The DayOne project: how far can a robot develop in 24 hours?

Paul Fitzpatrick
MIT CSAIL,
Cambridge, Massachusetts, USA

Abstract

What could a robot learn in one day? This paper describes the `DayOne` project, an endeavor to build an epigenetic robot that can bootstrap from a very rudimentary state to relatively sophisticated perception of objects and activities in a matter of hours. The project is inspired by the astonishingly rapidity with which many animals such as foals and lambs adapt to their surroundings on the first day of their life. While such plasticity may not be a sufficient basis for long-term cognitive development, it may be at least necessary, and share underlying infrastructure. This paper shows how a sufficiently flexible perceptual system begins to look and act like it contains cognitive structures.

1. Introduction

Sometimes development is a rapid process. Consider the first day in the life of a foal, which can typically trot, gallop, groom itself, follow and feed from its mare, all within hours of birth (McCusker, 2003). Such precociousness is a common pattern for ungulates that evolved in habitats with sparse cover, where the newborn needs to (almost literally) hit the ground running or risk becoming a sitting target for predators.

In epigenetic robotics, we seek to create a "prolonged epigenetic developmental process through which increasingly more complex cognitive structures emerge in the system as a result of interactions with the physical and social environment" (Zlatev and Balkenius, 2001). Should the rapid development of the young of many species give us hope that this process could be much faster than we imagine? Perhaps not, since there is a difference between the development of perceptual and motor skills and the development of *cognitive* structures. Cognitive structures exhibit at least some flexibility of use and reuse, whereas perceptual and motor structures are closely tied to immediate sensing and actuation. But for those who see value in embodied, situated cognition, this distinction may seem unconvincing. Explicit in the work of Brooks was the suspicion that techniques for dealing with the uncertainty and ambiguity of perception and the subtleties of appropriate actuation are the work-horses of intelligence, and are presumably then key to cognitive structures: "This abstraction process is the essence of intelligence and the hard part of the problem being solved" (Brooks, 1991).

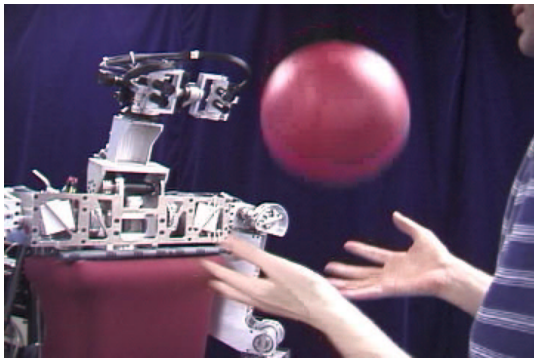


Figure 1: The humanoid robot Cog (Brooks et al., 1999).

Work on the humanoid robot Cog (see Figure 1) has focused very much on rapid perceptual development. This paper describes the `DayOne` project, which was an attempt to integrate much of that work into a single, continuously running system. We hope to demonstrate that sufficiently advanced perceptual structures begin to look a lot like cognitive structures, since there is much flexibility in how they are constructed and used.

2. The stages of DayOne

The robot, upon startup, has the innate ability to turn towards and track movements, and to reach towards close objects, as described in earlier work (Metta and Fitzpatrick, 2003) – other research has shown the feasibility of developing such behavior through experience (Metta, 2000, Fitzpatrick et al., 2003). Development of new

perceptual skills begins in earnest right from the beginning, in the following stages :-

Low-level vision – The robot's low-level vision system is not complete upon startup. It has a filter which, by its construction, is fated to develop into an edge orientation detector, but to do so requires visual experience (Fitzpatrick, 2003b). This is an alternative to using carefully constructed model-based filters such as those developed in (Chen et al., 2000).

Mid-level vision – Once the low-level filters have stabilized, the robot learns to differentiate objects in its immediate surroundings. Again, the object recognition modules involved are fated to perform this task by their construction, but the actual set of objects that the robot learns to recognize is dependent on the contents of its environment (Fitzpatrick, 2003b, Metta and Fitzpatrick, 2003).

