Intrinsic Representation: Bootstrapping Symbols From Experience

Stephen David Larson

MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)
200 Technology Square, Cambridge, MA 02139, USA,
stelar@alum.mit.edu

Abstract. If we are to understand human-level intelligence, we need to understand how meanings can be learned without explicit instruction. I take a step toward that understanding by showing how symbols can emerge from a system that looks for regularity in the experiences of its visual and proprioceptive sensory systems. More specifically, the implemented system builds descriptions up from low-level perceptual information and, without supervision, discovers regularities in that information. Then, the system, with supervision, associates the regularity with symbolic tags. Experiments conducted with the implementation shows that it successfully learns symbols corresponding to blocks in a simple 2D blocks world, and learns to associate the position of its eye with the position of its arm.

In the course of this work, I propose a model of an adaptive knowledge representation scheme that is intrinsic to the model and not parasitic on meanings captured in some external system, such as the head of a human investigator.

1 Introduction

In 1998, a group of MIT AI Lab faculty came together to form The Human Intelligence Enterprise, a group committed to the pursuit of understanding human intelligence from a computational perspective. From this group, the Bridge Project was initiated to investigate an interface between language representation and visual representation. The Genesis group, the name of the student group spawned by the Human Intelligence Enterprise to work on the Bridge Project, took an approach to investigating human intelligence that is very timely: to investigate the processes and representations that contribute to general human-level intelligence.

Inspired by this pursuit, I have asked the following questions: 1) How can we design representations for intelligent systems that bridge the gap between symbols and their subsymbolic referents, also known as the Symbol Grounding Problem? (Harnad 1990) 2) How can we design representations that are not limited to learning only within specific domains? This paper takes a step toward answering those questions.

Taking inspirations from the cerebral cortex and drawing from some older technologies (Kohonen's self-organizing maps, clustering algorithms, Hebbian association), I have fashioned a model that puts these elements together in a new way to form a self-organizing representation capable of representing the statistically salient features of an information space. The model can be used to train a system to associate the movement of its arm with the movement of its eye, as well as to associate colored blocks with linguistic symbols representing utterances.

2 Related Work

(Agre and Chapman 1987) investigated the instantiation of general symbols using indexical-functional aspects. Aspects were intended to provide a meaning for objects in a world relative to their usefulness for an agent. (Drescher 1991) looked at how a system can build knowledge on top of knowledge through interaction in a simple micro-world with a hand, an eye, and objects. The system in (Beule, Looveren, and Zuidema 2002) is given information about objects such as their positions and their constituent properties and returns syntactic structures describing notable objects in a world such as "the red square moving to the right". (Roy et al. 2003) describes a physically instantiated robot arm that can pick up objects by associating sensorimotor primitives together.

What the model of Intrinsic Representation shares with the previous works is the desire to associate symbols with subsymbolic descriptions. The following points outline the ways in which the model differs from the approach of traditional symbol systems. The same points explain the features of the model not shared by the previous works.

Symbols' descriptions are discovered from the statistical processing of experience. In almost every traditional symbol system, the symbols *and* their descriptions must be provided before-hand. In the model of Intrinsic Representation, the descriptions of symbols are discovered by processing sensory data in an unsupervised manner.

Symbols' descriptions are equivalent to statistical regularities found in information spaces. The nature of the symbols in traditional symbol systems are as tags whose descriptions are provided by a human designer. Descriptions formed from statistical regularities do not have to be provided by a human designer. As a result, symbols will represent certain classes of things in the world not because a designer finds them useful, but because those things are the most statistically salient in the given information space.

Symbols' descriptions carry their context with them by being situated in information spaces. Traditional symbol systems require context as an extra parameter to make sense of a symbol. In such systems, symbols may mean different things in different contexts. This is because a symbol is thought to be a separable part of a system rather than an intrinsic part of a system. Because symbols within an Intrinsic Representation are derived from information spaces they are inextricably linked to them and carry no separate meaning. For example, a cluster in an information space of visual information cannot be transported into an information space of any other kind of information, its data would not make sense in a different context. As a result, you cannot have a symbol that is not bound to its context.

