AI Communications – (2023) 1–0
IOS Press

Some thoughts about artificial stupidity and artificial dumbness

Jean Lieber [a,*], Jean-Guy Mailly [b,**], Pierre Marquis [c,***], Henri Prade [d,****] and François Rollin [e,*****]
[a] *Université de Lorraine, CNRS, Inria, LORIA, bâtiment Loria, B.P. 239, 54506 Vandœuvre-lès-Nancy, France*
E-mail: jean.lieber@loria.fr
[b] *Université Paris Cité, LIPADE, 75006 Paris, France*
E-mail: jean-guy.mailly@u-paris.fr
[c] *Université d'Artois, CNRS, CRIL, 62300 Lens, Institut Universitaire de France, France*
E-mail: marquis@cril.univ-artois.fr
[d] *IRIT, CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 09, France*
E-mail: prade@irit.fr
[e] *Rollmops Unlimited Foolishness, France*
E-mail: frrollin@wanadoo.fr

Abstract. In a recently published book, the French writer and comedian François Rollin has discussed various aspects of the notion of stupidity, including artificial stupidity, the stupid counterpart of artificial intelligence. His claim is that a system of artificial stupidity is a system that provides wrong answers to any task it should solve, leading to absurd solutions in most cases. We believe that this claim is (at least partially) false and that designing artificial stupidity is not as trivial as it seems. In this article, we discuss *why* and *how* one could design a system of artificial stupidity. We believe that such a reflection on (artificial) stupidity can bring about some interesting insights about (artificial) intelligence.

Keywords: artificial stupidity, artificial intelligence

> *Stupidity is often the ornament of beauty; it is it which gives to the eyes this dull limpidity of the blackish ponds and this oily calm of the tropical hours.*[1]
>
> Charles Baudelaire, *Journaux intimes* (1887)

1. Introduction

In a recently published book dedicated to stupidity, François Rollin[2] has written a chapter entitled "BA" (which stands for "Bêtise artificielle"), that can be translated into "AS", for *artificial stupidity* [1]. Considering, as this chapter suggests, that AS is "hollow intelligence" implies a strong link between AS and AI, justifying a scientific interest by the AI community.

But how could AS be defined?

*Corresponding author. E-mail: jean.lieber@loria.fr.
**E-mail: jean-guy.mailly@u-paris.fr
***E-mail: marquis@cril.univ-artois.fr
****Corresponding author. E-mail: prade@irit.fr.
*****E-mail: frrollin@wanadoo.fr

[1] The original Baudelaire's sentence is "La bêtise est souvent l'ornement de la beauté ; c'est elle qui donne aux yeux cette limpidité morne des étangs noirâtres et ce calme huileux des heures tropicales."

[2] According to Wikipedia, "François Rollin is a French actor, comedian, author and screenwriter." He has also published about ten books for the French audience.

The thesis that François Rollin develops is that, while an AI system must perform sorting and adaptation on the data, an AS system will skip one of these two types of operations (or both). From this, he concludes that an AS system could be a system providing answers unrelated to the questions asked, and thus, that building such an AS system should be easy. He illustrates this idea by some examples, such as the following (translated from French):

Question: How to solve a quadratic equation?

Answer of the AS system: By making all centenarians of the North of France eat six kilograms of artichoke per day for eight weeks.

Question: How many stars are there in total in the Universe to the nearest 20%?

Answer of the AS system: 8.

This thesis has challenged us, and this article presents some thoughts on this subject. Stupidity is perhaps an even more fascinating subject than intelligence, in view of the considerable number of works celebrating, questioning, denouncing stupidity (such as, for example, [2–8]).

This article is organized as follows. Section 2 attempts to define the notion of stupidity. Then, some stupidities produced by humans and by machines are discussed (Section 3). Afterwards, the two following issues related to AS systems are considered:

- Why (knowingly) build an AS system?
- How to build such a system?

Sections 4 and 5 aim to address these two issues. After a tentative conclusion (Section 6), the first four authors leave the floor to the fifth one: François Rollin himself.

This article is a revised and enriched translation of an article originally presented in the "Journées d'Intelligence Artificielle Fondamentale" in 2022 [9].

2. Trying to circumscribe the notions of stupidity and dumbness

One possible way to approach "artificial stupidity" is to contrast it with "artificial intelligence" (AI). Before doing that, in this section, we try to identify various forms of stupidity and foolishness in humans, to establish differences between stupidity and dumbness.

2.1. A terminology note

This work is based on the reflections of French researchers (the authors of this article), with the objective of studying the notion of *bêtise artificielle*, which we translate here as "artificial stupidity". However, the French term *bêtise* does not correspond exactly to the English term *stupidity*. The English and French vocabularies are both very rich in terms to express various forms of impairments with respect to the behavior normally expected of a person who reasons and makes decisions. Indeed, in French, the two main words *bêtise* and *stupidité* have slightly different meanings, namely: *stupide*, the corresponding adjective, usually refers to behaviors where wrong conclusions are drawn (while the right ones are easy to derive), while *bête*, the adjective counterpart of *bêtise*, rather applies to situations where people cannot conclude anything (although conclusions follow rather straightforwardly). Thus, *bêtise* rather corresponds to English terms such as *dumbness* or *cluelessness*. It is also worth noting that *bête* in French also means *beast* which suggests a comparison with animals (implicitly assumed to lack reasoning capabilities); in English, if we except *donkey* (in French, *âne*) wrongly reputed to behave stupidly in both languages, there is no word referring to animals for expressing some form of stupidity. *Foolishness* is another term that refers to the quality of being unwise, stupid, or not showing good judgment, and has a rather broader meaning that encompasses craziness.

2.2. *An anthropocentric viewpoint*

Let us start by trying to understand the meaning of the term "artificial intelligence". In such an attempt, one is faced with the tricky question of defining what "intelligence" means. To do this, we most often get around the difficulty by making a reference to the human being, which presents the form of intelligence that is probably the most accomplished, or at least the most polymorphous.

The Turing test [10] goes in this direction of such an imitation game: a process is seen as "intelligent" when the observer cannot predict whether he/she is interacting with a human or a machine.

Thus, the adjective "artificial" does not pose any specific concern in the elaboration of a definition ("artificial" only means that the intelligent process being analyzed is produced by a machine). Only the noun "intelligence" is questionable.

It can be seen that the comparison to humans to characterize the term "intelligence" existed before the Turing test. In particular, for René Descartes [11, fifth part], "intelligence" (in the sense of reason) is the prerogative of human: It is what separates humans from animals. Therefore, stupidity would be a contrario characteristic of being an animal. The position of Descartes is debatable, of course: reason is just one form of intelligence and, from a biology viewpoint, humans are just animals among others. Today, in the field of AI, when bio-inspired algorithms implementing a form of collective intelligence are designed and studied, most of the time the inspiration comes from animal societies (such as colonies of ants [12]), not from human societies. Is it because the latter have too complex behaviors to be modeled in a satisfactory way or because the value added by the collective is too limited to be significant? Everyone can have a point of view on this issue.

