DOI: http://dx.doi.org/10.7551/978-0-262-31050-5-ch040

# Second Order Learning and the Evolution of Mental Representation

Solvi Arnold[1], Reiji Suzuki[1] and Takaya Arita[1]

[1]Graduate School of Information Science, Nagoya University, Aichi 464-8601, Japan
solvi@alife.cs.is.nagoya-u.ac.jp

**Abstract**

Mental representation is a fundamental aspect of advanced cognition. An understanding of the evolution of mental representation is essential to an understanding of the evolution of mind. However, being a decidedly mental phenomenon, its evolution is difficult to study. We hypothesize how interactions between adaptation levels may cause emergence of isomorphism between a cognitive system and its environment, and that mental representation may be understood as an instance of this effect. Specifically, we propose that selection for second order learning translates into selection for isomorphism-based implementation of first order learning ability, and that mental representation is (an aspect of) the environment-cognition isomorphism produced by such learning ability. We then give a reformulation of cognitive map ability, a paradigm case of mental representation, in terms of our hypothesis and explore it computationally by evolving a neural network species with the neural basics for second order plasticity (the basis for second order learning) in an environment composed of randomly generated maze tasks, including tasks generally believed to require mental representation (in the form of cognitive maps). The model is shown capable of evolving nets that solve these tasks, providing preliminary support for our hypothesis.

## Introduction

Mental representation (MR for short) is, abstractly put, the ability to simulate or reconstruct in the mind aspects of the environment that lie outside the scope of one's current perception. The type of MR we focus on in this paper is the ability to navigate complex environments using "cognitive maps": mental representations of the spatial layout of an environment (see Tolman 1948). Cognitive maps aid navigation, as typically only a limited part of the area to be navigated is directly perceptually accessible. Other types of MR are "mental time-travel" and "theory of mind" (see Takano & Arita 2006, Minoya et al., 2011 for computational approaches to the latter). There too, inaccessible aspects of the environment (respectively: future and past, other minds) are mentally simulated or reconstructed.

The evolution of MR is not well-understood. MR is a highly structured and organized form of cognition, and already in the early decades of connectionism, it has become clear that (contrary to common intuition) adaptive processes such as evolution or learning do not, in general, produce such structured or organized AI (see e.g. Fodor & Pylyshyn, 1988). If our simulated adaptation processes do well at producing non-representational cognition, but fail to produce representational cognition, then this raises the question how MR can have evolved in biological cognitions. The question seems particularly important since there appears to be a tight conceptual link between representation and intelligence. Non-representational cognition can be attained via what we might call blind adaptation, be it mutation and selection fashioning fit but fixed innate behaviour, or trial-and-error learning chasing a reward-signal. Representational cognition, or at least the behaviour we recognize it by, is characterized by more advanced forms of adaptation. We recognize intelligence by the *absence of trial and error:* a solution is mentally represented, then executed. *Insight* crucially depends on representation.

We propose the following explanation of evolution of mental representation ability: As learning ability evolves, the need for trial-and-error is reduced. This reduction is attained by adapting to the environment the process that adapts behaviour to the environment, that is, by second order learning. Selection pressure on second order learning translates into selection pressure on isomorphism-based implementation of first order learning. Mental representation is part of this isomorphism.

The idea of a central role for isomorphism in the evolution of cognition is not new: Herbert Spencer viewed the evolution of mind as ever expanding correspondence between the internal and external (Spencer 1855, see also Godfrey-Smith 1996). Our contribution is a hypothesis on how evolution and the orders of learning interact to produce such correspondence.

We provide a proof of concept for our hypothesis in the form of a computational model in which a neural network species with the basic constituents for second order plasticity (the neural basis for second order learning) is evolved in an environment containing maze tasks generally believed to demand cognitive map ability. The model is shown capable of evolving nets that solve these tasks, providing preliminary support for our hypothesis.

## Isomorphism & Learning

Our theory explains MR as an instance of a more general organization effect. In this section we explain this effect, and in the following sections we discuss how it applies to MR. We first define our main terms:

*Behaviour*: a mapping from stimuli ($S$) to responses ($R$).

$$\boldsymbol{L_0: S \rightarrow R}$$

We denote behaviour as $L_0$ because in our theoretical framework it occupies the position of zero-order learning.

*1st order learning*: given the current behaviour and a stimulus, updates the behaviour.

$\boldsymbol{L_1: (S, L_0) \rightarrow L_0}$, i.e.:

$\boldsymbol{L_1: (S, (S \rightarrow R)) \rightarrow (S \rightarrow R)}$

*2nd order learning*: given the current 1st order learning mapping and a stimulus, updates the 1st order learning mapping.

$\boldsymbol{L_2: (S, L_1) \rightarrow L_1}$, i.e.:

$\boldsymbol{L_2: (S, (S, (S \rightarrow R)) \rightarrow (S \rightarrow R)) \rightarrow ((S, (S \rightarrow R)) \rightarrow (S \rightarrow R))}$

And so on for higher orders, though we do not concern ourselves with anything above $L_2$ here.

*Environment*: a mapping from responses to stimuli:

$\boldsymbol{E: R \rightarrow S}$

Note that an environment is much like an inverse behaviour (mapping responses to stimuli instead of stimuli to responses).

Each mapping may additionally update internal states, but for simplicity we leave this out in notation as we do not need it for this explanation.

