Stupidity and the Ouroboros Model

Knud Thomsen[1]

[1]Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland

knud.thomsen@psi.ch

Abstract. Since decades Artificial Intelligence is striving for an understanding and the production of general intelligence. In an attempt to put the relevant issues into a wider frame, the question shall be asked whether something could be learned from agents or situations characterized by an obvious lack of intelligence, i.e. "stupidity". The Ouroboros Model is a novel proposal for a biologically inspired cognitive architecture. It has earlier been proposed how the Ouroboros Model can shed light on selected cognitive functions including (human) reasoning, learning and emotions. In this short note, implications of the hypothesized structures, relations and processes shall be scrutinized with respect to their possible value for illuminating stupidity and dullness, - and in the end again, natural and artificial general intelligence.

Keywords: Algorithm, Iterative, Recursive, Schema, Process, Consumption Analysis, Limits, Consistency, Intelligence, Stupidity.

1 Introduction

In a series of recent papers the Ouroboros Model has been introduced as an attempt to explain a wide range of findings pertaining to cognition and consciousness of natural and also artificial agents [1-4]. It has been suggested how within a single approach centered around a principal algorithmic process on a suitably structured memory one can explain human cognitive performance and also formulate prescriptions of how to arrive at comparable capabilities for artificial agents implemented in hard- or software following a similar self-steered evolutionary program.

2 The Ouroboros Model in a Nutshell

2.1 Action and Memory Structure

Minds are seen as primarily data processing entities. The Ouroboros Model holds that memory entries are organized into (non-strict) hierarchies of schemata. Memory is made up of meaningful junks, combinations of features and concepts belonging together [1]. In brains, neural assemblies are permanently linked together when once co-activated in a specific manner. Later activation of a feature promotes the selected

concept and leads to graded activation for each of the associated constituents, which are usually active in the same context. Activation at a time of part of a schema biases the whole structure with all relevant slots and, in particular, empty slots, i.e. concurrently missing features.

2.2 Principal Algorithmic Backbone

At the core of the Ouroboros Model lies a self-referential recursive process with alternating phases of data-acquisition and -evaluation. A monitor process termed 'consumption analysis' is checking how well expectations triggered at one point in time fit with successive activations; three interweaved principal stages are identified:

- ... anticipation,
- action / perception,
- evaluation,
- anticipation, ...

These steps are concatenated into a full repeating circle, and the activity continues at its former end, like the old alchemists' tail-devouring serpent called the Ouroboros.

2.3 Consumption Analysis

Any occurring, e.g. sensory, activation excites associated schemata. The one with the highest activation is selected first, and other, possibly also applicable, schemata are inhibited, suppressed. Taking the first selected schema and ensuing anticipations active at that time as reference and basis, consumption analysis checks how successive input fits into this activated frame structure, i.e. how well lower level perceptual data are "consumed" by the chosen schema. Features are assigned to slots / attributes are 'explained away' [5].

If everything fits perfectly the process comes to a momentary partly standstill and continues with new input data. If discrepancies surface they have an even more immediate impact on the following elicited actions [2]. Attention is directed to and by highlighted dissonances. The actual appropriateness of a schema can vary over a wide range. In any case, consumption analysis delivers a gradual measure for the goodness of fit between expectations and actual inputs, in sum, the acceptability of an interpretation. Thresholds for this signal are set in terms of approval levels, depending in turn also on relevant experience in a context. There ensues a constraint and a trade-off: time is short, in the real world not everything can always be perfect, approximations and shortcuts often are good enough, and a wrong schema has to be abandoned at some point and another, new, conceptual frame is tried.

New schemata are preferentially laid down for concepts and episodes which are marked by the output of consumption analysis as deviating significantly from expectations derived from previous experience. Thus self-steered expansion and refinement over time guarantees the gradual and stepwise elaboration of useful hierarchically structured knowledge and behavior, especially in areas where the need surfaced [4].

3 Stupidity

Dozens of different definitions of intelligence have been suggested, and thus it is not immediately clear, with which one a model of general intelligence as outlined in the Ouroboros Model should comply [6]. This is the venture point for attempting to approach the issue from the other side, i.e. looking, what might be learned from situations or constellations, which are characterized by an apparent lack of intelligence, i.e. "stupidity".

