Information Trade-Offs and the Evolution of Sensory Layouts

Lars Olsson, Chrystopher L. Nehaniv, Daniel Polani
Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
College Lane, Hatfield Herts AL10 9AB
United Kingdom
{L.A.Olsson, C.L.Nehaniv, D.Polani}@herts.ac.uk

Abstract

In nature, sensors evolve to capture relevant information needed for organisms of a particular species to survive and reproduce. In this paper we study how sensor layouts may evolve in different environments and under pressure of different informational constraints. To do this we evolve sensor layouts for different environments and constraints using a fitness measure with weighted terms for *redundancy* and *novelty*, using, respectively, mutual information and Crutchfield's information metric. The results show how different sensor layouts evolve depending on the structure and complexity of the environment but also how selective pressure for redundancy or novelty might affect the design.

Introduction

Nature has produced a wide variety of sensory organs that are well adapted to the specific animals and their respective environments (Dusenbery, 1992): For example, consider the amazing echolocation capabilities exhibited by bats, dolphins, and some species of whales. Sound waves are emitted by the animal and it then listens for the echo, which is used to perceive predators, prey, and objects in the environment. Another example is some birds and bees that probably can use the horizontal component of the magnetic force around the earth to determine their direction. The senses and their interpretation in the brain can also develop during the lifetime of an individual. One example is humans that for some reason go blind and then learn to use other senses to navigate in place of vision.

But how are the different sensoric channels used by a particular species selected for to begin with and how do they evolve over time? These are some of the questions pondered in the field of *sensor evolution* (Dautenhahn et al., 2001). In contrast with natural systems, sensors of artificial systems are often, due to practical and historical reasons, seen as something that is "given" and fixed. Only recently has there been a strong research focus on building artificial systems where the sensors can evolve and adapt, be combined, or the sensory channels created or selected, as this may lead to more autonomous, adaptive, and powerful systems as well as better understanding of how sensors evolve in nature (Cariani, 1998; Dautenhahn et al., 2001; Nehaniv et al., 2002).

In this paper we discuss how the positioning and *informational coverage* of similar sensors, for example vision sensors, can evolve and adapt depending on the environment. This has been studied in for example (Stryker et al., 1978) where kittens were restricted to seeing either only vertical or horizontal lines. It was found that the cortical cells selective for orientation in the kittens preferred to fire mostly in response to the orientation that they had been exposed to. Evolution of sensory layout in an artificial agent was considered in (Jung et al., 2001) in relation to learning, where it was shown that longer learning periods lead to better suited sensors. In this paper we also consider the selective pressure posed by the trade-off between redundancy and novelty in sensoric input. If two sensors of the visual system of an agent are completely uncorrelated it is hard to find structure that for example can be used to compute optic flow (Gibson, 1986). However, if the sensors are completely correlated the information they transmit is redundant and one sensor is enough, unless they are used in an noisy environment where the redundancy may be used to filter noisy input, and thus provide robustness. Thus, there is a trade-off between similarity of information and the novelty of information between pairs and groups of sensors. To study the environmental impact on sensor layout but also the trade-off between redundancy and novelty, we apply evolution to sensor layouts where the redundancy and novelty can be weighted to be more or less important. In the experiment we consider the evolution of layout of visual sensors in different environments, ranging from an environment consisting of only vertical lines to a complex environment and a random environment without structure. The results show that the layout of the sensors depends on the environment in which the agent is situated in but also the selective pressure for redundancy or novelty.

The remainder of this paper is organized as follows. The next section discusses methods for computing differences between sensors and especially information-theoretic measures. One example with real world data from a robot using the described methods is also discussed. Then we describe our experimental model, the results, and their interpretation. Finally we summarize the paper and discuss possible applications and future directions of the presented work.

Information Distance between Sensors

In order to discuss the effectiveness and layout of sensors it is fruitful to be able to quantify the functional and informational distances between sensors. To do this a number of different methods can used, e.g., the Hamming distance and frequency distribution distance (Pierce and Kuipers, 1997). In (Olsson et al., 2004) these distance metrics are compared with the *information metric*, which was defined and proved to be a metric in (Crutchfield, 1990). The distance between two information sources is there defined in the sense of classical information theory (Shannon, 1948) in terms of conditional entropies. To understand what the information metric means we need some definitions from information theory.

