Projections as Concepts

Paul R. Cohen
Department of Computer Science
Lederle Graduate Research Center
University of Massachusetts, Amherst MA 01003
cohen@cs.umass.edu

Abstract

What do the first concepts look like? I propose that the earliest concepts learned by infants are abstractions of activities. The semantics of these concepts are predictive—a good abstraction is one that will help the infant predict reward. This idea has been implemented in several programs, in particular, as *fluents* in the Baby simulator and *preimages* in Coelho and Grupen's robotics work. Additional examples and a longer version of this paper can be found at `http://eksl-www.cs.umass.edu /research/conceptual-systems/index.html`

Keywords: Conceptual development

One of the great mysteries of human development is how a sensorimotor neonate becomes a thinking child. Piaget documented the stages of infancy that lead to conceptual knowledge and he sketched mechanisms—assimilation and accommodation—that drive development [13]. Today his account is challenged partly because assimilation and accommodation are mechanisms only in a very weak and general sense, a bit like saying, "to win football games you should move the ball downfield." Today scientists have a higher standard: It should be possible to implement the process by which a sensorimotor agent becomes a conceptual agent in a robot, and test the predictions of this model in human infants. This is what my colleagues and I are trying to do. It should be possible to explain the emergence of particular categories and concepts, and not others, from sensorimotor interaction with the environment. Ultimately we seek conceptual underpinnings for higher cognitive functions such as planning and language.

What is the primitive stuff out of which concepts develop? Nativists give one kind of answer, empiricists another (the full paper reviews several approaches to the question, particularly those of Piaget [13], E. Gibson [7], Spelke [15], Carey [2], Baillargeon [1], J. Mandler [11.12], M. Johnson [8], Drescher [6], and Lakoff [9,10]). Here, I give a simple account of conceptual primitives as projections of experience.

An ideal candidate for a conceptual primitive would be something that is just a small step away from sensorimotor experience, yet is an abstraction of experience (concepts are abstractions). If we regard experiences as trajectories through a space S of very high dimension, then experiences may be similar or different in many ways, each corresponding to a projection of S. The process of abstraction that produces concepts is identical with finding a projection of S; they are different ways of talking about the same thing. The criterion for a good abstraction (projection) is predictiveness, thus concepts have a well-defined semantics: the meaning of a concept is what it entails or predicts about what may be experienced, next. One sees immediately the relationship between concepts and attention: If experience is a trajectory through a high dimensional space, and if concepts are projections of the space, then a concept selects aspects of experience (called the scope of the concept) and thus "conceptual thought" involves attending to those aspects of S within the scope of a concept.

Our challenge has been to build programs that acquire concepts—that is, abstractions of experience with predictive semantics—by interacting with the world, and we are trying hard to be stingy with the innate endowment. I'll illustrate our work by describing two projects, from my lab and from the Laboratory for Perceptual Robotics at the University of Massachusetts. The full paper will elaborate on these examples.

The Babyworld simulator contains a highly stylized "baby" with eyes, ears, a mouth, one arm and no legs [4,5]. Baby's experience is implemented in 26 streams which code what it's seeing, hearing, and mouthing, as well as internal states such as hunger, pain and alertness. Some streams code for initiating action (e.g., DO-ARM) and others code for the action itself (ARM move...). Baby's behavior is simple and probabilistic; for instance, it gets hungry sometime after it eats, determined by a probability distribution. Babyworld is unrealistic in many respects. Notably, we have finessed all perceptual issues by placing "percepts" directly in Baby's streams; for example, when Baby's eyes are pointing at the green rattle, the token "green" is placed in the sight-color stream and the token "rattle-shaped" is placed in the sight-shape stream. Shortly I'll describe a project that doesn't finesse perception.

Baby learns concepts of the sort I've been talking about—predictive abstractions of experience, or if you prefer, predictive projections of its 26-dimensional space of experience. Baby's learning involves counting co-occurrences of events relative to the frequencies with which the events do not co-occur. These frequencies are sufficient to establish whether events are statistically dependent, that is, whether the occurrence of one predicts the occurrence of another. When a dependency between events is sufficiently strong, Baby creats an object that represents the joint event. The details aren't important. What's important is that this business of counting cooccurrences yields two kinds of predictiveness. First, if the events are lagged in time, then one event predicts the other. Second, if the events are simultaneous, then Baby has learned a two-dimensional projection of its 26-dimensional space in which events happen.

