Infomax Control as a Model of Real Time Behavior: Theory and Application to the Detection of Social Contingency

MPLab TR 2005-1

Javier R. Movellan

University of California San Diego & ATR, Kyoto

Abstract

I present a model of behavior for situations in which organisms react to the environment in a manner that maximizes information gain. I call the approach "Infomax control" for it combines the theory of optimal control with information maximization models of perception. The approach is not cognitivist, in that it is better described as a continuous "dance" of actions and reactions with the world, rather than a turn-taking inferential process like chess-playing. The approach however is intelligent in that it produces behaviors that optimize long-term information gain. I illustrate how Infomax control can be used to understand the detection of social contingency in 10 month old infants. The results suggest that, while lacking language, by this age infants actively "ask questions" to the environment, i.e., schedule their actions in a manner that maximizes the expected information return. A real time Infomax controller was implemented on a humanoid robot to detect people using contingency information. The system worked robustly requiring little bandwidth and computational cost. This suggest that contingency is indeed a reliable source of information to detect the presence of humans and that the infant brain is likely to capitalize on it to solve this task.

Key words: Information Maximization, Contingency Detection, Social Contingency, Control Theory,
PACS:

I present a model of behavior for situations in which organisms behave to maximize the expected information gain about events of interest. I call the approach "Infomax control" for it combines the theory of optimal control with information maximization models of sensory processing in the brain [5, 1]. Contrary

Email address: movellan@mplab.ucsd.edu (Javier R. Movellan).
URL: mplab.ucsd.edu (Javier R. Movellan).

5 May 2005

to cognitivist approaches the system presented here is reactive and designed for continuous real-time interaction with the world, where timing is critical. Indeed one of the goals of the approach is to provide an explanation for why organisms time their behaviors in the way they do. Contrary to classic behaviorist approaches, the Infomax framework views behavior as a goal-oriented activity designed to maximize the gathering of information about events of interest. Under this framework the value of responses and stimuli is measured by the number of bits of information they are expected to provide in the long run. Reinforcement is not an attribute of the stimulus per se but instead it is determined by the internal beliefs and goals of the organism. In this sense Infomax control can be seen as an example of self-supervised learning.

In this document I apply Infomax optimal control to the problem of detecting the presence or absence of responsive human beings. I show that the approach helps understand how infants use contingency information to detect social agents. The approach also works well in a robot implementation that recognizes the presence of people via contingency analysis.

1 Contingency Detection and Social Development

John Watson proposed that contingency detection plays a crucial role in the social and emotional development of infants [13, 14]. In his view contingency is perceived by the human brain in a direct manner, the same way we perceive other primitives like color or motion. In particular he proposed that early in infancy contingency is a fundamental source of information in the definition and recognition of caregivers. This view originated from an experiment [13] in which 2-month-old infants learned to move their heads on a pressure sensitive pillow to activate a mobile above their cribs. After four 10 minute daily sessions of exposure to this controllable mobile, and an average of approximately 200 responses, infants exhibited significantly higher response rates in the contingent group that in a control group in which the mobile moved in a non-contingent manner. More importantly, at about the same time, infants in the contingent group started displaying a number of responses that are typically directed towards caregivers. These included intense social smiles, cooing, and positive affect towards the mobile. Watson proposed that contingency was being used as a cue to define and identify con-specifics and that this cue was more important than other perceptual cues, like the visual features of the human face.

In 1986 [12, 8] Movellan and Watson conducted an experiment to test whether 10 month old infants use contingency information to detect novel social agents. Infants were seated in front of a robot that did not look particularly human. The “head” was a rectangular prism whose sides contained geometric patterns

(See left side of Figure 1). The robot's head could flash lights on its surface, make sounds, and rotate to "face" right or left. Infants were randomly assigned to an experimental group or a control group. In the experimental group the robot was programmed to respond to the environment in a manner that simulated the contingency properties of human beings. Each infant in the control group was matched to an infant in the experimental group and was presented the same temporal distribution of lights, sounds and turns of the central robot as was experienced by his/her matched participant. However, in the control group the robot was not responsive to the infant's behavior or to any other events in the room.

Fig. 1. *Left: Schematic of the robot head used in [12]. Right: Baby-9. The image of the robot is seen reflected on a mirror positioned behind the baby.*

1.1 Forty Three Seconds of an Infant's Day

In that study we found evidence that the infants in the experimental group treated the robot as if it were a social agent: For example, infants in the experimental group exhibited 5 times more vocalizations than infants in the control group. Moreover they followed the "line of regard" of the robot when it rotated, showing some evidence for shared attention [8]. I was however particularly surprised and unprepared for the quality of the interactions that occurred between some infants and the robot, and the speed with which these interactions developed. Some infants appeared to actively "decide" in a few trials, and a matter of seconds whether or not the robot was responsive.

Particularly telling were the first 43 seconds of the experiment with one of the infants in the experimental group. I will refer to him as Baby-9 (see right side of Figure 1). He was 10 months old on 7/14/1986, when the study was run at UC Berkeley's Institute for Human Development. The video of these 43 seconds is available at *http://mplab.ucsd.edu*. During this time Baby-9 produced 7 vocalizations, each of which was followed by a combination of sounds and lights from the robot. Most people that see this video agree that by the third or fourth vocalization (25 seconds into the experiment) the baby has clearly detected the fact that the robot was responsive to him. Thus it took in the order of 20 to 30 seconds and 3 to 4 responses for the infant to decide that the robot was responsive. Most importantly, by watching the video, most people feel that the infant is actively querying the robot to test whether or not it is responding to him. This brings some interesting questions that we will address formally in this document:

(1) What does it mean to "ask questions" for an organism that does not have language?
(2) Was it smart for Baby-9 to schedule his vocalizations in the way he did? Why did he not vocalize, for example, at a much rapid or a much slower rate?
(3) Was it smart for Baby-9 to decide within 3 to four responses and 20-30 seconds into the experiment that the robot was responsive? Why not more or less time and responses?