Let $\mathcal{X}$ be the alphabet of values of a discrete random variable (information source, in this case a sensor) X with a probability mass function $p(x)$, where $x \in \mathcal{X}$. Then the entropy, or uncertainty associated with X is

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{1}$$

and the conditional entropy

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 p(y|x) \tag{2}$$

is the uncertainty associated with the discrete random variable Y if we know the value of X. In other words, how much more information do we need to fully predict Y once we know X.

The *mutual information* is the information shared between the two random variables X and Y and is defined as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{3}$$

To measure the dissimilarity in the information in two sources Crutchfield's *information metric* (Crutchfield, 1990) can be used. The information metric is the sum of two conditional entropies, or formally

$$d(X,Y) = H(X|Y) + H(Y|X). \tag{4}$$

Note that X and Y in our system are information sources whose $H(Y|X)$ and $H(X|Y)$ are estimated from the time series of two sensors as described in the next section using (2).

It is worth noting that two sensors do not need to be identical to have a distance of 0.0 using the information metric. What an information distance of 0.0 means is that the sensors are completely correlated. As an example, consider two sine-curves where one is the additive inverse of the other. Even though they have different values in almost every point is the distance 0.0 since the value of one is completely predictable from the other. In this case, the mutual information, on the other hand, will be equal to the entropy of either one of the sensors.

Redundancy and Novelty

We will in this paper use mutual information as a measure of redundancy and the information metric to measure novelty between pairs of sensors. Redundancy can be seen as robustness to noise while novelty is important to capture as much different information as possible about the environment. Thus, there is a trade-off between capturing redundant and novel information.

An agent may optimize the relation between redundant sensors and sensors that detect novel things, differently according to selective pressure for redundancy or novelty and the environment. The *informational coverage* achieved by a set of sensors $\mathcal{S}$ can be used to calculate how sensors should be selected and is defined as

$$ic(\mathcal{S}) = \sum_{X \in \mathcal{S}} \sum_{Y \in \mathcal{S}} (w_{mi} I(X;Y) + w_{im}(H(X|Y) + H(Y|X))) \tag{5}$$

where $X \in \mathcal{S}$ and $Y \in \mathcal{S}$ are the sensors of individual $\mathcal{S}$, w_{mi} a weight associated with the mutual information and w_{im} a weight associated with the information metric. Obviously, if $w_{im} = w_{mi} = 1$, then $ic(\mathcal{S}) = \sum_{X \in \mathcal{S}} \sum_{Y \in \mathcal{S}} H(X,Y)$, since $H(X) - H(X|Y) + H(X|Y) + H(Y|X) = H(X) + H(Y|X) = H(X,Y)$, which is the joint entropy. The weights w_{mi} and w_{im} are used to specify how much emphasis that is put on redundancy and novelty.

As an example, consider Figure 1, which is a scatter plot with mutual information on the y-axis and the information distance on the x-axis for all pairs of sensors including the visual sensors of a SONY AIBO[1] robot dog.

The upper left corner of Figure 1 contains sensors with large mutual information, and hence redundancy, but a small informational distance. The lower right corner, on the other hand, is where we find pairs of sensors with a large informational distance but little redundancy. Finally, the upper right corner is the interesting part where pairs of sensors both share a large amount of mutual information but also have a large informational distance. This is implies that these sensors must have a high individual entropy. In Figure 1 the vision sensors are the sensors in the diagonal cloud where the upper left corner consists of vision sensors that are physically close together on the AIBO, since neighbouring pixels have much redundancy. One application of a scatter plot like this is that related sensors can be grouped together from raw uninterpreted sensor data, in a similar way to the sensory reconstruction method developed in (Pierce and Kuipers, 1997) and extended in (Olsson et al., 2004).

[1] AIBO is a registered trademark of SONY Corporation.