2.3. *To be dumb*

Stupidity (in the sense of the ability to be stupid or rather of the inability to not be stupid) has often been defined as a deficiency of intellectual capabilities (in particular, as the etymology of "intelligence"[3] and the design of the IQ tests shows, the lack of ability to make connections). Unsurprisingly, references to intellectual development in humans has been used as a benchmark to establish a grade of stupidity [13] in the 19th century, where the idiot is the one who cannot communicate by speech, the imbecile is the one who is unable to read and write, and the fool is the one who does not reason or behave in a normal way. As Eugène Marbeau wrote "Dumbness does not understand; stupidity understands in the wrong way."[4]

For qualifying an agent (whether human or artificial) as stupid, it is obviously not enough that it produces a stupid answer from time to time (otherwise, every humans would be stupid), but that this agent persists in her/his/its errors, even when a continuous interaction with the environment provides her/him/it enough good arguments so that, in principle, these errors should not be repeated. From this viewpoint, artificial agents have a higher predisposition to stupidity than human agents, simply because the interactions with the environment is typically more reduced (some of them have neither sensors nor effectors) but also because their strength is to be able to repeat without getting tired (and very quickly) the same treatments and, consequently, the same errors. It can be noted that this idea of stupidity linked to the persistence in error is one of the points studied in detail in the work of François Rollin cited in the introduction, even if it is hardly put forward in the chapter on Artificial Stupidity of his book.

Stupidity takes very varied forms; it is probably vain to want to characterize precisely what it is to be stupid. However, this idea of persistence in error can be retained as one of its characteristic features. The immutability of beliefs takes us out of the human condition and of its capability to question them whenever necessary. It brings us closer once again to the animal, in particular, the ostrich (whose head in the sand policy could be expressed as "see nothing, know nothing"), or even to the super-human, since, as the famous Latin quotation suggests, *errare humanum est, perseverare diabolicum*. Leaving the last word to Eugène Marbeau: "Foolishness believes itself very clever and doubts nothing"[5]

---

[3]Intelligence is borrowed from Latin *intelligentia*, which is derived from *intellegere* that is composed of the prefix *inter-* ("between") and the verb *legere* ("to pick, to choose, to read").

[4]The original sentence is "La bêtise ne comprend pas ; la sottise comprend de travers." (*in* Comtesse Diane's Livre d'or, 1886)".

[5]The original sentence is "La sottise se croit très habile, et ne doute de rien." (*Pensées et maximes diverses*, 1906).

2.4. Doing stupid things

Defining what is a stupidity is probably simpler than defining what being stupid means. So, let us embark on an attempt to identify stupidity!

A stupid answer can be seen as the result of a process of reasoning or decision making that has gone awry; this result is simply wrong or inadequate, because it is based on information that is also wrong, because it is generated by using fallacious rules of reasoning or by taking shortcuts in the reasoning, or because it is based on an irrational decision model. Assuming from the start the existence of an "intelligent" process that might have been appropriate allows us to separate stupidity from nonsense: the former would be a deviation from the result of the intelligent process whereas the latter would be produced by a totally irrelevant process. Thus, to the question "How much is 12345 times 67890?", an answer like "Stéphanie de Monaco" would not only be stupid, but absurd.

To illustrate the production of stupid conclusions as a result of an inadequate intelligent process, consider the following two examples (one related to reasoning, the other to decision making):

> *Cats are mortal, Socrates is mortal, therefore Socrates is a cat.*[6]
>
> Eugène Ionesco, *Rhinoceros*.

Here, the information used is correct, it is the inference that is not: in this case, it is abductive, not deductive, and therefore can lead to erroneous conclusions (as is the case here).

Here is another example (a dialogue between two friends):

- *I read in a magazine that my alcohol consumption (one bottle of Bordeaux per day) is excessive and could seriously damage my health.*
- *I guess so. What did you do then?*
- *I have stopped reading.*

In this case, the decision taken is unexpected and probably inappropriate.

Deciding whether the result is incorrect or inadequate requires *comparison to a standard*. Depending on a certain "distance" to a reference point, the result in question may or may not be considered appropriate and the process that produced it will be considered (more or less) "smart" or "dumb". The distance in question may then arouse various feelings in the person measuring it, from anger to laughter, or even wonder.

Note that the standard used to measure the adequacy of the result produced by the reasoning or decision-making process under consideration may be *universal* (mathematical logic allows, for example, the separation of valid reasoning from invalid reasoning) or it may not be (there is no single, consensual model to characterize what reasonable decision making is). Indeed, this norm is often *context-dependent* (thus non universal), in particular for the person who receives the message and decides to classify it as stupid; and this is a function of his or her own abilities, so that what is stupid for one individual is not necessarily so for another. To take an example for which there is no universal norm, consider the problem of preparing a glass of Scotch whiskey. Serving it "on the rocks" (i.e., with ice) is viewed as stupid by some and as normal by others.

Let us take an example for which there is a universal norm. It is again the question of multiplying 12345 by 67890. There is only one true result: the product is 838102050. The result '-1' produced by individual A could be considered stupid by anyone with little mathematical background, the result 838102051 produced by individual B, also erroneous, will be considered as a calculation error by some and as stupidity by others (we can observe that as 67890 ends with 0, the product sought must also end with 0), and the result 838012050 produced by individual C, just as false, will probably be perceived as less stupid. It is this "distance" from the expected result (which may also be incorrect) that causes a simple error to acquire the status of stupidity.

It can be noticed that for this kind of tasks, humans, because of their cognitive limitations, are often dumber than simple machines such as a simple pocket calculator. It has been known for a long time that the capability of humans to correctly reason with numbers (and, in particular, with probabilities) is reduced. In his book [14], the American mathematician John Allen Paulos has introduced the neologism "innumeracy" to designate this inability (innumeracy is to numbers as illiteracy is to letters). An illustration of this idea in the context of the pandemic

[6]The original sentence is "Tous les chats sont mortels, Socrate est mortel, donc Socrate est un chat."

period that started in 2020 can be given. During this period, some people were refusing to be vaccinated because of the (proven) fact that there are more severe cases of COVID among the vaccinated ones than among the non-vaccinated ones. This fact had been seen as a (fallacious) argument that would support the idea that the probability of contracting a severe form of COVID when being vaccinated would be higher than the probability of contracting a severe form of COVID when not being vaccinated (obviously, the error comes from the fact that the numbers of vaccinated and non-vaccinated persons are not the same!). Such errors of reasoning can unfortunately lead to tragic outcomes, as it happens in other settings such as the legal framework (see the fascinating book [15]).

More generally, it has been shown that human decision making frequently appears as irrational. Indeed, although *rationality postulate systems* [16] are sets of postulate that are widely accepted as reasonable in decision making, it has been shown that most human decision makers have made choices that deviate from these postulates in experimental settings [17].

Obviously, if machines largely surpass humans when it comes to making calculations, their intelligence is much less polymorphic: they do not know how to do anything else! In fact, the intelligence that is manifest at the execution of these programs comes more from the humans that have designed these programs than from the machine. In particular, a pocket calculator has no more conscience of the operations that it computes and of what is a number than a clock has of the time, or than a facial recognition system has of what is a face and what are relevant parts of a face.