3 Moving Forward

Representation has been at the center of the field of Artificial Intelligence from its inception. Formal symbolic systems capitalized on the ability to store information about the world in computer memory. The great successes of this approach, chess-playing computers and expert systems to name a few, relied on AI researchers' ability to invent ways of cleverly modeling the outside world.

On the other side of the tracks, the parallel distributed processing approach paved the way for a different approach to representation (Rumelhart and McClelland 1986). Rather than symbols, the values stored in the hidden nodes of pattern-analyzing networks provided a new perspective on what a representation could be. A great emphasis was placed on the learning of patterns. The ability for a network to produce the appropriate input-output behavior was taken as the proof that the representation formed by the hidden nodes was appropriate.

The rise of nouveau AI and the pursuit of insect-like intelligent robots ushered in an era that sought to dismiss internal representations, opting instead for using the world as its own best representation (Brooks 1991). The early success of this approach demonstrated a third perspective of representation that turned the classical view on its head.

The proper recognition of representation as the cornerstone of Artificial Intelligence helps put the road ahead into greater focus. As we understand more about the brain we gain an increased appreciation for the hard problem of representation that evolution has had to solve. This is the problem of the appropriate organization of information, a problem common to both organisms and man-made intelligent systems. While some might be hesitant to label the recognition of stimuli by simple animals as representation (as Brooks might), I would respond that there is much profit to be gained from broadening the definition. Doing so immediately leads us to the conclusion that simple organisms provide us with "page one" of the evolutionary history of representation. Our simpler evolutionary ancestors did not have the luxury of high-level concepts as humans do to help them craft clever taxonomies. Their only view of the world were the low-level patterns of activation they received from arrays of sensor cells. How can we model the processes that enabled them to distinguish food from poison and mate from attacker? For such simple creatures, we must boil the definition of representation down to its most basic and pure: representation is about grouping similar things together and placing dissimilar things farther apart.

From this viewpoint, it becomes clear that the two strategies of constructing representations, man-made and evolutionary, differ in an important way. The groupings of man-made representations are based on human choice of what is similar and what is dissimilar. The sensory groupings created by the nervous systems of simple organisms are certainly not. But what, if not the design of a creator, could be responsible for the appropriate groupings of stimuli that simple organisms exhibit?

In order to answer that question, we are forced to speak the language of the neuron– statistics and information theory. Biological systems have some help in making sense of the world from its sensory information: the world is not filled with Gaussian white noise. Natural environments contain patterns that are identifiable after the right processing has been carried out.

The information that an organism receives from its sensor cells can be thought of as forming a high-dimensional vector space of information, or "information space" (Larson 2003). For each organism, it is most advantageous to be able to process and react to sensory data within the those subspaces most relevant to survival. Seen from this angle we might explain the emergence of biological representation as the result of the natural selection of those organisms best able to model those subspaces.

If this is indeed the strategy that Mother Nature has exploited to enable simple organisms with the ability to represent the world, could we follow in her footsteps? Could a representation be built up from low-level sensory information by making statistical distinctions between similar and dissimilar input? Could such a representation self-organize from the statistical regularity of its input the ability to tell the difference between sensory patterns rather than relying on a magic "if" statement which, in turn, relies on human-designed assumptions of similarity? Such a representation is the goal of the model of Intrinsic Representation.

4 The Model

Figure 1 is the key diagram for understanding the model of Intrinsic Representation. Two information spaces, i and j, are shown side by side. At the bottom of the diagram, sensor arrays receive streams of data from the outside world. Such arrays could be imagined as a patch of skin or light-sensitive cells in the retina. As data comes into the system through these sensors, they travel to a subsystem devoted to organizing and storing the regularities in the data. This subsystem arranges these regularities with respect to their similarity, placing similar regularities in proximity, and dissimilar regularities farther apart. After a critical period, the regularities are grouped into clusters of high similarity. Once grouped, a cluster gains the ability to act as a unit that can be activated and deactivated. Data in the incoming stream can be treated as the trigger for the activation of the cluster of which they are a member. As clusters are activated by incoming data, they are associated together by their frequency of coincidence. The more often two clusters are active simultaneously, the more associated they

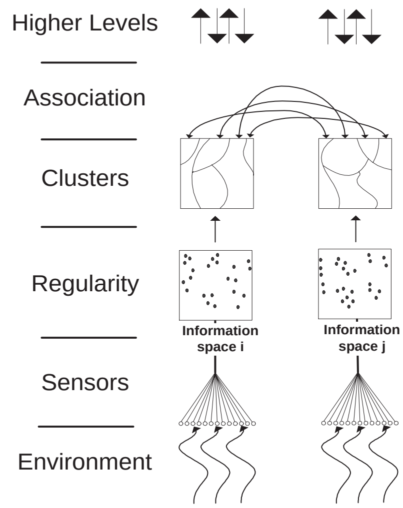

Fig. 1. The model of intrinsic representation.

will become. The resulting trained system treats incoming patterns as members of a class, and can react to the activation of that class.