We have defined mappings, but organisms are physical objects, not mathematical objects. In order for these mappings to exist in the physical world they must have implementations. For each of the mappings defined above, we let its lowercase partner denote its implementation: $l_0$, $l_1$, $l_2$, $e$. Implementation of the environment mapping here should be understood as the actual physical reality of the environment.

An organism's fitness depends on the mappings it implements, but generally not on *how* it implements those mappings. Any given mapping can be implemented in infinitely many ways (indirect fitness effects such as energy cost may weed out overly unwieldy implementations, but still leave many viable options). This poses a problem for understanding the evolution of mind: mental phenomena are part of the implementation of our mappings, but evolution does not generally care how we arrive at our responses as long as they fit the stimuli that triggered them. Why *this* implementation, and not another?

How real of a problem this is can be clearly seen in the history of the philosophy of artificial intelligence: the problem of connectionist systematicity may be interpreted as the problem that artificial adaptation processes typically fail to pick a systematic implementation from the set of viable implementations that solve the problem set they run their adaptation process on. Highly diffuse and unorganized implementations are viable for surprisingly complex tasks.

Yet we feel quite sure that cognition is the product of evolution, and given how systematic and seemingly organized it is, it seems unsatisfactory to appeal to coincidence as cause for this particular implementation. The question of what factors guide implementation choice in evolution is essential to an understanding of the evolution of mind, but remains largely unanswered.

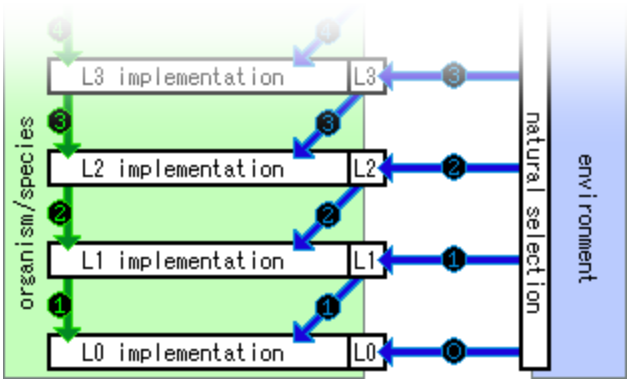


**Fig. 1.** Evolution and the orders of learning. Green arrows indicate adaptation on within-lifetime timescales (i.e. learning), blue arrows indicate adaptation on evolutionary timescales. As indicated by the diagonal blue arrows, selection operates only indirectly on the implementation structure of any given adaptation level, via the effect that implementation structure has on the feasibility of the adaptation level above it. In general, direct selection for adaptation level *i* converts into indirect selection for isomorphism-based implementation of adaptation level *i-1*.

Figure 1 expresses the relations between evolution and the orders of learning (including behaviour, as $L_0$). Each order of learning $L_i$ ($i > 0$) adapts $L_{i-1}$ (the green downward arrows) on a within-lifetime timescale. We pick any two adjacent orders of learning ($L_0$ and $L_1$ in (Arnold, 2011), $L_1$ and $L_2$ in most of this paper). If the environment has the sort of dynamics to which $L_i$ is applicable, then there is selection pressure on evolution of $L_i$. Different implementations of $L_{i-1}$ call for different implementations of $L_i$. For example in the highly unnatural case that $l_{i-1}$ would take the form of a table defining an output for each possible input independently, then $l_i$ would operate by rewriting entries of this table. So whether and how feasible evolution of $L_i$ is strongly depends on $l_{i-1}$. If there is selection pressure on $L_i$, then mutations in $l_{i-1}$ that are beneficial to $L_i$ are beneficial mutations (even if they have no effect whatsoever on $L_{i-1}$). As an extreme scenario, we could imagine $L_{i-1}$ remaining stable while $l_{i-1}$ evolves to facilitate $L_i$. This possibility shows that there is a fundamental difference between selection for a specific mapping and selection for a specific implementation of that mapping.

So while evolution working on $L_{i-1}$ alone does not care much about the structure of $l_{i-1}$, "co-evolution" (if we may abuse the term a little) of $L_i$ and $l_{i-1}$ *does* care about the structure of $l_{i-1}$. Along the horizontal blue arrows in figure 1, evolution treats its objects as black boxes (selecting on input-output relations alone), but indirectly through the neighbouring learning order above it (diagonal blue arrows), it peeks inside and selects for implementation structure.

$L_i$ constrains $l_{i-1}$, but we haven't said anything yet about what sort of $l_{i-1}$ is favoured by $L_i$. We will claim that $L_i$ benefits most from $l_{i-1}$s that are in some sense isomorphic with the environment. For the simplest case, $L_0$ and $L_1$, the basic idea is as follows: If the environment and (consequently) the optimal behaviour are static, then difference in the structure of their implementation poses no problem. But if the environment and (consequently) the optimal behaviour may change (by means of $L_1$), then the more the structure of $l_0$ and $e$ differ, the harder it is for $L_1$ to update $L_0$ in sync with $E$. The implementations