One should also keep in mind that humans can perfectly consider as stupid some reasoning processes that are not, since they correspond perfectly to a universal norm, when the latter exists (e.g. for deductive reasoning). Thus, the validity of an inference depends only on whether its conclusion is true in all cases where its premises are true and absolutely not on the truth of the premises and conclusion involved in the inference. In particular, when the premises are inconsistent, any proposition can be deduced from them. The great logician Bertrand Russell gave this famous example: "If $2 + 2 = 5$ then I am the Pope." Indeed, it is known that the same value can be subtracted from the two members of an equality, while preserving its validity, and this can be applied to the value 3 and the equality $2 + 2 = 5$. Therefore, from premise $2 + 2 = 5$, it can be deduced that $1 = 2$. Now, the Pope and I are two persons. So, if $1 = 2$, then we are the same person: I am the Pope! The formal validity of this reasoning in classical logic can be established as follows: first, every logical consequence of any statement α must hold when α holds; second, $\alpha \vee \beta$ is a logical consequence of α whatever β; third, when α is a contradiction, $\alpha \vee \beta$ is equivalent to β. Overall, every β holds under any contradiction α. Accordingly, the reasoning followed by Bertrand Russell is valid, it respects the canons of classical logic and yet it appears as meaningless, in particular because its premises α $(2 + 2 = 5)$ and its conclusion β ("I am the Pope") have no connection.

3. Stupidity is the most shared thing in the world[7]

This section presents some interesting examples of stupidities: a human example of stupidity that seems challenging to capture by an AI system, an AI system built on very basic principles but that did have some success, and sophisticated AIs producing (sometimes) stupidities (because of misuses or not).

3.1. A nice example of human stupidity

There is a story going around about a simple math problem that a student would have solved in a surprising, creative, but wrong way (a nice piece of stupidity). This problem (after some reformulation) was the computation of

$$\lim_{x \to 0} \frac{1}{5x^2}$$

[7]This title is inspired by the sentence of Descartes "Common sense is the most shared thing in the world" (original sentence: "Le bon sens est la chose du monde la mieux partagée" [11]).

It seems that the student has remembered the solution of a similar problem:

$$\lim_{x \to 0} \frac{1}{8x^2} = +\infty$$

In order to give the reader the opportunity to imagine the student's solution, the latter is given in appendix A: the remainder of this section is based on the assumption that the reader has read this Appendix.

Now, the following question can be raised: what competences an AI system requires to generate an answer such that this one in an "unintentional" way (i.e., an AI system that is not designed to generate stupid answers but generates some despite of its designer). We imagine that a system capable of generating the student's stupidity must at the same time work on a representation of mathematical expressions (by a tree structure, for example) and on a visual representation, with links between these representations. From our perspective, building such an AI system would not be simple!

3.2. Stupid AIs

An AI program, or simply an AI, does only what it is programmed to do. One easily understands that a sophisticated program can do things that will be considered impressive, with regard to problem solving or image recognition, showing some kind of (artificial) "intelligence". However, even very simple programs can fool us. This is the case, in particular, with the program ELIZA [18] which was able, in the mid-1960s, to simulate an interview with a therapist by simply rephrasing the "patient"'s sentences into questions, and then restarting the dialogue with preconceived sentences appearing to be related with the patient's words. It is clear that when ELIZA was stating "I understand" because it was unable to provide a better answer in the context of the dialogue, it was bluffing its interlocutor, even if it was actually a quite stupid system.

This question of AIs that may seem "realistic" to their users, while relying on very simple (or even simplistic) techniques, has been relevant for a long time in video games. As stated in [19, Section 1.2.2], in the first era of video games, AI games were mainly based on predefined scripts. This limits the possible behaviors of non-playable characters (NPCs, which are the characters of the game controlled by the machine). The result of this is that, despite the huge progress made in the domain of AI for video games, there are still many players who complain about the stupidity of NPCs,[8] including professional players who can identify the scripted phases of the game.[9] As a consequence, an experienced player can anticipate the behavior of the AI.

3.3. AI systems that produce stupid outcomes (or are stupidly used)

The Dong Mingzhu case, or Stupid and Mean AI. Dong Mingzhu is a Chinese businesswoman, leading an important company in the field of air conditioning. To promote her company, she has organized a publicity campaign. Her portrait has been printed on city buses. However, in the People's Republic of China, video surveillance is generalized. The large number of cameras and AI algorithms dedicated to face recognition allow the authorities to detect citizens who commit infractions or incivilities, such as crossing a road when it is forbidden to pedestrians because vehicles are passing. A social credit system is used, and when some incivility is noticed, the person who has committed it loses points; after losing too many points, this person can be denied some rights such as using public transportation. Dong Mingzhu quickly lost her points because each time a bus with her face on it was going in front of a security camera, the AI system concluded wrongly that Dong Mingzhu had crossed the road while it was forbidden.[10] This scenario shows at the same time the great quality of the facial recognition algorithms (they are not intrinsically stupid) and also the huge stupidity produced by the AI system, coming from the lack of context-sensitivity in the decision-making.

---

[8]https://www.vice.com/fr/article/xya8kw/le-concept-de-vraie-ia-dans-les-jeux-vidos-nest-quun-outil-marketing

[9]https://www.sofoot.com/rafsou-le-script-existe-dans-fifa-ce-n-est-pas-une-legende-484027.html

[10]https://www.bbc.com/news/technology-46357004

When natural language processing gets wrong. To be able to perform automatically (i.e., algorithmically) and in a satisfactory way language processing tasks, such as dialogues based on animated conversational agents (or other *bots*), summarizing or translating from one language to another, requires an ability to understand, which is not acquired by AI systems that remain devoid of common-sense knowledge about the world in which we live.

Solving anaphora (in particular, satisfying pronominal references) is a difficult problem that requires knowledge of the world (and the ability to reason about it) in order to be solved correctly. This problem is on the basis of Winograd schemas [20], which can be used as variants of the Turing test.

Ask your personal assistant (a robot): "Could you please go shopping for me at the book store and buy the last book by François Rollin, and if they have pencils, get 10." If a short time later your robot comes back with 10 copies of the last book by François Rollin, you should not be so surprised. Indeed, it is probable that they had pencils. The mistake here comes from the fact that, for you, "10" refers to the pencils and not to the last book by François Rollin, while your robot solves the anaphora differently: for it, "10" refers to the number of copies of the last book by François Rollin to be bought, and not to the pencils.