Mid-level audition – In parallel with visual development, the robot learns to differentiate utterances. This case is analogous to object differentiation. The actual set of utterances that the robot learns to recognize depends on what the humans in its environment choose to say (Fitzpatrick, 2003a).

High-level perception – As soon as the robot is familiar with some objects and utterances, it can begin to learn the causal structure of simple activities. The modules involved are fated to perform this task, but the activities, the utterances, and the objects involved are all a function of the environment. Together they cover a very rich space of structure, and the robot's ability to 'tune in' to this structure and use it for prediction and further learning begins to take on a cognitive richness (Fitzpatrick, 2003a).

This last stage is the one this paper addresses, along with the problem of integrating all the other stages.

3. Coupled development

At any moment in time, Cog's sensory input is distilled into a distributed set of percepts. In lower-level modules, these percepts are quantitative in nature, and closely tied to the details of the immediate sensor input – for example, the output of the edge orientation detector. In higher level modules, the percepts become more qualitative in nature, and less sensitive to accidental or irrelevant details – for example, the output of object recognition. Still higher-level percepts are even more qualitative, such as a percept that corresponds to seeing a familiar object, or hearing a familiar sound.

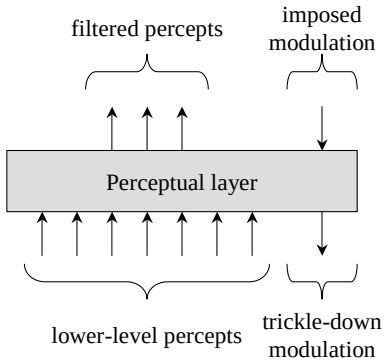


Figure 2: Each successively higher-level perceptual layer filters the one below it. For example, an object recognition layer finds clusters in the feature space presented to it and presents the clusters themselves as the features passed on to the next level. Modulation signals flow in the opposite direction to perception. They indicate when the output of the layer is overly detailed, or on the contrary insufficiently nuanced – for example, when a distinction needed for a task is not being made. If there is no way for the layer to make such a distinction, it passes the request on as 'trickle-down' modulation.

Figure 2 shows an abstract view of each perceptual layer. The primary direction of information flow is from lower levels to higher levels, with details being dropped along the way. A layer is useful if it drops irrelevant details; each layer has its own heuristics about what is relevant. For example, the object recognition module attempts to minimize the effects of pose. Of course, these heuristics will not always be appropriate, and only the overall task can determine that. Hence there is a *modulation* signal that operates in the reverse direction. It can request that more or less detail be supplied for recently activated percepts, or provide a training signal to drive differentiation. This is somewhat analogous to the behavior of neural networks.

The contract between each perceptual layer is as follows :-

- ▷ The "semantics" of what activation means for each line projecting to a higher layer will be preserved as much as possible over time. In other words, an output line will be activated for the same situations in the future as it was in the past.
- ▷ An important exception is that the semantics of an output line may change due to attempts to refine or purify it (so that it is less affected by noise, for example, or responds to the same basic property in an extended range of situations).

