Modifying representations through joint attention

Society of Mind – Final Project

Paul Fitzpatrick
paulfitz@ai.mit.edu

1 Introduction

In section 22.4 of the Society of Mind [9], Prof. Minsky discusses how a teacher can "lead a child to build large, complicated processes from smaller ones by laying out sequences of steps." I am interested in such sequences of steps. In particular, I am interested in how we can communicate new representations to each other, rather than simply communicating facts. One dictionary definition of communication is "a process by which information is exchanged between individuals through a common system of symbols, signs, or behavior"(Merriam-Webster). I will look at instances of communication where the emphasis is on building that common system.

For simplicity, in this paper I will consider the case of a single teacher and a single student, interacting in a series of tutorials through which the teacher endeavors to give the student new representations for some domain of interest. We can consider two general possibilities :-

1. The student already has representations for the domain of interest; just not the ones the teacher wishes the student to have. All that is missing in this case is *shared* symbols and representations.
2. The student does not have representations for the domain communication is desired within; shared representations have to be constructed from scratch, rather than just recognized or remapped.

Here are a few specific example situations that will be important for this paper.

- ▷ A care-giver playing games with an infant.
- ▷ A graduate student playing games with a robot.
- ▷ Actual formal teacher/student interactions.
- ▷ Communicating with an alien race.

I will look first at this last example since it is the one most amenable to formalization, and then work back towards the more interactive situations.

2 Cosmic intercourse

In Hans Freudenthal's book, Lincos [7], he gives a sequence of tutorials for a "language for cosmic intercourse" that might be understood by an alien race sharing none of our culture, and none of our languages. Here is what he *does* expect of his putative students:

> ... I shall suppose that the person who is to receive my message is human or at least humanlike as to his mental state and experiences. I should not know how to communicate with an individual who does not fulfil these requirements. Yet I shall not suppose that the receivers of my messages must be humans or humanlike in the sense of anatomy and physiology.

So he expects his students to already have experience and human-like representations of some common domains of interest, and the task is mostly getting them to recognize these within his formalism. He limits himself to linear sequences of symbols (transmitted as distinctive radio signals). Within this restriction, he presents a series of tutorials which in theory would give a diligent alien student a shared vocabulary with him for topics in mathematics, physics, and behavior, including concepts such as complex numbers, good, bad, birth, death, courtesy, rotation, liquids, and relativity. Freudenthal makes some revealing choices in the order in which he presents his tutorials :-

- ▷ First he introduces mathematics, as an abstract topic ideally suited to presentation within his constraints.

- ▷ Then time is introduced, also relatively easy given the nature of the assumed transmission medium.
- ▷ Then he introduces behavior through a series of "morality plays" between imaginary characters. The characters typically discuss mathematics, and other characters then comment about the nature of these discussions.
- ▷ Physics – space, motion, mass – is introduced in terms of behavior, almost animistically.

There is a marked boot-strapping effect here. First, the teacher introduces vocabulary and representations for some domain for which the student is expected to have experience. In this case, that domain is mathematics. Then, a means of referring to parts of the tutorial is introduced in the form of a representation for time. Time intervals can now be used as "pointers" to the tutorial fragment that occurred within them. Freudenthal now has the tools to discuss communicative acts, using mathematics as the somewhat arbitrary domain of discourse, and time-stamps as an initial means to refer to and annotate the communications. This acts as a gateway to discussing human behavior in general. Finally there is a further boot-strap from human behavior to "everything else".

Fans of SETI make much of mathematics as a universal language, so it is easy to assume that this is how Freudenthal is using it. In fact, it is to some extent an arbitrary domain choice (at least up to the point where physics is introduced). It could be replaced with any domain of common experience between the teacher and student.

The use of time as a pointing mechanism is also somewhat arbitrary, although very natural given the linear nature of the communication medium. Of course, time is a useful concept in its own right and is used for purposes other than indexical; however, my reading of the book suggests that time could have been introduced after rather than before behavior if some other indexical means had been available.

So a characterization of Freudenthal's overall scheme is: develop representations for an arbitrary topic, develop means to refer to parts of the communications about this topic, then develop shared representations of the communication and the communicators.