2 Probabilistic Functionalism

I respectfully believe that the cognitivist revolution is showing its age and that the scientific study of human nature is in dire need for an alternative. By definition cognitivism focuses on the study of cognition, and as such it conceives humans as factories of thoughts and concepts. This view necessarily misses the point about what makes the human condition particularly interesting: love, art, feelings, jazz, sex, emotion, religion, parenthood. The tragic part of this is that while these aspects of human life are critical to understand what it means to be human, cognitivism has relegated them outside the realm of science. Connectionism and the PDP revolution exposed some of the hidden biases and assumptions of cognitivism. However, in my view, it did not go far enough: It proposed alternative solutions to the problems set by the cognitivist agenda but it seldom questioned the agenda itself.

I have been promoting a functionalist framework to the study of human nature that is directly inspired by the work of David Marr, the founder of computational neuroscience [6]. Marr emphasized the need to understand the nature of the problems faced by organisms when operating in the world, propose possible solutions to these problems, and compare these solutions with the solutions found by the brain [4].

Within this framework neuroscience, machine learning, machine perception, and robotics are key tools to the scientific study of human nature. In order to understand how humans are "put together" scientist must attempt to put together systems that operate in everyday life, and learn from the problems encountered when doing so. I also believe that scientific progress requires mathematical formalization and that probability theory is key to this formalization. I intend for this paper to illustrate the functionalist approach I have been trying to promote.

3 Detecting Social Contingency: Causal Model

Notational conventions: *Unless otherwise stated, capital letters are used for random variables, small letters for specific values taken by random variables, and Greek letters for fixed parameters. I leave implicit the probability space* $(\Omega, \mathcal{F}, P)$ *in which the random variables are defined and assume it supports the propositions being made. When the context makes it clear, I identify probability functions by their arguments: e.g.,* $p(x, y)$ *is shorthand for the joint probability mass or joint probability density that the random variable* X *takes the specific value* x *and the random variable* Y *takes the value* y*. I use subscripted colons to indicate sequences: e.g.,* $X_{1:t} \stackrel{def}{=} \{X_1 \cdots X_t\}$*. The symbol* $\sim$ *to indicate the distribution of random variables. For example* $X \sim Poisson(\lambda)$ *indicates that* X *has a Poisson distribution with parameter* λ*. The notation* $Y \in \sigma\{X\}$ *means that the random variable* Y *is measurable by the sigma-algebra induced by the random variable* X*. Intuitively this means that* X *contains all the information needed to determine the value of* Y*. We use* E *for expected value, and Var for covariance matrix. We use* $\delta(\cdot, \cdot)$ *for the Kronecker delta function, which takes value 1 if its two arguments are equal, otherwise it takes value 0.* $\mathcal{N} = \{0, 1, 2, \cdots\}$ *represents the natural numbers,* $\Re$ *the real numbers.*

Our goal is to gain a better understanding of the problem of active contingency detection in simple social interactions between infants and caregivers. These are characterized by the existence of self-feedback (e.g., infants can hear themselves), significant delays and uncertainty in the caregiver's responses, and significant levels of background activity. We will investigate the problem from the point of view of a bare-bones *"social robot"* endowed with a single binary sensor (e.g., a sound detector) and a single binary actuator (See Figure 2). There will be two players: (1) A *social agent*, which plays the role of the caregiver, and (2) A *robot*, which plays the role of the infant. Agent and robot are in an environment which may have random background activity. The role of the robot is to discover as soon as possible and as accurately as possible the presence of *responsive social agents*.

We will develop a discrete-time model of the problem. The parameter $\Delta t \in \Re$ will represent the sampling period, i.e., the time between time steps, in seconds. Choosing a specific value of Δt is equivalent to assuming that the relevant information about social contingency occurs at temporal frequency bands lower than $0.5/\Delta t$ Hertz, i.e., that we do not loose any relevant information by discretizing what is inherently a continuous time process.

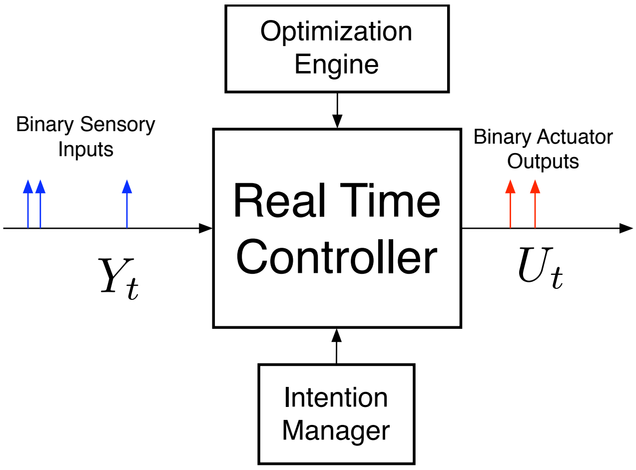

Fig. 2. *A Bare-bones social robot*

The activity of the robot's actuator is represented by the binary random process $\{U_t\}$. The variable U_t takes value 1 when the robot's actuator is active at time t, and zero otherwise. The presence or absence of responsive social agents is indicated by the random variable H. We refer to $\{H = 0\}$, the absence of a responsive agent, as the *" null hypothesis"*, and $\{H = 1\}$, the presence of a responsive agent, as the *"alternative hypothesis"*. The parameter π represents the prior probability of the alternative hypothesis, i.e., the robot's initial belief about the presence of a social agent, prior to the gathering of sensory information.

3.1 Modeling the Social Agent

We want to capture the essence of the behavior of agents in very simple social interactions. The model described below has the advantage of being mathematically tractable while maintaining two essential properties: (1) Different agents have different levels of responsiveness, and (2) Social agents respond with significant delays and levels of uncertainty in these delays.