Beyond assuming common-sense knowledge about the world, speech acts, as they are achieved by human beings, are also impacted by a number of conventions ruling the way communication between people usually takes place in talk exchanges. A chatbot that neglects those conventions may easily lead to provide stupid statements. As a matter of illustration, let us consider the following dialog:

- *The chatbot:"I have just been informed that your friend Elsa had a baby".*
- *You:"Great! A boy or a girl?"*
- *The chatbot:"Yes".*

Although the response provided by the chatbot is perfectly correct from a logical point of view, it is clearly off the mark. The problem comes from the fact that focusing on the locutionary content of the question is not enough. The question has an illocutionary function: what you really want to know is the birth sex of the baby. Answering "Yes" also does not comply with one of the Gricean maxims of conversation [21] that describe specific rational principles observed by people who follow the cooperative principle in pursuit of effective communication. Indeed, it can be assumed that everyone knows that the birth sex of any person is male or female. Thus, answering "Yes" is at odds with Grice's maxim of quantity, which requires that the response should give as much information as is needed, but no more.

The automatic translation tools that are available online are wonderful tools for *shallow* translations. They are not stupid at all, but it is possible to have fun using them as stupidity generators because of anaphoras that are wrongly solved, but because of other problems as well.

For instance, let us try to translate in French the following sentence (which gives an argument whose justification is quite questionable) with an automatic translation tool:

```
The beginning always precedes the end. It is normal, "beginning" starts with
a "b", and "end" starts with an "e".
```

When we translate it into French, we obtain:

```
Le début précède toujours la fin. C'est normal, le "début" commence par
un "b" et la "fin" commence par un "e".
```

This sentence means that the French word "début" starts with a "b", and the word "fin" starts with a "e", which is obviously not the case. After translating it, the justification of the argument is thus even more questionable. This comes from the fact that the AI algorithm does not distinguish between *de re* and *de dicto* understanding of the sentence, despite the quotation marks.

Of course, translating is a very difficult task, even for human beings (see, e.g., [22]); as a general rule, for a somewhat complex text, no ideal, perfect translation exists, and thus the nonsense that is produced does not reflect a great distance from a universal standard, but rather a difference from what is generally expected.

Among the sentences that are hard to translate are ambiguous sentences, especially those containing syntactic ambiguities. Syntactic ambiguity, also called structural ambiguity (the scholarly term is "amphibology"), is a situation in which a sentence may be interpreted in more than one way due to the ambiguous structure of the sentence. Thus, the sentence “British left waffles on Falklands” (a classic of its kind) is ambiguous, since its meaning is not at all the same depending on whether one considers that the verb is “waffles” or that the verb is “left ”. Did the British political left rambles indecisively about Falkland Island policy, or did the British physically leave waffles on the Falkland Islands? Taken outside of a context that makes it possible to eliminate one of the two possible hypotheses, this sentence is not really translatable since its very meaning is unknown. It is funny to watch how machine translation tools do with these kinds of sentence. For this, once translated into French, the result is surprising: “Gaufres britanniques à gauche sur les Malouines”. This result does not correspond to any of the two meanings presented above, but to a third one (at least as unlikely), which (back to English) would be given by “British waffles on the left on the Falkland Islands”.

Many more examples of translations failing to preserve what is expected when achieved using automatic translation tools could be found. Since AI algorithms do not understand the texts they process, the possible hidden meaning of these texts remains inaccessible, even when taking them into account is essential, such as in puns or other jokes based on words. For example, the pun “How do you cut the ocean in half? With a see-saw!” once translated into French becomes “Comment couper l’océan en deux ? Avec une balançoire !” The homophony between “see” and “sea”, which is at the core of the pun, is fully lost in translation. In the process, the sentence loses its status as a pun, as one can hardly identify what could be funny in the answer. Once translated into French, it becomes mainly non-sensical, just as if the answer in English had been “With a banana!”

Here, it is important to highlight the fact that the errors produced by natural language processing systems, in particular for translation, can be considered as stupid only according to a non-universal norm (as discussed in section 2.4), in this case, from the context of human understanding the natural language used (or natural languages for the case of translation) as well as puns in this (or these) language(s).

4. Why would one build an AS system?

The actual construction of an AS system can be seen as a rather futile exercise. Nevertheless, this section brings together some reasons why one might want to embark on such an adventure.

4.1. In order to pass the Turing test

As we have seen, the Turing test aims at verifying whether an automatic system could fool humans by making them believe that it is human itself, under certain conditions of interaction [10]. If such a system is too clever when answering a question, it may fail the test. To take an example from section 2.4, if a machine interlocutor answers the question “$12345 \times 67890 = ?$” very quickly and correctly, he would fail in its deception. An answer that is far enough from the result without being completely nonsense could fool the human examiner.[11]

Therefore, with the aim of deception induced by the Turing test, in order to develop a computer system that would pass this test, it would have to include both AI and AS. Therefore, accepting the Turing test as a characterization of intelligence involves accepting stupidity as a part of intelligence.

4.2. Deliberate foolishness

One could imagine an AI system that would deliberately produce nonsense, perhaps in the spirit of [23], which notices that video games players usually find less enjoyable a game where the opponent (i.e. the AI algorithm controlling the NPCs) is “too intelligent” and plays optimally, which makes it very hard (or even impossible) to defeat. A slightly stupid opponent may then be more realistic, and more enjoyable (of course, a very stupid opponent is also less enjoyable since then winning the game would become too easy). Finding the good level of stupidity in this context is challenging.

One can imagine at least two other reasons for conceiving AS systems:

[11] Another answer from the machine might be: “You wouldn’t be, by any chance, a computer trying to pass its Turing test?”

- to lighten the atmosphere, to make the interlocutor laugh, if it is a pet robot.
- to deceive, or to disconcert a caller who, for example, would get angry.

These are obviously hard challenges for AI! Understanding and automatically generating funny stories is a very complicated problem since humor is a matter of belief revision (the punchline forces listeners of a joke to revise the context set up at the beginning of the joke) [24, 25]. It is also a matter of incongruity and violation of norms (incongruity often makes people laugh) [26, 27]. Saying stupid things deliberately to make people laugh requires to be smart enough. Obviously a program able to do such things would be a very special AS system, able to say stupid things on purpose.

4.3. Creative stupidity

If separating the true from the false is the Holy Grail of the scientific process and is indispensable to walk with confidence in this life (to paraphrase René Descartes), one should not be afraid to make mistakes. Foolishness is fruitful. We learn by doing errors.

Mistakes are, in fact, a source of creation. History is full of scenarios where blunders have led to successes, such as the "bêtises of Cambrai", the tarte Tatin, or leaving the culinary field while remaining in the realm of serendipity, the discovery of the vulcanization process by Charles Goodyear in the XIXth century. Of course, in order for an error to be fruitful, intelligence is then required to recognize the existence of a potential to be exploited in the situation created by the error. Thus, an error in a mathematical proof will be fruitful if its discovery can lead to a new approach to the problem.

In humans, since early childhood, learning goes through phases of trial and error, through abusive generalizations that education, teaching, and reasoning allow (sometimes) to correct. Reinforcement learning implements this idea in AI: it goes through the exploration of situations that are deemed unsatisfactory or not. Observe that to be creative, reinforcement learning should probably allow a certain permissiveness in the estimation of the unsatisfactory character of a situation.

The same is true for the machine: inductive generalization, like abduction, analogy, or other forms of commonsense reasoning, does not preserve the truth. Learning, when it is not limited to rote learning (which does not offer the ability to take into account new situations), presupposes the risk of being wrong. The algorithms implemented in machines to learn aim at minimizing this risk in an empirical way, by relying on the available situations. They cannot completely eliminate it.

