2019 Isonomia – Epistemologica
Rivista online di Filosofia
Università degli Studi di Urbino Carlo Bo

Bounded Rationality and Heuristics in Humans and in Artificial Cognitive Systems

Antonio Lieto
Università di Torino, Dipartimento di Informatica
ICAR-CNR, Palermo
lieto@di.unito.it

Abstract

In this paper I will present an analysis of the impact that the notion of "bounded rationality", introduced by Herbert Simon in his book *Administrative Behavior*, produced in the field of Artificial Intelligence (AI). In particular, by focusing on the field of Automated Decision Making (ADM), I will show how the introduction of the cognitive dimension into the study of choice of a rational (natural) agent, indirectly determined - in the AI field - the development of a line of research aiming at the realisation of artificial systems whose decisions are based on the adoption of powerful shortcut strategies (known as *heuristics*) based on "*satisficing*"- i.e. non optimal - solutions to problem solving. I will show how the "heuristic approach" to problem solving allowed, in AI, to face problems of combinatorial complexity in real-life situations and still represents an important strategy for the design and implementation of intelligent systems.

Keywords: Bounded rationality, Heuristics, Decision-making, Commonsense reasoning, Artificial Intelligence

1. Introduction

The notion of bounded rationality was introduced by Herbert Simon[1] in his book *Administrative Behavior* as a constitutive element of a plausible model

[1] Herbert Simon was one of the founding fathers of the AI and, along with people like Marvin Minsky, John McCarthy, Allen Newell, Nathaniel Rochester and many others, was one of the participants to the Darmouth Workshop providing the foundation of the

of human problem solving and decision making. This notion proposed, for the first time, a radically different and alternative view with respect to the idealised model of "*homo oeconomicus*" which was dominant, in that times, within the classical economic theory of decision making. Roughly speaking, the notion of bounded rationality refers to the cognitive limits (e.g. perceptual, of memory, reasoning, processing etc.) that human decision makers have, and have to face, when they have to make decisions in real life situations. Such limits lead humans to adopt "bounded", i.e. non optimal (but "*satisficing*"), solution strategies that can be cognitively dealt with. As mentioned, this approach diverged from the classical theory of decision making according to which humans were seen as perfectly rational decision makers, able to make optimal decisions via a maximisation of their expected utility in any give situation[2]. As I will argue in this paper, the introduction of this notion influenced not only economic model of decision making but also the field of the early Artificial Intelligence, an area in which Simon was directly involved since its foundations.

In the following, I will present two case studies in the field of automated reasoning with the intent of showing, at least in part, the impact that such notion had in this field of AI research. Before presenting these particular case studies, however, I will propose a brief historical digression concerning the influence that such idea had in the development of the early AI systems and formalisms.

2. The Early days of AI

In the early days of Artificial Intelligence, whose date of birth is historically associated to the summer of 1956, during the so called "Dartmouth

discipline. Simon was trained as a psychologist and awarded by the Nobel Prize in Economy for his studies on the "bounded rationality" in human decision making. He was one of the main scholars pointing out, in both cognitive psychology and AI, the role played by the heuristics as decisional shortcuts to solve complex problems (for this line of research he was also awarded, with Allen Newell, by the Turing Award in 1975; see Newell & Simon, 1976). On the particular meanings attributed to the term "heuristics" in the AI research, we remind the reader to the following sections.

[2] The theory of the expected utility was introduced in game theory by Von Neumann and Morgenstern (see Morgenstern and Von Neumann, 1953). In general, in this setting, preference relations are usually modelled by means of utility functions defined on a set of alternatives with values in a suitable set of numbers (usually, real numbers). The decision model of this theory assumes an optimal (i.e. unbounded) decision maker able to calculate and chose, in each phase of problem solving, the move that maximise the utility function.

Workshop”[3], the research on intelligent machines was strongly and explicitly inspired by that one coming from the experimental research in Psychology[4]. A manifestation of this connection is represented by the early AI systems/frameworks in AI developed until the ‘80s of the last century: most of the them were developed according to a “cognitively-oriented” inspiration. In the following we just mention few examples of such systems. and formalisms.

3. From the General Problem Solver to the Cognitive Architectures

One of the first AI systems, developed from the end of the ‘50s of the last century, is represented by the pioneering work of Herbert Simon, John Clifford Shaw and Allen Newell on the *General Problem Solver (GPS):* a system able to demonstrate simple logic theorems whose decision strategies were explicitly inspired by human verbal protocols[5]. The underlying idea of this approach was that, the computer system had to approximate the decision operations described by the humans in their verbal descriptions as closely as possible. In this way, when the program ran on the computer, it would have been possible to identify its problems, compare them with the description of the human verbalisation and modify them to improve its performance. In particular, the GPS system was able to implement a key mechanism in human problem solving: the so called means-ends analysis (or **M-E heuristic**). In M-E analysis the problem solver compares the current situation with a goal situation; computes the difference between the two states; finds in memory an operator that experience has taught reduces differences of this kind; and applies the operator to change the current situation. Repeating this process, the goal may gradually be attained via a reduction of the search space(although, computationally, there are generally

[3] This event was formally organised by John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon. The workshop run for different weeks and saw the participation of many researchers interested in this emerging discipline. An archive of the Darthmouth workshop, based on the notes of Ray Solomonoff (one the participant), is available on line at the following url: http://raysolomonoff.com/dartmouth/.