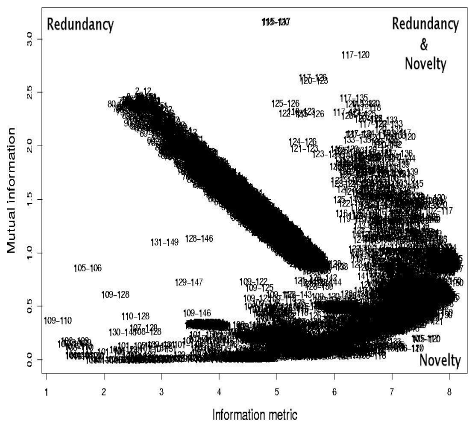


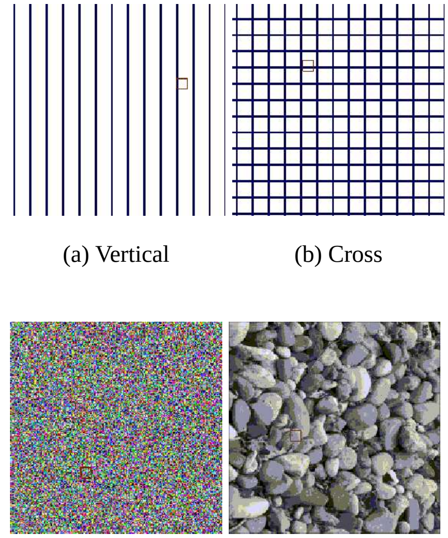
Figure 1: Scatter plot of all sensors and 100 vision sensors (pixels) from a SONY AIBO. The pixels are the diagonal linear cloud.

Experimental Model

An individual will have to select and layout sensory channels so as to maximize robustness and novelty of the information they provide. On a 200 x 200 pixel environment a 10 x 10 individual can move a maximum of one pixel per time step in the x-direction and a maximum of 1 pixel in the y-direction and hence dx and $dy \in \{-1, 0, 1\}$, but not both 0 at the same time. This means that there are 8 possible directions. Each time step there is a 15% probability that either dx or dy, or both, change value by -1 or 1. The 10 x 10 body has 100 possible positions to place sensors, numbered starting with 1 in the upper left corner and 100 in the right lower corner.

Each sensor saves for each time step the current value of the underlying pixel, where the value is in the range $[0, \ldots, 255]$ and then normalized in the range $[0.0, .., 1.0]$. Thus each sensor has an associated time series, with one element of data for each time step.

In the experiments four different environments were used, see Figure 2. The first environment with only vertical stripes in Figure 2(a) is similar to the experience of the kittens in (Stryker et al., 1978). The one in Figure 2(b) has stripes in two perpendicular orientations, while in the third one in Figure 2(c), each pixel has an equal probability of having any value in $[0, \ldots, 255]$. Finally, Figure 2(d) is a realistic image of some rocks on a beach. This image will give more piecewise smooth effects of movement on sensory features than the other images, something that is important for biological systems (Gibson, 1986) as well useful when building models of sensory features in artificial systems (Pierce and Kuipers, 1997; Olsson et al., 2004).



(a) Vertical (b) Cross

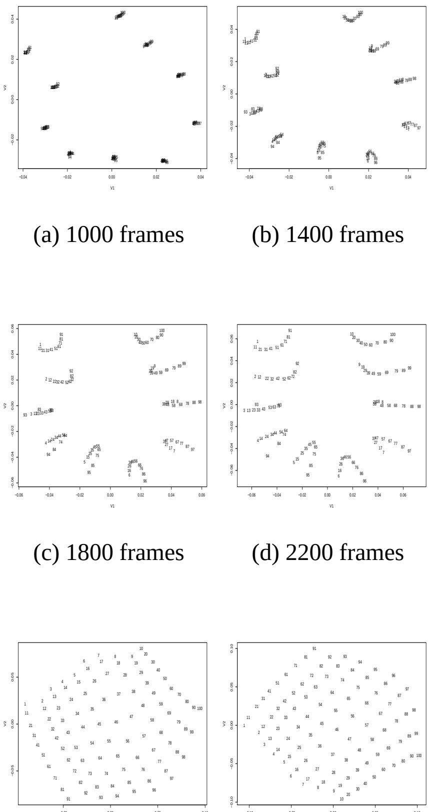
(c) Random (d) Rocks

Figure 2: The different environments used to evolve the sensor layouts.

