signs on the basis of their meaning. The possibility of recalling sensory impressions to consciousness is thus transformed, with the help of signs, into the possibility of manipulating abstract ideas on the basis of an intuitive grasp. The understanding of meaning and, correlatively, the act of aiming imply, as we have seen, in the internal semantic system of consciousness, the existence of a level higher than that of the object of thought that accounts for the driving role of the conscious activity of the mind. If this is right, signs do not simply play a computational or syntactic role, but a primordial semantic role through the meanings they awaken in the mind and which involve a whole system where they are integrated at different possible levels of abstraction, so that the objects of thought are apprehended from more abstract meanings.

One criterion that allows us to characterize the meta-analytical capacity of the human mind, in accordance with the works evoked previously from the social sciences to Gödel, could then be simply defined as *the capacity to make meaningful use of signs*. This characterization is comparable to the specific feature of human reason according to Gödel, which relies on the *use of abstract terms on the basis of their meaning*. By referring here to (artificial) signs rather than to abstract terms, one allows the inclusion of other possible auxiliary tools of thought that have the same purpose as mathematical signs or words of ordinary language. Moreover, by referring to the meaningful use of these artificial signs, one removes any ambiguity about the reference to the subjective character of the meaning of signs in their rational use.[[15]](#footnote-15)

4.3 The Capacity to Make Meaningful Use of Signs and the Logical Abilities

In the following I argue that the meta-analytical capacity of the human mind, conceived as the capacity to make meaningful use of signs, is at the fundament of human rationality, so that logical ability derives from this. The logical capacity as such is, on the contrary, usually held as being at the fundament of rationality, which justifies the pre-eminence given to normative and instrumental conceptions so that the subjective appears as derived and the meta-analytical is denied any role. In Jean Piaget for instance, logical ability relies on the development of general structures of thought and is based on a progress of logical-mathematical skills developed by the subjects’ interaction with their environment (see for instance Piaget 1967/1971). A consequence is that the childish thought is described as syncretic, without coherence, without the capacity of deduction, or insensitive to contradiction.

Alternatively, if the fundamental rational capacity, as defined here, is of a semantic order, it has an immediate universal scope. The nature of the objects of thought accounts for the form of the reasonings carried out, not of the rational ability of the subjects. Especially, carrying particular logical links may necessitate a hierarchical system of relationships between concepts, with more abstract terms including less abstract ones. For instance, it has been noted that the Piagetian stages of intellectual development may or may not be respected by children or adults in specific areas according to whether they have acquired adequate systems of concepts or not. Since every concept is in some way linked to others, the total body of concepts acquired during a lifetime influences the acquisition and use of other concepts, and this also explains why most children are unable to engage in general abstract reasoning before the age of eleven or twelve (Novak 1977, p. 122).

On these bases, a basic condition for human beings to develop rational forms of thought is to have meaningfully interiorized systems of conventional signs. That is made possible thanks to language development, in particular. Their capacity to make meaningful use of signs is in a latent state before that. Without language or other signs at their disposal in mind, human beings are thus rational beings in capacity.[[16]](#footnote-16)

Vygotsky argues that children develop true logical thinking, though unacceptable to adults, because their thinking tools are structured differently. Words do not refer to hierarchically organized systems of concepts until later. Before that, they are organized in “complexes”, which act as family names. Complexes are not based on generic traits but on various relationships that link beings or things while preserving their individuality. Therefore, when a word intervenes as a family name, it designates groups of concrete objects brought together by a specific relationship (of a functional, experiential or whatever nature), so that one object can have multiple links of this kind, just as one person can have different “family groups”. Even adults have daily recourse to pseudo-conceptual forms of thought, by constantly passing from conceptual forms to concrete, complexes-based forms of thought (Vygotsky, 1934/1986, p.134, 140).