($e$) of environments that cognition evolves in are composed of distinct aspects (food sources, temperatures, other agents, spatial layouts, etc. etc.) that act and interact to give rise to $E$. Let's call a change in one such aspect a *simple* change. Simple changes in $e$ often lead to *complex* changes in $E$: multiple input-output pairs change. Consequently a complex update of $L_0$ is required. If $l_0$ contains an aspect corresponding to the changed aspect of $e$, in a functionally similar position, then the required complex change in $L_0$ can be realized by a simple change in $l_0$. This makes $L_1$ quite feasible. If no such corresponding aspect exists, a complex implementation update is required. In this case no straight-forward relation exists between the environmental change and the appropriate behaviour change, making $L_1$'s work difficult or infeasible.[1]

So the organization that evolves in $l_0$ to facilitate $L_1$ should in one form or another capture the variable aspects of the environment along with their functional roles therein. This correspondence is what we mean by isomorphism. Note that we do not claim that $L_i$ is strictly impossible without isomorphism between $e$ and $l_{i-1}$, nor that such isomorphism cannot occur in absence of $L_i$. What we claim is that selection pressure on $L_i$ translates into selection pressure on isomorphism at $l_{i-1}$, and that this selection pressure conversion is an organizing factor in the evolution of cognition.

Hopefully the argument for the case of $L_1$ and $l_0$ is clear now, but the focus of this paper is the case of $L_2$ and $l_1$. The two cases differ most importantly across the type/token distinction between species/specimen, that is, the timescale on which isomorphism is acquired. Without $L_1$, $L_0$ is static per specimen (for convenience we ignore other factors that modify behaviour). It is a given individual's innate behaviour. With $L_1$ redirecting selection onto $l_0$, we should see *evolution* of isomorphism in $l_0$, i.e. isomorphism in the innate organization of cognition (*innate isomorphism*). This effect was demonstrated in (Arnold, 2011). The topic of this paper, MR, may be described as isomorphism, but clearly it is not innate isomorphism. Mental representations are acquired on the within-lifetime scale, the timescale of learning. Mental representations are a form of *acquired isomorphism*, the result not of evolution processes but of learning processes.

What is the function of acquired isomorphism? Looking at Figure 1, we may hypothesize that, just as isomorphism evolved at $l_0$ benefits $L_1$, so does isomorphism acquired at $l_1$ benefit $L_2$. Evolution of isomorphism-acquisition at $l_1$ (which should include MR), then, would be a consequence of selection for $L_2$.

[1] One might object that reinforcement learning algorithms manage to learn just fine without dependence on such "corresponding aspects". However, such algorithms depend on a reinforcement signal. We cannot in general assume such a signal to be available. Stimuli delivered by the environment may convey information about the fitness effects of a given response, but in natural settings the signal is more often than not incomplete, extremely noisy, or absent altogether. Even when a clear signal is available, a reinforcement learning algorithm must still depend on extensive trial-and-error learning to adapt to a complex change in E, even if the change in $e$ is simple. Learning in biological species is routinely seen to do better than that, on basis of less information (e.g. first language acquisition).

This somewhat cryptic proposition will become clearer in the next sections, where we apply our framework to a well-known experiment from cognitive psychology in which the role of acquired isomorphism is intuitively clear, and show how that role can be understood in terms of second order learning.

## Tolman's Detour Mazes

In experimental psychology, MR ability in biological species is often studied using Tolman's detour maze (Tolman & Honzik, 1930). These mazes have multiple paths (typically three) from their start to their goal, varying in length (see Figure 2). The shorter two paths join some distance before the goal position. The experiment runs as follows: a rat is fed to satiation, then placed at the start of the maze. A food reward is placed at the goal position. The rat explores the maze, and eventually finds the food reward, but, being satiated, does not eat it. After the rat has thoroughly explored the maze, it is taken out. We call this the exploration phase. Later, once the rat is hungry, it is placed again at the start position in the maze. The rat will now typically try to run the shortest path to the goal position and eat the reward. We call this the exploitation phase. In this phase, MR ability can be revealed by blocking the shortest path and observing the rat's reaction. If the shortest path is blocked such that the medium path is still open (in Figure 2: blocked at a cell with only a green dot) then the rat would ideally choose the medium path. If the shortest path is blocked such that the medium path is blocked too (in Figure 2: blocked at a cell with both a green and an orange dot), then the long path is the correct choice. If the rat, upon encountering the blockage, backtracks to the start position and then picks the new optimal path, then this taken as evidence of MR ability: If the rat had merely learned to solve the maze using action-sequences or state-action pairs, then finding one path blocked would tell it nothing about the viability of the other paths. So if it can pick the correct path right away, then it must also have grasped the spatial relations between the paths. That is, it must have a spatial representation of the maze. Note that we recognize MR here by the *absence of trial-and-error*: we would *not* ascribe MR ability to the rat if it would need to try the other two paths to figure out which choice is now optimal.

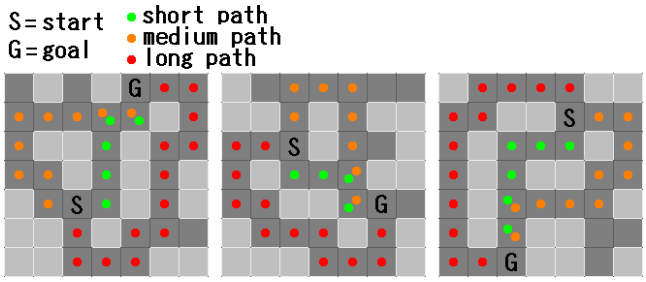


**Fig. 2.** Randomly generated detour mazes on a 7x7 grid. Dot colours indicate path lengths. Blockage on a cell with only a green dot obstructs only the short path, while blockage on a cell with both a green and an orange dot blocks both the short and medium path.