Inspired by careful observation of real human behavior, no formal account, e.g. of bounded rationality, is adopted here but the following provocative working definition of real stupidity is suggested:

An agent will be called stupid, if he unwittingly works against his (important) goals and self-interest, not considering information, which is (easily) available.

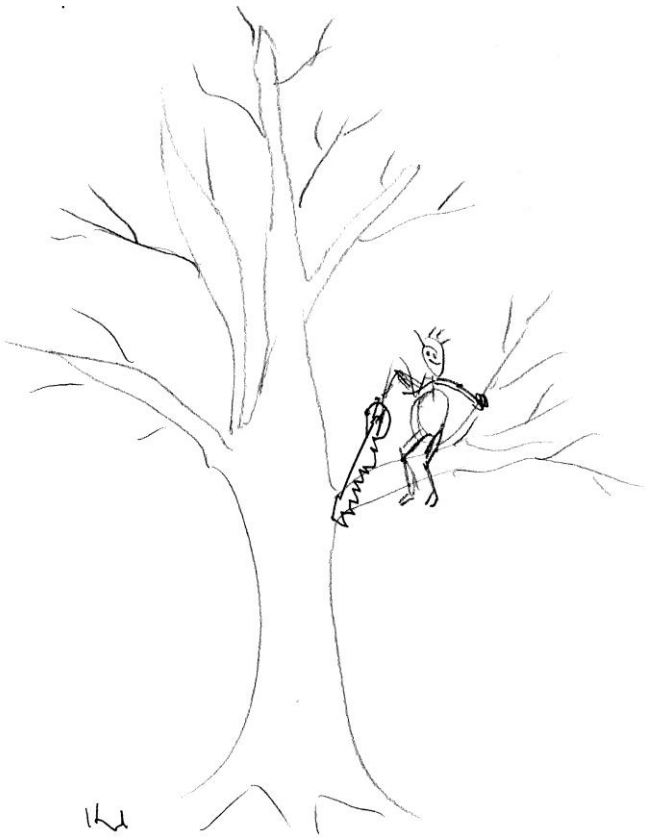

Fig. 1. A precursor of this figure was given in [3] with the caption "A lack in self-awareness can easily become costly to an agent in the real world".

Guided by the classical example as depicted in Fig. 1, some main characteristics of not so clever behavior can be identified and highlighted:

- Stupidity basically is a feature ascribed to an actor by others observing his behavior from the outside; we see the poor guy on the tree and his way of cutting is deemed to cause him a problem.
- A stupid person is focusing on some minor detail while neglecting much more important aspects, which are directly relevant to the goals and the self-interests of the observed; holding fast to that twig certainly is not going to help for long.
- To an outside observer this neglect is hard to understand as a more complete picture and fitting behavior appear to be obvious and easily available to the agent; the guy knows he is cutting the branch he is sitting on, he ought to realize that he is going to fall down and harm himself. Watching this is itself hurtful to the viewer.

The observed points and more can be distilled from the examples of manifestly stupid (human) behaviors collected and analyzed over time by many authors [7,8,9]. As their first common point it shall be emphasized here that natural stupidity is a feature ascribed to an agent from the outside by another knowledgeable agent.

The more the resulting effects of behavior run counter to the (assumed) intentions of the observed actor, the stronger the impression of his stupidity. The same holds true with respect to the availability of alternatives: the easier different behavior, which would lead to success, can be found and performed, the sillier one must be to neglect or discard that information.

Stupidity is individual, and here also the border runs between simple innocent dumbness or blunt incapability and more sophisticated forms of stupidity. If an agent is charged beyond his capabilities, his behavior will not be optimum, but we would be much more willing to excuse his failure than if he could have managed if not too lazy or ignorant and arrogant. The same performance, which might be rated respectable for a student or novice, could fall under the rubric of stupid reaction if it came from a Nobel laureate in his claimed field of expertise.

Quite general, being an accomplished expert in one restricted field does not necessarily imply overarching intelligence. The Ouroboros Model stresses the 'local' nature of schemata and effective processes, and it identifies this at the same time as inherent constraint and principal limitation for performance.