[4] It must be noted that, at those times, there was not a “Cognitive Science” or a “Cognitive Psychology” field. However, all the disciplines (philosophy, psychology, computer science, anthropology, linguistics and neurophysiology) and the cultural elements that would have been later formed the interdisciplinary field of “Cognitive Science” were already present.

[5] Newell *et al*. (1959). This technique in also known as “thinking aloud" protocol in the psychological literature and consists in recording the verbal explanations provided by the people while executing a given laboratory task.

no guarantees that the process will succeed). This kind of heuristic was used to solve, also in the decades to come, problems in a number of AI domains (from planning to diagnostics). In order to be executed, in fact, it “only” requireda domain representation of the problem to solve (a problem space), operators to move through the space, and information about which operators were relevant for reducing which differences[6].

In the decade following the development of the GPS, i.e. in the ‘60s, another influential contribution in AI of cognitive inspiration was due to Ross Quillian (a Ph.D. student of Herbert Simon at the Carnegie Institute of Technology, now Carnegie Mellon University) that invented the *Semantic Networks,* a psychologically plausible model of human semantic memory implemented in a computer system. The idea of Quillian was that human memory was associative in nature and that concepts were represented as sort of nodes in graphs and activated through a mechanism of “spreading activation”, allowing to propagate information through the network to determine relationships between objects. In this setting, the highest was the activation of a node in the network, the more contextually relevant that node/concept was assumed to be for the task in focus (e.g. retrieval, reasoning, recall, planning etc.). Also in this case, the cognitive limit concerning the storage capacity of our memory triggered the inquiry about the underlying mechanisms governing our memory system. As a consequence, this led to the development of a cognitively inspired model of knowledge representation and retrieval. Evolution of the semantic networks are still nowadays widely used in the subfield of AI known as Knowledge Representation (and in particular in systems known as computational ontologies, that will be better introduced later in the paper).

In the ‘70s, another well know example of cognitively-inspired AI framework was represented by the introduction of the *Frames* operated by Marvin Minsky[7]. This framework was used for representing, in artificial systems, commonsense knowledge (including *default* knowledge) about the external world. This type of knowledge organization proposed in the Frames enabled the first AI systems to extend their automated reasoning abilities from the classical deduction to more complicated forms of commonsense

[6] The ingredients required for the execution of this kind of heuristic strategy - essentially based on a “search space” approach to problem solving - explicitly supported the so called “symbolic approach” to the study, the analysis, the execution and the replication of intelligent behaviour.

[7] Minsky (1975).

and defeasible reasoning[8]. In this case, the idea of the Frames was directly inspired by the work of the psychologist Eleanor Rosh about the organization of conceptual information in humans known as prototype theory[9].

In the ‘80s, finally, other notable examples of the so called “cognitivist” tradition in Artificial Intelligence were represented by: i) the invention of the notion of *Script (Shank and Abelson)* as a data structure for representing knowledge of common sequences of events (e.g. the sequence of events used to go out for a dinner) and used in natural language processing systems as way to enable intelligent to answer questions about simple stories[10] or by ii) the work operated by Allen Newell and colleagues at Carnegie Mellon University on the first cognitive architecture for general intelligence: SOAR[11].

This list of examples is, of course, not exhaustive but it can be useful to indicate the main assumptions made by the early AI researchers and synthesized by Pat Langley[12] in this way: “(early) AI aimed at understanding andreproducing in computational systems the full range of intelligent behavior observed by humans”[13]. The underlying assumption of that period was that, in order to reach this broad goal, the only way to proceed in the analysis of intelligent systems (natural and artificial) was that

---

[8] All the forms of commonsense reasoning can be seen as abounded rationality phenomenon since they represent a plethora of shortcuts allowing us to make decisions in an environment with incomplete and uncertain information.

[9] According to the Rosh’s theory, concepts, the building block of our thoughts, are organised in our mind in terms of prototypes (i.e. in terms of representative elements of the category) and such organization explain many types of so called typicality effects (i.e. of commonsense inferences) that we naturally perform in our everyday reasoning.

[10] A classical example to explain the notion of “Script” is the so called “restaurant situation”. Let us consider to model a situation of an agent going out to a restaurant for dinner. A script representing the restaurant situation is a data structure that would record the typical events associated to this scenario: e.g. entering in the restaurant, asking for a table, sitting down, consulting a menu, eating the food, pay the check etc. This kind of representational structure enabled early AI system to answer questions to simple stories. For a story like this, for example: “Mary went to a restaurant, ordered a salmon. When she was paying she noticed that she was late for the next appointment”, computerised systems were able to answer to a question like: “Did Mary eat dinner last night?” in a positive way (as we do). It is worth-noticing that this information is not explicitly provided in the story. Answering to these types of questions was possible through the use of a script of the restaurant situation.

[11] Laird *et al*. (1987). On the role of the Cognitive architectures for general intelligent systems we remind to (Lieto *et al*., 2018).

[12] Langley (2012).

[13] Ibid.

one of discovering the heuristic strategies adopted by humans (or other animals) that could have been later operationalised and replicated in artificial systems.

It is worth-noticing, however, that this kind of approach to the "study of the artificial" did not came out *ex-abrupto.* It borrowed its original inspiration –from a historical perspective–from the methodological apparatus developed by the scholars in Cybernetics[14].