Many other species (as well as standard reinforcement learning algorithms) are quite capable of learning the shortest path in a maze, but have to re-learn whenever the layout of the maze changes. In this case, we may assume that no representations but simple action-sequences or state-action pairs were learned.

## Second Order Learning

Here we place the detour maze task in the theoretical framework introduced above. The maze task is composed of paths (or "accessible space", to be more precise), walls ("inaccessible space"), and a food reward. These are the aspects of $e$, implementing $E$. We see that a simple change in $e$ (replacing one piece of accessible space with inaccessible space) calls for complex changes in $E$ and consequently for complex changes in $L_0$ (running a different path altogether). We also find ourselves strongly inclined to ascribe the ability to mentally represent spatial layouts to an animal if it can make this complex update of $L_0$ in an instant (without further exploration) upon observing the blockage. We know that when we ourselves update our behaviour in such manner, we do so using our mental representation ability.

We said that mental representation is a form of acquired isomorphism. We saw that our framework explains the evolution of isomorphism-acquisition at $l_1$ as a consequence of the evolution of $L_2$. Can we recognize $L_2$ in the detour maze experiment?

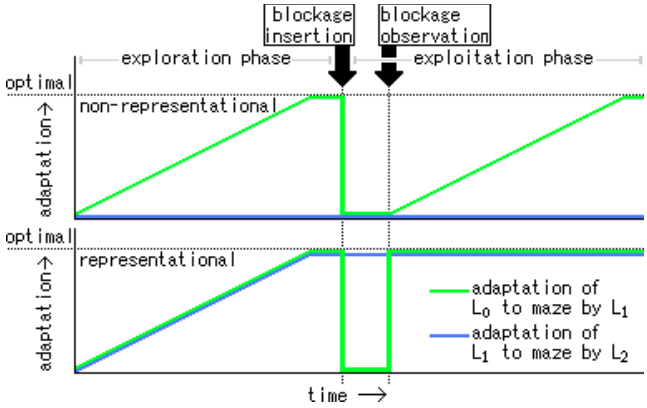


**Fig. 3.** Adaptation of $L_0$ and $L_1$ by $L_1$ and $L_2$ in the detour maze task. When a blockage is inserted, behaviour ($L_0$) inevitable becomes outdated. However, if first order learning ($L_1$) has been adapted (by $L_2$) to the maze, then optimality of behaviour can be restored with minimal information (observation of blockage alone).

When after blockage of the shortest path a rat infers the new optimal path without additional exploration, we can view this inference as a split-second $L_1$ process: a stimulus (observation of the location of the blockage) produced a change in behaviour (the subject abandons the blocked path and switches to the new optimal path). For $L_1$ to produce such a fast and effective behaviour-update, $L_1$ itself must have been adapted to the maze (the update cannot be the result of fixed pre-existing learning ability, as the information in the observation alone does not suffice to explain the update without reference to the specific layout of this maze). In other words, an $L_2$ process must have optimized $L_1$ to the current environment: the optimal update in behaviour has come to be causable by minimal information.

$L_2$ can be said to pre-emptively associate future stimuli with suitable behaviour updates. This would be infeasible for almost all $l_1$, but if $l_1$ employs isomorphism (here: the isomorphism between the cognitive map and the environment), then $L_2$ becomes quite feasible.

Hopefully it is clear how this is a concrete instance of the effect abstractly hypothesized earlier on. We suggest that equivalent reformulations can be given for many or all other scenarios that we take to indicate MR. We omit detailed examples here, but the general form is as follows: Consider an environmental object X to be represented. We (should) perceive our subject as representing X if and only if it can pre-emptively adjusts its behaviour so as to avoid or bring about specific unseen situations involving X after some period of observation of X. In all such cases, observation affects future changes in behaviour. To the extent such pre-emptive adaptation characterizes MR, explanation in terms of second-order learning should be applicable. In the focal case of cognitive maps, X is the maze, but the general scheme may equally well describe a scenario of spontaneous novel tool use (a scenario generally recognized as involving MR), with X being the tool.

We hypothesized that evolution of second order learning causes evolution of mental representation, but we haven't said anything yet about what it takes to evolve second order learning. We know that the neural basis for learning ability is neural plasticity. Would second order learning require second order plasticity? We would need neural circuitry that can not only change its input-output relation in response to stimulation, but also the way the input-output relation changes in response to stimulation. Such second order plasticity can quite simply be achieved by stringing two plasticity loci together on a neural pathway (examples are given below in the next section). So second order learning should be evolvable from standard neural plasticity, but only if we allow for multiple independent plasticity loci to exist along neural paths between input and output neurons[2].

Given that we said that MR is characterized by second order learning, and that second order learning depends on second order neural plasticity, we see that our hypothesis makes two predictions.

**P1.** In principle, the abilities that characterize mental representation ability can evolve from second order neural plasticity.

**P2.** It is impossible to evolve the abilities that characterize mental representation in a species restricted to first order neural plasticity.

