Marvinian Hope and Disappointment

Walter Ratjen & Otto E. Rössler, Tübingen

Marvin, the paranoid android of Douglas Adam's "A Hitchhiker's Guide to the Universe", while hoping for the best, always prepares for the worst. We describe the fundamental mathematics which he functions by.

1. Introduction

Defining consciousness implies differentiating between phenomenal ("3rd person account") and introspective ("1st person account") consciousness. Proving the latter means probing the former. To design an appropriate test we endow the system to be tested with values that range from plus to minus unlimited on scales of intensity.

Values ranging without limits are equivalent to addable potentials. Aristotle seems to have been aware of this when he taught that the falling stone accelerates because it looks forward to getting home and to being at rest. Let us assume that the stone contains magnetite. Then we translate "house" and "rest" into "gravity" and "magnetic field." The resulting model is a special case of a Newtonian system of multiple potentials. The gradients of these are finite, and thus are computable.

We assume that potentials with infinite gradients are exclusive to motivation and emotion. Equations of potentials of this kind, therefore, describe brains. They are brain equations. To make the brain complete we need a cognitive system to complement the motivational part. It comprises the abilities of spatial and temporal anticipation, potencies capable of being formally defined. Thus, building a brain rests on formalizing the in principle infinite potency equations of motivation and emotion.

Our model given in chapter 4 is complete in this respect. Realized, it would give rise as well to phenomenal as to introspective consciousness. Furthermore, we hypothetize that the "social" interaction of merely two of these brains, provided they are "complete" in the way described, inevitably leads to sacrificial behaviour, i.e. behaviour that implies own disadvantage to the benefit of the other. This would give rise to phenomenal personship.

2. The motivational subsystem

The Newtonian model mentioned above consists of potential equations of the form

$$P_{\Sigma} = \sum_i P_i(x) \qquad (1)$$

The gradients of these describe forces:

$$F_i = \frac{\delta P_i}{\delta x} \qquad (2)$$

The sum total of these is given by

$$F_{\Sigma} = \sum_i F_i \qquad (3)$$

Forces can be positive, and in this case attract to their source, or can be negative, in which case they repel from their source. Positive forces are produced by increasing potentials, while negative forces are produced by decreasing potentials.

If a positive potential Pi+ increases while the distance between the system and the source of the potential decreases, i.e. shows a positive gradient, then we call it an **attraction potential**

If a negative potential Pi- decreases while the distance between the system and the source of the potential decreases, i.e. shows a negative gradient, then we call it a **repulsion potential**. The pertaining forces Fi+ and Fi- vary accordingly.

Both potentials are equivalent in the sense of physics. Increase at approach and decrease at increasing distance are the same numerically. The difference is a physiological one in that it is relevant to the behaviour of either approaching or avoiding a potential. This very behaviour decides the definition of the potential as such as being positive or negative. The strength of the behaviour mirrors the force of the potential. The behaviour, lead by these forces, follows a trajectory in space and / or time, the dynamics of which differ under the two conditions of either attraction or repulsion.

The trajectory may contain locally positive gradients. This would lead to an increase in attractivity of the source of the potential. This increase will boost the force Fi+, which is directed towards the positive potential.

If the trajectory contains locally negative gradients, this would lead to a decrease in repulsivity of the source of the potential. This decrease will inhibit the force Fi-, which is directed away from the negative potential.

Thus, local conditions may cause either an autocatalytic reinforcement of an attraction potential, or an autocatalytic inhibition of a repulsive potential.

Thus, the local movement of the system may be caused by either the decrease of a repulsive potential, or by the increase of an attractive potential, or by a combination of both.

This leads to potential equation (4)

$$P_{\Sigma} = \sum_i P_i^{+}(x) - \sum_i P_i^{-}(x) \qquad (4)$$

which is complemented by force equation (5)

$$F_{\Sigma} = \sum_i F_i^{+}(x) - \sum_i F_i^{-}(x) \qquad (5)$$

The trajectory the system follows results from controversing infinite forces.

3. An example

First case, called "generic":

A 2-dimensional field may be situated between a source of positive forces on top, and one of negative forces at the bottom. The trajectory of the system may depart from a point nearer to the bottom than to the top, but more or less in the center between the left and right edges of the plane. Its first move will be to the right to increase the distance between itself and the negativity as fast as possible. Thereafter it will describe an elegant curve to the left towards the positive force at the top.

Second case, called "nongeneric":

Let us assume the same 2-dimensional plane as before, but now with 2 positive sources at the top, one to the left, the other one to the right. The point of departure may be the same as before. The trajectory, at first, will aim towards that positive source that happens to be the closer one of the two. Let us consider the xtreme case of the system being situated at a point of departure that is located near the bottom at exactly equal distances towards either one of the two positive sources. The trajectory will proceed straight upwards, then halt at the saddle point at the top equidistant to the two sources. Undisturbed, the system will stay forever at that location.

This reminds of the ass of Buridan which starved to death between two stacks of hay, because it couldn't make up its mind as to which stack to eat first, with both stacks looking equally delicious. Our model is immune against this case, because we included spatial as well as temporal dynamics of forces as well as of potentials.