Effects of Environments on Information Distances between Sensors

To illustrate the effects of environments on the information distances between sensors we first consider an individual with 100 sensors placed in a 10 x 10 grid that moves around in a simple environment like Figure 2(a). After 1000 time steps, or frames, a metric projection (Olsson et al., 2004) is created using the information metric to compute the distance between all pairs of sensors. A metric projection is a map that represents, in a small number of most relevant dimensions (in this case two), the distances between sensors so that the distance between two sensors is an approximation of their informational distance. The metric projection is shown in Figure 3(a). As we can see there are 10 groups of sensors with 10 sensors in each group. This is due to the fact that in a simple vertical environment it is not possible to detect vertical movement and since there are ten vertical lines in the vision layout all sensors in one line will be grouped together since they always have the same input. Now assume that after 1000 time steps the agent moves to the more rich environment of Figure 2(d). In Figure 3(b) to Figure 3(f) metric projections are shown after 1400 to 3000 time steps. As the individual experience more of the complex environment the informational distance of the sensors increases. After 3000 time steps the real layout of the sensors has been found[2].

[2] Since the raw input data contains no directional information it is impossible to find the absolute layout, and only the relative positions can be computed, see (Olsson et al., 2004).



(a) 1000 frames (b) 1400 frames

(c) 1800 frames (d) 2200 frames

(e) 2600 frames (f) 3000 frames

Figure 3: Metric projections of informational distances between visual sensors. After 1000 time steps the agent moves to a richer environment from the vertical environment and the distances between the sensors increase until their layout has been found.

Evolution of Sensor Layouts

Suppose that an individual in an evolving population has enough resources to select only ten sensors from amongst the 100 possible to lay them out as to maximize informational coverage, where the layout is genetically determined by the genome. To evolve the layout of the 10 sensors a Microbial genetic algorithm (GA) (Harvey, 2001) was used, with a population size of 20, a mutation rate of 5%, and crossover probability of 50%. The Microbial genetic algorithm uses tournament selection and was chosen for its extreme simplicity, speed, and the fact that it seems to perform just as well as many other GAs without much tweaking. Each individual's body is a 10 x 10 square with 10 sensors placed somewhere on it. The genome thus encodes a list of 10 positions within that square. The number of possible genomes with 10 genes in the range [1,..,100] is

$$\frac{n!}{k!(n-k)!} = \frac{100!}{10!(100-10)!} \approx 6.2 * 10^{19}.$$

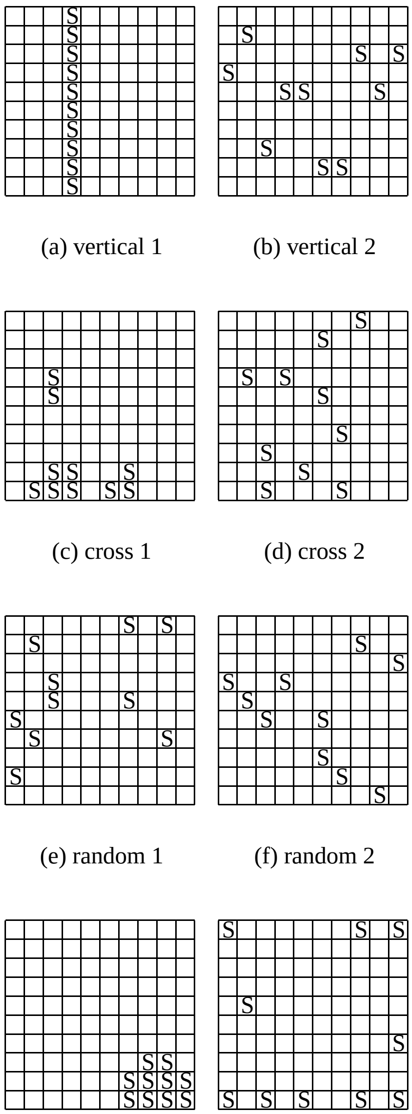
Each position can only be used once in each agent and thus care is taken during mutation and crossover not to let the same position occur twice. If this occur during crossover that particular position is not copied, while in mutation a new random position is selected.