In our model we will let the behavior of the social agent depend on two auxiliary processes: A timer $\{Z_t\}$ and an indicator $\{I_t\}$. The timer takes values in $\{0, \cdots, \tau_2^a\}$ where $\tau_2^a \in \mathcal{N}$ is a parameter of the model, whose meaning will be explained below. The timer keeps track, up to τ_2^a, of the number of time steps since the last robot action (See Figure 3), i.e.,

$$Z_1 = \tau_2^a + 1 \tag{1}$$

$$Z_t = h(Z_{t-1}, U_t) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } U_t = 1 \\ Z_{t-1} & \text{if } U_t = 0 \text{ and } Z_{t-1} = \tau_2^a + 1 \\ 1 + Z_{t-1} & \text{else} \end{cases} \tag{2}$$

for $t = 2, 3, \cdots$ The indicator vector $I_t = (I_{1,t}, I_{2,t}, I_{3,t})^T$ consists of three binary variables that indicate whether or not time t belongs to the following categories: (1) *"Self Period"*, indicated by $I_{1,t}$; (2) *"Agent Period"*, indicated by $I_{2,t}$, and (3) *"Background Period"*, indicated by $I_{3,t}$. The meaning of these three periods is explained below.

The reaction times of social agents is bounded by the parameters $0 \leq \tau_1^a \leq \tau_2^a$, i.e., it takes agents anything from τ_1^a to τ_2^a time steps to respond to an action from the robot. "Agent periods", which are designated by the indicator process $\{I_{2,t}\}$ are periods of time for which responses of agents to previous robot actions are possible if an agent were to be present. Thus,

$$I_{2,t} = \begin{cases} 1 & \text{if } Z_t \in [\tau_1^a, \tau_2^a] \\ 0 & \text{else} \end{cases} \tag{3}$$

During agent periods, the robot's sensor is driven by the Poisson process $\{D_{2,t}\}$ which has rate R_2. The distribution of R_2 depends on whether or not a responsive agent is present in a manner that will be specified below.

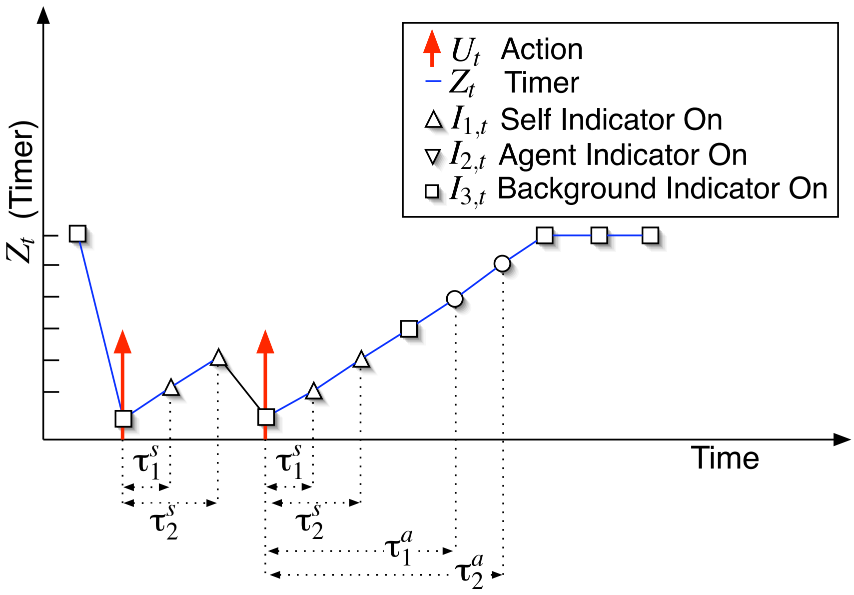

Fig. 3. *Graphical representation of the dynamics of the timer and the indicator variables. The time delay parameters were* $\tau_1^s = 1, \tau_2^s = 2, \tau_1^a = 4, \tau_2^a = 5$.

3.2 Modeling Self-Feedback and Background Processes

We allow for the robot sensor to respond to the robot actuators, e.g., the robot can hear its own vocalizations, and allow for delays and uncertainty in this self-feedback loop. In particular we let the distribution of self-feedback reaction time be uniform with parameters $\tau_1^s \leq \tau_2^s$, where $\tau_1^a > \tau_2^s$. The indicator variable for self-feedback period is thus defined as follows:

$$I_{1,t} = \begin{cases} 1 & \text{if } Z_t \in [\tau_1^s, \tau_2^s] \\ 0 & \text{else} \end{cases} \tag{4}$$

During Self periods, the activation of the sensor is driven by the Poisson process $\{D_{1,t}\}$ with rate R_1.

With regard to the background, we model it as a Poisson process $\{D_{3,t}\}$ with rate R_3. The background process is responsible for driving the sensor's activity that is not due to self-feedback and is not due to social agent responses to the robot's behaviors. Note this can include, among other things, the actions from external social agents who are not responding to the robot (e.g., two social agents may be talking to each other thus activating the robot's sound sensor). We endow the background rate R_3 with an uninformative prior Beta distribution. This reflects the fact that the background activity may change dramatically from situation to situation in ways that are not known to the robot:

$$R_3 \sim \text{Beta}(1,1) \tag{5}$$

The background indicator keeps track of periods for which self-feedback or responsive actions from a social agent may not happen, i.e.,

$$I_{3,t} = (1 - I_{1,t})(1 - I_{2,t}) \tag{6}$$

3.3 Modeling the Robot's Sensor

The activity of the sensor is a switched Poisson process: during self-feedback periods it is driven by the Poisson process $\{D_{1,t}\}$, during agent periods it is driven by $\{D_{2,t}\}$ and during background periods it is driven by $\{D_{3,t}\}$, i.e.,

$$Y_t = I_t \cdot D_t = \sum_{i=1}^{3} I_{i,t}\, D_{i,t} \tag{7}$$

We still need to specify the distribution of the response rate R_2 during agent periods. If an agent is present, i.e., $H = 1$, we let R_2 be independent of R_1 and R_3 and endow it with an uninformative Beta prior distribution. This reflects the fact that different agents respond at different rates in ways that the robot does not know apriori. If an agent is not present, i.e., $H = 0$, then the response rate during agent periods is not different from the response rate during background periods, i.e., $R_2 = R_3$.