In the same spirit, in any society, individuals demonstrate a genuine logical ability if one takes into account the meaning they lend to their words. In this respect, Vygotsky offers a key to understanding the so-called “pre-logical” nature of “primitive mentality”, as the anthropologist Lucien Levy-Bruhl (1910/1966) characterized it. He observes that the members of a tribe in Brazil, the Bororos, claimed that they were Araras, which means parrots, that they therefore claimed to be both human and non-human, and more generally, that objects, beings, and phenomena were supposed to be both themselves and something other than themselves. This seemed to go against the fundamental logical principle of the excluded third. But on the basis of a complexes-based categorization, the law of the excluded third is not faulted (Vygotsky & Luria, 1930/1993, p. 87; Luria 1976). Bororos’ logical thinking is not concept-based but complex-based, and logically, the product of thought by complexes has to be what Lévy-Bruhl called participation, identifying beings and things by specific relationships with other beings and things that appear, for conceptual thought, impossible and unthinkable. Especially, the functional use of the word as a sign or family name does not imply, as is the case in its conceptual use, the identity of the designated beings. The word “Arara” is a general name for a given complex including birds as well as men (Vygotsky, 1934/1986, pp. 128-130).

Human logical ability might thus derive from a more fundamental capacity of a semantic type conceived as the capacity to make meaningful use of signs. Such semantic capacity entails the possibility to maintain internal relations of meaning. Therefore, the membership, or non-membership, in a set, whether defined on a conceptual basis, on a complex basis, or on any other basis, can be maintained, so that the excluded third party can be respected. As Ian Hacking (1982, p. 56) states in this regard: “Logic is the preservation of truth while a style of reasoning is what brings the possibility of truth or falsehood”.

4.4 The Capacity to Make Meaningful Use of Signs and the Meta-Analytical Abilities

The capacity to make meaningful use of signs underpins the possibility of evolution of the structures of meaning which we have seen the role it plays in the meta-analytical conceptions of rationality. Already, perceptual awareness, whatever we do, tries to offer a “single coherent interpretation at any given moment”. This highlights the fact that the need for understanding involves an internal activity of meaning ascription, and this activity may be reiterated at all levels of consciousness. Conscious thought tends to develop a unitary grasp of reality on the basis of information from separate sources and on ideas from separate domains of knowledge (see Baars 1997, p. 87, pp. 163-164).[[17]](#footnote-17) The genuine, meta-analytical role of meaning understanding thus underlies an internal evolution of meanings in order to make new information consistent with the structures of meaning involved in interpretation.[[18]](#footnote-18) In this respect, the logician Jean-Blaise Grize (1963) explains that contradiction does not have the same status in formal and non-formal thought. In the former, contradiction is to be avoided. In the latter, it is situated within a totality in the making, it is to be overcome, and is the very motor of meaning evolution.

The meaningful use of signs assumes that understanding is the fruit of an empowerment of the mind through its recursive capacities of abstraction, so that it can abstract, represent symbolically, generalize, and develop the structures of meaning from which it apprehends reality. On this subject, there exists a form of “trade-off”, noted by different authors, between such an understanding power of mind and its computational capacities. Penrose (1994, p. 399) opposes “genuine understanding”, which involves conscious awareness, to memory and calculational powers, while Ausubel (2000, p. 136) opposes the limited human conscious ability to remember and manipulate only a few discrete units at a time to the computers’ capability of storing and interrelating vast quantities of discrete units of information. But he specifies that this limitation of the mind is compensated by the assimilation of these dissociated units by the wider meaning of more stable and inclusive “subsumers” within the cognitive structure.[[19]](#footnote-19)

Finally, the ability to make meaningful use of signs may help in understanding and circumscribe conceptions of rationality used to serve specific scientific purposes. For example, the ability to respond appropriately to a subjectively perceived situation, as Popper characterizes basic rational capacity, can be derived from the meta-analytic capacity of the human mind, knowing that, for Popper, sign systems and other possible elements of the world 3 are part of the subjects’ situation and must be grasped by them. The meta-analytical point of view even allows us to delimit reason and madness in a more general and less axiological way than the condition of readiness to correct one’s beliefs conceived as a personal attitude in Popper. Reason and madness can be demarcated on the basis of man’s very semantic capacity to make meaningful use of signs, for example by the deterioration of the subjects’ mind’s sensitivity ‒whatever its origin ‒ to evidence from the external world that, according to the tools they have at hand for understanding reality, should lead them to revise their beliefs.

**5. Towards a Renewal of the Rationality Concept**

I have defended here that the rational qualities of human thought are rooted in its capacity to make meaningful use of signs. This interpretation of the specific rational faculty of the mind is in accordance with the assumptions of psychology conceived of as a science of consciousness, from Vygotsky to Ausubel. This interpretation of rationality is also in accordance with the approaches to human action in other branches of the social sciences where the subjective reasons are conceived of as causes, so that the meaning that individuals give to their action is an integral part of the explanation of social phenomena, from Weber to Boudon. On this subject, the limitations of formalism highlight the impossibility of totally assimilating rational knowledge and formal knowledge, thought and calculation. By doing so, they reveal the existence of the mind’s own causal power of a semantic and meta-analytical order.