Each individual i moves around the image for t time steps, in our experiments $t = 4000$, and is then evaluated according to its informational coverage, Equation (5). Note that the fitness of one sensor position is dependant on the other positions in the genome, since the fitness is calculated over all pairs of sensors in the genome. In the first experiment the weights were $w_{mi} = 2$ and $w_{im} = 1$, which means that the theoretical maximum value for the mutual information term and the information metric is the same. In the second experiment $w_{mi} = 1$ and the information metric weight $w_{im} = 4$, to reward novelty more than redundancy.

Figure 4 shows the evolved sensors after 10000 generations of evolution of a typical run using the Microbial GA. First consider the case where the mutual information and information metric is given equal importance, i.e., $w_{mi} = 2$ and $w_{im} = 1$, displayed in Figure 4(a), 4(c), 4(e), and 4(g). In Figure 4(a) all sensors are aligned in a single vertical line. This is an example of maximizing the mutual information since all sensors in every time step will extract the same value from the vertical environment. In general for simple environments with variation in only one dimension, like this one, a sensor layout that maximizes the mutual information between sensors will be a spatial representation of the environment.

In Figure 4(c) and Figure 4(g), we find that the sensors are grouped together in both dimensions. This is due to the fact that these environments are two-dimensional, where the environment with only vertical lines informationally can be seen as a one-dimensional environment. In Figure 4(e) and Figure 4(f), the random environments, we find that there is no structure in the sensor layout. An analysis of the fitness landscapes for this environment reveals that there is no structure, and almost all configurations have exactly the same fitness. This is an obvious result since each pixel has the same probability of being in any state, and thus prediction or finding structure is, at least in practice, impossible.

Now consider the case where there is a strong selective pressure for maximizing the distance between sensors (novelty), the case where $w_{mi} = 1$ and $w_{im} = 4$. In Figure 4(b) the layout is completely different from Figure 4(a). Now the sensors are all placed in different horizontal positions, with some redundancy in the vertical positions. By placing each sensor in a different column the distance between the sensors has been maximized, and the fact is that the position in different rows does not make a difference at all in this en-



(a) vertical 1 (b) vertical 2

(c) cross 1 (d) cross 2

(e) random 1 (f) random 2

(g) rocks 1 (h) rocks 2

Figure 4: Evolved sensor layouts for the different environments. In Figure 4(a), 4(c), 4(e), and 4(g) the weights in the fitness function are $w_{mi} = 2$ and $w_{im} = 1$. In Figure 4(b), 4(d), 4(f), and 4(h) $w_{mi} = 1$ and $w_{im} = 4$.

vironment, so that coverage of the dimension giving rise to novelty is complete. In Figure 4(d) we find that the number of collinear sensor pairs and the number of horizontal and vertical lines is maximized by maximizing information coverage (with a high weight for novelty). In Figure 4(h) we find that the sensors are spread out over the grid, which maximizes the informational distance between the sensors.

What would happen if several sensors were allowed in the same position? In the second case where the informational distance is important nothing would change. But, if we consider the case where the mutual information is more important, we would find that all sensors would be placed on the same position. This would be the position with the highest individual entropy, since the mutual information between this position and any other position must be lower, due to the fact that $I(X;Y) \leq H(X)$.

Conclusions

This paper has presented some initial results regarding some of the trade-offs associated with evolving sensor layouts and sensory channel selection that an agent can face. We use formal, information-theoretic, measures of redundancy (mutual information) and novelty (information distance) between pairs of sensors. One trade-off is between redundancy, something that might be important in noisy and/or static environments, and novelty, which is more important in dynamic and complex environments. Some redundancy is necessary in vision since vision is a spatial sense, and without any relation between different vision sensors (eye cells) no spatial information can be found. To study this problem four visual environments were used. These range in complexity from a very simple one with only vertical stripes to a realistic image of rocks and a completely random image. In these environments agents evolve their sensor layout where the fitness of the agent depends on the weights associated with mutual information (redundancy) and the informational distance (novelty) between the sensors. The fitness associated with one sensor's position depends on the positions of the other nine sensors. The results show that the evolved layouts depend both on the environment and the kind of coverage rewarded (in our studies, a weighted combination of redundancy and novelty).