The origins of cybernetics are usually traced back to the middle of the 1940s, with the release of the 1948 book of Norbert Wiener entitled *Cybernetics: Or Control and Communication in the Animal and the Machine*. One of underlying idea of cybernetics was that one of building mechanical models to simulate the adaptive behavior of natural systems. As indicated in Cordeschi[15]: "the fundamental insight of cybernetics was in the proposal of a unified study of organisms and machines". In this perspective, the computational simulation of biological processes was assumed to play a central epistemological role in the development and refinement of theories about the elements characterizing the nature of intelligent behaviour in natural and artificial systems. Such simulative approach, as mentioned, was inherited by the early AI research and the adoption of such perspective was crucial for the development of both intelligent solutions inspired by human processes and heuristics[16] and for the realisation of computational models of cognition realized with the aim of providing a deeper understanding of human thinking, as originally suggested in the manifesto of the Information Processing Psychology (IPP)[17]. These two sides of the cognitivist tradition are nowadays still alive. They correspond, roughly, to the research areas of Cognitive Systems or cognitively-inspired AI (Cognitive-AI) and to that one of cognitive modelling (or Computational Cognitive Science) respectively.

4. Heuristics and AI Eras

The notion of heuristics deserves, in this historical digression, and in the light of the purposes of this paper, a special attention. It was used, in fact, with two different meanings since the times of the first AI researches. In its first sense, the term referred to the most detailed simulation possible of human cognitive processes, and it characterized the above mentioned

[14] Cordeschi (1991).
[15] Cordeschi (2002).
[16] Newell & Simon (1976), Gigerenzer & Todd (1999).
[17] Newell & Simon (1972).

Information Processing Psychology (IPP) introduced by Newell and Simon. In this view, a computer program was considered as a model providing a test of the hypothesis that the mind is an information-processing system. More precisely, “the program was considered to be a highly specific behavioral theory, concerning the behavior of an individual human problem-solver: a microtheory”[18].

In another sense, the term referred to the possibility of obtaining the most efficient (and efficacious) performance possible from computer programs, by allowing also for typically non-human procedures, such as those where the computer can excel. Before the introduction of the term “heuristics” in AI, operated by Newell, Shaw and Simon, there were already available algorithmic procedures which might have been defined as heuristic in the second of these senses had already been tried out experimentally. The first among them were the procedures that allowed the program developed by Arthur Samuelto play chess despite the combinatory explosion of moves[19].

The fact that these two tendencies, reflected in this double meaning of the term “heuristics”, coexisted in AI was immediately clear. As reported in Cordeschi (2002: 190), in 1961, while discussing a presentation of GPS given by Simon during a seminar at MIT, Minsky drew a distinction in AI research between those who were willing to use "non-human techniques" in constructing intelligent programs and those, like the Carnegie-Mellon group, who were interested in simulating human cognitive processes[20]. This distinction is crucial, since outlines the raise of different research agendas already in those years. This divergence started to be very significant in the mid’80s of the last century. After decades of pioneering collaborations, in fact, several sub-fields, each with its own goals, methods and evaluation criteria were produced. On the one hand this fragmentation led AI to reach remarkable results in a variety of narrow fields by focussing on quantitative results and metrics of performance, and on a machine-oriented approach to

[18] Cordeschi (2002: 182). In this view, the general theory of human information-processing was assumed to be derivable from a body of qualitative generalizations coming from the study of individual simulative programs, or microtheories.

[19] Samuel (1959).

[20] As reported in Cordeschi (2002), Minsky emphasized that these two tendencies were distinguished “in methods and goals” from a third tendency, that which “has a physiological orientation and alleges to be based on an imitation of the brain,” i.e. neural net and self-organizing system approaches. This tendency would have lead to the development of the neural and brain inspired methods which are the basis of the so-called “connectionist agenda” (nowadays wide-spreading due to the recent success of deep learning).

the intelligent behaviour (i.e., without taking into account human-inspired heuristics). On the other hand, however, it significantly inhibited the cross-field collaborations and the research efforts targeted at investigating a more general picture of what natural and artificial intelligence is, and how intelligent artifacts can be designed by taking into account the insights coming from human cognition.

In the last few years, however, the cognitive approach to AI gained a renewed consideration, both from academia and industry, in wide research areas such as Knowledge Representation and Reasoning, Robotics, Machine Learning, Bio-Inspired Cognitive Computing, Computational Creativity and further research fields that aspire to Human Level Intelligence. Nowadays, in fact, artificial systems endowed with human-like and human-level intelligence[21] are still far from being achieved and, as I will try to show in the following section, the adoption of cognitively-inspired heuristics (all directly based on bounded rationality assumption) seems to be a suitable way to handle problems that non-human oriented algorithmic processes are not able to deal with.

In the next section, I will decline this argument by presenting two case studies in the field of knowledge representation and automated reasoning.

5. Cognitive heuristics in modern AI: two case studies in Automatic Reasoning

The representation of conceptual information, and the corresponding reasoning mechanisms that can be built on such representations, is a crucial problem to deal with in the subfield of AI known as Knowledge Representation (KR). In this field, one of the most widespread system currently used are the so called computational ontologies, defined as "an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words"[22].

The main building blocks of ontological models are, therefore, concepts (or classes), roles (or properties), and individuals describing a given domain. In other words: ontologies provide an explicit and axiomized reference domain model. Such model is used to interpret and organise the information and to perform simple forms of automatic reasoning.

[21] McCarthy (2007).
[22] Guarino (1998).