In fact, Baby counts several kinds of events. The simplest is a change in the value of a stream; for example, when Baby's eyes shift from the green rattle to the red keys, the SIGHT-COLOR and SIGHT-SHAPE streams both change value simultaneously. Over time, Baby learns scopes, which are pairs of streams in which tokens tend to change simultaneously. Scopes are extraordinarily primitive concepts in the sense of being predictive abstractions of experience. The abstraction is just, "here are two streams in which, in my experience, things tend to change simultaneously," and it is predictive in the weak sense that if one stream changes, the other probably does, too. A stronger notion of predictiveness comes when Baby starts to use scopes to focus its attention in learning. Once a scope has been learned, Baby uses it as a template for simple fluents, which are constructed from token values that start and stop within a scope. For example, within the (SIGHT-SHAPE, SIGHT-COLOR) scope, Baby might relatively frequently see the simultaneous occurrence of (SIGHT-SHAPE RATTLE-SHAPE) and (SIGHT-COLOR GREEN), in which case it would create a conjunctive fluent. An even stronger notion of predictiveness comes from having Baby count cooccurrences not of token values but of fluents. In particular, Baby learns that some fluents tend to start after others start. These produce what we call context fluents.

Here is a context fluent that Baby learned. It says, "in the context of having no tactile sensation in the mouth and crying, grasp something with a wooden texture":

```
(CONTEXT
 ((tactile-mouth none) (voice cry))
 ((tactile-hand wood) (hand close))
```

And here is another context fluent that says, in the context of grasping something with wooden texture, the experience of feeling something wooden in the mouth and mouthing it, begins:

```
(CONTEXT
 ((tactile-hand wood)(hand close))
 ((tactile-mouth wood)(do-mouth mouth)))
```

As the first component of this fluent is identical to the second component of the previous one, Baby may form a chain fluent:

```
(CHAIN
```

```
 ((tactile-mouth none) (voice cry))
 ((tactile-hand wood) (hand close))
 ((tactile-mouth wood)(do-mouth mouth)))
```

Here is another chain fluent learned by Baby:

```
(CHAIN
 ((tactile-mouth none) (voice cry))
 ((tactile-hand plastic) (hand close))
 ((tactile-mouth plastic)(do-mouth mouth)))
```

The structure of these fluents is almost identical. Each describes an episode in which Baby is holding nothing and crying, then holding an object (either wood or plastic), then mouthing the object. Although Baby cannot yet form equivalence classes of objects, I think these chain fluents provide the evidence for a class of objects, some wooden, some plastic, that play a particular role in the activity represented by the chains. In other words, I think Baby is poised ready to learn to categorize objects based on their roles.

Returning to my original claim, the concept to be learned is a projection of Baby's experience (which encompasses a 26-dimensional space) onto five dimensions: TACTILE-HAND, TACTILE-MOUTH, HAND, VOICE, DO-MOUTH. The chain provide expectations of tokens in these five dimensions; indeed, chains aren't learned unless they do a good job predictively. Thus the chain is a concept, as I described it earlier—a predictive projection—and it is an abstraction in the sense that the other 21 dimensions are ignored. But when we compare multiple chains, another kind of predictive projection can be found. The TACTILE-HAND and TACTILE-MOUTH dimensions can contain something that is either wooden or plastic. This disjunction is also an abstraction, and it corresponds to the two kinds of rattles that Baby has in its environment.

The second project I want to mention was done by Jefferson Coelho and Rod Grupen at the Laboratory for Perceptual Robotics [3]. They showed how a projection of the state information encountered while grasping an object could be used to identify the object. Very roughly speaking, a Salisbury hand attempts to grasp a prismatic object by activating controllers that run to convergence, that is, the controllers attempt to minimize force residuals and wrench residuals. The net effect of running these controllers is that the fingers of the hand migrate over the surface of the object until they achieve a stable grasp. It turns out that these migrations describe trajectories in the two-dimensional error space. The union of all trajectories for a given object is called a preimage, and the shape of the preimage depends on the object geometry. In consequence, one can imagine that as the hand attempts to grasp the object, it is "considering" all object preimages simultaneously. It rules some out when the hand enters a configuration that is inconsistent with that object.