Once we characterize human rationality as the capacity to make meaningful use of signs, which are internalized instruments for evoking ideas, concepts etc., we accept some important premises about the functioning of the mind. Not only are signs supposed to be necessary to the conduct of rational thought,[[20]](#footnote-20) but their meaningful use implies the existence of a genuine power of conscious processes in the form of the direction and mastery of thought. This power of the mind to act upon itself tends to be denied by scientific thought, as the modern history of the conceptions of the human reason testifies. The meaningful use of signs has no meaning for 20th century positivisms and pragmatisms. Whether knowledge is assumed to be ultimately composed of sense data or schemes of action, it is translatable into mechanistic or computational processes. This is why the sense of “meaning” varies completely according to the conceptions of knowledge. In behaviorism and neobehaviorism, meaning is not something developed in consciousness but corresponds to a disposition to act in certain ways in certain input situations. In functionalism, which represents the dominant approach of contemporary philosophy of mind and cognitive science, mental states are identified with the causal role they play within the cognitive system, independently of any specific substrate, so that the psychological processes modelled are relational and abstract, and can be described as computational (see Block 1978, Polger 2012, Richardson 1981).

Philosophers since Plato and Aristotle had agreed to hold rationality, linked to the use of a language that is not natural but symbolic, as a demarcation line between human beings and animals. The human rational specificity has nevertheless tended to be doubted with the rise of naturalist thought, and the success of Darwinian evolutionism which revealed an intrinsic continuity of human beings with the rest of nature. The computational-instrumental model of the mind agrees well with this continuity, but it cannot account for the phenomenon of consciousness and thus leads to ignore the specific empowerment of the human mind by its capacity to make meaningful use of signs. Conceiving of rationality as the expression of this capacity could open the way to a renewal of scientific psychology, and more generally of the social sciences, based on the genuine meta-analytical role that meaning plays in human thought.

Nathalie Bulle is a sociologist, research director at the National Center for Scientific Research (CNRS) in France. Her interest in the analysis of human thought, in its common or scientific form, is at the core of her works applied to teaching and pedagogy and to the development of ideas and ideologies. http://www.nathaliebulle.com/

**References**

Agassi, J. & Jarvie, I. C. (1967). “The problem of the rationality of magic.” *British Journal of Sociology* xviii(1), 55-74.

Agassi, J. & Jarvie, I. (1980). “The Rationality of Irrationalism.” *Metaphilosophy* 11(2): 127-133.

Ausubel, D. P. (2000). *The Acquisition and Retention of Knowledge: A Cognitive View*. New York: Springer.

Baars, B. J. (1997). *In the Theater of Consciousness*. Oxford: Oxford University Press.

Block, N. (1978). “Troubles with Functionalism.” *Minnesota Studies in the Philosophy of Science* 9: 261-325.

Boudon, R. (1989). *The Analysis of Ideology* (originally published in 1986 with the title *L’idéologie*). New York: Polity Press.

Boudon, R. (2003). “Beyond Rational Choice Theory.” *Annual Review of Sociology* 29: 1-21. https://doi.org/10.1146/annurev.soc.29.010202.100213

Boudon, R. (2007). *Essais sur la théorie générale de la rationalité*. Paris : PUF.

Boudon, R. (2008). “Mais où sont les théories générales d’antan ? ” *Revue européenne des sciences sociales*. xlvi-140 : 31-50. https://doi.org/10.4000/ress.166

Bronner, G. & Di Iorio, F. (Eds). (2018). *The Mystery of Rationality: Mind, Beliefs and the Social Sciences*. New York: Springer.