I will see how these ideas apply to other less constrained instances of communication, when the teacher and student are interacting, rather than thousands of light-years apart.

3 Interactive communication

I will now consider communicating with specific Earth-bound aliens, namely robots and human infants.

Freudenthal's tutorial is a linear sequence of symbols presented in an environment that is idealized almost to the point of non-existence, and without interaction between the teacher and student. When we move away from passing a linear sequence from the teacher to the student, what exactly is the tutorial? Formally it would be the entire history of the environment, including the student and teacher, throughout their period of interaction. But this is too vague and structureless to be a useful concept. We can retain some structure by focusing on the student's perception of the environment. If the student's perception is human-like in the sense of being strongly mediated by attentive processes, only a limited slice of the environment may be attended to at any particular time. And this is the interesting part of the environment, for teaching.

> The tutorial is whatever the student pays attention to. It is the time sequence of the student's locus or loci of attention.

So we assume some constraint on attention – which seems reasonable from computational arguments – and use that constraint as a source of structure. Notice that it is now impossible to specify the tutorial without first specifying the student. We need to characterize the student to a far greater degree than Freudenthal does.

What part of the world is the student attending to now? The question can be broken down as follows :-

- ▷ Objects of attention – What parts of the world is it possible for the student to attend to? This could be quite limited, at least to begin with. It is intimately connected with the student's representational abilities, and so will change as these change.
- ▷ Saliency – What biases and influences are there on what the student chooses to attend to, in terms of the properties of potential objects of attention? These are prejudices about what the student finds more intrinsically interesting.
- ▷ Motivation – What internal influences are there on what the student chooses to attend to? I will lump goals, emotions, drives etc. under this heading. I will also not require a clear distinction between motivation and saliency.

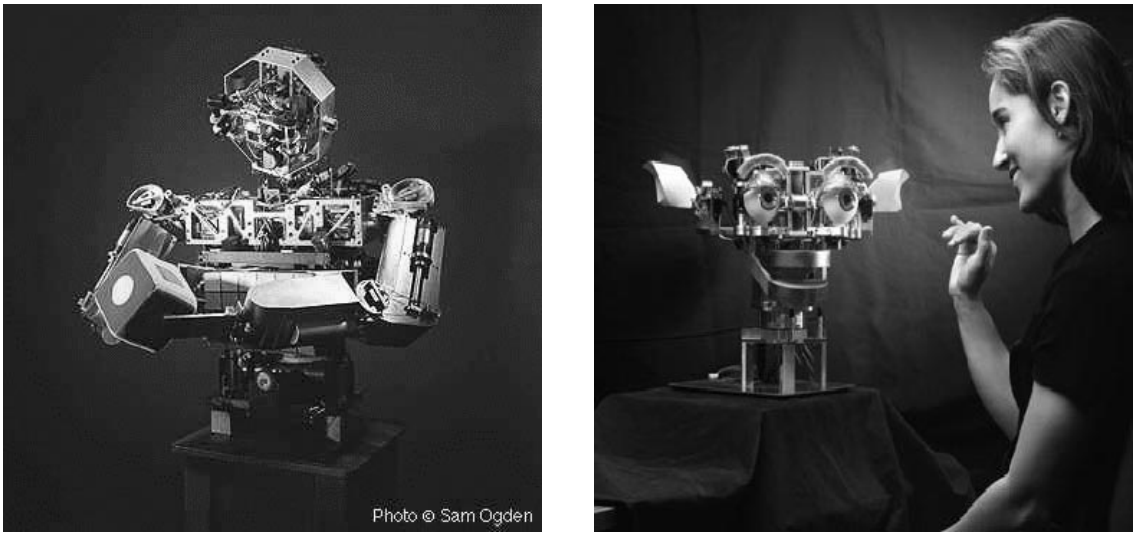

Figure 1: Cog (left) and Kismet (right), two humanoid robots.