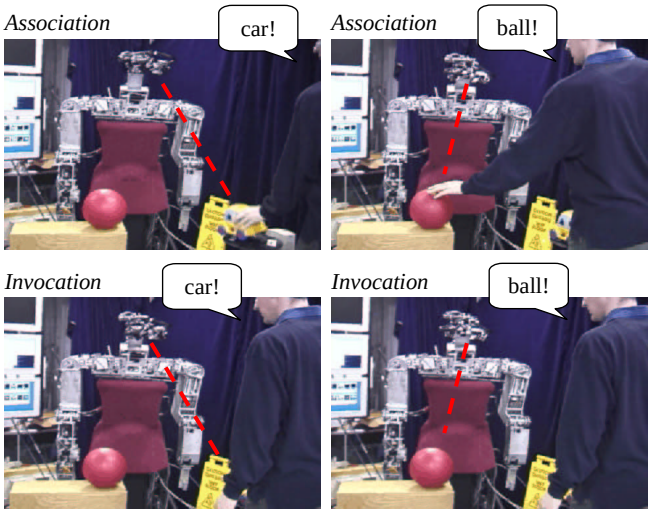


Figure 3: Association and invocation via the egocentric map (Fitzpatrick, 2003a). When the robot looks at an object and recognizes it, its head rolls into an inquisitive look. If a word is spoken at this point (e.g. "car!" or "ball!" in top two frames – note that the human is bringing the robot's attention to an object with his hand) then that word is associated with the object the robot is viewing. If that word is spoken again later (as in the lower frames – note that the human is standing back, only interacting through speech), then the robot queries the egocentric map for the last known location of the associated object, turns there, and looks for the object.

- ▷ Requests for refinement are handled locally if possible, otherwise passed back to a lower layer.
- ▷ Input lines with very similar activation should be detected and merged.

The contract is important because in actual implementation, layers change in both an incremental and batch manner, and this requires careful regulation to stay consistent over time. For example, the object recognition layer quickly creates a new output line when the robot appears to experience a novel object; periodically, all object clusters are examined and optimized using an off-the-shelf clustering algorithm in MATLAB, and the new output line may turn out to be redundant. The output of this clustering is mapped to the current output lines in such a way as to maximally preserve their semantics. Excess lines are never removed, but simply made identical, so that there are no abrupt changes in semantics.

The incremental version of hierachical discriminant regression could be an alternative to this approach (Hwang and Weng, 2000, Weng and Hwang, 2000).

4. Generalization of percepts

Cog has a primitive innate facility for associating cues in different modalities via an egocentric map (similar to the system of (Peters et al., 2001)). Figure 3 shows an example for associating utterances to objects. More generally, Cog also continually searches for useful new ways to perceive the world, where being 'useful' means having predictive power. This search is performed by considering combinations of existing percepts, when heuristics suggest that such combinations may be fruitful. There are three categories of combinations :-

- ▷ **Conjunctions:** if two percepts are noted to occur frequently together, and rarely occur without each other, a composite percept called their conjunction is formed. From then on, this percept is activated whenever the two component percepts do in fact occur together in future.
- ▷ **Disjunctions:** if two percepts are noted to occur frequently together, but also occur independently in other situations, a composite percept called their disjunction is formed. This percept is activated whenever one or both of the two component percepts occur.
- ▷ **Implications:** Causal versions of the above composite percepts, which are sensitive to event order and timing, are also considered.

These composite percepts are intended to enable the robot to make meaningful generalizations, by allowing the same physical event to be viewed in ways that are sensitive to past history. Figure 4 demonstrates the use of such generalizations to link an object with its name through an extended search activity. This is a simplified version of an experiment carried out on human infants by Tomasello (Tomasello, 1997), which in combination with other experiments seeks to rule out many heuristics proposed for fast word learning in the infant development literature (Markman, 1989). A human and the robot engage in a simple search activity, where the human goes looking for an object, which they fail to find immediately. The robot is then tested to see if it can associate the object eventually found with its name, which is given at the start of the search and never mentioned in the presence of its referent.

Searches are presented to the robot as a game following a fairly strict script: first the word 'find' is uttered, then the name of the object to search for is mentioned. Then a series of objects are fixated. The word 'no' is uttered if the object is not

the target of the search. The word 'yes' indicates that the search has succeeded, and the object currently fixated is the target of the search. The meaning of these words is initially entirely in the mind of the human. But the robot can discover them using event generalization, if it experiences a number of searches for objects whose name it already knows.

The word spoken after 'find' gets a special composite implication percept associated with it, let us call it `word-after-find` (of course, no such symbols are used internally, and the word 'find' initially has no special significance – it could be replaced with any other word, such as 'seek,' 'cherchez,' or 'fizzle-tizzle'). When the search is for an object whose name the robot knows (through a pre-established disjunction) that is also noted as a simultaneous event with `word-after-find`. The object seen when 'yes' (`object-with-yes`) is said matches this and an implication is formed between the two. This implication is sufficient to link an *unknown* word following 'find' with the object seen when 'yes' is said, via the `word-after-find` and `object-with-yes` generalizations (again, the choice of the word 'yes' has no special significance, and could be replaced with 'frob').