Buechner, J. (2010). “Are the Gödel incompleteness theorems limitative results for the neurosciences?” *Journal of Biological Physics* 36(1): 23-44. https://doi.org/10.1007/s10867-009-9160-1

Bulle, N. (2021). “Vygotsky versus Dewey on Mental Causation: The Core of Two Divergent Conceptions of Human Thought.” *New Ideas in Psychology* 63. https://doi.org/10.1016/j.newideapsych.2021.100898

Bulle, N. (2019). “Student’s Activity and Development: Disentangling Secondary Issues from the Heart of the Matter.” Educational Research Review 27: 56-70. https://doi.org/10.1016/j.edurev.2018.12.004

Cohen, L. J. (1981). “Can human irrationality be experimentally demonstrated?” *The Behavioral and Brain Sciences* 4: 317-370 https://doi.org/10.1017/S0140525X00009092

Condillac, Etienne Bonnot de. (1746/2014). *Essai sur l’origine des connaissances humaines.* Paris : Vrin.

Corballis, M. C. (2011). *The Recursive Mind. The origins of Human Language, Thought and Civilization*. Princeton: Princeton University Press.

Dehaene, S. (2014). *Consciousness and the Brain*. London: Penguin Books.

Dehaene, S., Lau, H., & Kouider, S. (2017). “What is consciousness, and could machines have it?” *Science* 358: 486–492.

Dennett, D. C. (1971). “Intentional Systems.” *The Journal of Philosophy* 68(4): 87-106.

Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Henry Holt & Company.

Di Iorio F. (2016). “World 3 and Methodological individualism in Popper’s Thought.” *Philosophy of the Social Sciences* 46(4): 352-374. https://doi.org/10.1177/0048393116642992

Di Nuoscio, E. (2018). “On the Explanation of Human Action: ‘Good Reasons’, Critical Rationalism and Argumentation Theory”. In G. Bronner & F. Di Iorio (Eds). *The Mystery of Rationality: Mind, Beliefs and the Social Sciences* (pp. 37-52). New York: Springer.

Dufrenne, M. (1979). “La raison aujourd’hui” In T. F. Geraets (Ed.) *Rationality To-Day. La Rationalité Aujourd’hui* (pp. 15-25). Ottawa : University of Ottawa Press.

Feferman, S. (2009). “Gödel, Nagel, Minds, and Machines.” *The Journal of Philosophy* 106(4): 201-219.

Hacking, I. (1982). “Language, Truth and Reason”. In M. Hollis and S. Lukes (Eds). *Rationality and Relativism* (pp. 48–66). New York: Blackwell

Hardcastle, J. (2009). “Vygotsky’s Enlightenment precursors.” *Educational Review* 61(2): 181-195.

Gigerenzer, G. (1991). “How to make cognitive illusions disappear: Beyond ‘heuristics and biases’”. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 83–115). Chichester: Wiley.

Gintis, H. (2007). “A Framework for the Unification of the Behavioral Sciences.” *Behavioral and Brain Sciences* 30(1): 1-16.

Gödel, K. (1931/1964). “On Formally Undecidable Propositions of the Principia Mathematica and Related Systems”. In M. Davis (Ed.) *The Undecidable* (pp. 4-38). New York: Dover Publications.

Gödel, K. (193?/1995). “Undecidable Diophantine Propositions”. In S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, R. N. Soloway (Eds), *K. Gödel, Collected Works III* (pp. 164–175). Oxford: Oxford University Press.

Gödel, K. (1933/1995). “The Present Situation in the Foundations of Mathematics”. In S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, R. N. Soloway (Eds), *K. Gödel, Collected Works III* (pp. 45–53). Oxford: Oxford University Press.

Gödel, K. (1934/1986). “On Undecidable Propositions of Formal Mathematical Systems”. In S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, R. N. Soloway (Eds), K. Gödel, *Collected Works I* (pp. 346–371). Oxford: Oxford University Press.

Gödel, K. (1951/1995). “Some Basic Theorems on the Foundations of Mathematics and Their Implications”. In S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, R. N. Soloway (Eds), *K. Gödel, Collected Works III* (pp. 304–323). Oxford: Oxford University Press.

Gödel, K. (1958/1990). “Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes.” Translated and reprinted in Solomon Feferman et al. (Eds) *Gödel,* *Collected Works*, vol. II (pp. 240-251). Oxford: Oxford University Press.

Granger, G.-G. (1985). « Les deux niveaux de la rationalité. » *Dialectica* 39(4) : 355-363.

Grize, J.-B. (1962). “Remarques sur les limitations des formalismes. ” *Etudes d’Epistémologie génétique* xvi : 103-127.