- ▷ Joint attention – How can a teacher use knowledge of all the above to direct the attention of the student? I am using joint attention in a more asymmetric and manipulative a sense than is usual. I would like to emphasize that I am deliberately taking a narrow simplistic viewpoint, for a constrained social relationship.
- ▷ Internalizing – How does what the student attends to come to induce long-term changes in the student, and what is the nature of these changes? This is a form of learning, but I will talk about "internalizing" the tutorial instead because there will be learning involved on many other levels as well.

The teacher needs to have a theory for each of the above. Freudenthal was able to skip to the last of these steps, internalizing, and for that to simply assume "human-like" abilities. I will discuss the earlier steps in some detail, using as a reference current work being done for the Cog project at the MIT AI lab. Cog is an anthropomorphic robot, and the Cog project explores the impact of various aspects of the human condition on human intelligence [3]. Of interest for this paper is work done on attention, saliency, motivation, and joint attention. I will use examples from Cog and Kismet, a related robot (see Figure 1).

3.1 Object saliency

Some objects are more interesting than others. For a human infant, faces are particularly interesting. Kismet likes faces too, and brightly colored, moving objects [5]. It is important that these prejudices on the part of the student exist, so the teacher can use them as leverage to direct and predict the student's attention. For Kismet, being able to attend to something means two things :-

- ▷ Kismet can bias its perception to filter for the object (face or toy).
- ▷ Kismet has a representation for the object.

Just as Freudenthal's initial domain of discourse was arbitrary, the nature of the objects that can be attended to initially is not crucial from the point of view of the teacher.

3.2 Motivation

Saliency does not uniquely define what the student will pay attention to. For example, whether Kismet will attend to a toy or a person is a complex function of how its "social drive" and "stimulation drive" evolve over time [6]. This "motivation" is a complex dynamic that needs to be understood by the teacher. I will assume it is strongly social, or in other words well engineered to allow the teacher to manipulate the student's locus of attention.

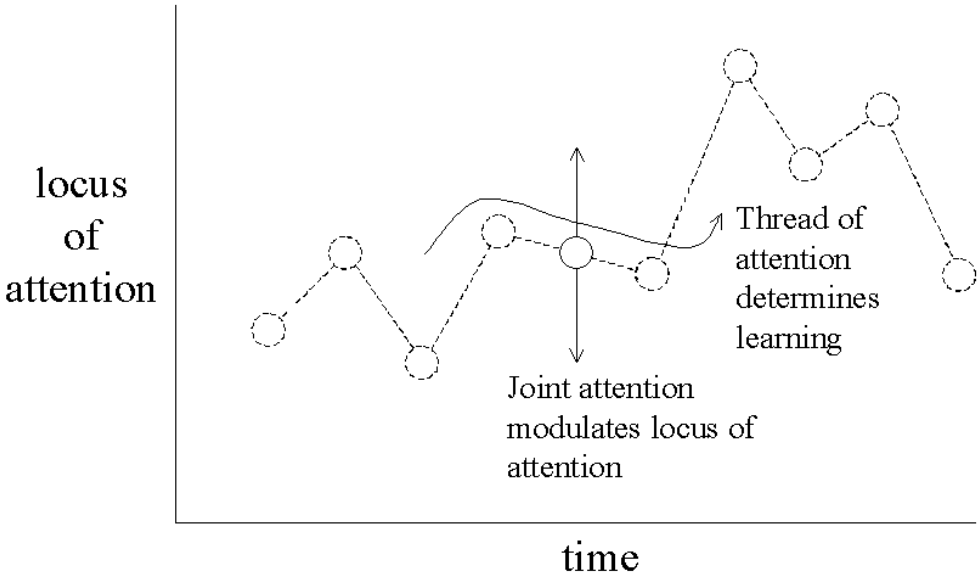

Figure 2: Locus of attention

3.3 Joint attention

Joint attention refers to the influence the student and the teacher have on each other's locus of attention, and their understanding of the locus of attention of the other . I will view this in an asymmetrical way, asking the questions :-

- ▷ How can the teacher determine the locus of attention of the student?
- ▷ How can the teacher influence the locus of attention of the student? This includes manipulating any ability the student has to determine and respond to the locus of attention of the teacher.