The above description omitted many other composite events which were created, but served no purpose.

When the generalization mechanism adescribed above was integrated with the full perceptual and motor system of Cog, then the search activity became much simpler to learn, requiring less generalization. This is because the egocentric map has internal state to track when the robot perceives something such as an utterance that is strongly associated with an object it is not looking at (so that it can then direct its gaze towards a remembered prior location of that object). With this structure built in, the robot simply has to map the search activity on to it, which it can do with just two observations (the details of the composite percepts involved are now omitted) :-

- ▷ 'Find' is followed by utterance associated with an absent object.
- ▷ 'Yes' is said when a previously absent object is in view.

Of course, there are many limitations to this generalization mechanism, including :-

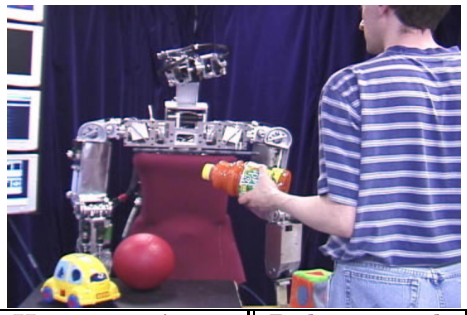
- ▷ The cues the robot is sensitive to are very impoverished, relative to what a human infant can perceive. For example, there is no direct representation of the teacher, and no perception of prosody or non-verbal cues.
- ▷ If there are multiple activities that overlap in some respects, there is the potential for interference between them. The issue of capturing the overall activity context has not been addressed.
- ▷ The basic events used are word and object occurrences, which do not begin to capture the kind of real world events that are possible. So the robot could not respond to non-speech sounds, or changes in distance, or any of the infinite possible events that are not simply word/object appearances.

To begin to deal with this last point, a simple mechanism was developed to get the robot's attention to an unnamed feature or feature combination (as opposed to simply an object) using periodicity detection. All perceptual features on Cog are monitored over a sixty second time window to detect the occurrence of periodicity. Hence if it is desired that the robot attend to the color of objects as opposed to their identity or size, for example, then objects of contrasting colors can simply be shown to the robot. The periodic signal oscillation increased the salience of a percept in a manner similar to the behavioral influences used on Kismet (Breazeal and Scassellati, 1999). But this is a rather limited extension. The next section describes ongoing research into potentially more powerful ways to enhance Cog's ability to generalize percepts.

5. Early generalization

There are many improvements possible on the mechanism for association and generalization described so far. For example, in (Arsenio and Fitzpatrick, 2003), periodic change is used to spot associations between visual and auditory percepts. Suppose two objects in the scene are moving, and a periodic sound is heard, the sound will be bound to the object that could plausibly generate it based on whether the phase of the sound appears locked to the motion of that object.

Is it possible to have a mechanism for association that requires not just temporal synchrony, or periodicity, but can tease out more complex relationships? This is just what the generalization method described in the previous section did, but it relies on producing many composite percepts – which means it can only operate when the number of percepts is relatively low, or in other words at a high level of abstraction. Ongoing work on Cog seeks to give the robot the ability to rapidly perceive non-trivial repeated patterns in its sen-



Human speech	Human action	Robot speech	Robot action
...	...	...	...
say	[shows ball]	say	[looks at ball]
beh		ball	
say	[shows car]	say	[looks at car]
keh		car	
say	[shows cube]	say	[looks at cube]
keh		cube	
say		say	
	[waits]	cube	
	[shows ball]		[looks at ball]
say		say	
	[waits]	ball	
...	...	...	...
	[attracts attention]		[looks at person]
find		find	
ball		ball	
no	[shows cube]	no	[looks at cube]
no	[shows car]	no	[looks at car]
yes	[shows ball]	yes	[looks at ball]
...	...	...	...
	[attracts attention]		[looks at person]
find		find	
toma		toma	
no	[shows ball]	no	[looks at ball]
no	[shows cube]	no	[looks at cube]
yes	[shows bottle]	yes	[looks at bottle]
say	[shows cube]	say	[looks at cube]
		cube	
say	[shows bottle]	say	[looks at bottle]
		toma	
...	...	...	...