Ontological models are built on by using a class of logical formalisms that are known as Description Logics (DLs)[23]. Technically, DLs are decidable subsets of first-order logic in which a standard Tarskian semantics is directly associated to the syntax of the language. The decidability is usually obtained by restricting the set of variables and operators that are allowed inside formulae. In DL systems, inferential knowledge is formally expressed by means of “terminological axioms”, which are simply notational variants of classical meaning postulates - i.e. universally quantified (bi)conditional statements that constrain the extensions of the constant that appear in the antecedent (e.g. see the figure below).

$$
\begin{aligned}
\mathit{Woman} &\equiv \mathit{Person} \sqcap \mathit{Female} \\
\mathit{Man} &\equiv \mathit{Person} \sqcap \neg \mathit{Woman} \\
\mathit{Mother} &\equiv \mathit{Woman} \sqcap \exists \mathit{hasChild}.\mathit{Person} \\
\mathit{Father} &\equiv \mathit{Man} \sqcap \exists \mathit{hasChild}.\mathit{Person} \\
\mathit{Parent} &\equiv \mathit{Father} \sqcup \mathit{Mother} \\
\mathit{Grandmother} &\equiv \mathit{Mother} \sqcap \exists \mathit{hasChild}.\mathit{Parent} \\
\mathit{MotherWithoutDaughter} &\equiv \mathit{Mother} \sqcap \forall \mathit{hasChild}.\neg \mathit{Woman} \\
\mathit{MotherInTrouble} &\equiv \mathit{Mother} \sqcap \geq 10\, \mathit{hasChild}
\end{aligned}
$$

Figure 1. Examples of terminological axioms in Description Logics

The meaning postulates approach assumed in the tarskian semantics of Description Logics, however, is not immune from problems. For instance, it is known that meaning postulates can formally express only rigid relations between concepts, i.e. relations that do not allow for exceptions (e.g. if x is a bachelor, than x is not married). As a consequence, standard DL system and ontologies are not capable to model prototypical knowledge and defeasible inference[24], which - on the other hand - represents the way in which the humans encode and reason on conceptual information.

In the field of logic-oriented KR, various non-monotonic extensions of DL formalisms have been designed to deal with some aspects of typicality and automated commonsense reasoning[25]. A proposed solution has been that one of integrating a set of meaning postulates with a non-monotonic logic

[23] Badeer *et al*. (2010).
[24] Lieto (2012).
[25] E.g. Bonatti *et al*. (2006), Lukasiewicz & Straccia (2008), Giordano *et al*. (2013).

like Default Logics[26]. This is the solution adopted by Baader & Hollunder (1995); the same authors, however, point out both the computational difficulties of this integration. Current literature also offers several attempts to formulate DLs extensions based on fuzzy logic[27] and some "probabilistic" versions of DL languages have been also proposed, which try to combine the DL language with probabilistic theories based on Bayesian formalisms[28]. Despite these computational advancements, however, the use of non-monotonic and, in general, non-classical logics in the field is not universally accepted due to the high computational complexity, and therefore their practical intractability, intrinsic to such formalisms. In other words: the use of non-human inspired procedures does not seem to provide any practical advantage for modelling the problem of commonsense reasoning related to typicality.

An alternative move was proposed by adopting a cognitively-inspired system explicitly implementing a plethora of cognitive assumptions and heuristics for commonsense reasoning. This has recently led to the development of a system called DUAL-PECCS (Dual Prototype and Exemplars Based Conceptual Categorization System)[29], able to combine on a large scale, both commonsense representation and reasoning with standard ontological DL-based semantics. The main merit of such proposal lies in the adoption of the representational component of Conceptual Spaces (a geometric and cognitively inspired framework for concept representation and formation, see Gärdenfors, 2000) that has been successfully integrated with ontological formalisms[30].

DUAL PECCS, explicitly implement the following cognitive assumptions: i) the fact that conceptual representations are heterogeneous co-referring representational structures demanded to different computational frameworks and ii) the fact that the different reasoning strategies executed on such representations are harmonised according to the dual process theory or reasoning and rationality[31].

---

[26] Reiter (1980).
[27] E.g., Caligari *et al*. (2007).
[28] Ding *et al* (2006).
[29] Lieto *et al*. (2015), Lieto *et al.* (2017).
[30] The solution adopted in DUAL-PECCS is one of the three proposals currently suggested to overcome the current problems affecting the knowledge level of cognitive artificial systems. Alternative proposals are represented by the Semantic Pointers (Eliasmith *et al*., 2012) and by the neuro-symbolic approaches à la ACT-R (Anderson *et al*., 2004). A detailed comparative analysis among the different approaches is in Lieto *et al*. (2018b).
[31] For the details see Lieto (2014) and Lieto (2019).

The heterogeneous conceptual architecture of DUAL PECCS currently includes "typical" commonsense representations (i.e. prototypes and exemplars) and classical "rigid" representations allowed by the Description Logics Semantics. All these different bodies of knowledge point to the same conceptual entity[32]. An example of the heterogeneous conceptual architecture of DUAL PECCS is provided in the figure 2 below. Such figure shows how it is represented the concept DOG. In this case, the prototypical representation grasps information such as that dogs are usually conceptualized as domestic animals, with typically four legs, a tail etc.; the exemplar-based representations grasp information on individuals. For example, it is represented the individual of Lessie, which is a particular exemplar of dog with white and brown fur and with a less domestic attitude w.r.t. the prototypical dog (e.g. its typical location is lawn).