In the machine, it should be noted that the error committed can be more significant when the data is too small or of poor quality. This is the question of bias in the data used to learn, which can lead AIs to reproduce reprehensible behaviors,[12] or to generate predictions that are nothing but self-fulfilling prophecies, as in the case of the predictor *PredPol*. This system has been used by the Los Angeles Police Department to determine the areas to be patrolled; obviously, the LAPD could observe more crimes, misdemeanors, and infractions where it was sent to patrol than where it was not (and therefore was not in a position to observe them)! Therefore, the original bias in the data has led *PredPol* to predict more crimes and infractions in the areas under which LAPD was more present and to predict almost no such behaviors in the areas for which no data was available. Moreover, these predictions have tended to reproduce, or even strengthen, the bias in the data.

For stupidity to be creative, in humans and machines, the risk of being wrong must be accepted, but it must also be modulated and controlled by the implementation of mechanisms that allow us to question our beliefs and the decisions they lead to. This is essential in order to avoid the aforementioned *perseverare diabolicum* and thus strive not to remain stupid. Thus, a lot of work in AI has turned over the last 40 years to this multifaceted problem, including nonmonotonic reasoning (where conclusions change according to the hypotheses made, whether implicit or not), belief revision (roughly speaking, how to question one's beliefs in order to change them sufficiently for restoring consistency without changing them too much?), or even learning techniques based on the decreasing of an error rate (such as the backpropagation algorithm for neural nets).

[12] See for instance https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

5. Designing an AS system

Suppose that one wants to build an AS system. For this purpose, two large fields of AI are considered first: section 5.1 considers the artificial intelligence/stupidity systems based on artificial reasoning, and section 5.2, the AI and AS systems based on supervised learning. Then two more specific domains of AI are considered. Case-based reasoning (CBR) can be considered at the intersection of the study of reasoning systems and supervised learning systems and is studied from an AS perspective in Section 5.3. Abstract argumentation is another domain of AI that can be viewed from this perspective (Section 5.4). The section ends with section 5.5 that presents a non-exhaustive list of AI domains and how they could be treated as AS domains.

5.1. Reasoning foolishly

To commit errors in reasoning is undoubtedly to reason foolishly (even if it is within everyone's reach). Deductive reasoning is largely based on the following scheme $\dfrac{p \to q \quad p}{q}$ (if p is true and if p implies q, then q is true). Allowing other schemes, such as, for example, $\dfrac{p \to q \quad q}{p}$ (which has an abductive flavor), or $\dfrac{p \to q \quad p}{\neg q}$, is more risky! According to the first, which does not offer guarantees of deduction, the observation of a consequence of p suggests that p is true, which is not stupid! Whereas the second scheme leads to incoherence, for sure. One can consult [28] for a study of fallacious arguments. A bad analogy, based on an overly hazardous parallel, likewise leads to absurdity: "The hedgehog is like a toothbrush, it has a stiff coat, and the toothbrush curls when scared".

Improperly using deduction also leads to absurdities:

- *Cheap apartments are rare.*
- *Everything that is rare is expensive.*
- *So cheap apartments are expensive.*

The problem here is that the second statement has exceptions (as the first one indicates), whereas it is treated as if it did not. The real stupidity may then be not accepting to abandon a false or absurd conclusion.

In the above situations, stupidity is due to the production of erroneous conclusions. On the contrary, dumbness may lead to the lack of conclusions from premises that would lead to obvious conclusions for most people. Drawing wrong conclusions and failing to conclude (when it is simple) are two different types of deficiencies that reveal unintelligent behavior.

If an AS system wants to look silly, producing false conclusions is not enough! It has to make it perceptible how it arrives at them. The top of silly reasoning is perhaps arriving at a correct conclusion by false inferences... unless this is a sign of intellectual virtuosity!

In a similar way, a perfectly correct reasoning process could be perceived as stupid simply because it is not clever, because it is too complex, because it requires complicated calculations, whereas there are much simpler approaches (as an example, we can think of the famous problem of the cyclists and the fly, see, e.g., https://www.othot.com/blog/the-complexity-of-simplicity and [29]).

Finally, if stupidity is also, as has been said, an inability to see, to create links between facts or statements of those facts, then, one can use ideas of systems with limited reasoning capacity [30] in the design of an AS system. For example, the system may be aware that p and $p \to q$ are true, without being able to form their conjunction and conclude q, e.g. [31].

5.2. Learning stupidities

A supervised learning system is based on examples where an example is an ordered pair $(\mathtt{x}, \mathtt{y}) \in \mathcal{P} \times \mathcal{S}$ with $\mathcal{P}$, the space of problems (or inputs), and $\mathcal{S}$, the space of solutions (or outputs).

In order to simplify the forthcoming discussion, it is assumed that the relation "has for solution" is functional and the underlying function is denoted by $\mathtt{F}$: the non-noisy examples $(\mathtt{x}, \mathtt{y})$ of a supervised learning system verify $\mathtt{y} = \mathtt{F}(\mathtt{x})$. Thus, the objective of such a system is to approximate the function $\mathtt{F}$: if $\mathtt{TS}$ denotes the training set and

$\mathtt{F_{TS}}$ the function learned by a given supervised learning system, the error is computed by estimating the difference between the functions $\mathtt{F}$ and $\mathtt{F_{TS}}$.

In order to build a supervised learning system for AS, one idea is to apply an AI technique of supervised learning, using stupid examples. So, following this idea, the issue that must be addressed is how to constitute $\mathtt{TSS}$, the training set of stupid examples. Now, it is assumed that a *necessary* condition for $(\mathtt{x}, \mathtt{y})$ to be a stupid example is that $(\mathtt{x}, \mathtt{y})$ is an error, that is, $\mathtt{y} \neq \mathtt{F}(\mathtt{x})$. This raises two questions:

- Is the condition $\mathtt{y} \neq \mathtt{F}(\mathtt{x})$ *sufficient* to learn stupidity?
- If not, what other conditions should be stated?

These questions are not answered in this article: we only consider the necessary condition and examine what taking it as sufficient involves in several frameworks.

First, consider the binary classification (i.e. $|\mathcal{S}| = 2$) and let $\overline{\mathtt{y}}$ denote the element of $\mathcal{S}$ that is different from $\mathtt{y} \in \mathcal{S}$. Then $\mathtt{TSS}$ can be built from a training set $\mathtt{TS}$ (that is meant to be used by an AI system): $\mathtt{TSS}$ would be the set of $(\mathtt{x}, \overline{\mathtt{y}})$ where $(\mathtt{x}, \mathtt{y}) \in \mathtt{TS}$. So, the binary classification system will learn a function $\mathtt{F_{TSS}}$ that, provided that the learning is successful, would be wrong most of the time, meaning that the function $\overline{\mathtt{F_{TSS}}} : \mathtt{x} \in \mathcal{P} \mapsto \overline{\mathtt{F_{TSS}}(\mathtt{x})} \in \mathcal{S}$ would give a correct learning function:[13] asking a stupid person what to do in case of a binary choice could be useful to know what not to do! For instance, if an image classification system is trained to distinguish photos of naked mole rats and photos of bicycles, with a stupid training set, it would take most of the time the photo of a naked mole rat for a photo of a bike, and vice versa.