If true, Prediction 1 should be confirmable empirically by taking a suitable artificial species with second order plasticity, evolving it in an environment composed of tasks requiring mental representation, and observing whether it evolves to

[2] This may sound like a weak requirement, but note that error back-propagation neural networks do not meet it. It follows that such networks are incapable of implementing $L_2$, and therefore their $l_1$ cannot be exposed to selection for isomorphism, making them unsuitable for evolution of representational cognition.

solve those tasks. Note that failure of such a species to evolve MR would not disconfirm our hypothesis: prediction 1 states merely a possibility, not a necessity. Prediction 2, on the other hand, cannot feasibly be confirmed empirically, as we would have to test every possible first order learning species. However, even a single counter-example against prediction 2 would disconfirm our hypothesis, so any computational successes should be analyzed to verify that evolved solutions use at least second order plasticity.

## Model

We test the hypothesis using a model in which neural nets with the basic elements for second order neural plasticity are evolved in an environment containing detour mazes.

### Environment

The task environment is composed of detour mazes and various simpler maze tasks. An environment composed of detour mazes alone was found ineffective. This is unsurprising: the evaluation criteria of the detour task evaluate MR ability only, i.e. the ability to walk the correct path from a choice of paths after the preferred path has become blocked, but our species starts out unable to walk any path at all (the initial generations spend most of their time bumping into walls helplessly). Inclusion of simpler tasks facilitates evolution of the sub-skills necessary for the detour task. Each task has an exploration phase in which the agent should locate the target, and one or more exploitation phases in which it should run to target in as few steps as possible. The full set of tasks is as follows:

1. An open field. Here there are no walls (aside from the edges of the grid-world). The start position differs between exploration and exploitation phase. This task facilitates evolution of the ability to memorize a location by (geocentric) coordinates (a skill called "place learning" by Tolman, 1948). Exploration time: 200 steps.
2. A "maze" with just a single path from start to finish. Simply following the path leads to reward. This task facilitates evolution of the ability to walk a path. Exploration time: 100 steps.
3. A "dark" version of task 2. Here no visual input (i.e. wall perception) is given during the exploitation phase. This task facilitates evolution of the ability to memorize a sequence of actions (the shape of the path). Exploration time: 100 steps.
4. A two-path maze (one short path, one long path) with dynamic path-blocking. This task has three exploitation phases. In the first, the agent has to pick the short path. In the second, the short path is blocked. The agent is expected to try the short path, find it blocked, then backtrack and pick the long path. In the third, the agent should remember that the short path is blocked, and pick the long path straight away. Exploration time: 150 steps.
5. Detour mazes, as described above. Here too there are three exploitation phases, handled just like in task 4, but now with the added difficulty of having to pick the correct path out of the two paths that remain after the short path is blocked. Each agent is evaluated in two detour mazes, one in which the medium path is the correct choice and one in which the long path is the correct choice (exposing each agent to both gives a more representative fitness signal than when this aspect is randomized. The same could be achieved by exposing each agent to a large number of detour mazes, but this gets computationally expensive. Agents are reset to their innate phenotype between tasks, so no inference about path choice can be made from prior tasks). Exploration time: 200 steps.

Tasks 4 and 5 are further complicated by the presence of arbitrary dead ends (as seen in Figure 1: cells without any coloured dots are dead ends). New mazes are generated continuously over the course of the experiments, to avoid overfitting to any given maze-set. In tasks 1, 2, 3 and 4, fitness is awarded for proximity to the target at the end of the exploitation phase, by the following fitness function:

$$f = 1 - \left(\frac{d_t}{d_s}\right)^p \quad (1)$$

Where $d_t$ is the distance to the goal at the end of the exploitation phase, $d_s$ the distance from the start to the goal, and $p$ a parameter controlling stringency of the fitness function, set to the experiments discussed here. The detour mazes have more stringent evaluation: only actually reaching the target yields a fitness reward (this prevents asymmetrical fitness reward for erroneously picking the medium path and erroneously picking the long path).

### Network species

In the environment described above, a population of 100 neural networks is evolved, using a genetic algorithm with mutation but no crossover. Both connection weights (as well as connection types, see below) and network architecture is evolved. Our network species distinguishes itself from standard neural networks by the use of neural grid structures, neuromodulators, and neurotransmitters. We briefly describe these features here.

**Neural Grids.** Informed by what's known about the neurology of spatial representation (See Moser et al., 2008, for a review), we let the genotype encode not only single neurons, but also neuron grids. We use square grids of three sizes: 1x1 (single neuron), 3x3, and $W$x$W$, where $W$ is the size of the world (7 for our 7x7 world). Given the setup of the model, sizes larger than $W$ offer no additional functionality (i.e. $W$x$W$ is functionally equivalent to an infinite grid).

The nets have one 3x3 grid and a number of 1x1 grids receiving input. The 3x3 grid encodes for each of the four cardinal directions whether there is a wall in that direction (on the 4 neurons adjacent to the middle neuron). The 1x1 grids encode whether the current position is the start position, whether the current position is the goal position, and the current phase (exploration or exploitation). Additionally, there are input neurons for bias (always 1.0) and noise (random real numbers from [0,1]). Output is read from two 3x3 grids. From the four neurons corresponding to the cardinal directions, the one with the highest activation is selected, and movement in that direction is performed (if possible). One set is read during

exploration and the other during exploitation (so that the nets can easily evolve specialized behaviour per phase). Connectivity is defined on two levels: inter-grid and intra-grid.