4.1 Architecture and Implementation

This section discusses the architecture and the implementation of a computer program demonstrating the model of Intrinsic Representation.

The architecture of the program is divided into four major areas:

1. A Blocks World Simulation
2. A Self Organizing Map
3. A Cluster Set
4. Cluster Associations

Each of these blocks feeds into the next in succession. Data flows into the system from a simple 2D blocks world that provides an environment for the present system. In the blocks world, there is an arm with a simple grip useful for grasping objects, much like the arcade novelty that invites players to catch prizes by controlling a robot arm. There is also an eye, whose focus is represented by a square that moves around the environment. The eye and the arm are the sources of sensory and motor interaction with the world. There are also blocks in the world that can be picked up and stacked on top of one another. Simple physics modeling enables such constraints as gravity and friction.

Both the eye and the arm are equipped with sensors. Every sensor is normalized, and thus reads a real value between zero and one. The eye has retina sensors arranged in a 2D array that register the red-green-blue value of any spot they are over. The eye also has sensors that report its horizontal and vertical orientation, mimicking the information the brain receives from the muscles that position the eyes.

The arm has proprioceptive sensors and tactile sensors on its grip. The proprioceptive sensors report how far it is extended both horizontally and vertically, mimicking feedback from muscles and joints. The tactile sensors report collisions with objects. A vector of sensor values is read from the eye and arm each time their values change (see figure 2).

As data are output from a given device, they are received by a self-organizing map (Kohonen 2001). For the experiments conducted, two self-organizing maps are used; one for the eye and one for the arm. The map is specialized to its input only in the respect that its dimensionality must match. As vectors are received, the map's self-organizing process iterates following a growing Kohonen map algorithm (Dittenbach, Merkl, and Rauber 2000).

The maps undergo a self-organizing process that is driven by the visual or proprioceptive information that they are set to receive. The maps are allowed to self-organize and are later prevented from acquiring new data (see figure 3). At this point, a clustering algorithm is used on the data held by the map to separate the major clusters in the space. This is visually represented as finding the discrete sets on the map which correspond to the groupings of the most similar cells (see figure 4).

Once clusters have been established, each of the clusters on the map is given its own unique identifier, and a new data object is created to represent it. This data object keeps track of the similarity measure of the cells inside the cluster, as well as the indices of the cells it contains.

This data object also keeps track of any associations that a cluster has made with other clusters, even those outside of the current map. In this sense, the cluster acts like a single unit whose activation can become associated with the activation of other units. The association process is a simplified version of Hebbian learning. When clusters are activated by sensory data simultaneously, a counter labeled with the names of those clusters is incremented. Clusters are active when input matches most closely with a cell within the cluster. The cluster association system identifies the clusters most associated with any particular cluster. Clusters are never associated with clusters from the same map.

Once the system has been trained, the associations can be used to activate units between modalities. This allows it to exhibit behavior that neither modality separately would be able to accomplish as easily by itself.

4.2 Three Levels of Representations

The architecture I have just described takes advantage of three different levels of representation arranged in a hierarchy. The following describes these representations in greater detail.

Map Cells A map cell is a single cell in a self-organizing map. Each cell keeps track of a subset of information that enters from the environment during the initial self-organizing process. Each cell in the map stores an n-dimensional vector. The data contained in that vector corresponds to a point in an n-dimensional information space. In the context of a self-organizing map, a map cell represents a point in an n-dimensional information space that is representative of a class of points that the map has observed.

Clusters The cluster is one representational level higher than a map cell. It is a collection of statistically similar map cells in a self-organizing map. Because cells represent vectors in an n-dimensional information space, a cluster therefore is a collection of statistically similar vectors.