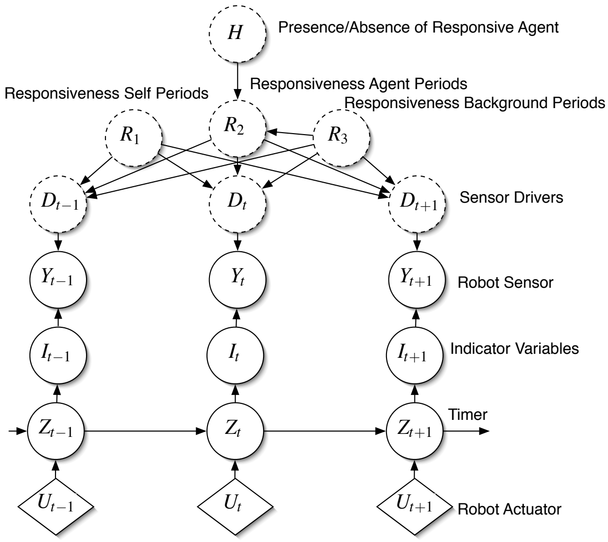

Fig. 4. *Graphical Representation of the Causal Model. Arrows represent dependency relationships between variables. Dotted figures indicate unobservable variables, continuous figures indicate observable variables. Diamonds indicate control variables.*

3.4 Auxiliary Processes

We will use the processes $\{O_t, Q_t\}$ to register the sensor activity and lack-thereof up to time t during self, agent and background periods. In particular for $t = 1, 2, \cdots$

$$O_{i,t} = \sum_{s=1}^{t} I_{i,t} Y_t, \text{ for } i = 1, 2, 3 \tag{8}$$

$$Q_{i,t} = \sum_{s=1}^{t} I_{i,t}(1 - Y_t), \text{ for } i = 1, 2, 3 \tag{9}$$

3.5 Probabilistic Constraints

Appendix I contains a summary of the parameters, random variables, and stochastic processes that specify the model. Figure 4 display Markovian constraints in the joint distribution of the different variables involved in the model. An arrow from variable X to variable Y indicates that X is a “parent” of Y. The probability of a random variable is conditionally independent of all the other variables given the parent variables. Dotted figures indicate unobservable variables, continuous figures indicate observable variables. Diamonds indicate control variables.

4 Development and Learning. Inference, and Control

I will refer to *"development"* as the problem of discovering the causal structures and parameter underlying social interaction, i.e., discovering a model of the kind displayed in Figure 4. This is a difficult problem that may require large amounts of data gathered over months or years. I will refer to *"learning"* as the problem of discovering contingencies, i.e., making inferences about unobservable variables of a given model. This is a process that in general requires less data than model development and may occur within seconds, minutes or hours.

Development and learning rely on two basic processes: inference and control. Inference refers to the problem of combining prior information with sensor data in a principled manner. Control refers to the problem of scheduling the behavior in real time to achieve the goals of the organism.

4.1 Development

In practice the model we have developed so far simply says that when interacting with the world the robot may encounter two "causal clusters" (See Figure 5):

- When in "Cluster 1" the sensor activity tends to change, with respect to the background activity, during the period $[\tau_1^s, , \tau_s^2]$ following an action. This is due to the effect of self-feedback.
- When in "Cluster 2", the sensor activity tends to change during the period $[\tau_1^s, , \tau_2^s]$ but it also changes during the period $[\tau_1^a, \tau_2^a]$ following an action. The second change in activity is due to the presence of a responsive social agent. Again we let the particular levels of background and agent related activity change from situation to situation.

While in this paper we developed the causal model by hand, I believe current unsupervised learning methods could be used for data-driven discovery of causal clusters, like the ones implicit in our model. We will tackle the model development problem in future documents, and here we will simply assume that after 10 months of interaction with the social world, infants have discovered, one way or another, these causal clusters.