Sophistication is possible in countless directions; as one commonplace and important example, an arrogant agent might be guided by a severely distorted self-assessment and be wrongly convinced that he knows very well what to do, even actively rejecting any help. ("Pride and ignorance are akin" the proverb knows).

Judging the motivation behind any non-trivial behavior from the outside is intrinsically difficult as it amounts to guessing some only personally accessible parts of the total frame relevant to the actor. Sometimes it thus might be hard to tell whether an act belongs to the category of stupidity or even obstruction and sabotage.

As a preliminary summary one might say that in any variant of stupidity information directly relevant to a behavior is mistakenly not taken into account in any appropriate and apparently obvious way. In terms of the Ouroboros Model this means that the applied schemata are a poor representation of the relevant features of reality

or that the processes working on the basis of these schemata are faulty, e.g. perceptions are distorted or discrepancies are ignored.

4 General Intelligence

In the light of the claimed characteristics of naturally occurring stupidity one is tempted to see a reason for the obvious difficulty for defining intelligence in an objective and unquestionable way in the non-existence of context-free intelligence. Some statements concerning intelligence, inverse to the above hypotheses referring to stupidity, suggest themselves and seem intuitively right:

- Intelligence is a label that humans grant to other (rational) agents.
- An intelligent person (agent) takes into systematic account all conceivably relevant data, possibilities and circumstances, reaching with coherent and correct reasoning much beyond the directly obvious.
- The more unexpected and sophisticated a decisive combination or insight is, the higher we appreciate the feat and the apparent cognitive skill.

So, what would be an appropriate counterpart to Fig. 1? A modest attempt to provide one example is sketched in Fig. 2.

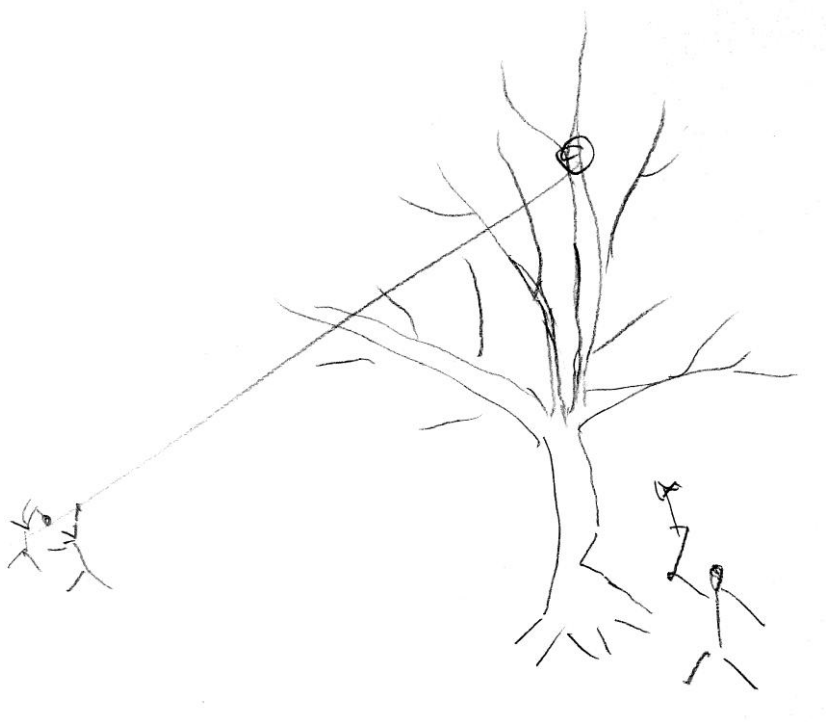

Fig. 2. Clever is, who applies an understanding as wide as possible, chooses appropriate tools as available and accepts help from friends.

Importantly, in contrast to the gist of what is usually understood as intelligence, the sketch of a definition as given in the caption of Fig. 2 does not mention success as a criterion for a person or an action to qualify as intelligent. Adaption and success, no doubt, in the long run are more often the consequences resulting from intelligent rather than from stupid actions but in an open and complex world there is no simple one-to-one correlation. There are cases, e.g. of limited resources, in which no measure of intelligence can guarantee a favorable outcome; and it might not even be possible in hindsight to judge. An agent might have failed and still, she might have acted as intelligent and efficient as possible at all. Many a success is the child of mere luck.