In the context of a self-organizing map, a cluster represents a collection of cells that have been arranged near each other by the SOM process (Kohonen 2001). The SOM process also arranges cells such that dissimilar cells are distant in the space. As a result, the cells inside a cluster are similar to each other, but different from other major clusters in the map.

Putting this together with what we know about a map cell, a cluster represents a subspace of an n-dimensional information space by storing a collection of n-dimensional vectors. While these vectors are not a basis for that subspace as in linear algebra, they are a collection of statistically interesting vectors that are representative of a class of points in that subspace.

Associations/Symbols Associations are a representational level higher than clusters. In general, they represent a collection of clusters, usually two. They are discovered between different maps through the process of Hebbian learning. This creates connections between clusters that are active simultaneously in different maps.

5 Experiments

Learning About Blocks In this section I demonstrate how the system learns about blocks in a simple 2D blocks world at all three of the representational levels just described. In a nutshell, the system discovers implicit representations of blocks in its world in an unsupervised manner. Then it assigns explicit symbols to those implicit representations in a supervised manner.

The first experiment involves two self-organizing maps that read data from a blocks world. One map reads data from the eye in the blocks world, and the other map reads data from the arm. Figure 2 illustrates the features that make up the input data vectors.

Additionally, there is a third fixed map that contains linguistic symbols that both the eye map and the arm map are allowed to associate with. This linguistic symbols map is artificially manipulated by the experiment to act as supervisor. When the eye or the arm come in contact with a block of a particular color, this linguistic symbols map will activate a cluster that corresponds to the appropriate symbol, in an attempt to model a distinguishable utterance identifying the block. The task before the learner, then, is to build up a representation that matches the correct utterances with the things it sees and feels in the world.

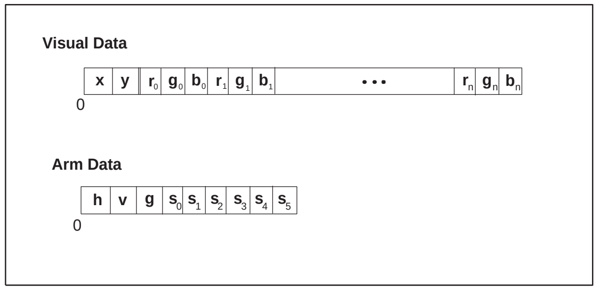

Fig. 2. A guide to the features stored in the visual and arm data vectors that arrive at the eye map and the arm map respectively. The visual data includes x and y coordinates for the eye position, followed by r, g, and b which store the red, green and blue values for each pixel position in the eye. The arm data includes h, v, and g which store the horizontal and vertical components, the width of the grip, and 6 sensor values that are read from sensors on the gripper/hand.

Training The Maps First, the eye map is trained on the objects in the scene separately. To accomplish this, the eye is shifted between the objects in the scene several times. The scene consists of a red, a green, and a blue block, plus the arm. While viewing a single object, the eye is allowed to scan by taking a circular path around it. As the eye moves from object to object, the eye map grows to capture more of the views it is receiving. Figure 3 illustrates the progress of the eye map in this phase of training. The result of this process is a map that captures views of the objects in the scene.

Once the eye map has been trained with the stationary objects, the arm is trained on the blocks in the scene by being moved to each block in succession, grasping it, picking it up, moving it around, lifting it up, and eventually dropping it. The result of this process is shown in figure 5 and illustrates a distinction between the arm with something in its grasp, and the arm without something in its grasp.

While the arm is being trained, the eye is focused on the actions of the arm, thus updating the eye map to include the interactions between the arm and the blocks. Figure 4 illustrates the resulting map.

Clusters Following the training of the maps, Figure 4 shows the eye map segmented into clusters, and figure 5 shows the arm map segmented into clusters.

Associations Once the maps have been segmented into clusters, their clusters are associated together. The training process is a repetition of the arm training—the arm travels to each block, grips it, picks it up, and moves it around, while the eye looks on.