The positive forces are given by equation (6)

$$F_i^{+} = \frac{t_i \cdot (1 - t_i) \cdot 0.01}{(r_i + 0.01) \cdot \cos \varphi_i^{+}} \qquad (6)$$

and by equation (7)

$$F_i^{-} = \frac{-0.01\, t_i^{2} \cdot \dfrac{1 - t_i}{1 - (t_i - r_i)}}{(r_i + 0.01)\,(1 - \cos \varphi_i^{-})} \qquad (7)$$

for negative forces, respectively.

The time elapsed since visiting a source of type i is denoted t-i. The distance between the system and source i is denoted r-i. The azimut of the trajectory is denoted phi-i. For details, see [1].

The forces as well as their integrals, the potentials, are time dependent. This defines a special case of the motivational system, a case known as "traveling salesman problem." A motivational system that comprises and solves this very problem is called a "**directional optimizer**." - Note that the seemingly cognitive problem of the traveling saleman is a special case of the motivational subsystem of the brain, not of the cognitive subsystem that we have not dealt with at all, so far.

In simulating this system, 2 parameters need to be predefined, namely average distance and permitted error ratio.

Ofcourse, the strongest potential obtains priority. With the potentials being time-dependent, though, the behaviour of the system is unpredictable due to an environment of sources of potentials and voids of potentials. It is even chaotic, because the trajectory is sensitive to minute differences in initial conditions.

4. Simulating the system

The motivational system is devoid of any anticipation. In following gradients its actions are local. This changes as soon as we construct the system in such a way as to be capable of deciding between internal and external space. This splitting of space is a necessary consequence of the very model elaborated so far. When talking about sources of potentials we implied that gradients can somehow be recognized and measured. Input devices, though, have to be assumed as being non-panoramic. This limitation leads to the trajectory being locally determined and proceding in a sequential mode.

The first step in establishing space by constructing an internal space is to alternate between a phase of exploration and a subsequent phase of evaluation. The latter process needs an internal working memory.

The system generates the source related forces by computing sum gradients, or sum vectors, then to follow these forces mechanically. This process we call<B> "optimizing."</B> As the system itself computes the forces it is subject to we call it <B>"autonomous."</B> An autonomous optimizer is a motivational system, which itself is one of two subsystems of a "brain."

Before a system can start to optimize, i.e. chart its course, it needs to identify specific sources and to acquire and store the relevant information about them. Identifying the specifications of sources is an optimizing process of its own. It is governed by a scalar potential of success. - With this, the machine possesses the following **modes of operation**:

FIRST MODE: Optimizing in real space.

SECOND MODE: Exploration in real space. Potentials are set to scalar values.

THIRD MODE: Optimizing in real space. Potentials are set to vector values.

FOURTH MODE: Optimizing in imagined space. Potentials are vectors, but sensorimotor input / output is inhibited.

FIFTH MODE: Restart of mode 3 with results won in mode 4.

Non-panoramicity of sensors is sufficient to produce phenomenally intelligent behaviour, because it requires phases of explorational behaviour.

5. An illustration

We may conceive of the exploring subsystem in analogy to a sonar scanner similar to the ones applied in pre-natal diagnostics. The subsystem projects spatial information onto a screen. The total system we then call a **universal simulator** (cf. [2], [3]). It is a genralization of the simulator proposed by Kosslyn and Schwartz in [4].

Mode no. 5 is of special interest. The results of mode 4 modify the environment of spatially distributed potentials. A source that may have appeared to possess a strong attraction potential under mode 3 may now be revealed as being juxtapositioned to a strong, though short-ranged repulsive potential. The autonomous optimizer acts in imagined space (cf. [5]). Imagined space modifies real space. A respective system is called a "**path optimize**r."

6. Discussion

The motivational system functions in a continuous way, as opposed to a discrete mode. Its anticipatory faculty is defined as the ability to compromise between forces of attraction and repulsion in accordance with the principle of maximum pay-off. Without a RAM based learning ability it yet is capable of producing best possible solutions under given conditions.

A naive observer who has no knowledge about the constructive details of the system yet will intuitively interpret its actions. If the system is in mode 5 he may use anthropomorphic metaphors like "hope" and "disappointment" to describe the acts of the system.

The strongest part of the Turing test is the test for altruism, which therefore is included in principle. While humans possess human rights because of being humans, regardless of whether they pass the Turing test or fail, computers need to pass in order to acquire personality rights. We suggest they be granted a legal status equal to those pieces of art that rank as heritage of mankind.

LITERATURE

[1] ROSSLER, O. E.: Adequate locomotion strategies for an abstract organism in an abstract environment. in: Lecture notes in Biomath 4, 342-369, 1974.

[2] ROSSLER, O. E.: A universal simulator. in: Systems Science and Science. (Banathy, B.H., ed.) 143-145. San Francisco: A.A.A.S. 1980.

[3] ROSSLER, O. E.: An artificial cognitive map system. in: BioSystems 13, 203-209, 1981.

[4] KOSSLYN, S.M. and S.P. SCHWARTZ: A simulation of visual imagery. in: Cognitive Science 1, 265-295, 1978.

[5] LORENZ, K.: Behind the Mirror. (German ed. Munich 1973)

[6] ROSSLER, O. E. and R. ROSSLER: The lion and the princess. How an artificial brain changes the attractivity of locations through simulated encounters. Biophys. J. 78, 253A, 2000.