Cog's locus of visual attention can be determined by an untrained human because of its familiar body plan. Similarly, Kismet's motivational state can be estimated from its facial expression by a naive observer, which gives a measure that is correlated with attention. A knowledge of what Kismet finds salient gives the teacher some simple ways to influence its locus of attention – for example, shaking a toy to increase saliency through motion. Knowledge of motivation can be used in a similarly manipulative way.

4 Internalizing the tutorial

The various components of joint attention are very active areas of robotics research – gesture recognition, gaze and pose estimation, etc. Hope is often expressed that once robots are sufficiently "social", they will be able to learn through interaction with humans. Imitation is generally the mechanism suggested for learning. I will take a different perspective, suggested by the tutorial structure.

One of the conceptually simplest way to learn associations is through a coincidence detector, where events that coincide repeatedly within some temporal window become grouped in terms of representation. Regardless of what form of learning the student actually has, it would be expected to make this kind of association – so I will model it this way.

Whatever the representations the student currently has, the coincidence detector will lead to the generation of representations for sequences, cycles, sets, or tentative amorphous links between these representations. At heart, it is just a temporal window for spotting correlations between when representations become active in close temporal proximity (with the proximity measure being on a scale related to the natural temporal scale of the representations). Such a detector is useful in the absence of a teacher, in purely exploratory learning – for instance, to detect causal structure. But it truly comes into its own when a teacher is present.

We defined the tutorial as whatever the student pays attention to. The teacher exercises control over the tutorial through the mechanisms of joint attention. She can control the order in which the student attends to different objects. Therefore, she can control the associations the student makes (see Figure 2). In general there is no guarantee that

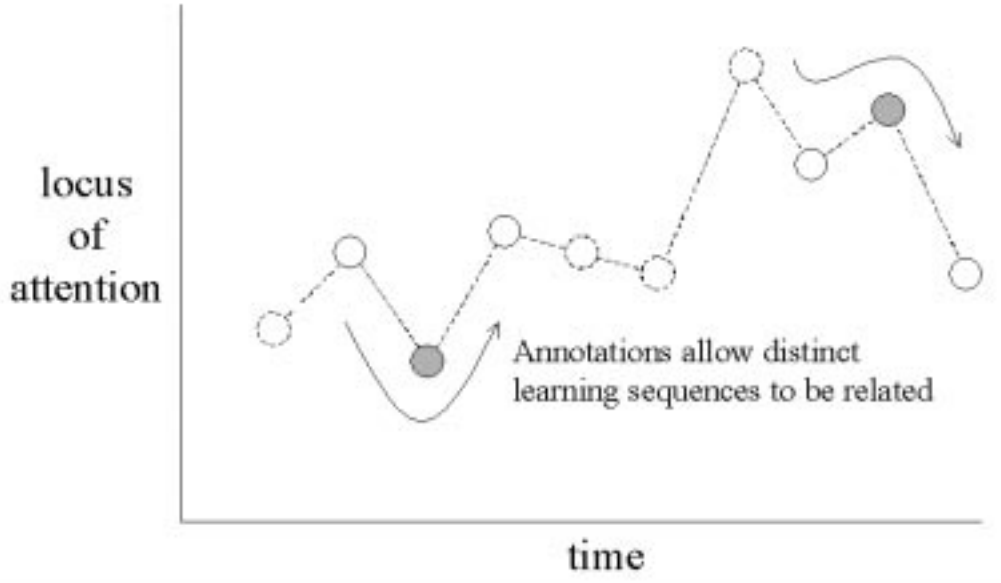

Figure 3: Annotations

objects or events that are related in interesting ways will in fact come to the student's attention within some temporal window – but the teacher can choose to make this happen. In effect, the teacher's knowledge that the student has a coincidence detector allows the teacher to partially control the representations it constructs, by careful ordering and repetition.

The coincidence detector is effectively being used as an indirect basis for communication of representations. Any mechanism that alters the student's representations, and which can be influenced by actions of the teacher, can be used to support communication – even if the mechanism's nominal purpose has nothing to do with communication.