Remarkably, Coelho and Grupen's robot can identify objects by "feel" as it attempts to gain a stable grasp around them. This is a very clear example of a conceptual activity—classifying objects—that arises out of a purely sensorimotor activity. Preimages work the trick, and preimages are just predictive projections of experience. Of all the features of the robot's experience that might have been measured, Coelho and Grupen chose to measure force error and wrench error, and the resulting projections were sufficient to discriminate objects. By the way, preimages are bona fide predictive projections: The robot uses the predictions to adjust its grasp.

Concepts for Baby and the grasping robot are grounded in activities. For example, two chains represent identical activities, hence the different participants in those activities (wooden and plastic rattles) form an equivalence class with respect to the activities. Following Lakoff, Johnson, Mandler and others I suggest that classification grounded in activity is the only way to go. The objectivist position—that the world "has" categories, and it is the infant's job to learn them—must be wrong. For one thing, categories change as the infant develops, as do the criteria for categorization. Initially, children categorize thematically (i.e., in an activity-oriented way), only later do they attend to so-called objective properties such as shape and color.

Objective properties have gotten a bad name because they appear inadequate to explain the psychology of categories [14,9]. Lakoff and others suggest that categorization might instead be based on interactional properties [9,8]. For example, one such category for Baby is, "something I can grasp and mouth, either wooden or plastic," based on interactional properties "graspable," "mouthable," and "texture." One is tempted by the conjecture that although categories cannot be defined in terms of necessary and sufficient objective features, they might be defined in terms of necessary and sufficient interactional features. However, I believe categories are best defined in terms of activities, and the apparent superiority of interactional features is due to them describing activities better than objective features.

Consider a conceptual activity such as judging whether a cup and a ladle are similar or different. We immediately want to ask, "Similar or different in what context?" As devices for transferring liquid, cups and ladles are similar; as containers to drink from, they require different motor schemas; as something to serve coffee in at an elegant dinner, they aren't similar. Note that the required context in which we judge similarity is often an activity, often purposeful, and often agent-centered. We are tempted to say the interactional properties of cups and ladles are a better basis for judgments of similarity than their objective properties. However, objective properties might be relevant to judgments of similarity; for instance, the material of which ladles and cups are made is relevant in the context of dropping them on the floor or taking them backpacking. Evidently, what's central to judging the similarity of objects is knowing which activities the objects are involved in. Activities seem to select which features of objects are relevant to judgments of similarity; these features will sometimes be objective, often interactional.

REFERENCES

1. R.Baillargeon. How do infants learn about the physical world? Current Directions in Psychological Science, 3:133–140, 1994.

2. S.Carey and R. Gelman. The epigenesis of mind: Essays on biology and cognition. Hillsdale NJ: Erlbaum, 1991.

3. Coelho, J., Grupen, R., "A Control Basis for Multifingered Grasps," submitted to the IEEE Journal of Robotic Systems, August, 1995.

4. Paul R. Cohen, Tim Oates, and Marc S. Atkin. Preliminary evidence that conceptual structure can be learned by interacting with an environment. 1996.

5. Paul R. Cohen, Tim Oates, Marc S. Atkin, and Carole R. Beal. Building a baby. To appear in Proceedings of the Cognitive Science Conference, 1996.

6. Gary L. Drescher. Made-Up Minds. MIT Press, 1991.

7. Eleanor J. Gibson. An Odyssey in Learning and Perception. MIT Press, 1990.

8. Mark Johnson. The Body in the Mind. University of Chicago Press, 1987.

9. George Lakoff. Women, Fire, and Dangerous Things. University of Chicago Press, 1984.

10. George Lakoff and Mark Johnson. Metaphors We Live By. University of Chicago Press, 1980.

11. Jean M. Mandler. How to build a baby: On the development of an accessible representational system. Cognitive Development, 3:113–136, 1988.

12. Jean M. Mandler. How to build a baby: Ii. conceptual primitives. Psychological Review, 99(4):587–604, 1992.

13. Jean Piaget. The Origins of Intelligence in Childhood. International Universities Press, 1952.

14. E. Rosch and C. B. Mervis. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7:573–605, 1975.

15. E. S. Spelke, K. Breinlinger, J. Macomber, and K. Jacobson. Origins of knowledge.. Psych logical Review, 99:605–632, 1992.

16. E. Thelen and L. Smith. A Dynamic Systems Approach to the Development of Cognition and Action. MIT Press, Cambridge, MA, 1994.