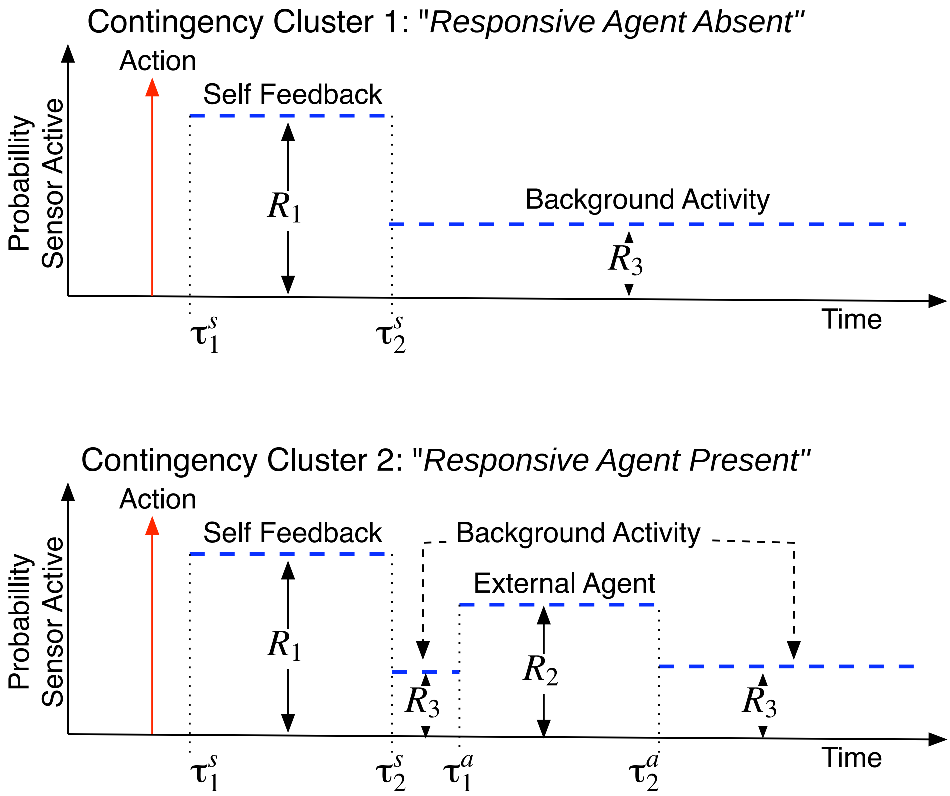

Fig. 5. *Illustration of two contingency clusters produced by the model. The variable H indicates which of the two clusters is active in the current situation.*

4.2 Learning: Inference

In this section we assume the robot has already developed a causal model and focus on how to make inferences about the presence or absence of a social agent based on a given sequence of sensor activities $y_{1:t}$ and actions $u_{1:t}$. Let $(y_{1:t}, u_{1:t}, o_t, q_t, z_t)$ be an arbitrary sample from $(Y_{1:t}, U_{1:t}, O_t, Q_t, Z_t)$. Then

$$p(y_{1:t} \mid r, u_{1:t}, h) = \prod_{i=1}^{3} (r_i)^{o_{i,t}} (1 - r_i)^{q_{i,t}} \tag{10}$$

Note the rate variables R_1, R_2, R_3 are independent under the prior distribution. Moreover if $H = 1$ they affect the sensor at non intersecting sets of time. It follows that the rate variables are also independent under the posterior distribution. In particular

$$p(r \mid y_{1:t}, u_{1:t}, H = 1) = \prod_{i=1}^{3} \text{Beta}(r_i; 1 + o_{i,t}, 1 + q_{i,t}) \tag{11}$$

Under the null hypothesis $R_2 = R_3$, i.e., the sensor activity does not change during the "agent periods". Moreover the set of times for which the sensor's activity depends on R_2, R_3 does not intersect with the set of times for which it depends on R_1. Thus R_1 will be independent of R_2, R_3 under the posterior distribution:

$$p(r \mid y_{1:t}, u_{1:t}, H = 1) = \text{Beta}(r_1; 1 + o_{1,t}, 1 + q_{1,t}) \\ \text{Beta}(r_2; 1 + o_{2,t} + o_{3,t}, 1 + q_{2,t} + q_{3,t})\, \delta(r_2, r_3) \tag{12}$$

Note

$$p(y_{1:t} \mid u_{1:t}, h) = p(r \mid u_{1:t}, h)\, \frac{p(y_{1:t} \mid r, u_{o:t}, h)}{p(r \mid y_{1:t}, u_{1:t}, h)} \tag{13}$$

Thus

$$p(y_{1:t} \mid u_{1:t}, H = 1) = \prod_{i=1}^{3} \frac{\text{Beta}(r_i; 1, 1)\ (r_i)^{o_{i,t}} (1 - r_i)^{q_{i,t}}}{\text{Beta}(r_i; 1 + o_{1,t}, 1 + q_{i,t})} \tag{14}$$

$$= \prod_{i=1}^{3} \frac{\Gamma(1 + o_{i,t})\, \Gamma(1 + q_{i,t})}{\Gamma(2 + o_{i,t} + q_{i,t})} \tag{15}$$

and

$$p(y_{1:t} \mid u_{1:t}, H = 0) = \frac{\left(\prod_{i=1}^{2} \text{Beta}(r_i; 1, 1)\right)\ \prod_{i=1}^{3} (r_i)^{o_{i,t}} (1 - r_i)^{q_{i,t}}}{\text{Beta}(r_1; 1 + o_{1,t}, 1 + q_{1,t}) \text{Beta}(r_2; 1 + o_{2,t} + o_{3,t}, 1 + q_{2,t} + q_{3,t})} \tag{16}$$

$$= \frac{\Gamma(1 + o_{1,t})\, \Gamma(1 + q_{1,t})}{\Gamma(2 + o_{1,t} + q_{1,t})} \frac{\Gamma(1 + o_{2,t} + o_{3,t})\, \Gamma(1 + q_{2,t} + q_{3,t})}{\Gamma(2 + o_{2,t} + o_{3,t} + q_{2,t} + q_{3,t})} \tag{17}$$

where we used the fact that $r_2 = r_3$ with probability one under $H = 0$. Thus the log-likelihood ratio between the two hypothesis is as follows:

$$\log \frac{p(y_{1:t} \mid u_{1:t}, H = 1)}{p(y_{1:t} \mid u_{1:t}, H = 0)} = \log \frac{\Gamma(2 + o_{2,t} + o_{3,t} + q_{2,t} + q_{3,t})}{\Gamma(1 + o_{2,t} + o_{3,t})\, \Gamma(1 + q_{2,t} + o_{3,t})} \tag{18}$$

$$+ \sum_{i=2}^{3} \log \frac{\Gamma(1 + o_{i,t})\, \Gamma(1 + q_{i,t})}{\Gamma(2 + o_{i,t} + q_{i,t})} \stackrel{\text{def}}{=} f(o_{2,t}, o_{3,t}, q_{2,t}, q_{3,t}) \tag{19}$$

and the posterior distribution about the hypothesis of interest is as follows:

$$p(H = 1 \mid y_{1:t}, u_{1:t}) = \text{logistic}\left(\log \frac{\pi}{1 - \pi} + f(o_{2,t}, o_{3,t}, q_{2,t}, q_{3,t})\right) \tag{20}$$

This posterior distribution, contains all the information available to the robot about the presence of a responsive agent. It has two important properties: (1) It does not depend on $o_{1,t}, q_{1,t}$, i.e., the self-periods are uninformative about the hypothesis, and (2) If $o_{1,t} + q_{1,t} = 0$ or $o_{2,t} + q_{2,t} = 0$ the log-likelihood ratio is 0. In other words, if no data has been gathered in either the agent or the background condition then we have gained no information about H. Thus in order to gain information about H the robot must use its actuator at least once and not use it at least once.