During *this* training phase, both maps are given the opportunity to form associations with the special utterance map, which contains symbols that are activated or deactivated during training. When the eye and the arm are interacting with a block, the appropriate symbol is activated. When the eye and

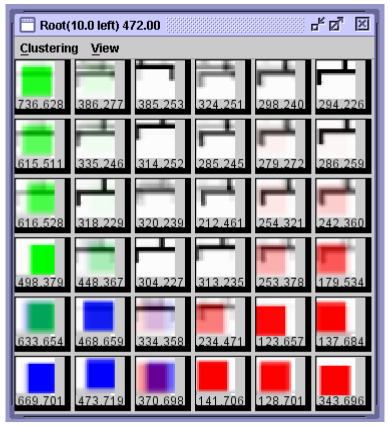

Fig. 3. A visual representation of the eye map partially trained on the four objects in the simple 2D blocks world. Each cell is a bitmap representation of the rgb values of underlying visual input data vectors. In the bottom left, the map shows a few views of the blue block. In the bottom right several cells show views of the red block. In the upper right, middle and center, several views of the gripper/hand can be seen. The cells along the upper and middle left show views of the green block. The "smeared" cells are the product of different inputs that came soon after one another and thus created cells that are on a cluster border.

the arm leave the area of the blocks, no symbols are asserted. The symbols are `(Red_Block)`, `(Blue_Block)`, and `(Green_Block)`. These symbols are undivided; while to us they imply a combination between a color and a shape to a human reader, to the system, they are only treated as distinguishable atomic tags. This is consistent with the notion that symbols by themselves are just tags until associated with sub-symbolic meaning.

The left half of table 1 shows the results of the association between these symbols and the clusters in the eye map. Higher frequency corresponds to greater association between the symbol and the cluster. The cluster IDs are the same as the cluster IDs in figure 4 to provide the reader a sense of the clusters that the symbols are being associated with. The right half of table 1 shows corresponding results for association between the symbols and the clusters in the arm map. These cluster IDs correspond to the ones in figure 5.

Generally, the associations tend to fall into the largest clusters that, to the eye, appear to be the most representative of the block in the symbol. In this case, however, table 1 shows that the symbol `(Blue_Block)` is instead most associated with a smaller cluster, though the largest cluster corresponding to the symbol `(Blue_Block)` appears farther down on its list. This demonstrates how even small clusters can represent important parts of the information space.

One of the interesting results of table 1 is that the `(Blue_Block)` symbol and the `(Green_Block)` symbol are both most highly associated with cluster 166. This cluster corresponds to a state where the arm is gripping something. The

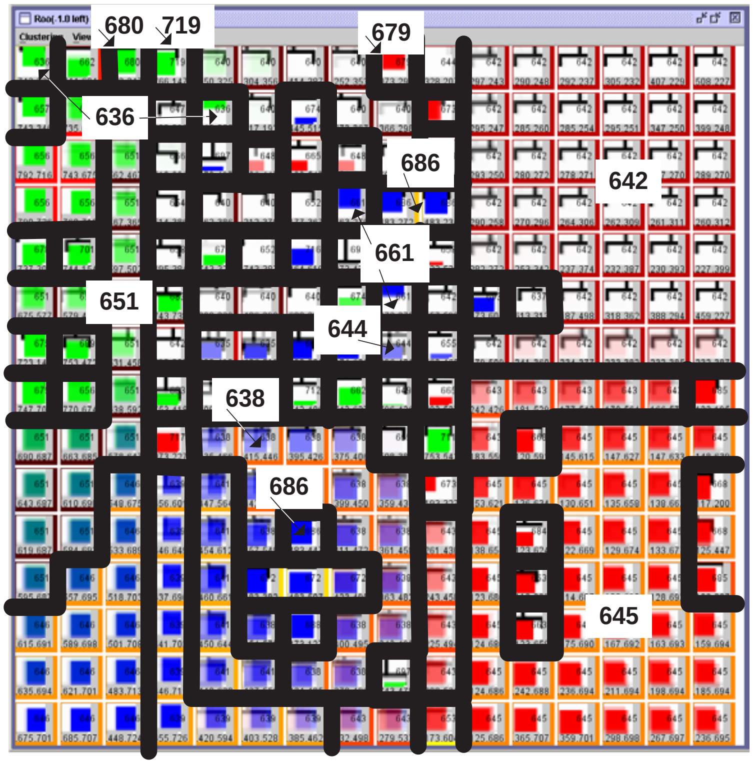

Fig. 4. A trained eye map with clusters overlaid. Here, the images visually represent the values that the arm sensors can take on. The cluster IDs of relevant clusters are highlighted.