**Inter-grid Connectivity.** If the genotype defines a connection between two grids, then the phenotype gets uniform connectivity between the neurons in the two grids. If the grids are equal in size, connectivity is one-to-one, otherwise all-to-all. This leads to a highly symmetrical connectivity, which by itself would cause the activation within a grid to remain uniform and redundant. This symmetry is broken by our neurotransmitter logic. We label this neurotransmitter nt-B to distinguish it from our other neurotransmitter (see below).

There are two global nt-B values, nt-Bx and nt-By. These dynamically control (in two dimensions, as the neuron grids are 2D) which connection subsets of an all-to-all projection can transmit activation. When both are zero, then this set comprises connections linking corresponding neurons in the grids, relative to the grid centre (e.g. the centre neuron in the pre-synaptic grid to the centre neuron in the post-synaptic grid, the neuron left of the centre neuron in the pre-synaptic grid to the neuron left of the centre neuron in the post-synaptic grid, etc.). Non-zero nt-B values cause simple offsets, as illustrated in Fig. 4. Currently, nt-Bx and nt-By values are hard-wired to reflect the agent's current x-coordinate and y-coordinate, so signal transfer can shift along with position in space. This makes it relatively easy for evolution to devise nets that store information in different locations in a grid depending on their own position in space: if a smaller grid projects to a larger grid, then the activation pattern on the smaller grid affects only a sub-region of the larger grid. We will call this sub-region the *focal area* of the smaller grid on the larger grid. Nt-B does not correspond directly to any biological neurotransmitter, but can be reduced to the species' biologically plausible neurotransmitter (see below) via a trivial network transformation (which, however, increases network size dramatically, so this transformation is not performed).

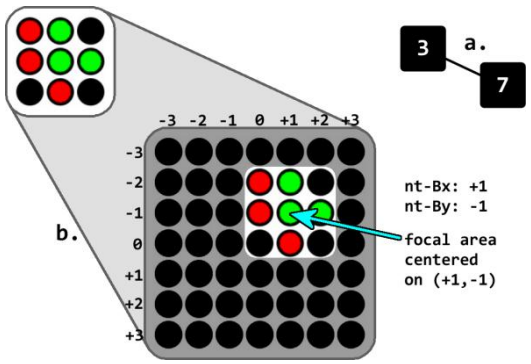


**Fig. 4**. Neural grids and nt-B. a. Genotype encoding a 3x3 grid, a 7x7 grid, and their connection. b. The corresponding phenotype. The 3x3 grid projects into the focal area of the 7x7 grid. The position of focal areas for projections between unequally sized grids is dynamically controlled by the global neurotransmitter values nt-Bx and nt-By. This mechanism lets the nets conveniently allocate neurons and circuits to specific spatial locations.

Inclusion of coordinates in the input is unnatural, but preliminary experiments focusing on the place learning task (task 1) have shown it quite possible with our model to evolve agents that keep track of their own coordinates without these inputs. Cognitively interpreted, the coordinates in the input and their linkage to the nt-B values make it fairly easy to evolve an innate sense of space as an extended medium in which movement predictably changes one's position. Construction of the ability to represent the volatile and non-uniform *contents* of space, however, is left to evolution.

**Intra-grid Connectivity.** As the neurons within a grid are not individually represented in the genotype, their connectivity is uniform. Two uniform intra-grid connection patterns are provided: neighbourhood connections (each neuron linking to its four neighbours, with innately identical connections) and reflexive connections (each neuron linking to itself). Neighbourhood connections allow for activation to diffuse over a grid. As neighbourhood connectivity leads to an abundance of loops, linear propagation order cannot be established, so instead we divide each time-step into smaller time-steps in which the activation pattern on grids with neighbourhood connectivity is updated iteratively. Reflexive connections allow for activation patterns to be retained over time (to be precise, a reflexive connection projects from a neuron to its future self, in the next time-step). Reflexive connections are a possible basis for learning, as retention of activation patterns allows acquired activation patterns to influence the behaviour indefinitely. Having multiple reflexive connections in a neural circuit allows for second order learning: if the activation pattern on some grid $g_x$ permanently affects the activation patterns on grid $g_y$ and the activation pattern on $g_y$ permanently affects the formation of the activation patterns on some grid $g_z$, then $g_x$ has a second order effect on $g_z$. Such second order effects provide a possible basis for second order learning.

**Neuromodulation.** Another possible mechanism for both first and second order learning ability is neuromodulation, which we also include in our model. We adopt the variation on the neuromodulation concept from Soltoggio et al., 2008. Neuromodulation provides an evolvable basis for learning ability by making it possible to let networks control their own weight update dynamics. It works as follows: In addition to standard activatory connections, there are modulatory connections. If there is a modulatory connection from neuron $X$ to neuron $Y$, then activation of $X$ causes modulation of $Y$. A neuron's modulation value affects the weight updates of its connections. Weights of modulated connections are updated each time-step, using the following update rule:

$$\boldsymbol{W_{xy} \leftarrow Gr \cdot W_{xy} + A_x^{Gxa} \cdot A_y^{Gya} \cdot M_x^{Gxm} \cdot M_y^{Gym}} \qquad \textbf{(2)}$$

Where $A_x$ is activation of neuron $X$ and $M_x$ is modulation of neuron $X$. *Gr* is a binary gene determining whether the previous value of the weight is included in the update. *Gxa*, *Gya*, *Gxm* and *Gym* are binary genes controlling for the corresponding pre- and post-synaptic activation and modulation values whether or not they affect connection weight updates. Connection weight values are clipped to the range [-1, +1]. Neuromodulation supports second order learning much like reflexive connections do: If there is a modulated connection on the

path from grid $g_x$ to grid $g_y$, and a modulated connection on the path from grid $g_y$ to grid $g_z$ then $g_x$ can have a second order modulatory effect on $g_z$.