Quite commonly, it is often not so clear, how to qualify an action or omission. Intelligence and stupidity are individual also in the sense that different observers might arrive at contradicting conclusions referring to one and the same instance. On top of observers' personal limitations principal restrictions due to bounded resources apply; this results in a grey scale comprising fundamental uncertainty.

The idea that intelligence is to be attributed by judges in a particular situation is not new; it lies at the heart of Turing's famous test [10]. The first chapter-title in his groundbreaking paper gives also a hint on how to achieve intelligence, and at the same time it names a limitation with the keyword of "imitation". Observing others, remembering and copying their (successful) behavior certainly can help to avoid some mistakes; on the other hand, simple imitation lacks the features of novelty, understanding or creativity, of which at least the first two are most often considered hallmarks of intelligence. Teachers are familiar with the related distinction between learning something by rote and true comprehension.

At a somewhat abstracted level, the cognitive skills of a person can be determined rather reliably and reproducibly. Interestingly, also the best available quantitative measure of general intelligence, i.e. the IQ of an individual as derived from his performance in a series of well defined and validated tests, is relative; it refers in its definition to a comparison to an ensemble comprising many other persons.

5 Intelligence according to the Ouroboros Model

It is hypothesized by the Ouroboros Model that not only representational capacity, but the total potential mental processing power of an agent is ground-laid as well as limited by structured knowledge. i.e. the number, complexity and elaboration of the concepts at her disposition [4].

Differentiated schemata, their numbers of slots, the level of detail, the depth of hierarchies, the degrees of connection and interdependence of the building blocks, and the width, i.e. the extent of main schemata and their total coverage from a bodily grounding level to the most abstract summits, determine what can be done or thought of efficiently. Sheer performance at a single point in time arises as a result of the optimum interplay between these structured data and the effective and systematic execution of the processing steps, in particular, self-referential consumption analysis.

Most important, efficient (long-term and working-) memory is seen as a mandatory prerequisite of intelligence. Understanding working memory as temporary common

activation and binding of current input with material laid down in long term storage makes this proposal consistent with vast evidence, which indicates that working memory capacity is a valid predictor of general intelligence [11.12].

The pursuit of consistency as the fundamental basis for rational behavior and the efficient self-steering by consumption analysis entails a principal difficulty for differentiating between cunningness and foolishness or dullness [4]. If something is marked as strongly discordant with hitherto experience, it might be because it is novel and brilliant, or, it might simply be wrong, even stupid. Only the intelligent (!) embedding of the issue in question in the widest possible frame and taking into account all conceivably relevant information can offer a chance of meaningful and fair assessment. Stupidity can sometimes be temptingly simple and beautiful, while thoughtful elaborations, often tedious, might be deterrent.

Appraisal of any elaborate and complex behavior cannot be collapsed onto one single scale running between stupid and intelligent; there are much more dimensions to this topic than can be indicated with a few opposites. As just one example, a related distinction could be between rational and irrational. In the light of the Ouroboros Model, rational behavior takes all relevant information into account in a systematic manner, it can be comprehended in detail, and it necessarily entails consequent and traceable reasoning. Irrational actions cannot be fully understood by a spectator, at least not on the basis of the knowledge applicable for him at that point in time. Creativity would in this frame still lie somewhere in an orthogonal direction emphasizing the feature of unprecedentedness while relaxing requirements for stringency, "usefulness" and "grounding".

6 Conclusions

The Ouroboros Model holds that intelligence is conceded by a spectator to an actor whenever the actor consistently brings to bear all information considered to be really relevant by the spectator.

There is nothing like absolute stupidity or intelligence, no such observer-independent objective entities do exist. Whether a behavior is called clever or silly depends on the contexts prevalent for the involved agents at that particular time in question. The employed frames of reference can make a decisive difference.

Both figures in this paper are misleading; no adequate picture of stupidity or intelligence can be drawn in black and white. Any qualification of cognitive performance is possible only in a wider context admitting the agents, (self-) involved as actors and observers. The described mandatory inclusion of a human (even consciousness-) component into definitions of intelligence and stupidity can explain why so far no consensus has been reached among researchers on the essential abstract characteristics or a unique definition of intelligence.