Table 1. The major associations between symbols and clusters in the eye map and the arm map

SYMBOL	Eye Map Cluster ID	Frequency	Arm Map Cluster ID	Frequency
`(Red_Block)`	645	69	173	78
			169	35
`(Blue_Block)`	686	35	166	63
	661	24	165	22
	638	10	170	14
`(Green_Block)`	651	80	166	100
	680	26	168	10
	636	23		
	719	20		

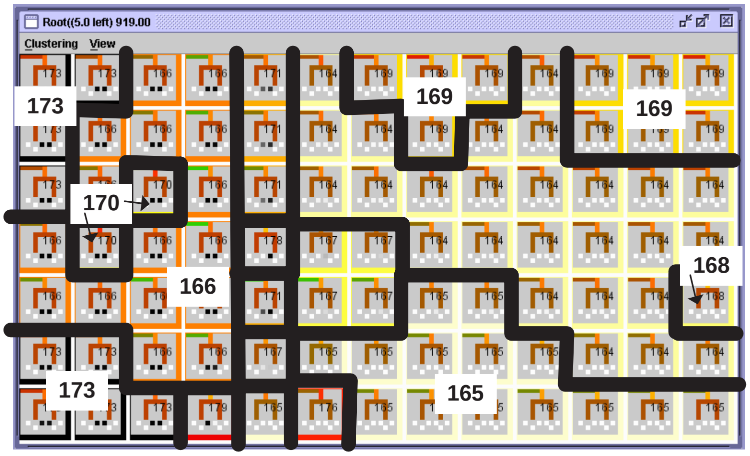

Fig. 5. A trained arm map with clusters overlaid. The cluster IDs of relevant clusters are highlighted.

associations further down the list, which match to different clusters, dissociate the two symbols so that they can be independently identified.

This experiment demonstrates the ability of the system to ground symbols in sensory data. The system has acquired sensory data and organized them into classes in an unsupervised fashion. Later, this data was associated with symbols in a supervised fashion. Thus, the symbolic information stored in the special utterance map has been grounded by the subsymbolic data from the environment.

Associating Hand with Eye In this section I demonstrate how the system can use associations formed between the position of the eye and the position of the arm to enable the arm to move to the current position of the eye.

The training of the eye map and the arm map follows a simpler process than in the previous blocks-learning situation. There is no third map whose activity is managed by the experimenter. In this experiment, the two maps are allowed to associate directly with one another. Training proceeds by fixing the position of the eye to that of the gripper/hand, and moving them together to random points in the space. The space does not have any blocks into it.

What we expect from this training is for the eye map to have stored very similar images since it views the same part of the gripper for all time. Because of this, the x and y components of the eye data should become more significant. The resulting eye map should have an even distribution of x and y positions across it. The arm map will emphasize the horizontal and vertical components as they are the only ones that are varied.

Figure 6 shows the trained eye map after the clustering process has run. Notable clusters are highlighted. Our expectation of a smooth and continuous map of x and y values has been met. Figure 7 shows the trained arm map after the clustering process has run, with notable clusters highlighted. Table 2 shows the data from the cluster associations.

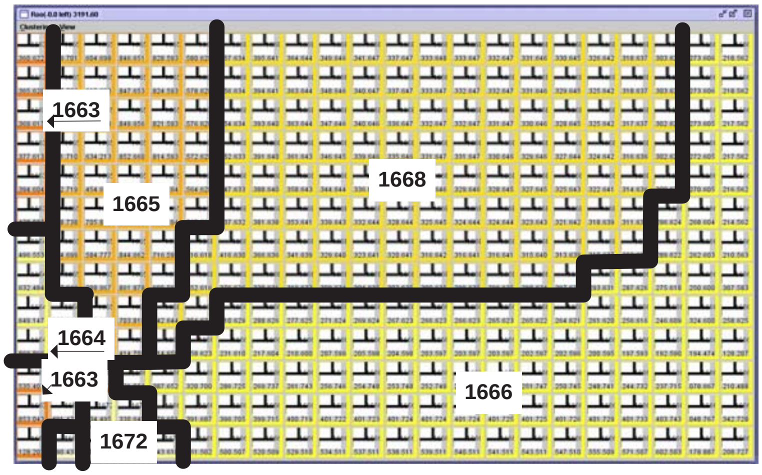

Fig. 6. A trained eye map with clusters overlaid. Relevant clusters are highlighted. All the cells show closely similar views of the world, but the x and y components vary across the map.