Kahneman D. & Tversky, A. (1974). “Judgment under uncertainty: Heuristics and biases.” *Science* 185(4157): 1124-1131.

Kenny, A. (1975). *Will, Freedom and Power*. New York: Basil Blackwell.

Koch, C. (2019). *The Feeling of Life Itself Why Consciousness Is Widespread but Can’t Be Computed*. Cambridge: MIT Press.

Koertge, N. (1979). “The methodological status of Popper’s rationality principle.” *Theory and Decision* 10: 83-95.

Krajewski, J. (2020). “On the anti-mechanist arguments based on Gödel’s theorem.” *Studia Semiotyczne* (Semiotic Studies) xxxiv(1): 9–56. https://doi.org/10.26333/sts.xxxiv1.02

Kuczynsk, J.-M. (2006). “ Two Concepts of ‘Form’ and the So-Called Computational Theory of Mind.” *Philosophical Psychology* 19(6): 795–821. <https://doi.org/10.1080/09515080601001887>

Ladrière, J. (1960). “Les limitations des formalismes et leur signification philosophique.” *Dialectica* 14(4): 279-320.

Levy-Bruhl, L. (1910). *Les fonctions mentales dans les sociétés inférieures*. Paris: Felix Alcan

Lindström, P. (2001). “Penrose’s New Argument.” *Journal of Philosophical Logic* xxx: 241-250.

Lucas, J. R. (1961). “Minds, Machines and Gödel.” *Philosophy* 36(137): 112-127.

Luria, A. R. (1976). *Cognitive Development, Its Cultural and Social Foundations*. Boston: Harvard University Press.

Müller, F. M. (1887). *The Science of Thought*. New York: Scribner.

Newel A. & Simon, H. A. (1972). *Human Problem Solving*. Hoboken: Prentice-Hall.

Novak, J. D. (1977). *A Theory of Education.* Ithaca: Cornell University Press.

Penrose, R. (1989). *The Emperor’s New Mind*. Oxford: Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford: Oxford University Press.

Piaget, J. (1967/1971). *Biology and Knowledge. An Essay on the Relations between Organic Regulations and Cognitive Processes*. Chicago: The University of Chicago Press.

Polger, T. W. (2012). “Functionalism as a philosophical theory of the cognitive sciences.” *WIREs Cognitive Science* 3: 337–348.

Popper, K. R. (1972). *Objective Knowledge*. Oxford: Oxford University Press.

Popper, K. R. (1978). “Three Worlds, The Tanner Lecture of Human Values”. Delivered at The University of Michigan, April 7, 1978.

Popper, K. R. (1994). “The Myth of the Framework”. In M. A. Notturna (Ed.) *In Defence of Science and Rationality* (pp. 179-181). London: Taylor and Francis.

Post, E. (1924/1965). “Appendix.” In M. Davis (Ed.) *The Undecidable* (pp. 394-406). New York: Raven Press.

Putnam, H. (1995). “Review of The Shadows of the Mind.” *Bulletin of the American Mathematical Society* 32(3): 370-373.

Richardson, R. C. (1981). “Internal representation: Prologue to a theory of intentionality.” *Philosophical Topics* 12(1): 171-211

Searle, J. (1997). *The Mystery of Consciousness*. New York: New York Review Books.

Shagrir, O. (2006). “Gödel on Turing on computability”. In A. Olszewski, J. Wolenski, & R. Janusz (Eds.), *Church’s thesis after 70 years* (pp. 393–419). Frankfurt: Ontos Verlag.

Shapiro, S. (1998). “Incompleteness, Mechanism, and Optimism.” *Bulletin of Symbolic Logic* iv: 273-302. https://doi.org/10.2307/421032

Simon, H. A. (1957). *Models of Man*. New York: John Wiley.

Simon, H. A. (1976). “From substantive to procedural rationality”. In S. J. Latsis (Ed.), *Method and Appraisal in Economics* (pp. 129-148). Cambridge: Cambridge University Press. Reprinted in Models of Bounded rationality. 2 vols. Cambridge: MIT Press, 1982.

Simon, H. A. (1978). “Rationality as process and product of thought.” *American Economic Review* 68: 1-16. Reprinted in Models of Bounded rationality. 2 vols. Cambridge: MIT Press, 1982.

Simon, H. A. (1982). *Models of Bounded rationality*. 2 vols. Cambridge: MIT Press.