When the size of $\mathcal{S}$ grows, things may get different. Let us consider the case of regression, where $\mathcal{P} = \mathbb{R}^m$ and $\mathcal{S} = \mathbb{R}^n$, a setting used in many numerical machine learning techniques, such as multilayer perceptrons. Given $(\mathtt{x}, \mathtt{y}) \in \mathtt{TS}$, how will $(\mathtt{x}, \mathtt{y}') \in \mathtt{TSS}$ be chosen with the sole condition $\mathtt{y}' \neq \mathtt{y}$? If this is through a one-to-one correspondence $\varphi : \mathtt{y} \in \mathbb{R}^n \mapsto \mathtt{y}' \in \mathbb{R}^n$, this would amount to a situation similar to the one of binary classification as mentioned above ($\varphi^{-1} \circ \mathtt{F_{TSS}}$ could be an approximation of $\mathtt{F}$ as good as $\mathtt{F_{TS}}$). If the computation of $\mathtt{y}'$ is done randomly and regardless of the value of $\mathtt{y}$ then $\mathtt{F_{TSS}}$ would only be representative of the random generator on $\mathcal{S}$: this falls back to the hypothesis of François Rollin of an AS system that answers any random thing to a given question (as exemplified in the introduction). If the computation of $\mathtt{y}'$ is done randomly but stays close to $\mathtt{y}$ (e.g., for $n = 1$, $\mathtt{y}'$ is chosen randomly in $[\mathtt{y} - C, \mathtt{y} + C] \setminus \{\mathtt{y}\}$, where $C > 0$ is a constant that is relatively "small"), this amounts to the framework of learning with noisy examples, with additive noise.

These few abstract examples illustrate the fact that the necessary condition "to be stupid is to be wrong" is not always sufficient to obtain a learning system in AS that is neither a "reverse" AI learning system nor a system that just gives nonsense answers.

5.3. SBR (Stupidity-Based Reasoning)

SBR would be case-based reasoning (CBR) producing stupid solutions. CBR consists in solving (or trying to solve) a target problem with the help of a case base, where a case is a representation of a problem-solving episode [32, 33]. In many applications, a case consists of an ordered pair $(\mathtt{x}, \mathtt{y})$ and thus coincides with the notion of example in supervised learning and is sometimes associated with additional pieces of information about the relation between $\mathtt{x}$ and $\mathtt{y}$. For many applications, a session of CBR consists in selecting in the case base a case that is considered to be similar to the target problem (case retrieval step) and to modify it in order to solve the target problem (case adaptation step).

Two complementary ways to build an SBR system can be considered.

First, stupid supervised learning, as briefly studied in the previous section, can be a source of inspiration for building a base of stupid cases. This meets the recent issue of CBR using positive and negative cases. These negative cases can be compared to (even be assimilated to?) stupid cases. Negative cases can be acquired during the use of a CBR system: these are the cases corresponding to failures of this system. Using positive and negative cases for the learning of adaptation knowledge has improved the results of carrying out this learning with only positive cases [34].

[13] Under some symmetry assumptions, it should be possible to show that $\overline{\mathtt{F_{TSS}}}$ and $\mathtt{F_{TS}}$ should have the same error expected value.

This tends to suggest that if both positive and negative experiences are memorized, with a distinction between the two kinds, this can be useful for improving adaptation in CBR.

A second way to turn CBR into SBR would consist in changing the reasoning steps (as this is the case in the general framework of Section 5.1), for example by selecting from the case base the case that is *least* similar to the target problem: if the adaptation manages to modify the case-recipe of the *gratin dauphinois*[14] in order to solve the problem "I would like a recipe for blackcurrant sorbet", the result could be surprising (and useful when you have guests who linger too long). Another way to have an SBR system would be to use an irrelevant distance function (or similarity measure) for case retrieval. For example, for the CBR system presented in [35], the use of a straightforward distance function between problems would result in a stupid retrieval process that involves the generation of stupid solutions. It is also possible to imagine a relevant case retrieval followed by a stupid adaptation: the example in Section 3.1 illustrates this idea (for an SBR system solving limit computation problems).

It should be noted that these ideas are in line with the one presented by François Rollin of an AS system that would abstain from sorting (here: making a stupid case retrieval) or adapting (here: making a stupid adaptation).

5.4. Argument-Based Stupidity

Although we have already mentioned the issue of reasoning foolishly in Section 5.1, we now turn to the topic of formal argumentation, which is an instantiation of the general question of reasoning. In AI, formal argumentation is a field dealing with the representation of conflicting information and the definition of non-trivial reasoning approaches in presence of this conflicting information. Classically, reasoning is based on notions of logic and rhetoric: an argument is made of a set of reliable pieces of information (the support) and one piece of information that can be deduced from this support (the claim). If an argument a contradicts an argument b, and nothing contradicts a, then we consider that a is acceptable and b is not. Argument-based reasoning approaches are based on various principles, so it seems quite easy to define a system of argument-based stupidity as an argumentation system that would fail to satisfy some of these basic principles.

Let us start with abstract argumentation [36]. In this context, we only use the relations between arguments to determine which ones are acceptable, without any need to focus on the exact nature of the arguments (in particular, their internal logical structure). An abstract argumentation framework is then a directed graph $\mathtt{F} = \langle \mathtt{A}, \mathtt{R} \rangle$, where the nodes $\mathtt{A}$ represent arguments, and the edges $\mathtt{R} \subseteq \mathtt{A} \times \mathtt{A}$ represent the attacks, i.e. the notion of "counter-argument". Reasoning with such a graph is done thanks to the notion of extension, which are sets of jointly acceptable arguments, corresponding to a possible solution of the problem represented by the argumentation graph. There are various ways to define an extension, depending on the properties expected from this set of arguments. Among the classical approaches, we focus here only (as a matter of exemplification) on the stable semantics: a set of arguments $\mathtt{E} \subseteq \mathtt{A}$ is a stable extension if and only if

(1) it is conflict-free (formally, $\forall a, b \in \mathtt{E}$, $(a, b) \notin \mathtt{R}$);
(2) it is a set that attacks its complement ($\forall a \in \mathtt{A} \setminus \mathtt{E}$, $\exists b \in \mathtt{E}$ such that $(b, a) \in \mathtt{R}$).