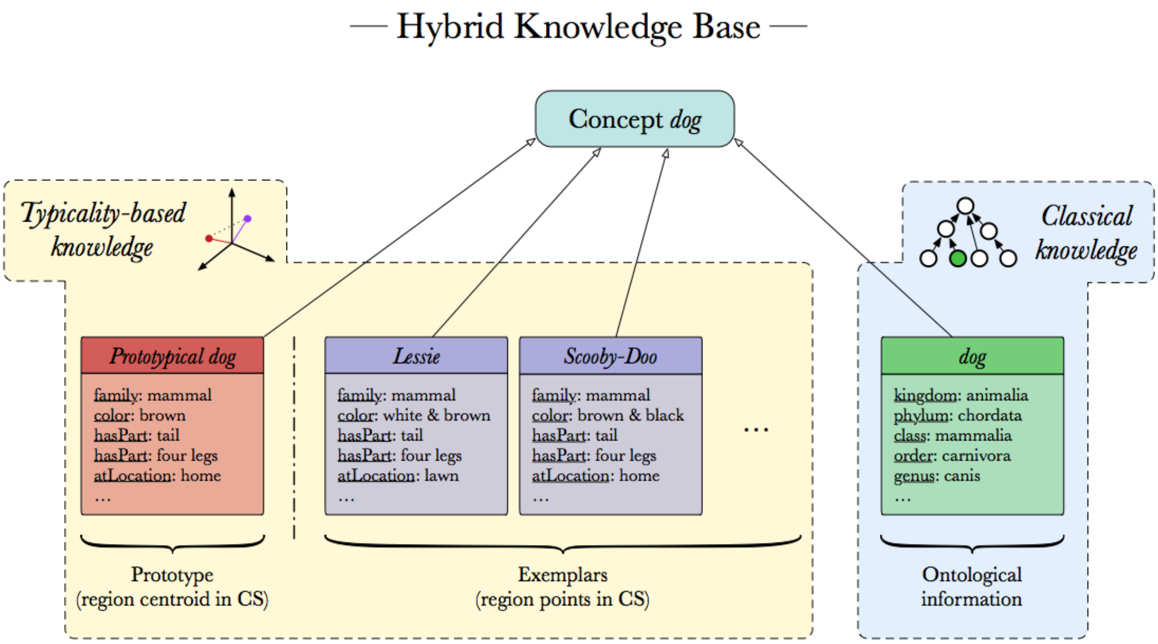

Figure 2. An example of the hybrid conceptual architecture in DUAL PECCS.

Within the system, both these two types of typicality based representations are represented by using the framework of the conceptual spaces proposed by Gärdenfors[33]. Such framework allows to adopt standard

---

[32] The procedures for the automatic anchoring for these different types of representations are described in Lieto *et al*. (2016).

[33] Gärdenfors (2000).

similarity metrics to determine the distance between instances and concepts within the space. The representation of classical information (e.g. the fact that Dog $\sqsubseteq$Animal , that is to say that "Dogs are also Animals") is, on the other hand, demanded to standard ontological formalisms.

From a reasoning perspective, in DUAL-PECCS, different kinds of reasoning strategies are associated to these different bodies of knowledge. In particular, the system combines non-monotonic commonsense reasoning (executed on conceptual spaces representations and based on measures of "semantic similarity") and standard monotonic categorization procedures (executed on the classical, ontological, body of knowledge). These different types of reasoning are harmonized according to the theoretical tenets coming from the dual process theories of reasoning and rationality according to which fast - non monotonic - reasoning is executed first and more logical - deliberative - reasoning is executed later.

Given this explicit cognitive assumptions, DUAL PECCS has been tested in conceptual categorisation of commonsense linguistic descriptions similar to riddles (i.e. of the form "which is the big feline with yellow fur and black stripes?" etc.).

This kind of task is a very difficult one from an AI perspective since, in commonsense reasoning, also state-of-the-art AI systems like IBM Watson obtain very poor results[34]. On the other hand, the results obtained by DUAL PECCS are promising both when compared with human performances (with an overlapping of the 89% of the responses, see Lieto *et al*., 2017 for the results in details) and when compared to other artificial systems like Google, Bing or Wolphram Alpha[35]. An additional advantage of the proposed approach lies in the fact that it limits the computational complexity of the commonsense reasoning strategy to linear time (i.e. O(n)), since the mechanism of typicality based categorisation technically corresponds to measuring the semantic relatedness of a query vector - in which the linguistic input is transformed - with respect to a vectorial knowledge-base of conceptual spaces (and this process can be solved in linear time with respect to the size of the knowledge base).

As a consequence, this case study seems to suggest that, also in the modern AI, the adoption of bounded rationality heuristics in the design and implementation of intelligent systems can represent an important aspect to consider in order to progress forward more human-like and human-level AI systems.

[34] See Davis & Marcus (2015).

[35] See Lieto *et al*. (2017b) for a detailed analysis of the obtained results.

As a second case study, which also related to the above mentioned problem of commonsense representation and reasoning in AI systems, let us consider the problem of obtaining, via compositionality heuristics, a commonsense compound concept obtained by combining the typical knowledge of the composing concepts.

Dealing with such ability requires, from an AI perspective, the harmonization of two conflicting requirements that are hardly accommodated in symbolic systems: the need of a syntactic compositionality (typical of logical systems) and that one concerning the exhibition of typicality effects[36]. According to a well-known argument[37], in fact, prototypical concepts are not compositional. The argument runs as follows: consider a concept like PET FISH. It results from the composition of the concept pet and of the concept fish. However, the prototype of pet fish cannot result from the composition of the prototypes of a pet and a fish: e.g. a typical pet is furry and warm, a typical fish is grayish, but a typical pet fish is neither furry and warm nor grayish (typically, it is red).