Simon, H. A. (1983). *Reason in Human Affairs*. Stanford: Stanford University Press.

Sinha, C. (1989). “Evolution, development and the social production of mind”. *Cultural Dynamics.* *2*(2), 188-208.

Slezak, P. (1982). “Gödel’s Theorem and the Mind.” *British Journal of Philosophy of Science* 33: 41-52.

Smedslund, J. (1989). “A critique of Tversky and Kahneman's distinction between fallacy and misunderstanding.” *Scandinavian Journal of Psychology* 31:110-120. https://doi.org/10.1111/j.1467-9450.1990.tb00822.x

Tieszen, R. (2006). “After Gödel: Mechanism, Reason, and Realism in the Philosophy of Mathematics.” *Philosophia Mathematica* iii(14): 229–254. https://doi.org/10.1093/philmat/nkj008

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). “Integrated information theory: From consciousness to its physical substrate.” *Nature Reviews Neuroscience* 17: 450–461. https://doi.org/10.1038/nrn.2016.44

Turing, A. M. (1936/1965). “On Computable Numbers, with an Application to the Entscheidungsproblem”. In M. Davis (Ed.) *The Undecidable* (pp. 115-154). New York: Dover Publications.

Turing, A. M. (1950). “Computing Machinery and Intelligence.” *Mind* lix(236): 433-460.

Vygotsky, L. S. & Luria A. R. (1930/1993). *Studies on the History of Behavior, Ape, Primitive, and Child*. Mahwah: Lawrence Erlbaum associates.

Vygotsky, L. S. (1930-1933/1999). “Scientific Legacy”. In R. W. Rieber & J. Wollock (Eds.) *The Collected Works of L. S. Vygotsky*. New York: Springer.

Vygotsky, L. S. (1933/2016). “Play and Its Role in the Mental Development of the Child.” *International Research in Early Childhood Education* 7(2): 3-25.

Vygotsky, L. S. (1934/1986). *Thought and language*. Cambridge: MIT Press.

Weber, M. (1905-1906/1974). “Subjectivity and Determinism”. In A. Giddens (Ed.) *Positivism and Sociology* (pp.23-32). London: Heinemann Educational Books.

Weber, M. (1921/1968). *Economy and Society. An Outline of Interpretive Sociology*. Berkeley: University of California Press.

Wettersten, J. (1995). “Styles of Rationality.” *Philosophy of the Social Sciences* 25(1): 69-98.

Zavershneva, E. I. (2010). “The Vygotsky family archive (1912-1934).” *Journal of Russian and East European Psychology* 48(1): 3-96. https://doi.org/10.2753/RPO1061-0405480101

Zavershneva, E. (2016). “Vygotsky the unpublished: An overview of the personal archive (1912–1934)”. In A. Yasnitsky & R. van der Veer (Eds) *Revisionist Revolution in Vygotsky Studies* (pp. 94-126). London: Routledge.