These associations allow the arm to move into a region near the eye's view. To accomplish this, the currently active cluster in the eye map is identified using the current state of the eye (see figure 8). Then, the arm cluster most highly associated with the currently active eye cluster is selected from the arm map. The vectors from the cells within this cluster are averaged together, and the arm is driven toward this average vector. A feedback process is used whereby the arm continues to move in the direction of the average vector so long as its current state vector does not match it. Note that while this experiment focused on moving the arm to the eye, the inverse could also be accomplished with the same trained maps by simply reversing the process just described.

This experiment demonstrates a deeper use of the model of Intrinsic Representation; to enable action. Through an unsupervised process, the different coordinate systems of the eye and the arm are trained to be compatible, and to allow one to "speak the same language" as the other.

This experiment also shows how exchanges can be made between sensory systems without involving any linguistic symbols whatsoever. This observation expands what we mean by a symbol. The experiment demonstrates the equivalence of 1) associations made with clusters that store linguistic symbols and 2) associations made with clusters that store other kinds of sensory data. Of course,

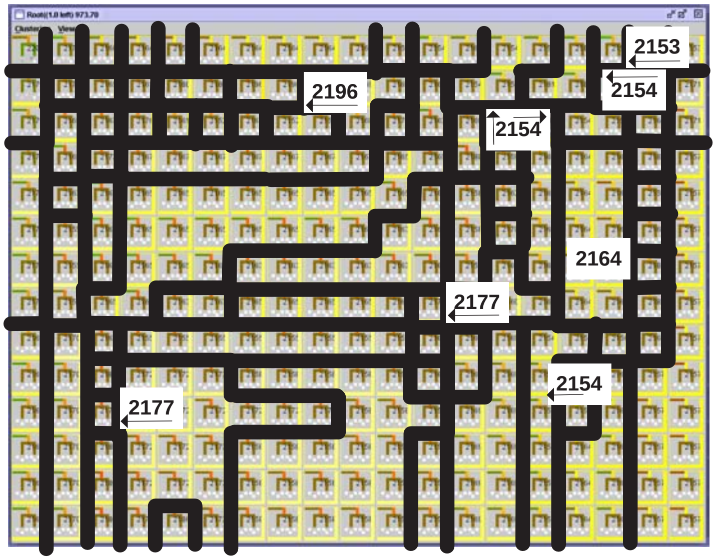

Fig. 7. A trained arm map with clusters overlaid. Relevant clusters are highlighted. While the gripper sensors and grasp remain constant, the map self-organizes using the horizontal and vertical components of the data alone.

Table 2. Associations between the eye map and the arm map.

Eye Map Cluster ID	Arm Map Cluster ID	Frequency	Eye Map Cluster ID	Arm Map Cluster ID	Frequency
1663	2177	152	1663	2186	49
1663	2154	140	1663	2176	45
1663	2164	136	1663	2182	45
1663	2153	127	1664	2153	27
1663	2181	125	1664	2181	24
1663	2196	125	1664	2152	20
1663	2169	101	1664	2187	20
1663	2152	92	1665	2196	20
1663	2156	81	1665	2164	19
1663	2155	78	1665	2181	15
1663	2166	72	1665	2153	15
1663	2187	63	1666	2177	36
1663	2163	58	1666	2154	33
1663	2180	57	1666	2181	29
1663	2167	56	1666	2180	19
1663	2157	55	1672	2196	17
1663	2165	52			