Recently, a logical framework able to account for this type of human-like concept combination has been proposed in the field of nonmonotonic Description Logics of typicality: $\mathbf{T}^{\mathbf{CL}}$ (Typicality-based Compositional Logic, for the details see Lieto & Pozzato, 2018).

This logic combines three main ingredients. The first one relies on the DL of typicality ALC + TR introduced in Giordano *et al*. (2015). In this logic, “typical” properties can be directly specified by means of a “typicality” operator T enriching the underlying DL, and a knowledge base (KB) can contain inclusions able to represent that “typical Cs are also Ds”. In ALC + TR one can consistently express exceptions and reason about defeasible inheritance as well.

A second ingredient is represented by a distributed semantics similar to the one of probabilistic DLs known as DISPONTE[38], allowing to label ontological axioms with degrees representing probabilities, but restricted to typicality inclusions. The basic idea is to label inclusions of the type “typical Cs are also Ds” with a real number between 0.5 and 1, representing its probability, assuming that each axiom is independent from each others (the actual probabilistic values are assumed to come from an application domain). The resulting knowledge base defines a probability distribution over scenarios.

---

[36] See Lieto (2012) and the references in Lieto *et al*. (2018) for a contextualized overview of such a problem.
[37] Osherson & Smith (1981).
[38] Riguzzi *et al*. (2015).

The third element of the proposed formalization is represented by the adoption of an heuristics method inspired by cognitive semantics[39] called HEAD-MODIFIER heuristics. Such heuristics allows to characterise a dominance effect between the concepts to be combined. In particular, for every combination, it allows to distinguish a HEAD, representing the stronger element of the combination, and one (or more) MODIFIERs. The basic idea is: given a KB and two concepts CH (HEAD) and CM (MODIFIER) occurring in it, only some scenarios are considered in order to define a revised knowledge base, enriched by typical properties of the combined concept.

The selection criteria of such scenario is obtained as follows: given a KB K and given two concepts CH and CM occurring in K, $\mathbf{T}^{\mathbf{CL}}$ allows defining the compound concept C as the combination of the HEAD (CH) and the MODIFIER (CM), where C $\sqsubseteq$ CH $\sqcap$ CM.

The typical properties of the form T(C) $\sqsubseteq$ D to ascribe to the concept C are obtained in the set of scenarios, obtained by applying the DISPONTE semantics, that: i) are consistent with respect to K; ii) are not trivial (i.e. those with the highest probability, in the sense that the scenarios considering all properties that can be consistently ascribed to C, or all the properties of the HEAD that can be consistently ascribed to C are discarded); iii) are those giving preference to the typical properties of the HEAD CH (with respect to those of the MODIFIER CM) with the highest probability.

An additional element in $\mathbf{T}^{\mathbf{CL}}$, inherited from the representational assumptions considered in DUAL PECCS, is that a KB in such a formalism combines both "rigid" and "typical" knowledge.

In this way, as shown in detail in Lieto & Pozzato (2018), it is possible to model phenomena like the above described PET FISH composition. In particular, this problem is modelled as follows: let K be a Knowledge base containing the rigid inclusion (*) Fish $\sqsubseteq \forall$ livesIn.Water and the following typical inclusions equipped with probabilities:

1. $0.9 :: \mathbf{T}(Pet) \sqsubseteq \forall livesIn.(\neg Water)$
2. $0.8 :: \mathbf{T}(Pet) \sqsubseteq Affectionate$
3. $0.7 :: \mathbf{T}(Fish) \sqsubseteq \neg Affectionate$
4. $0.8 :: \mathbf{T}(Pet) \sqsubseteq Warm$
5. $0.6 :: \mathbf{T}(Fish) \sqsubseteq Greyish$
6. $0.9 :: \mathbf{T}(Fish) \sqsubseteq Scaly$
7. $0.8 :: \mathbf{T}(Fish) \sqsubseteq \neg Warm$

[39] See Hampton (1987) for a review.

In this case, by applying the DISPONTE semantics, we have $2^7 = 128$ different scenarios. In $\mathbf{T}^{\mathbf{CL}}$, inconsistent scenarios are discarded along with those that are trivial and privilege the MODIFIER with respect to the HEAD. It turns out that the logic $\mathbf{T}^{\mathbf{CL}}$ is able to select the scenario with the following typical properties (which are those required for handling the PET FISH prototypical composition):

$$\begin{array}{l} 3.\ 0.7 :: \mathbf{T}(\mathit{Fish}) \sqsubseteq \neg \mathit{Affectionate} \\ 6.\ 0.9 :: \mathbf{T}(\mathit{Fish}) \sqsubseteq \mathit{Scaly} \\ 7.\ 0.8 :: \mathbf{T}(\mathit{Fish}) \sqsubseteq \neg \mathit{Warm} \end{array}$$

On the other hand, the composed concept PET FISH also inherits the rigid inclusion $\sqsubseteq \forall$ livesIn.Water from (*).

In addition, in such a logical framework, adding a new inclusion **T** (PET $\sqcap$ FISH) $\sqsubseteq$ Red, would not be problematic (i.e. this logic tackles the phenomenon of prototypical attributes emergence). The proposed logic has been recently applied to a number of cognitive phenomena including: conjunction fallacy, metaphors generation, and iterative conceptual combination[40].