4.3 Predictive Distributions

The predictive distribution of the sensor, which will be useful for the control problem presented in the next section, has the following form:

$$p(Y_{t+1} = 1 \mid y_{1:t}, u_{1:t+1}) = \sum_h p(h \mid o_t, q_t)\, p(Y_{t+1} = 1 \mid y_{1:t}, u_{1:t+1}, h) \tag{21}$$

$$= \sum_h p(h \mid o_t, q_t)\, \mathrm{E}[R \cdot I_{t+1} \mid y_{1:t}, u_{1:t+1}, h] \tag{22}$$

Note I_t is a 3 dimensional vector with two 0s and one 1. The position of the 1 is determined by Z_t. Thus $\mathrm{E}[R \cdot I_{t+1} \mid o_t, q_t, z_t, u_{t+1}, h]$ will the expected value of the rate of the Poisson process picked by z_t. From (12) we get that for $i = 1, 2, 3$

$$p(r_i \mid y_{1:t}, u_{1:t}, H = 1) = \mathrm{Beta}(r_i \mid 1 + o_{i,t}, 1 + q_{i,t}) \tag{23}$$

and therefore, using (69)

$$\mathrm{E}[R_i \mid y_{1:t}, u_{1:t+1}, H = 1] = \frac{1 + o_{i,t}}{2 + o_{i,t} + q_{i,t}} \tag{24}$$

$$\mathrm{E}[R_1 \mid y_{1:t}, u_{1:t+1}, H = 0] = \frac{1 + o_{1,t}}{2 + o_{1,t} + q_{1,t}} \tag{25}$$

$$\mathrm{E}[R_2 \mid y_{1:t}, u_{1:t+1}, H = 0] = \mathrm{E}[R_3 \mid y_{1:t}, u_{1:t+1}, H = 0] \tag{26}$$

$$= \frac{1 + o_{2,t} + o_{3,t}}{2 + o_{2,t} + o_{3,t} + q_{2,t} + q_{t,e}}$$

This shows that the predictive distribution is a function of (o_t, q_t, z_t, u_{t+1})

4.4 Learning: Infomax Control

In this section we focus on how to schedule the behavior of the robot's sensor in real time in order to maximize the information received about the presence or absence of social agents. Let t represent the current time and suppose by time t we have observed $y_{1:t}, u_{1:t}$. For a future time $s > t$ let us consider the mutual information between the observable variables and the hypothesis of interest H

$$\mathrm{I}(H, Y_{t+1,s}, U_{t+1:s} \mid y_{1:t}, u_{1:t}) \tag{27}$$

$$= \mathcal{H}(H \mid y_{1:t}, u_{1:t}) - \mathcal{H}(H \mid Y_{t+1:s}, U_{t+1:s}, y_{1:t}, u_{1:t}) \tag{28}$$

where $\mathcal{H}$ stands for entropy. The equation tells us that the information about H provided by the observable processes $Y_{t+1:s}, U_{t+1:s}$ equals the reduction of uncertainty about H provided by those observables. Since the term $\mathcal{H}(H|y_{1:t}, u_{1:t})$

does not depend on future actions then maximizing the information return provided by future actions is equivalent to minimizing the future entropy of H. We will define W_s as the negative entropy of H, the certainty about the state of interest.

$$W_s \stackrel{\text{def}}{=} -\mathcal{H}(H \mid Y_{t+1:s}, U_{t+1:s}, y_{1:t}, u_{1:t}) \tag{29}$$

An Infomax controller is an open-loop controller that maximizes the expected value of W at future times $t+1, \cdots, T$. An open-loop controller $\{C_t\}$ is a collection of functions that map sequences of observations into actions, i.e.,

$$U_{t+1} = C_{t+1}(Y_{1:t}, U_{1:t}) \tag{30}$$

Let the *expected return* of the the observed sequence $y_{1:t}, u_{1:t}$ given a controller $c_{t+1:T}$ be defined as follows:

$$V_t(y_{1:t}, u_{1:t} \mid c_{t+1:T}) \stackrel{\text{def}}{=} \sum_{s=t+1}^{T} = \mathrm{E}[W_s \mid y_{1:t}, u_{1:t}, c_{t+1:s}] \tag{31}$$

Our goal is to find a controller, $\widehat{c}_{t+1:T}$, that maximizes the expected return

$$\widehat{c}_{t+1:T}(y_{1:s}, u_{1:s}) \stackrel{\text{def}}{=} \operatorname*{argmax}_{c_{t+1:T}} V_t(y_{1:t}, u_{1:t} \mid c_{t+1:T}) \tag{32}$$

We define the optimal expected return as the expected return given an optimal controller

$$V_t(y_{1:t}, u_{1:t}) \stackrel{\text{def}}{=} V_t(y_{1:t}, u_{1:t} \mid \widehat{c}_{t+1:T}) \tag{33}$$