Figure 4: Extracts from a dialogue with Cog. First, the robot is taught to name the object it is looking at when the word 'say' is spoken. This is done by speaking the word, then prompting the robot with a short utterance (beh and keh in this example). Short utterances prompt the robot to take responsibility for saying what it sees. A link is formed between 'say' and prompting so that 'say' becomes an alternate way to prompt the robot. Then the robot is shown instances of searching for an object whose name it knows (in the one example given here, the ball is the target). Finally, the robot is shown an instance of searching where an unfamiliar object name is mentioned ('toma'). This allows it to demonstrate that it has learned the structure of the search task, by correctly linking the unfamiliar name ('toma') with the target of search (a bottle). This experiment is close to one considered by Tomasello for human infants (Tomasello, 1997). Ideally, to match Tomasello's experiment, all the objects in this search should be unfamiliar, but this was not done. In the infant case, this would leave open the possibility that the infant associated the unfamiliar word with the first unfamiliar object it saw. In the robot case, we have access to the internal operations, and know that this is not the cue being used.

sory input, at a very low level, with space and time costs that are consistent with massively parallel and *pre-attentive* application, analogous to early visual processing (Nothdurft, 1993). Real-time machine perception benefits greatly from heuristics for quickly filtering out irrelevant stimuli and thus focusing computational effort where it is most likely to pay off. The robots built by the Humanoid Robotics Group at MIT all use one form or another of such heuristics for visual perception, such as biases towards skin-colored regions, moving objects, and bright stimuli (Breazeal et al., 2000). More recently, as already mentioned, we have investigated the utility of periodicity as a perceptual bias, demonstrating cross-modal priming where visually periodic motion influenced the perception of the sound of tools and toys (Arsenio and Fitzpatrick, 2003). To extend this idea still further, to patterns, requires finding an extremely fast way to look at a sequence of percepts and spot the regularity.

5.1 Fast regularity detection

Perception involves many 'missing information' problems which are straightforward to model but difficult to invert. For example, transforming a 3D scene into a 2D view such as our eye might see is a much more tractable mathematical problem than that of recovering the 3D scene given just the 2D view. The basic difficulty is that many possible world states could have produced the same sensory impression, so there is a fundamental ambiguity to contend with. Of course, not all those world states are equally likely to occur, and this fact is explicitly or implicitly used in all computer vision algorithms to generate plausible interpretations of raw sensory input.

For application to robotics, which requires real-time parallel processing of sensory data, there is little time to weigh alternative hypotheses – either algorithms must be quite simple, or the results must be pre-computed. For pattern detection we make use of the second approach, where many possible interpretations of each possible percept sequence are considered, and a favored interpretation and measure of confidence is assigned off-line prior to operation.

5.2 Counting patterns

For an alphabet of k symbols, there are k^n possible sequences of length n. However, if we are concerned only with the *pattern* of symbol recurrence (that is, if we consider a sequence `abbbac` and `zdddza` to be the same pattern), then the number of possibilities is much, much less. The Bell numbers count these – see Table 1.

length n	**distinct sequences**	**distinct patterns**
5	3,125	52
6	46,656	203
7	823,543	877
8	16,777,216	4,140
9	387,420,489	21,147
10	10,000,000,000	115,975
11	285,311,670,611	678,570
12	8,916,100,448,256	4,213,597

Table 1: For sequences with at most n distinct symbols, the middle column of this table shows the number of distinct sequences of length n, while the column on the right shows the number of distinct *patterns* of the same length. This number is far smaller.

The numbers of patterns is smaller than one might expect. This is very important because it suggests that an exhaustive enumeration of patterns (not sequences) is practical, both for off-line evaluation and on-line storage in RAM, for non-trivial pattern lengths.