One interesting topic that we will consider in our future work is the relevance of information acquired by a certain sensor to the agent. This notion of relevant information was introduced in (Nehaniv, 1999) and formalized in relation to utility in (Polani et al., 2001) by associating the relevance of information with the utility for a particular agent acting in its environment. By measuring the relevant information acquired by different sensors it should be possible to compute the distance between sensors regarding their relevance to an agent performing a certain task. Equipped with the notion of relevant information and a way to measure the distance between sensors using a metric we expect it to be easier to design sensoric systems that can adapt to different conditions and tasks to be performed, and also to understand how the selection of sensory channels happens in nature. This is especially important since using each sensory and effector channel has an associated cost and this has to be traded against the utility of using that channel (Nehaniv, 1999; Polani, 2003). For example, in some cases it might be more effective from the agent's point of view to accept a slight decrease in overall utility depending on the expenditure for attaining the information. Thus, it is interesting and important to develop a predictive theory of how an agent can most effectively select, combine, or integrate channels to solve a certain task.

Finally, is it important to note that what you do in the world determines what you can distinguish (active perception). Conversely, the world also to some extent determines what you can know about your sensors.

References

Cariani, P. (1998). Epistemic autonomy through adaptive sensing. In *Proceedings of the 1998 IEEE ISIC/CIRA/ISAS Joint Conference*, pages 718–723.

Crutchfield, J. P. (1990). Information and its Metric. In Lam, L. and Morris, H. C., editors, *Nonlinear Structures in Physical Systems – Pattern Formation, Chaos and Waves*, pages 119–130. Springer Verlag.

Dautenhahn, K., Polani, D., and Uthmann, T. (2001). *Artificial Life (Special Issue on Sensor Evolution)*, 7(2).

Dusenbery, D. B. (1992). *Sensory Ecology*. WH Friedman & Co, first edition.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, first edition.

Harvey, I. (2001). Artificial evolution: A continuing saga. In Gomi, T., editor, *Evolutionary Robotics: From Intelligent Robots to Artificial Life: Proc. of 8th Intl. Symposium on Evolutionary Robotics (ER2001)*, pages 94–109. Springer Verlag.

Jung, T., Dauscher, P., and Uthmann, T. (2001). Some effects of individual learning on the evolution of sensors. In Kelemen, J. and Sosík, P., editors, *Advances in Artificial Life, 6th European Conference, ECAL 2001, Prague, Czech Republic, September 10-14, 2001, Proceedings*, Lecture Notes in Computer Science, pages 432–435. Springer Verlag.

Nehaniv, C., Polani, D., Dautenhahn, K., te Boekhorst R., and Cañamero, L. (2002). Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In Standish, R. K., Bedau, M. A., and Abbass, H. A., editors, *Artificial Life VIII: Proceedings*, pages 345–349. MIT Press.

Nehaniv, C. L. (1999). Meaning for observers and agents. In *IEEE International Symposium on Intelligent Control / Intelligent Systems and Semiotics, ISIC/ISIS'99*, pages 435–440.

Olsson, L., Nehaniv, C. L., and Polani, D. (in press, 2004). Sensory channel grouping and structure from uninterpreted sensor data. In *2004 NASA/DoD Conference on Evolvable Hardware June 24-26, 2004 Seattle, Washington, USA*. IEEE Computer Society Press.

Pierce, D. and Kuipers, B. (1997). Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–229.

Polani, D. (2003). Trade-offs in sensoric information acquisition and processing. In Miller, J., Polani, D., and Nehaniv, C., editors, *Evolvability and Sensor Evolution Symposium Abstracts*. EPSRC Network on Evolvability in Biological & Software Systems/University of Hertfordshire Technical Report 384.

Polani, D., Martinetz, T., and Kim, J. (2001). An information-theoretic approach for the quantification of relevance. In Kelemen, J. and Sosík, P., editors, *Advances in Artificial Life, 6th European Conference, ECAL 2001, Prague, Czech Republic, September 10-14, 2001, Proceedings*, Lecture Notes in Computer Science, pages 704–713. Springer Verlag.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656.

Stryker, M. P., Sherk, H., Leventhal, A. G., and Hirsch, H. V. (1978). Physiological consequences for the cat's visual cortex of effectively restricting early visual experience with oriented contours. *Journal of Neurophysiology*, 41(4):896–909.