An important element emerging from this particular application, and relevant with respect to the overall overview proposed in this paper, lies in the fact that it would have been not possible to model all such phenomena without considering all the 3 ingredients of such logic (including the HEAD-MODIFIER heuristics). In other words, all such elements are individually necessary but only jointly sufficient to tackle a complex problem like the one described above. This is a symptom of the fact that, the application (and integration) of cognitive heuristics in artificial systems and formalisms, can still play an important role in AI.

Concluding Remarks

In this paper I have argued that the notion of bounded rationality heuristics has influenced, directly or indirectly, the theoretical and applied research in the field of Artificial Intelligence. I have provided a brief historical contextualization to the rise, the fall and the recent renewed interest in the design approach based on cognitive heuristics and historically ascribable to

[40] Lieto & Pozzato (2019).

the cognitivist tradition in AI. Finally I have provided two simple case studies in the field of ADM by showing that the role that cognitive heuristics can play in the so-called "science of the artificial"[41] is still a relevant one in the years to come. In particular, a cognitive-based approach to development of AI systems seems to be still relevant in all those tasks that are easily solvable for humans but very hard to solve for machines (and the problem of commonsense reasoning is paradigmatic in this perspective). Other relevant research areas that may benefit from this kind of approach concern, for example, the problems of: analogical reasoning; learning from few examples (differently from what current deep learning techniques do); transfer learning; multimodal perception and attentive mechanisms; computational creativity and knowledge invention; narrative and story understanding; integration of mechanisms involving planning, acting, monitoring and goal reasoning; emotion modelling and the so called area of explainable AI or XAI (a novel name for a very old problem consisting in truing to provide human-understandable explanations of algorithmic decisions). As this illustrative lists shows, such problems involve both low-level (e.g. perceptual) and high level (e.g. reasoning) cognitive capacities (usually modelled in AI by adopting connectionist and symbolic approaches respectively). That is to say that the cognitive approach is agnostic with respect to the classes of formalisms applied to model a given phenomenon and can be applied to both the symbolic and the connectionist research agendas, as the recent history of Cognitive Science has successfully showed in the last 30 years.

Acknowledgements

The content of this paper has benefited from many discussions with Roberto Cordeschi and Marcello Frixione. Of course eventual errors can be ascribed only to myself. I am also indebted to the feedback and comments received by Amedeo Cesta, Antonio Chella, Fabio Paglieri, Oliviero Stock and Giuseppe Trautteur during the panel
"Can AI and Cognitive Science still live together happily ever after?"
(link: http://aiia2017.di.uniba.it/index.php/joint-panel-aiia-and-aisc/)
organised at the international conference AI*IA 2017 in Bari. The work on DUAL PECCS and on the $\mathbf{T}^{\mathbf{CL}}$ logic has been carried out with Daniele Radicioni, Valentina Rho and Gian Luca Pozzato.

[41] Simon (1981).

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004), "An Integrated Theory of the Mind", in *Psychological review*, 111(4), 1036.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (2010), *The Description Logic Handbook: Theory, Implementations and Applications* , 2nd edition, Cambridge, University Press.

Baader, F., & Hollunder, B. (1995), "Embedding Defaults into Terminological Knowledge Representation Formalisms", in *J. Autom. Reasoning* 14, 1: 149–180.

Bonatti, P. A., Lutz C., & Wolter, F. (2006), "Description Logics with Circumscription", in *Proceedings of KR*, pp. 400–410.

Calegari, S., & Ciucci, D. (2007), "Fuzzy Ontology, Fuzzy Description Logics and Fuzzy-OWL", in *Proc. WILF 2007* , LNCS, volume 4578.

Cordeschi, R. (1991), "The Discovery of the Artificial. Some Protocybernetic Developments", in 1930–1940. *AI & society*, 5 , 218–238.

Cordeschi, R. (2002), *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*, Springer Science & Business Media, Volume 28.

Davis E., & Marcus, G. (2015), "Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence", in *Communications of the ACM*, 58(9):92–103, 2015.

Ding, Z., Peng, Y., & Pan, R. (2006). "Bayes OWL: Uncertainty Modeling in Semantic Web Ontologies", in *Soft Computing in Ontologies and Semantic Web , Z. Ma* (ed.), Studies in Fuzziness and Soft Computing, Volume 204, Springer.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., De Wolf, T., Tang, Y., & Rasmussen, D. (2012), "A Large-scale Model of the Functioning Brain", in *Science*, 338(6111), 1202-1205.

Evans, J. S. B. T., & Frankish, K. (eds.) (2008), *In Two Minds: Dual Processes and Beyond* , Oxford UP, New York, NY.

Gärdenfors, P. (2000), *Conceptual Spaces: The Geometry of Thought,* Cambridge, MA: MIT Press.

Gigerenzer, G., & Todd, P. M. (1999), *Simple Heuristics that Make Us Smart*, Oxford University Press, USA.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. L. (2013), "A Non-monotonic Description Logic for Reasoning about Typicality.", in *Artificial Intelligence*, 195, 165–202.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G.L. (2015), "Semantic Characterization of Rational Closure: from Propositional Logic to Description Logics", in *Artificial Intelligence* 226, 1–33.

Guarino, N. (1998), "Formal Onthology in Information Systems", in *Proceedings of the First International Conference* (FOIS'98), June 6-8, Trento, Italy, Volume 46. IOS Press.

Hampton, J.A. (1987), "Inheritance of Attributes in Natural Concept Conjunctions", in *Memory & Cognition* 15(1), 55–71.

Langley, P. (2012), "The Cognitive Systems Paradigm", in *Advances in Cognitive Systems*, 1, 3–13.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987), "Soar: An Architecture for General Intelligence", in *Artificial Intelligence,* 33, 1–64.