Ideally, the interpretation of patterns should be driven by experience. Given the volume of sensed data of varying regularity, this seems quite possible. Currently, for short patterns, human expertise is captured directly by examining the patterns by hand (this was done for patterns of length 5 in less than an hour). For longer patterns, an automated evaluation process is used which exhaustively evaluates a set of models, and compares the probability of the patterns they generate to find the most plausible interpretations.

5.3 Progress

Tables have been built automatically for sequences of length up to 10, and by hand for sequences of length 5. Preliminary testing shows results superior to analytic methods previously used (primarily because those methods needed to be weakened to run in real-time, a trade-off not needed for off-line preparation). Some initial experiments have been done with noisy sequences – this requires longer to build tables, but has little impact on run-time operation. Currently the model of activity used is equivalent to regular expressions augmented with the ability to refer to previous sub-expressions. Models are compared based on their description length and specificity. Table 2 shows some sequences, their interpretation as a pattern (determined with a single look-

up), and predicted continuations of the sequence. This ability is suitably fast to allow large-scale comparisons of low-level percepts for regularity.

This work is motivated by recent advances in processor speed and cache size. Much previous work needs to be re-evaluated, to see what algorithms have input spaces that are small enough to allow them to be converted to look-up tables (and 'small enough' can now be quite large!) for fast real-time operation. Of course, this conversion is not always possible, especially if there is significant contextual information that needs to be factored into the interpretive process. But for pre-attentive biases, it seems to make sense.

6. Discussion and conclusions

This paper gave a snapshot of the current state of the `DayOne` project implemented on the humanoid robot Cog. This project focuses on the rapid development of perceptual skills, in an open-ended framework (see Figure 5). Interactions with the physical environment are incredibly rich in terms of sensory feedback – consider the amount of raw information flowing in every second from a robot's cameras, microphones, tactile sensors, etc. Given that wealth of information, we can expect that the developmental process could operate quite rapidly for the growth of appropriate perceptual abilities. And this has been the experience in the `DayOne` project.

There are many exciting research projects that continue to press the boundaries of what can be achieved through robot learning and development – (Weng et al., 2000, Metta, 2000, Roy and Pentland, 2002) etc. It seems that by its nature, the field of epigenetic robotics will advance by a combination of innovation, aggregation, and consolidation. In our own system, a promising direction of research seems to be to take the most 'cognitive-like' ability of the robot (understanding patterns of activity through composite percepts) and find ways to push something analogous back into the very lowest levels of perception. It is interesting to imagine what a robot of tomorrow could learn in a single hour, if everyone's most advanced methods of today become just a part of the smallest building blocks of tomorrow's systems!

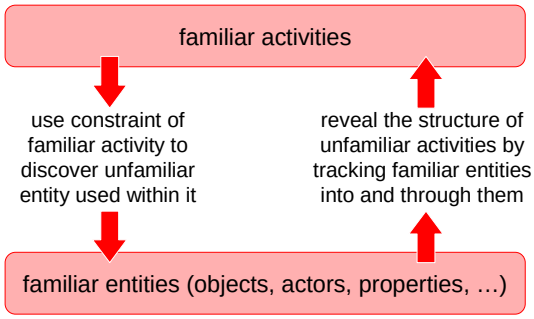


Figure 5: If the robot is engaged in a known activity (top), there may be sufficient constraint to identify novel elements within that activity (bottom). Similarly, if known elements take part in some unfamiliar activity, tracking those can help characterize that activity. Potentially, development is an open-ended loop of such discoveries. Familiar activities can be used to learn about components within those activities and then tracked out into novel activities; then when the robot is familiar with those activities it can turn around and use them for learning also.

Acknowledgements

Funds for this project were provided by DARPA as part of the "Natural Tasking of Robots Based on Human Interaction Cues" project under contract number DABT 63-00-C-10102, and by the Nippon Telegraph and Telephone Corporation as part of the NTT/MIT Collaboration Agreement.