1. See for instance Bulle (2021) on this subject. [↑](#footnote-ref-1)
2. According to Daniel Dennett (1971), intentional states (involving the ascription of beliefs and desires) do not exist ontologically, they only represent an interpreter’s tactic for predicting the behavior of a complex system, organism or machine. [↑](#footnote-ref-2)
3. Cf. Bulle (2019) for an overview in education. [↑](#footnote-ref-3)
4. Let us specify that the analytical and meta-analytical dimensions are differentiated like syntax from semantics, form from meaning, and not like the metacognitive (applied to the representations of mental contents) from the cognitive (applied to the representations of the world). [↑](#footnote-ref-4)
5. See especially Di Iorio (2016) and Koertge (1979) on world 3 and Popper’s situational logic. [↑](#footnote-ref-5)
6. It is deduced from the Pythagorean theorem that if the sides of a triangle are equal and measure one unit, then the square of the hypotenuse is equal to 2. It thus appeared to Greek mathematicians, against their beliefs, that it was impossible to relate such a measure to that of the side using a rational fraction. The theory developed by Eudoxus in the 4th century B.C. overcame this situation: Reasoning involving ratios that could not be expressed as rational fractions was then conducted in a qualitative manner, without numerical application. [↑](#footnote-ref-6)
7. The fact that people vote cannot be explained within the consequentialism framework since a vote has virtually no chance of influencing the outcome of a popular consultation. [↑](#footnote-ref-7)
8. I mainly rely on Gödel’s introductory presentation of his theorem. All the demonstrations establishing the limitation facts follow this same general scheme (Ladrière 1960). [↑](#footnote-ref-8)
9. Being governed by finite and mechanical procedures is what characterizes a formal system, while mind may involve finite but non-mechanical mental procedures according to Gödel. See Shagrir (2006) for an analysis of Gödel’s disagreement with Turing (1936/1965)’s argument on this subject. [↑](#footnote-ref-9)
10. See Feferman, 2009; Putnam, 1995; Searle, 1997; Shapiro, 1998; Slezak, 1982. [↑](#footnote-ref-10)
11. According to Searle, Penrose fails to make a good objection to such algorithms (applied to the simulation of the meta-analytical dimension of thought): They merely describe brain processes but do not explain them. However, according to Penrose, Searle’s argument according to which the inner aspects of consciousness are not encapsulated by computation is too feeble, because a computational description of higher mental processes is not even possible. [↑](#footnote-ref-11)
12. See Feferman, 2009; Krajewski, 2020; Kuczynski, 2006; Lindström, 2001; Shapiro, 1998. [↑](#footnote-ref-12)
13. The conscious understanding involved in Gödel’s notion of meaning may be highlighted by his statement that rules of inference in formal systems refer “only to the outward structure of the formulas, not to their meaning so that they could be applied by someone who knew nothing about mathematics, or by a machine” (Gödel, 1933/1995, p. 45). [↑](#footnote-ref-13)
14. Post (1924/1965), p. 399; see also Krajewski (2020), p. 26. [↑](#footnote-ref-14)
15. This conception of the basic rational powers of the human mind, which refers primarily to Vygotsky’s historical-cultural psychology here, can be found in other works in a similar or not too distant form, but which may assume different relations between language and thought. For instance, the philologist Friedrich Max Müller (1887, p. 200) develops the idea that human thought cannot exist without words, that is, signs for concepts involving forms of generalization. On these bases, he states that reason is *neither more nor less than the faculty, or if we dislike that word, the act of forming and handling concepts*, which implies a too narrow kinship of nature between thought and language. The philosopher of mind Anthony John Kenny (1975, pp. 3-5) writes on his part that to have a mind is to have the capacity to acquire the *ability to [self-consciously] operate with symbols in such a way that it is one’s own activity that makes them symbols and confer meaning on them,* which has the disadvantage of introducing an unclear logical link between meaning and activity. [↑](#footnote-ref-15)
16. In a similar way, according to Kenny’s (1975, p. 2) definition of mind as the ability to acquire intellectual capacities, babies have a mind even though they cannot yet create and use symbols, whereas even gifted chimpanzees do not have a mind in this sense. [↑](#footnote-ref-16)
17. The global workspace theory assumes that what we experience as consciousness is the global sharing of information (see Baars, 1997, Dehaene, 2014). Nevertheless, the computational perspective endorsed (Dehaene, Lau & Kouider 2017, p. 492) cannot account for the very role of meaning involved here in the evolution of meaning structures. [↑](#footnote-ref-17)
18. On this subject, the integrated information theory (IIT, see for instance Totoni et al., 2016) of consciousness initiates an interesting approach by starting from essential properties of the conscious experience, which entails that the physical substrate of consciousness has causal power upon itself (i.e., independent of extrinsic causes and effects), and disqualifies computational interpretations of consciousness that cannot simulate such causal power (Koch, 2019, p. 149). [↑](#footnote-ref-18)
19. A major implication of Ausubel’s theory of meaningful verbal learning is that “the learner’s acquisition of clear stable, and organized bodies of knowledge is not only the major long-term objective of education, but also that the learned properties of these bodies of knowledge, once acquired, constitute in their own right, and in their turn, the most significant independent variables influencing the meaningful learning and retention of new subject matter material.” (Ausubel 2000, p. 42). [↑](#footnote-ref-19)
20. The idea of the intrinsic role of signs in thought was defended by the Abbé Etienne Bonnot de Condillac, who is not well known for this essential contribution to human psychology, developed especially in his *Essay on the origin of human knowledge* (Condillac, 1746/2001). Its links with Vygotsky’s psychology are thus generally ignored (Hardcastle, 2009; Sinha, 1989). [↑](#footnote-ref-20)