Of course reflexive connections and modulatory connections can also be combined to form circuits with second order effects. As long as there are at least two points of lasting change on a path from input neurons to output neurons, there is potential for second order changes of the input-output mapping.

**Standard Neurotransmitter.** The network species has one more special connection type, which transmits a very simple, biologically plausible neurotransmitter, which we label nt-A. As part of the activation function, each neuron multiplies its activation value with its nt-A value. Neurons have an nt-A bias value (genetically defined per group) of 0 (needs to receive positive transmission to be excitable) or 1 (excitable by default, but propagation can be reduced or blocked by negative neurotransmission). Connections of this type only occur in between grids, not within grids. They can be susceptible to nt-B and/or neuromodulation, and have their own set of weight update rule genes.

## Results

The model is computationally expensive, and we have insufficient runs with the current version to make definitive statements about its success rates, but so far we have seen a number of successful runs, producing networks with near-optimal performance on all our maze tasks, including the our detour mazes. For our purpose, two aspects of the evolved networks are of particular interest: 1) whether or not they crucially rely on at least second order plasticity circuits, and 2) whether the way the nets solve the maze can be deemed representational in one way or another (i.e. whether the activation or weight patterns acquire any recognizable isomorphism with the maze being explored). We briefly discuss these points for one of the evolved solutions (Fig. 5). The performance of this network in detour mazes is 98% of the theoretical maximum (measured over 4000 maze trials). Observed failures are often the result of incomplete exploration.

Plasticity loci are found at numerous places in the network, but the functionally important ones appear to in the grids marked M2 and M3 in Figure 5. Grid D3 contains a (diffused) copy of the visual input pattern, and forwards this activation pattern via an nt-B controlled projection to M2. M2 uses retention to store the received activation patterns in spatially coherent fashion, forming an image of the maze with positive activation representing accessible positions and zero-activation representing walls. When a blockage is encountered, the negative activation of the D3 neuron for the direction where the blockage is seen knocks out the positive activation on the blockage-position in M2, effectively deleting that position from the image of the maze (the position of the blockage distinguishes itself from other inaccessible positions by its slightly negative value). The image in M2 is used to modify activation flow in M3.

Internally, M3 has positive neighbourhood connections and reflexive connections, however, it has an nt-A bias of zero, meaning that neurons can only activate if they receive positive nt-A from another grid. M2 has a 1:1 nt-A projection

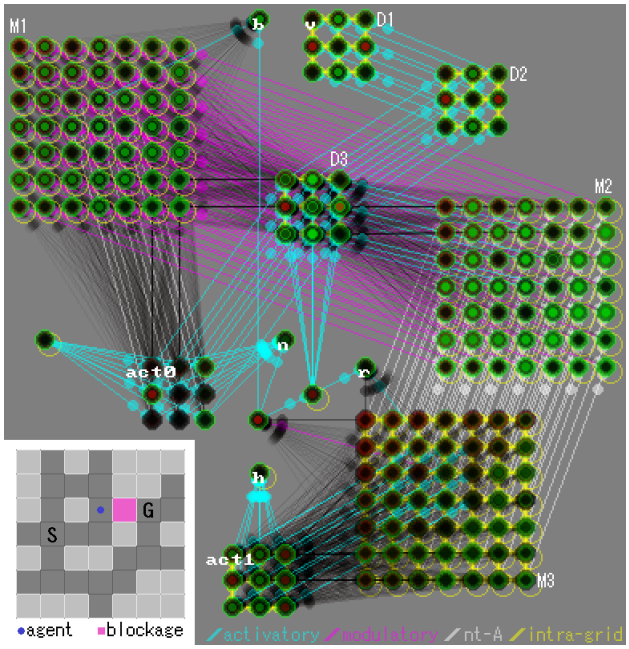


**Fig. 5.** An evolved solution. Connections run downward. Functionally irrelevant neurons are removed. D1: visual input. b: bias. n: noise. r: reward. h: home. act0: exploration phase output. act1: exploitation phase output. Grayed-out connections have their transmission blocked on this time-step by nt-B mismatch between their pre- and post-synaptic neurons. Snapshot of network state right after observing the blockage in the exploitation phase of a detour maze task (shown in inset). Activation patterns on grids M1 & M2 can be seen to encode the maze layout, but note that the blockage is only correctly reflected in M2. M3 encodes, at low activation, a gradient over the paths encoded on M2. Output grid act1 reads the activation pattern from its focal area in M3, causing the agent to climb up the gradient during the exploitation phase.

to M3, so the nt-A values on M3 replicate the activation on M2, which in turn replicates the maze layout. The result is that activation diffusion on M3 follows the shape of the maze. Reflexive connections on M3 are innately positive, but sensitive to modulation. At the focal position, modulation is received from a bias neuron, and activation from the reward neuron. The evolved update rule for this connection is absolute (i.e. $Gr = 0$) and takes into account modulation and activation of both the pre- and post-synaptic neuron (though in this case the pre- and post-synaptic neuron coincide, as the modulated connection is reflexive). When the reward neuron has is inactive, the result of modulation is that the reflexive connection's weight is set to zero. When the reward neuron is active, positivity of the reflexive connection is retained, and activation inserted at the focal position. The retained positive reflexive connection then ensures that at the neuron at this position retains this activation over time (though it drops off slowly), and the neighbourhood connections let it diffuse over the grid, following the shape of the maze. Note that, as the reward position is the only neuron that retains activation over time, the gradient is effectively recomputed every time-step. Consequently, when the activation pattern on M2 changes (e.g. when a blockage is detected), diffusion flow is instantly rerouted in accordance with the changed maze layout. The

output for exploitation simply reads out the local gradient on its focal area within M3. Optimal choice of path then follows naturally.