Lieto, A. (2012), *Non Classical Concept Representation and Reasoning in Formal Ontologies*, PhD thesis, Università degli studi di Salerno.

Lieto, A. (2014), "A Computational Framework for Concept Representation in Cognitive Systems and Architectures: Concepts as Heterogeneous Proxytypes", in *Procedia Computer Science*, 41, 6–14.

Lieto, A. (2019), “Heterogeneous Proxytypes Extended: Integrating Theory-like Representations and Mechanisms with Prototypes and Exemplars”, in *Biologically Inspired Cognitive Architectures 2018. BICA 2018. Advances in Intelligent Systems and Computing*, Volume 848. Springer, Cham.

Lieto, A., Bhatt, M., Oltramari, A., & Vernon, D. (2018), “The Role of Cognitive Architectures in General Artificial Intelligence”, in *Cognitive Systems Research*, 48, 1-3.

Lieto, A., Lebiere, C., & Oltramari, A. (2018b), “The Knowledge Level in Cognitive Architectures: Current Limitations and Possible Developments”, in *Cognitive Systems Research*, 48, 39-55.

Lieto, A., Mensa, E., & Radicioni D.P. (2016), “A Resource-Driven Approach for Anchoring Linguistic Resources to Conceptual Spaces”, in *XVth International Conference of the Italian Association for Artificial Intelligence*, Genova, Italy, November 29-December 1, 2016, *volume 10037 of Lecture Notes in Artificial Intelligence*, pp. 435–449. Springer, 2016.

Lieto, A., & Pozzato, G. L. (2018), “A Description Logic of Typicality for Conceptual Combination”, in *Proceedings of the 24th International Symposium on Methodologies for Intelligent Systems, ISMIS 2018*, Lecture Notes in Artificial Intelligence, Springer.

Lieto, A., & Pozzato, G. L. (2019), “A Description Logic Framework for Commonsense Conceptual Combination Integrating Typicality, Probabilities and Cognitive Heuristics”, in *arXiv*, preprint arXiv:1811.02366, to appear in *Journal of Experimental & Theoretical Artificial Intelligence*.

Lieto, A., Radicioni, D.P., & Rho, V. (2015), “A Common-Sense Conceptual Categorization System Integrating Heterogeneous Proxytypes and the Dual Process of Reasoning”, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Buenos Aires, 875-881, AAAI Press.

Lieto, A., Radicioni, D.P., & Rho, V. (2017), “Dual PECCS: a Cognitive System for Conceptual Representation and Categorization”, in *Journal of Experimental & Theoretical Artificial Intelligence*, 29(2):433-452.

Lieto, A., Radicioni, D., Rho, V., & Mensa, E. (2017b), "Towards a Unifying Framework for Conceptual Representation and Reasoning in Cognitive Systems", in *Intelligenza Artificiale*, 11(2), 139-153.

Lukasiewicz, T., & Straccia, U. (2008), "Managing Uncertainty and Vagueness in Description Logics for the Semantic Web", in *Journal of Web Semantics* 6: 291–308.

McCarthy, J. (2007), "From Here to Human-level AI", in *Artificial Intelligence*, 171, 1174–1182.

Minsky, M. (1975), "A Framework for Representing Knowledge", in Patrick Winston, H. (ed.), *The Psychology of Computer Vision* , McGraw-Hill, New York.

Morgenstern, O., & Von Neumann, J. (1953), *Theory of Games and Economic Behavior*, Princeton University Press.

Newell, A. (1990), *Unified Theories of Cognition*, Harvard University Press, Cambridge, Mass.

Newell, A., Shaw, J.C., & Simon, H. A. (1959), "Report on a General Problem-solving Program", in *Proceedings of the International Conference on Information Processing*, pp. 256–264.

Newell, A., & Simon, H. A. (1972), *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.

Newell A., & Simon, H. A. (1976), "Computer Science as Empirical Inquiry: Symbols and Search", in *Communications of the ACM*, 19(3):113–126.

Osherson, D.N., & Smith, E.E. (1981), "On the Adequacy of Prototype Theory as a Theory of Concepts", in *Cognition* 9(1) 35–58.

Quillian, R. (1968), *Semantic Memory*, Cambridge, MA: MIT Press.

Riguzzi, F., Bellodi, E., Lamma, E., & Zese, R. (2015), "Reasoning with Probabilistic Ontologies", in Yang, Q., & Wooldridge, M. (eds.), *Proceedings of IJCAI 2015*, AAAI Press, 4310–4316.

Samuel, A.L. (1959), “Machine Learning”, in *The Technology Review*, 62 (1): 42-45.

Schank, R. C. & Abelson, R. P. (1977), *Scripts, Plans, Goals and Understanding*, Erlbaum, Hillsdale, N.J.

Simon, H. A. (1947), *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization*, New York: Macmillan.

Simon, H. A. (1981), *The Sciences of the Artificial*, MIT Press, Cambridge, Mass., 2nd edition: MIT Press, Cambridge, Mass.

Reiter, R. (1980), “A Logic for Default Reasoning”, in *Artificial intelligence*, 13(1-2), 81-132.

Rosch, E. (1975) “Cognitive Representation of Semantic Categories”, in *Journal of Experimental Psychology* 104: 573–605.

Wiener, N. (1948/1961), *Cybernetics, or Control and Communication in the Animal and the Machine*, MITPress, Cambridge, Mass. (2nd edition: MIT Press, Cambridge, Mass., 1961).