Who Needs Time? Implicit Time Is Sufficient for Some HRI Tasks

Richard Veale and Matthias Scheutz
Human-Robot Interaction Laboratory
Cognitive Science Program
Indiana University
Bloomington, IN 47406, USA
{riveale,mscheutz}@indiana.edu

INTRODUCTION

We examine the temporal evolution of a neural-network model of word-referent co-learning engaged in an interactive learning task with a human teacher (Veale, Schermerhorn, and Scheutz in preparation). The observed naturally-timed interaction is used to argue that in at least some interesting interactive situations, explicit representation of or operation on time is not necessary. Observing that many interactive situations will be similar, we hypothesize that in fact *most* interactions will require no explicit representation or reasoning about time. Rather, natural timing of interaction falls out of dynamical properties of the coupled system involving the interactors' brains, bodies, and the shared environment. A change in internal dynamics in one interactor may be caused by its own internal dynamics (e.g. getting bored) or by cues from the other interactor (e.g. a word uttered), or by a combination of both. This may cause cause him to produce cues affecting the other interactor, completing the dynamical loop.

To contextualize our argument, we examine research showing that a variety of interesting (temporal) behaviors can be produced and explained by adopting a dynamical systems stance. None of the agents examined have any explicit representation or knowledge of time, yet still manage to engage in interesting temporally extended behaviors. We then present a more thorough analysis of the interaction in our experiments, investigating the internal dynamics of the robot model and the human interactor, how they produce what behaviors, and how all of this can be put together to build a convincing picture of the coupled interaction dynamics. We conclude that even in complex HRI interactions, explicit representation of time is not required to produce rich temporal dynamic interaction.

RELATED RESEARCH

It is not a new development that interesting (temporally-extended) behavior can be had from very simple dynamical systems. A concrete example is the walking behavior in a variety of simulated and real hexapod agents (e.g. (Beer 2008) for an overview). A neural network with no explicit representation of time is controlling the agent. A similar dynamical controller is used in (Williams, Beer, and Gasser 2008), where referential communication is evolved in simple agents engaged in a task where a sender knowledgable of the correct target location must guide an ignorant receiver to that location. This communication is accomplished via strategies which necessarily incorporate time. The interaction between the agents is naturally extended over time, yet in neither agent does any explicit representation of or reasoning about time occur. Kelso et al's Virtual Partner Interaction (Kelso et al. 2009) is a paradigm in which a virtual hand is guided by a dynamical system known to guide most human coordination. A human interacts with this on certain tasks (e.g. matching the rhythm of the hand moving back-and-forth), thus creating a coupled dynamical system which can be used to study interaction in novel ways. Interesting temporally extended interaction (such as regular motion, rhythm) is demonstrated both by each individual system, and by the coupled system as a whole. However, each system is guided by a dynamical system, with no explicit representation of time.

TIME-AGNOSTIC HRI

In (Veale, Schermerhorn, and Scheutz in preparation) we present an architecture and model of word-referent co-learning, as well as experiments involving interaction between a human and the model instantiated in a robot. The neural-network based model non-linearly and incrementally builds associations between words (as phoneme strings) on one side, and colors (as locations in 3-d RGB-feature space) on the other. Activation of one side will cause some activation of associated nodes on the other. The vision side of the network is connected to a rudimentary attention-guiding algorithm which is based on saliency. The saliency of an object is a combination of its novelty, size, local motion, and "context" given the state of other perceptual modalities. E.g., if a word associated with an object is uttered, that increases the "contextual" saliency of that object.

In the interactive experiment (figure 1), the robot is allowed to habituate to two blocks placed in its visual field. A human interactor then enters, and draws the robot's attention to each block in turn, several times, while uttering the associated word. To test association learning, a third, novel object is placed with the original two, drawing the robot's attention. The robot's interest in the salient novel object is

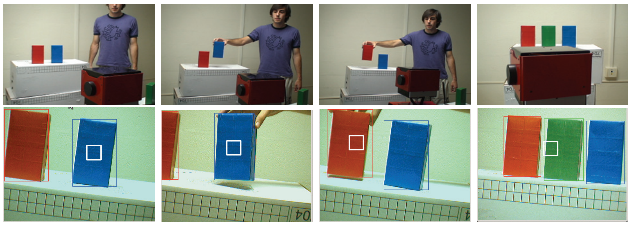

Figure 1: Above: steps of the experiment. Below: model's view/focus.