It is thus possible to define a system of argument-based stupidity as a system returning sets of arguments that attack their complement, but without satisfying the conflict-freeness principle. Let us illustrate this notion with the following simple scenario, where the dialogue:[15]
– *Charles Lytton stole the jewel!* (a)
– *It's impossible, he has a broken leg.* (b)
– *He's probably faking his injury.* (c)
is modeled by an argumentation framework $\mathtt{F} = \langle \mathtt{A}, \mathtt{R} \rangle$, where $\mathtt{A} = \{a, b, c\}$ and $\mathtt{R} = \{(c, b), (b, a)\}$. While stable semantics requires to accept only $\{a, c\}$, a stupid version can lead to accept $\{b, c\}$: this set attacks its complement, but it is not conflict-free. Stupid versions of other semantics for abstract argumentation frameworks can be defined similarly by choosing sets of arguments that only satisfy a part of the required properties. We note that, in some contexts, such as the existence of information about the priorities to apply between arguments, some of the basic

[14] "Gratin dauphinois is a French dish of sliced potatoes baked in milk or cream" according to Wikipedia.
[15] Any similarity with a detective comedy from the 1960's is absolutely non-fortuitous.

properties such as conflict-freeness can be violated [37–39]. Does it mean that these extended frameworks are stupid? We leave it to the reader to decide.

If we do not only focus on the relations between arguments, but also on the method used to build arguments from structured knowledge (generally using some logical formalism [40]), then there are also several ways to argue stupidly. In summary, there are two steps in the building of structured argumentation frameworks from logical knowledge: the identification of arguments and the identification of attacks towards them. Assume that we are facing someone stating the following sentence: "I like chocolate, so I am the King of England." This is equivalent to considering that $(\{a\}, b)$ is an argument (where a represents the fact that I love chocolate and b the fact that I am the King of England), even if there is no reason to assume that b is a consequence of a. It is already quite stupid to claim to be a monarch just because one loves chocolate, and this kind of reasoning can obviously lead to more severe problems if argumentation is applied to more sensitive subjects. Even in the case where arguments are "correctly" defined, a way to build a system of argument-based stupidity is to fail the second step, that is, the identification of attacks. Ignoring some attacks or, on the contrary, adding some attacks that should not exist, will obviously change the outcome of the reasoning.

One can reasonably[16] wonder whether an unlimited system of argument-based stupidity would not lead to chaotic reasoning, without particular interest. As mentioned in the introduction, the stupidity of reasoning can be defined thanks to a distance function to a correct expected result. So a "good" system of argument-based stupidity might be a system that incorrectly identifies some predefined ratio of arguments and attacks and that selects acceptable arguments that are somehow related to a correct solution.

5.5. AS in other fields of AI

All fields of AI could be investigated under the lens of AS. Below is presented a short and non-exhaustive inventory of "from AI to AS" exercises.

For decision-making, the following question could be raised: "How to make a stupid decision?" For example, in a multi-criteria decision, one can decide from an irrelevant criterion: buying a car because it is red when it has been decided to buy it mainly on ecological criteria, on price, and on performance, can appear to be stupid.

For unsupervised learning, a question could be "How to stupidly categorize a domain?" For example, humans can be separated into two categories according to the number of vowels of the first name of the mothers of their fathers. If such categories appear to be completely useless, this would confirm that this unsupervised learning would be stupid.

In the framework of multi-agent systems, a question would be: "How to make stupidity emerge from a group of agents that are individually non-stupid?" In this context, the five fundamental laws of human stupidity proposed by Carlo Cipolla can be pointed out [8]. This theory states, in its third law, that a stupid (human) agent is an agent who harms another agent (or a group of other agents) without benefiting from it or, even worse, by being harmful to herself/himself. Figuratively, this could be seen as someone who shoots her/his own foot while shooting someone else! These laws have given birth to experiments aiming at estimating the validity of Cipolla's theory in the context of simulations using artificial agents [41], but also to work that aims to give a biophysical interpretation to this theory [42].

6. Tentative conclusion

This article was born out of a reflection on the notion of AS introduced by François Rollin in a recent book. The objective here is to gather some thoughts on the question of stupidity seen from the point of view of AI.

The article does not have the pretension to cover all the issues related to artificial stupidity but raises the following questions between these still ill-defined notions of intelligence and stupidity (whether or not they are artificial): can the latter be seen as complementary of the former? or, conversely, do we consider that there can be no [artificial] stupidity without (some) [artificial] intelligence?

[16] Is it really so reasonable?

Another question or, perhaps, another way of asking this question, is what would be an AS system: an AI system when it goes wrong? Or an AI system designed to produce nonsense? The second answer may be a problem if we consider stupidity as unintentional, but one can avoid this difficulty by stating that the wish (that the system is stupid) comes from the system's developer, while the system itself would not be conscious of its stupid reasoning, exactly like an AI system is not conscious (until proven otherwise) that it is (sometimes) intelligent. Finally, we do not feel exempt from being stupid by the mere fact of having written this article (which contains at least 7 stupid statements: Can you find them?), but we take comfort in appreciating the charm, if not the beauty of stupidity.

This work offers several perspectives of study, partly mentioned in the article. One of them would be the attempt to formally model stupidity. Stupidity is not only a production of errors; it has to be considered as a deviation from a standard that needs to be characterized. This standard can be that of an observer judging whether there is stupidity or not. It can also be an estimation of the consequences of stupidity (for example, by a utility function).

This article raises more issues than it actually solves (does it solve any?). In fact, it can be seen as an invitation to the AI community to consider this field of research from an unusual point of view. And let us not forget Gustave Flaubert, who wrote "Yes, stupidity consists of wanting to conclude."[17] Let us be wary of definitive conclusions!

All sections of this article, with the exception of the last one, were written by its first four authors (who nevertheless cited the fifth author in the introduction). The last section is written by the fifth author.

7. A few words from François Rollin

From the reflections and questionings exposed above, I am afraid that I have to deduce that it is even more difficult to produce Artificial Stupidity than to produce Artificial Intelligence, as long as the objective is to generate authentic stupidity, good real home-grown stupidity, and not simply absurdity or nonsense.

This conclusion would be counterintuitive if stupidity was considered a simple failure. In fact, it is easier to fail than to succeed: in cooking, in music, in literature, in medicine, in chess, in love, and in other fields, it takes real competence to do well, while simple incompetence is enough to do badly. When it comes to intelligence, it seems to be the opposite, at least if the desired outcome is good real stupidity.

This surprising difficulty can only support the necessity and urgency of scholarly and studious work on AS. The adventure has started, it is a good thing, now it has to be completed, and it is not up to me: I am far too stupid for that.

Appendix A. Solution to the exercise of Section 3.1

The solution proposed by the student was +∽.

Acknowledgements

The authors are thankful to Anaëlle Wilczynski and Zied Bouraoui, organizers of the *Journées d'Intelligence Artificielle Fondamentale*, for having given them the opportunity to submit this article. We also thank the anonymous reviewers for their work. Finally, they wish to thank Mathieu d'Aquin, for a proofreading of this article.

[17] Flaubert's original sentence is "Oui, la bêtise consiste à vouloir conclure." (in a letter to Louis Bouilhet, September the 4th, 1850).

References

[1] F. Rollin, *Suis-je bête ! — L'héroïque Professeur Rollin foudroie la bêtise avec ruse et modestie*, Presses Universitaires de France, 2020, Illustrations by D. Goossens, foreword by A. Astier.