References

Arsenio, A. and Fitzpatrick, P. (2003). Exploiting cross-modal rhythm for robot perception of objects. In *Proceedings of the Second International Conference on Computational Intelligence, Robotics, and Autonomous Systems*, Singapore.

Breazeal, C., Edsinger, A., Fitzpatrick, P., Scassellati, B., and Varchavskaia, P. (2000). Social constraints on animate vision. *IEEE Intelligent Systems*, 15:32–37.

Breazeal, C. and Scassellati, B. (1999). A context-dependent attention system for a social robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 1146–1151, Stockholm, Sweden.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, 47:139–160. originally appeared as MIT AI Memo 899 in May 1986.

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M. M. (1999). The Cog project: Building a humanoid robot. In Nehaniv, C. L., (Ed.), *Computation for Metaphors, Analogy and Agents*, volume 1562 of *Springer Lecture Notes in Artificial Intelligence*. Springer-Verlag.

sequence	guessed pattern	predicted possibilities for sequence continuation
01010	(01)*	1010...
0101110	(01+)*	1010..., 1011..., 1101..., 1110..., 1111...
0120130120	(012013)*	1301...
0120130130	(01[23])*	1201..., 1301...
0011220011	(001122)*	2200...
0011221122	((.)\2)*	0000..., 0011..., 0022..., 0033..., 1100..., 1111..., 1122..., 1133..., 2200..., 2211..., 2222..., 2233..., 3300..., 3311..., 3322..., 3333..., 3344...

Table 2: On the left are some example sequences, where each number represents a percept. The center column shows the interpretation of the sequence as a pattern (written in Perl regular expression syntax). Testing shows that the system can make fairly subtle judgements even for short sequences. With larger sequences, where an approach based on look-up tables becomes impractical, there is enough data for raw statistics to be cues to regularity.

Chen, J., Sato, Y., and Tamura, S. (2000). Orientation space filtering for multiple orientation line segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):417–429.

Fitzpatrick, P. (2003a). *From First Contact to Close Encounters: A developmentally deep perceptual system for a humanoid robot*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering Computer Science, Cambridge, MA.

Fitzpatrick, P. (2003b). Object Lesson: Discovering and learning to recognize objects. In *Proceedings of the 3rd International IEEE/RAS Conference on Humanoid Robots*, Karlsruhe, Germany.

Fitzpatrick, P., Metta, G., Natale, L., Rao, S., and Sandini, G. (2003). What am I doing? Initial steps towards artificial cognition. In *IEEE International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan.

Hwang, W. and Weng, J. (2000). Hierarchical discriminant regression. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 22(11):1277–1293.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction*. MIT Press, Cambridge, Massachusetts.

McCusker, M. (2003). Investigation of the effects of social experience on snapping intensity in Equus caballus foals. Master's thesis, Virginia Polytechnic Institute and State University.

Metta, G. (2000). *Babybot: a study into sensorimotor development*. PhD thesis, LIRA-Lab, DIST.

Metta, G. and Fitzpatrick, P. (2003). Early integration of vision and manipulation. *Adaptive Behavior*, 11(2):109–128.

Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, 33:1937–1958.

Peters, R. A., Hambuchen, K. E., Kawamura, K., and Wilkes, D. M. (2001). The sensory ego-sphere as a short-term memory for humanoids. In *Proceedings of the 2001 IEEE-RAS International Conference on Humanoid Robots*, pages 451–459, Waseda University, Tokyo, Japan.

Roy, D. and Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, 26(1):113–146.

Tomasello, M. (1997). The pragmatics of word learning. *Japanese Journal of Cognitive Science*, 4:59–74.

Weng, J. and Hwang, W. (2000). An incremental learning algorithm with automatically derived discriminating features. In *Proceedings of the Asian Conference on Computer Vision*, pages 426–431, Taipei, Taiwan.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2000). Autonomous mental development by robots and animals. *Science*, 291(5504):599–600.

Zlatev, J. and Balkenius, C. (2001). Why "epigenetic robotics"? In *Proceedings of the First International Workshop on Epigenetic Robotics*, volume 85, pages 1–4. Lund University Cognitive Studies.