The actual result of an action cannot righteously be taken as touchstone for the intelligence leading to that outcome. The situation is a little like looking into a mirror: the distinction between up and down is easy as well-defined by an outside reference; judging intelligence versus stupidity, alas, corresponds more to telling left from right.

Self-reflecting, we emphasize as one defining feature of stupidity a deficiency in considering all relevant (in principal available) information, which is judged as essential for success by an observer. This evaluating instance need not necessarily be a different person; it can be the actor himself reflecting on his own behavior at some earlier point in time.

A special case worth while mentioning in this respect are situations and behavioral options for which short-term benefits to an actor entail long-term disaster [9]. It certainly is a sign of stupidity, if the temporal dimension is not carefully paid heed to, and, for example, an action is judged as ingenious because it had turned out a success, - it is well possible that at the time of taking the corresponding decision, this particular behavior could only have been condemned as stupid; - the same is also possible the other way round.

In a setting of intended collaboration, stupidity of an actor can still easily be topped by an observer, - if he clearly displays that he considers the first one stupid.

The Ouroboros Model stresses the indispensability of orderly processes and it highlights the preeminent importance of exploiting a knowledge base as vast as ever possible. There is no contradiction to efficiently employing well-tuned short cuts and heuristics [13]; - the latter can be seen as a means of taking the dimension of time and, in particular, time-limitations in a natural and dangerous world, into due account.

A strong link between intelligence and consciousness has been proposed, see Fig. 1, and it has been argued that at a certain level of intelligence consciousness emerges naturally, even inevitably [3,14].

The Ouroboros Model claims to offer a self-consistent and self-relational consistent approach for understanding and avoiding stupidity in natural and artificial agents and for fostering the self-steered growth of intelligence. The general advice for guarding against stupidity that could be drawn from it might best be summarized in a plea for aiming at all-embracing consistency in the widest possibly and applicable frame.

Acknowledgments

It might be a bit delicate for this given subject but it should not be concealed that the author is indebted to certain individuals for providing a wealth of inspiration.
In enjoyable contrast, responding to encouraging questions and suggestions posed by unknown reviewers was a true pleasure.

References

1. Thomsen, K.: The Ouroboros Model in the light of venerable criteria. *Neurocomputing 74*, 121-128 (2010)
2. Thomsen, K.: The Ouroboros Model, Selected Facets. In: C. Hernández et al. (eds.) From Brains to Systems. New York Dordrecht Heidelberg London: Springer, 239-250 (2011)
3. Thomsen, K.: Consciousness for the Ouroboros Model. *Journal for Machine Consciousness, 3*, 163-175 (2011)

4. Thomsen, K.: Knowledge as a Basis and a Constraint for the Performance of the Ouroboros Model, presented at a workshop at ZiF in Bielefeld, October 29-31, (2009)
5. Yuille, A. and Kersten, D.: Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Science 10*, 301-308 (2006)
6. Legg, S. and Hutter, M.: A Collection of Definitions of Intelligence. Proceedings of the 2007 conference on Advances in Artificial General Intelligence: Concepts, Architects and Algorithms (2007)
7. Geyer, H.: Über die Dummheit. VMA-Verlag Wiesbaden (1954)
8. Van Boxsel, M.: Die Enzyklopädie der Dummheit. Eichborn AG, Frankfurth am Main (2001)
9. Welles, J.F. : Understanding Stupidity. Mount Pleasant Press, NY (1995)
10. Turing, A.M.: Computing machinery and intelligence. Mind, 59, 433-460 (1950)
11. Ericsson, K.A. and Kintsch, W.: Longt-term working memory, *Psychological review 102*, 211-245 (1995)
12. Oberauer, K., Süss, H.-M., Wilhelm, O., and Wittmann, W.W.: Which working memory functions predict intelligence? *Intelligence 36*, 641-652 (2008)
13. Gigerenzer, G.: Peter M. Todd, ABC Research Group (Ed..): Simple heuristics that make us smart. Oxford University Press, New York (1999)
14. Sanz, R., López, I., Rodríguez, M., and Hernandéz, C.: Principles for consciousness in integrated cognitive control. *Neural Networks 20*, 938-946 (2007)