What order is this circuit's plasticity, and could it be reduced to 1st order? If we focus on M2 and M3, then we see with three crucial plasticity loci: retention on M2, retention on M3, and modulation on M3. The latter two might be deemed an ambiguous case, with modulation working on reflexive connections. If we consider those a single locus then we are left with two loci. Can we go down to one? No: the functionality of M2 and M3 is not collapsible. For instant adaptation to a layout change, it is crucial that M3 regenerates its activation pattern from scratch every time-step, remembering only the reward position. M2 on the other hand, must hold on to its content over time, because the limits of the net's perception imply that its information can only be gathered in bits and pieces. We conclude that this particular solution relies crucially on second or higher order neural plasticity.

As for the question of whether the solution is representational, we can conclude that the evolved approach clearly employs isomorphism: The maze layout is replicated in the activity patterns of M2 and M3 (as well as M1, although M1's activation pattern does not update in response to observation of a blockage). This solution may be deemed representational.

Different runs of the model produce different solutions. We have for example seen solutions where neuromodulation is used to encode the maze layout in the weights of neighbourhood connections on a 7x7 grid. However, all solutions analyzed so far employ circuits with at least second order plasticity and express the layout of the maze in connection weights and/or activation patterns. We need more successful runs and more extensive analysis before general claims can be made, but these results provide preliminary support for our hypothesis.

## Conclusions & Future Work

In this paper we introduced a general hypothesis about how cognitive architectures based on environment-cognition isomorphism may emerge as a consequence of the evolution of learning, and we showed how mental representation ability may be viewed as an instance of this effect. Specifically, we proposed that mental representation may be viewed as the ability for within lifetime acquisition of isomorphism that our hypothesis predicts should evolve under selection for second order learning ability. Given this evolutionary dependence on second order learning, we conjectured that evolution of mental representation requires second order plasticity. We evolved a neural network species that allows for second order plasticity, in an environment containing maze tasks generally believed to require mental representation ability. Successful runs of this model produced network that were found to crucially rely on second or higher order plasticity to solve these mazes, and made clear use of environment-cognition isomorphism, providing preliminary support for our hypothesis.

In this research we clearly used an operational definition of mental representation. We ascribed a species mental representation (in the form of cognitive maps) if it is capable of solving the detour maze task. We should expect the philosophically inclined to take issue with this, so let us state that our choice of definition is purely pragmatic. If it seems behaviouristic, this is only because evolution itself is a behaviourist. Any evolutionary explanation of a mental phenomenon must run via outward behaviour that can be selected on. We haven't touched upon the question of how or why the sort of representation we aim to explain is mental, and we acknowledge this explanatory gap. The objection might be raised that our work then pertains to neural representation only. However, while representation in our evolved networks is clearly neural, we note that our general hypothesis does not make specific claims about the nature of the isomorphism it predicts, requiring merely that it can causally affect behaviour. Depending on how one views the causal powers of mental phenomena, the hypothesis may be equally applicable to the representations we recognize as mental in ourselves.

Beyond improvement of the current model, future directions for our research are extension of this approach to other cognitive domains involving representation, using temporal and social scenarios.

## References


Arnold, S. (2011). Neuro-cognitive Organization as a Side-effect of the Evolution of Learning Ability. *Proceedings of the IEEE Symposium on Artificial Life*, (pp. 100-107).

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and Cognitive Architecture: a Critical Analysis. *Cognition* (28), 3-71.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature.* Cambridge University Press.

Minoya, K., & Arita, T. (2011). An artificial life approach for investigating the emergence of a Theory of Mind based on a functional model of the brain. *Proceedings of the 2011 IEEE Symposium on Artificial Life*, (pp. 108-115).

Moser, E. I., Kropff, E., & Moser, M. (2008). Place Cells, Grid Cells, and the Brain's Spatial Representation System. *Annual Review of Neuroscience* (31), 69-89.

Soltoggio, A., Bullinaria, J. A., Mattiussi, C., Dürr, P., & Floreano, D. (2008). Evolutionary Advantages of Neuromodulated Plasticity in Dynamic, Reward-based Scenarios. *Proceedings of Artificial Life XI* (pp. 569-576). MIT Press.

Spencer, H. (1855). *The Principles of Psychology.* New York: Appleton.

Takano, M., & Arita, T. (2006). Asymmetry between Even and Odd Levels of Recursion in a Theory of Mind. *Proceedings of ALIFE X*, (pp. 405-411).

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review , 55* (4), 189-208.

Tolman, E. C., & Honzik, C. H. (1930). "Insight" in rats. *University of California Publications in Psychology* (4), 215-232.