

Why know myself? Flexible behaviour and the need for self-modelling

Julian Hauser | University of Barcelona

The gist of it: *Flexible behaviour requires a certain form of representation, a self-model. Self-models are similar to centred maps, containing an objective model of the world and a 'you-are-here' marker. Such an architecture is at the root of subject/object differentiation. This research therefore sheds light on the evolutionary emergence of self-representation.*

What is flexible behaviour?

A system engages in *flexible behaviour* when its output – or behaviour – isn't simply a function of sensory input. The response function becomes: $O = f(I, x_1, x_2, \dots)$, where O = output, I = input and x_n = additional parameters.



Visual input on day 1



Visual input on day 2

Arsenius encounters a construction site on his way to work. The next day, he picks a different route *even if the sensory input on leaving the house is exactly the same*. Some parameter x must determine his response.

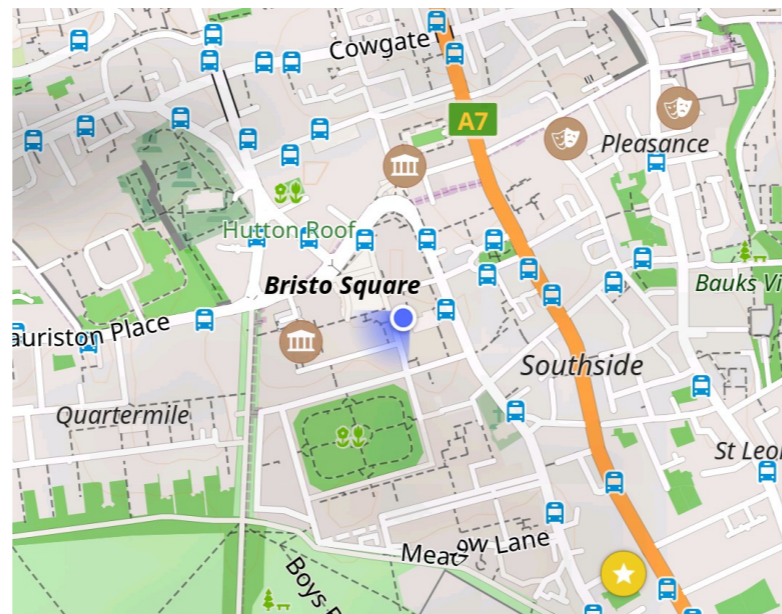
Forms of flexible behaviour:

- (1) Simple flexible behaviour (above)
- (2) Increase in flexible behaviour (adding parameters to response function)

- (3) Massive behavioural flexibility (massive number of parameters)
- (4) Sophisticated behavioural flexibility (parameters capture objective relations between objects or events)

Claim: Self-models are advantageous for (3) and necessary for (4).

Self-models

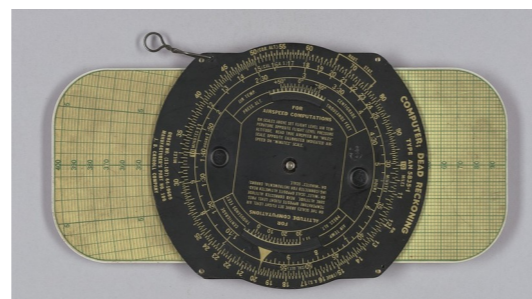


Self-models have two aspects:

1. Objective model of the world, i.e. the 'map'
2. Self-as-subject: i.e. the 'you-are-here dot'

This is the root of subject/object differentiation. Objective system properties are represented as relations to a subjective core.

Contrast: Non-self-modelled representation (e.g. dead reckoning)



Dead-reckoning computer

The values of the parameters with which a dead reckoning system (e.g. ant) represents its states capture divergencies from the

present state. The subjective and the objective component are run together.

Cognitive advantage of self-modelling

Advantage 1: Updating values after movement

As long as many state changes are due to the system itself, these can be represented simply by updating the position of the self-as-subject. A non-self-modeller, in contrast, will have to update each property individually. This means version (3) of flexible behaviour is computationally expensive for non-self-modellers.

Advantage 2: Representing relations *as* relations

A self-modeller can make its response dependent on relations between features in its objective model of the world. A non-self-modeller cannot even in principle represent such objective relations as all states are represented as differences to the systems actual state. This means that version (4) of flexible behaviour is impossible for non-self-modellers.

Where to now?

In the paper of this poster, a number of additional points are discussed:

- Not all self-models are map-like.
- Not only spatial properties can be represented this way.

Future research areas:

- Can the account solve puzzles in philosophy of selfhood?
- Are the findings born out by empirical research?
- How does subjectivity in the sense here integrate into an account of full-blown human self-representation?