[2] Jean Paul, *Das Lob der Dummheit*, 1782, *Eulogy of stupidity*, no English translation known; French and Spanish translations: Éloge de la Bêtise, Jose Corti, 1993, and Elogio de la estupidez, Sequitur, 2012.

[3] R. Musil, On stupidity, *PN Review* **5**(3) (1978), 5, *Über die Dummheit*, Reclam Verlag, 1937.

[4] L. Latzarus, *Éloge de la Bêtise* , Hachette, 1925.

[5] G. Bechtel and J. Carrière, *Dictionnaire de la Bêtise et des erreurs de jugement, suivi du Livre des Bizarres*, Robert Laffont, Bouquins, 1965, 1991, 2014.

[6] H. Frankfurt, *On Bullshit*, Princeton University Press, 2005.

[7] P. Denis, *Éloge de la Bêtise*, Presses Universitaires de France - PUF, 2001.

[8] C.M. Cipolla, *The basic laws of human stupidity*, Bologna: il Mulino, 2011.

[9] J. Lieber, J.- Mailly, P. Marquis, H. Prade and F. Rollin, Quelques réflexions autour de la notion de bêtise artificielle, in: *16èmes Journées d'Intelligence Artificielle Fondamentale, JIAF 2022, Saint-Étienne, France, June 30 - July 1, 2022*, Z. Bouraoui and A. Wilczynski, eds, 2022, pp. 50–60. https://hal.archives-ouvertes.fr/hal-03781382/document.

[10] A. Turing, Computing Machinery and Intelligence, *Mind* **LIX**(236) (1950), 433–460.

[11] R. Descartes, *Discours de la méthode*, 1637, Discourse on the Method, Optics, Geometry and Meteorology, translation by. P. J. Olscamp, Revised edition: Hackett, Indianapolis, 2001.

[12] E. Bonabeau, M. Dorigo and G. Théraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1999.

[13] C. Godin, Does stupidity exist?, *Le philosophoire* **42**(2) (2014), 27–38, The author develops his point in his *Encyclopédie Conceptuelle et Thématique de laPhilosophie* (Editions Champ Vallon, 2018) in chapter 96 on intelligence which includes a section III on stupidity and a section IV on artificial intelligence.

[14] J.A. Paulos, *Innumeracy: Mathematical Illiteracy and Its Consequences*, Hill and Wang, 1988.

[15] L. Schneps and C. Colmez, *Math on trial: how numbers get used and abused in the courtroom*, Basic Books New York, 2013.

[16] L.J. Savage, *The Foundations of Statistics*, John Wiley and sons Inc., New-York, 1954.

[17] D. Bouyssou, D. Dubois, M. Pirlot and H. Prade (eds), *Decision Making Process: Concepts and Methods*, Wiley, 2009.

[18] J. Weizenbaum, ELIZA - A computer program for the study of natural language communication between man and machine, *Communications of the ACM* **9**(1) (1966), 36–45.

[19] G.N. Yannakakis and J. Togelius, *Artificial Intelligence and Games*, Springer, 2018.

[20] T. Winograd, Understanding natural language, *Cognitive psychology* **3**(1) (1972), 1–191.

[21] P.H. Grice, *Studies in the Way of Words*, Harvard University Press, 1989.

[22] U. Eco, *Experiences in translation*, University of Toronto Press, 2008.

[23] L. Lidén, Artificial stupidity: The art of intentional mistakes, in: *AI Game Programming Wisdom*, Vol. 2, S. Rabin, ed., Charles River Media Rockland, MA, 2003, pp. 41–48.

[24] G. Ritchie, *The Comprehension of Jokes: A Cognitive Science Framework*, CRC Press, 2018.

[25] F. Dupin de Saint-Cyr and H. Prade, La compréhension des histoires drôles : une affaire de révision de croyances, *Revue Ouverte d'Intelligence Artificielle* **3**(3–4) (2022), 253–285.

[26] A. Nijholt, "All the world's a stage": incongruity humour revisited, *Ann. Math. Artif. Intell.* **88**(5–6) (2020), 405–438.

[27] F. Dupin de Saint-Cyr and H. Prade, Belief revision and incongruity: is it a joke?, *J. of Applied Non-Classical Logics* (2023), to appear.

[28] P. Bisquert, F. Dupin de Saint-Cyr and P. Besnard, Assessing arguments with schemes and fallacies, in: *Proc. 15th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'19), Philadelphia, June 3-7*, M. Balduccini, Y. Lierler and S. Woltran, eds, LNCS, Vol. 11481, Springer, 2019, pp. 61–74.

[29] P.R. Halmos, The legend of John von Neumann, *Amer. Math. Monthly* **80** (1973), 382–394.

[30] R. Fagin and J.Y. Halpern, Belief, Awareness, and Limited Reasoning., *Artif. Intell.* **34**(1) (1987), 39–76.

[31] H. Prade, Handling (un)awareness and related issues in possibilistic logic: A preliminary discussion, in: *Proc. of NMR 2006*, 2006, pp. 219–225.

[32] C.K. Riesbeck and R.C. Schank, *Inside Case-Based Reasoning*, Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1989.

[33] M.M. Richter and R.O. Weber, *Case-based reasoning, a textbook*, Springer, 2013.

[34] J. Lieber and E. Nauer, Adaptation knowledge discovery using positive and negative cases, in: *Case-Based Reasoning Research and Development, proceedings of the 29th ICCBR*, A.A. Sánchez-Ruiz and M.W. Floyd, eds, Springer, 2021, pp. 140–155.

[35] É. Astier, H. Iopeti, J. Lieber, H. Mathieu Steinbach and L. Yvoz, Case-based cleaning of text images, in: *Case-Based Reasoning Research and Development, 31st International ICCBR*, S. Massie and S. Chakraborti, eds, Aberdeen, UK, 2023, pp. 344–358.

[36] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intell.* **77**(2) (1995), 321–358.

[37] L. Amgoud and C. Cayrol, A Reasoning Model Based on the Production of Acceptable Arguments, *Ann. Math. Artif. Intell.* **34**(1–3) (2002), 197–215.

[38] K. Atkinson and T.J.M. Bench-Capon, Value-based Argumentation, *FLAP* **8**(6) (2021), 1543–1588.

[39] J. Rossit, J.-G. Mailly, Y. Dimopoulos and P. Moraitis, United we stand: Accruals in strength-based argumentation, *Argument Comput.* **12**(1) (2021), 87–113.

[40] P. Besnard and A. Hunter, *Elements of Argumentation*, MIT Press, 2008. ISBN 978-0-262-02643-7.
[41] A.G.B. Tettamanzi and C. da Costa Pereira, Testing Carlo Cipolla's Laws of Human Stupidity with Agent-Based Modeling, in: *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, Poland, August 11-14, 2014 - Volume III*, IEEE Computer Society, 2014, pp. 246–253. doi:10.1109/WI-IAT.2014.174.
[42] I. Perissi and U. Bardi, The Sixth Law of Stupidity: A Biophysical Interpretation of Carlo Cipolla's Stupidity Laws, *Systems* **9**(3) (2021), 57.