4.4.1 Bellman's Optimality Equation:

The causal model we are working with belongs to the family of partially observable Markov processes. Finding optimal open-loop controllers for these processes is in general difficult. In this case however the problem simplifies because it is possible to find a recursive statistic that summarizes the observable sequences without any loss of information about H. In particular we used the summary statistic $S_t \stackrel{\text{def}}{=} (O_t, Q_t, Z_t)$. It can be verified that the following conditions are satisfied:

$$p(y_{t+1} \mid y_{1:t}, u_{1:t+1}) = p(y_{t+1} \mid s_t, u + t + 1) \tag{34}$$

$$p(h \mid y_{1:t}, u_{1:t}) = p(h \mid s_t) \tag{35}$$

$$S_{t+1} \in \sigma\{S_t, Y_{t+1}, U_{t+1}\} \tag{36}$$

$$W_{t+1} \in \sigma\{S_{t+1}\} \tag{37}$$

Given these conditions it is possible to show that the optimal controller satisfies the following form of Bellman's optimality equation [7]:

$$C(y_{1:t}, u_{1:t}) = C'(s_t) \stackrel{\text{def}}{=} \operatorname*{argmax}_{u_{t+1}} N_t(s_t, u_{t+1}) + F_t(s_t, u_{t+1}) \tag{38}$$

$$V_t(y_{1:t}, u_{1:t}) = V'(s_t) \stackrel{\text{def}}{=} \max_{u_{t+1}} N_t(s_t, u_{t+1}) + F_t(s_t, u_{t+1}) \tag{39}$$

where C' is a controller that makes decisions based on s_t, and V' is the value of s_t

$$N_t(s_t, u_{t+1}) \stackrel{\text{def}}{=} \underbrace{\mathrm{E}\left[W_{t+1} \mid s_t, u_{t+1}\right]}_{\text{Next Step Expected Return}} \tag{40}$$

$$F_t(s_t, u_{t+1}) \stackrel{\text{def}}{=} \underbrace{\mathrm{E}[V_{t+1}(y_{1:t}, Y_{t+1}, u_{1:t+1}) \mid y_{1:t}, u_{1:t+1}]}_{\text{Future Expected Return}} \tag{41}$$

This equation can be solved using dynamic programming techniques [2, 7].

5 Analysis of the the optimal controller

The dynamic programming problem was solved on a cluster of 24 2.5Ghz PowerPC G5 CPUs. The computation time was in the order of 1 hour. The parameters of the model were set as follows: $T = 40, \tau_1^s = 0; \tau_2^s = 0; \tau_1^a = 1; \tau_2^a = 3; \pi = 0.5$. I then used logistic regression to model the behavior of the controller for times $15 < t < 25$, since these are times which are not too close to the beginning and end of the controller's window of interest, i.e., $t \in [1, 40]$. I did not expect for logistic regression to provide a perfect prediction since in some cases the value function is equal for both actions and in such occasions the optimal controller may arbitrarily chooses one action over the other. Surprisingly logistic regression approximated the optimal controller with 96.46 % accuracy over all possible conditions. The final model was as follows:

$$\hat{u}_t = \begin{cases} 1 & \text{if } I_{3,t} = 1 \text{ and } \frac{\mathrm{Var}(R_3 \mid y_{1:t}, u_{1:t}, H_t = 1)}{o_{3,t} + q_{3,t} + 3} > 9 \, \frac{\mathrm{Var}(R_2 \mid y_{1:t}, u_{1:t}, H_t = 1)}{o_{2,t} + q_{2,t} + 3} \\ 0 & \text{else} \end{cases} \tag{42}$$

Interpretation: While the derivation of the optimal controller was somewhat arduous the final product ends up being a simple reactive system that can easily operate in real time. What the derivations provided was a guarantee that this simple controller is optimal for the task at hand. No other control policy is better under the model. Note that "greedy" one-step controllers [10, 9] that ignore the future expected return would fail on this task. The reason is that when making a response the next time steps are occupied by self-feedback, that happens to be uninformative, thus a greedy controller ends up deciding to never act. Including future expected return allows the controller to implicitly look ahead and see that in the long run making an action can provide a better information return than being inactive.

The statistic

$$\frac{\mathrm{Var}(R_i \mid y_{1:t}, u_{1:t}, H_t = 1)}{o_{i,t} + q_{i,t} + 3} \tag{43}$$

is used by the controller to decide when to act. It is the expected reduction in the uncertainty about R_i provided by a new observation from the period under which R_i actively drives the sensor: a self-feedback period for R_1, an agent period for R_2 and a background period for R_3. The optimal controller thus appears to want to keep the uncertainty about R_3 and about R_2 within a fixed ratio. If R_2, the agent rate, is too uncertain, then the controller chooses to act. If R_3, the background rate, is too uncertain then the controller chooses to remain silent, thus gaining information about the background activity rate. It is interesting to note that actions occur when the variance about the background rate R_3 is at least 9 times larger than the variance about the agent rate R_2. The reason for this particular ratio may be due to the fact that actions are more costly, in terms of information return, than lack of action. If the robot acts at time t it gains no information during the times $[t+\tau_1^s, t+\tau_2^s]$ since self-feedback observations are not informative about H. Moreover during times $[t+\tau_1^a, t+\tau_2^a]$ the robot cannot act and thus can only get information about R_2, not R_3. By contrast if the robot does not act at time t no time will be wasted on self-feedback. Moreover the controller can still choose to act or not to act in the future without constraints. This may help explain why uncertainty about the agent activity rate R_2 needs to be 9 times larger than the uncertainty about the background activity rate, R_3, before an action occurs.

6 Understanding 43 Seconds of an Infant's Day

Here we examine whether the Infomax controller can provide a qualitative understanding of the first 43 seconds of the experimental session with Baby-9, as described in Section 1.1. During this time Baby-9 produced 7 vocalizations, which occurred at the following times in seconds from the start of the experiment: $\{5.58, 9.44, 20.12, 25.56, 32.1, 37.9, 41.7\}$. Each of these vocalizations were followed by a combination of sounds and lights from the robot. The intervals, in milliseconds, between the beginning of two consecutive infant vocalizations were as follows: $\{4.22, 10.32, 5.32, 6.14, 5.44, 3.56\}$. Most people agree that by the third or 4th vocalization the infant knows that there is a responsive agent in the room.