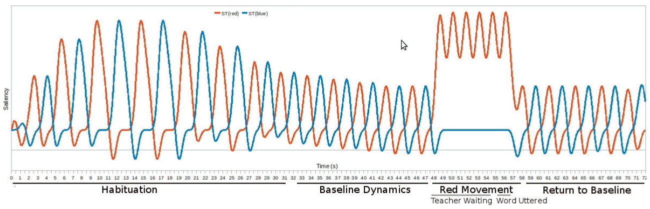

Figure 2: Evolution of the salience of two objects to the word-learning agent over the beginning of a typical experiment with a human. Qualitative events are indicated.

then "displaced" by one of the learned words, uttered by the human. The robot fixates on the object associated with the uttered word.

The graph (figure 2) shows the saliency of objects to the agent over the first part of a typical experimental run. The agent is presented with two objects in its visual field: a red block and a blue block. Initially one can see the fixation times on each object extend as habituation occurs. Finally, fixation times settle into a baseline dynamics. This is displaced when the human experimentor picks up and wiggles one of the boxes to draw the attention of the robot. The human waits until the robot is attending to the desired object (cued by a head movement). This takes several seconds. When he is convinced the robot is attending, he utters the word. He then ceases local motion of the object, allowing the robot to return to its natural baseline dynamics of switching focus between objects in the visual field.

Analysing the interactive dynamics between the human and the robot in this context, we see natural interaction as a human would expect when performing a similar task with a human infant, as intended. The robot moves naturally between several general states in response to cues and internal dynamics: habituating, baseline, when being provoked to attend to an object, and finally when a known word is uttered. We see natural temporal behavior in the fixation times of objects, and also in the lag between the initiation of an action by the teacher and the desired reaction in the robot. The human is also very much a part of this interaction. After he cues the robot by shaking the box, he waits until the behavior of the robot cues him that it is ready, so he can perform the next step in word-teaching. Though we cannot access the human's dynamical state as we can that of the robot, we can assume at the very least that he is in some state before shaking the object. Some change in internal dynamics (e.g. tired of waiting) causes a qualitative change in behavior: he begins to shake the object. Knowing his task, he waits for the learner robot to respond, so he waits and shakes the object until convinced. Becoming convinced represents yet another change in qualitative behavior, and he utters the word (a complex process in itself, involving the production of several sounds in order). Finishing this task, his behavior changes yet again: he returns to the initial do-nothing period (though the exactly dynamical state will have shifted slightly).

Over this whole evolution, the part of the robot's state involving attention is internally evolving (figure 2). This evolution causes changes in behaviors on the part of the robot, from "baseline state" to "focus on red" and then back again. Taken as a single system, the human and the robot engage in what looks to a third party to be natural interaction of adult-infant word-teaching. Indeed, up to the point of the infant being a robot, the system *is* doing so. Not only that, but it is doing so in a way sufficiently faithful to real interactions that the human behaves as if the robot is not a robot. Again, the robot has no explicit representation of time, and we might even postulate that the human does not explicitly reason about time during the interaction either. The evolution of the interaction and behavior of the agents is instead entirely driven by cues and the internal rules driving the two agents. Being unable to find the representation of time in either agent, one might suggest that it is rather the entire coupled system which is keeping track of things temporally, to synchronize the interaction. We hold that such a framing is so far removed from any normal take on the meaning of "representation" or "reasoning" that it does not detract from our conclusions. Indeed, this is exactly the view we wish to endorse — that "reasoning" and "representations" *always* happen at the level of the entire coupled system, and so things are best viewed from that standpoint.

References

Beer, R. 2008. The dynamics of brain-body-environment systems: A status report. In *P. Calvo and A. Gomila (Eds.), Handbook of Cognitive Science: An Embodied Approach*, 99–120.

Kelso, S.; de Guzman, G.; Reveley, C.; and Tognoli, E. 2009. Virtual partner interaction (vpi): Exploring novel behaviors via coordination dynamics. *PLoSONE* 4(6):e5749.

Veale, R.; Schermerhorn, P.; and Scheutz, M. in preparation. A situated embodied model of associative word-referent co-learning. In *ICDL*.

Williams, P.; Beer, R.; and Gasser, M. 2008. Evolving referential communication in embodied dynamical agents. In *Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*.