5 Annotations

A coincidence-detector mechanism is limited in the representations it can generate, since the associations it makes can only relate to objects or events that are attended to roughly contemporaneously. So if, for example, there were events that have the same effect but which are mutually exclusive and so cannot occur in close proximity in time, there is no way to group them. Or if the teacher wants to signal a grouping of events over which she has no control, there is no way to enforce this.

I could introduce new mechanisms for this. However, it is possible to stay with the same simple coincidence-detector mechanism, and instead use special objects or events with no intrinsic meaning to act as proxies for relating disparate events. I will call these annotations; they could be sounds, gestures, or arbitrary actions. It is important that they can be generated at will by the teacher, and that they have no intrinsic associations for the student. These annotations can be inserted by the teacher into the student's stream of attention so that they become associated with each of a set of events which – since they are then associated with the same annotation – are now associated with each other through that annotation. So annotations are formal signs that act as a nexus of associations that would not be possible with pure temporal coincidence. They serve to knit representations together (see Figure 3).

Annotations can also bias learning by implicitly solving some model selection problems, such as determining the number of classes or dimensions to use in a learning task.

Before Freudenthal starts teaching human behavior, he introduces a method for refering to parts of the tutorial itself and commenting about them. He uses time-stamping, which is a simple and unambiguous technique. Annotations can provide an approximation to this.

6 Conclusions

Learning by imitation is a popular notion in robotics, but is more problematic than is often allowed – as Prof. Minsky emphasizes with his cup-grasping example. I have suggested here that it may be useful to see mechanisms of joint

attention as tools for guiding the student through a linear experience of the world that encodes for representations relative to the student's learning mechanisms (assumed known, as they are in robotics). This makes clearer the questions that need to be answered :-

- ▷ What are the student's attention-mediated learning mechanisms? What are the theoretical limits of the representations they could generate?
- ▷ How much control can the teacher exercise over the input to the student's learning mechanisms? What are the practical limits of the representations that can be learned for the kind of inputs the teacher can arrange?

None of the ideas I presented in this paper are new in themselves. For example, annotations are just words – formal sounds or gestures. I do not claim to have invented language. My contribution has been to show a specific way that mechanisms of joint attention and annotations can be used as tools to manipulate a learner's built-in ability to detect coincidences in such a way that a teacher can cause the learner to detect and represent relationships they might never have experienced of their own accord. From this perspective, roboticist's work on joint attention becomes more fundamental to learning and communication than they themselves perhaps realize.

References

[1] S. Baron-Cohen and H. Ring. A model of the mindreading system: neuropsychological and neurobiological perspectives. In P. Mitchell and C. Lewis, editors, *Origins of an understanding of mind*. Lawrence Erlbaum Associates, 1994.

[2] B. Blumberg. *Old Tricks, New Dogs: Ethology and Interactive Creatures*. PhD thesis, MIT, 1996.

[3] R. A. Brooks, C. Breazeal, R. Irie, C. C. Kemp, M. Marjanovic, B. Scassellati, and M. Williamson. Alternate essences of intelligence. *AAAI*, 1998.

[4] G. Butterworth. The ontogeny and phylogeny of joint visual attention. In A. Whiten, editor, *Natural Theories of Mind*. Blackwell, 1991.

[5] C. Brazeal (Ferrell) and B. Scassellati. A context-dependent attention system for a social robot. *IJCAI99*.

[6] C. Brazeal (Ferrell) and B. Scassellati. Infant-like social interactions between a robot and a human caretaker. *Special issue of Adaptive Behavior on Simulation Models of Social Agents*, 1998.

[7] H. Freudenthal. *Lincos: design of a language for cosmic intercourse*. North-Holland Publishing Company, 1960.

[8] M. Minsky. Communication with alien intelligence. *Extraterrestrials: Science and Alien Intelligence*, 1985.

[9] M. Minsky. *Society of Mind*. Simon and Shuster, 1986.

[10] G. Rizzolatti and M. A. Arbib. Language within our grasp. *Trends in Neuroscience*, 1998.

[11] B. Scassellati. Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. *To appear as part of a Springer-Verlag series*, 1999.