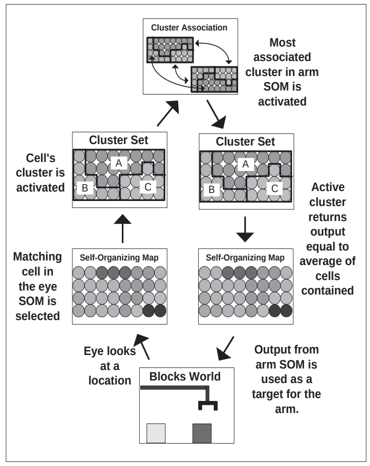

Fig. 8. A schematic showing how a trained representation can give rise to behavior. The self-organizing map and cluster set on the left store sensory information from the eye, while on the right they store sensory information from the arm. Starting at the blocks world at the bottom, we see that when the eye looks at a location, that information matches a cell in the eye map. This activates a cluster of cells in the eye's cluster set. The eye cluster set has been associated with the arm cluster set, and thus, a cluster in the arm map is activated. The average value of the cells in that cluster is outputted to the arm, causing it to move to a location.

in the brain, linguistic symbols *are* just another kind of sensory data, so this is just as it should be.

6 Contributions

In this paper, I have:

1. Outlined the differences between Intrinsic Representation and other approaches to the Symbol Grounding Problem.
2. Justified the need to pursue self-organizing representations as an attempt to recapitulate the evolution of biological representations.
3. Elucidated a hierarchical, self-organizing representational system called Intrinsic Representation that acquires symbols in an unsupervised manner by extracting statistically salient information through interaction with its environment.
4. Described the architecture of Intrinsic Representation through its three key levels of representation: the map cell, the cluster, and the association.
5. Provided the results of two experiments carried out using this representational system instantiated in a computer program.

References

[Agre and Chapman1987] Agre, P.E., and D. Chapman. 1987. "Pengi: An Implementation of a Theory of Activity." Edited by Morgan Kaufmann, *Proc. Sixth National ConferenceAmerican Association for Artificial Intelligence*, All ACM Conferences. American Association for Artificial Intelligence, 268–272.

[Beule, Looveren, and Zuidema2002] Beule, Joachim De, Joris Van Looveren, and Willem Zuidema. 2002. "Grounding Formal Syntax in an Almost Real World." Ai-memo 02-03, Vrije Universiteit Brussel, Artificial Intelligence Laboratory.

[Bickhard and Terveen1995] Bickhard, Mark H., and Loren Terveen, eds. 1995. *Foundational Issues in Artificial Intelligence and Cognitive Science*. Advances In Psychology no. 109. North-Holland.

[Brooks1991] Brooks, Rodney A. 1991. "Intelligence without representation." *Artificial Intelligence*, no. 47:139–159.

[Dittenbach, Merkl, and Rauber2000] Dittenbach, M., D. Merkl, and A. Rauber. 2000, July 24. – 27. "The Growing Hierarchical Self-Organizing Map." Edited by S. Amari, C. L. Giles, M. Gori, and V. Puri, *Proc of the International Joint Conference on Neural Networks (IJCNN 2000)*, Volume VI. Como, Italy: IEEE Computer Society, 15 – 19.

[Drescher1991] Drescher, Gary L. 1991. *Made-up Minds*. Cambridge, Massachusetts: MIT Press.

[Fuster2003] Fuster, Joaquin M. 2003. *Cortex and Mind: Unifying Cognition*. Oxford University Press.

[Harnad1990] Harnad, Stephen. 1990. "The Symbol Grounding Problem." *Physica D* 42:335–346.

[Kohonen2001] Kohonen, Teuvo. 2001. *Self-Organizing Maps*. Third. Springer Series in Information Science no. 30. Springer.

[Larson2003] Larson, Stephen David. 2003, September. "Intrinsic Representation: Bootstrapping Symbols From Experience." Master's thesis, Massachusetts Institute of Technology.

[Minsky1988] Minsky, Marvin. 1988. *The Society of Mind*. New York: Simon and Schuster.

[Roy et al.2003] Roy, Deb, Kai-Yuh Hsiao, Nikolaos Mavridis, and Peter Gorniak. 2003. "Ripley, Hand Me The Cup! (Sensorimotor representations for grounding word meaning)." *Int. Conf. of Automatic Speech Recognition and Understanding*.

[Rumelhart and McClelland1986] Rumelhart, David E., and James L. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Volume 1. MIT Press.

[Winograd1971] Winograd, Terry. 1971, Feb. "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language." Mit ai technical report 235, MIT.

[Winston1993] Winston, Patrick. H. 1993. *Artificial Intelligence*. Third. Reading, Massachusetts: Addison-Wesley.