The Infomax control model presented in Section 4.4 requires setting five pa-

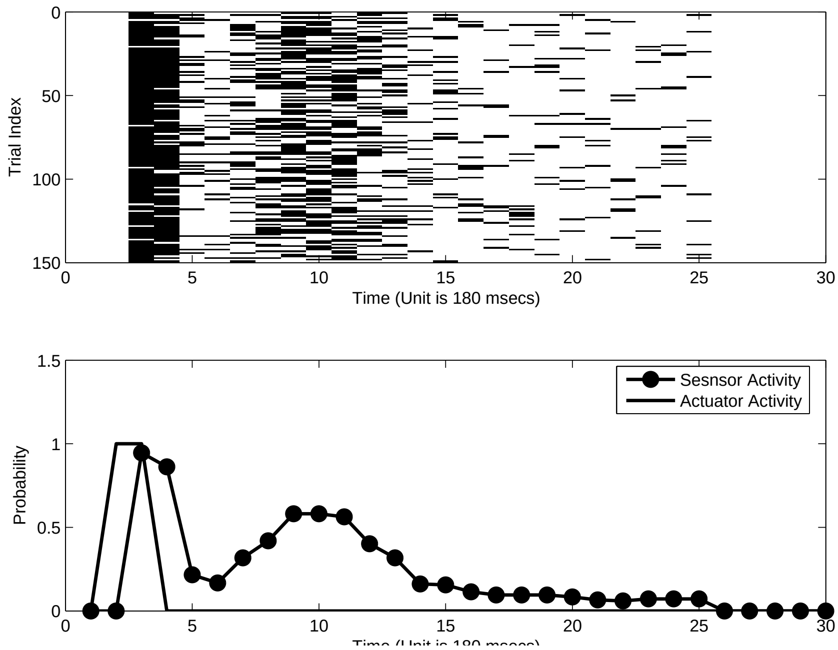

Fig. 6. *Above: Raster plot of 150 trials. On each trial an animated character made a sound and subjects were asked to talk back to the character and let it know that they were listening. Dark indicates that the audio sensor was active. Below: Probability of the audio sensor being active as a function of time. The probabilities is estimated by averaging across the 150 trials in the figure above.*

rameters: The sampling period for the time discretization, the self-delay parameters, and the agent delay parameters. We run an experiment to get rough estimates of these parameters. For the agent latency parameters τ_1^a, τ_2^a, we asked 4 people, unaware of the purpose of the study, to talk to a computer animated character. The ages of the 4 participants were $4, 6, 24$ and 35 years. We used an optimal encoder to binarize the activity of an auditory sensor and plotted the probability of activation of this binary sensor as a function of time over 150 trials. Each trial started with a vocalization of the animated character and ended 4 seconds later. The results are displayed on Figure 6. The top graph in the figure shows the activity of the acoustic sensor as a function of time from the beginning of the character's vocalization over 150 trials. Each horizontal line is a different trial. The first vertical bar is due to self-feedback from the character. By about 1200 to 1440 msec after the end of the vocalization from the animated character there is another peak of activity in the sensor, which is now caused by the vocalizations of the human participants. The lower graph of the Figure shows the probability of sensor activity as a function of time collapsed across trials. Note the first peak in activity due to self-feedback, and the gradual raise and fall in sensor activity due to the human response. Based on this graph I run a simulation of the optimal controller with the following parameters: $\Delta t = 800$ msec, $\tau_1^s = \tau_2^s = 0$, $\tau_1^a = 1; \tau_2^a = 3$. In other words, we let self-delay to be negligible with respect to the expected delays in human responses, and we bracket the human activity to occur within 800 to 2400 milliseconds. I let $\pi = 0.01$ to simulate a worst case scenario, thus requiring more data to decide that there is a responsive system.

Figure 7 shows the results of the simulation. The horizontal axis in all the graphs is time, measured in seconds. The top graph shows the vocalizations of the optimal controller, which now plays the role of Baby-9. The controller produced 6 vocalizations over a period of 43 seconds. The average interval between vocalizations was 5.92 seconds, compared to 5.833 secs for Baby-9. The difference is not significant using a standard T-test ($T(9) = 0.08$, $p = 0.94$).

The second graph from the top of Figure 7 shows the system's belief's about the presence of a responsive agent. By the fourth response, thirty seconds into the experiment, this probability passes the 0.5 level. The third graph shows the posterior probability distributions about the the agent and background response rates by the end of the 43 second period. Finally the last graph shows the ratio between the uncertainty about the sensor rate during agent periods and the rate during background periods. Note when this ratio reaches the value of 9, the simulated baby makes a response.

The model thus shows that Baby-9 scheduled his responses and made decisions about the responsiveness of social agents in an optimal manner, given the statistics of times delays and levels of uncertainty typically found in social interactions. The model also is consistent with the idea that Baby-9 was "asking questions" to the robot, in the sense that his vocalizations were scheduled in a manner that maximized the information returned about the responsiveness of the robot. Another point of interest is that the optimal controller exhibits turn-taking, i.e., after an action is produced the controller waits for a period of time, an average of 5.92 seconds, before vocalizing again. The period between vocalizations is not fixed and depends on the relative uncertainty about the levels of responsiveness of the agent and the background. For example, if unexpected background activity occurs, the controller automatically increases the period between vocalizations to better "understand" the changes in background activity. If unexpected agent activity occurs, the controller increases the response rate accelerating the gathering of information about agent periods.

7 Real Time Robot Implementation

One of the primary motivations for the current work was to investigate whether or not contingency can be used as a reliable source of information in robots designed to interact with humans. It is our feeling that in order to understand

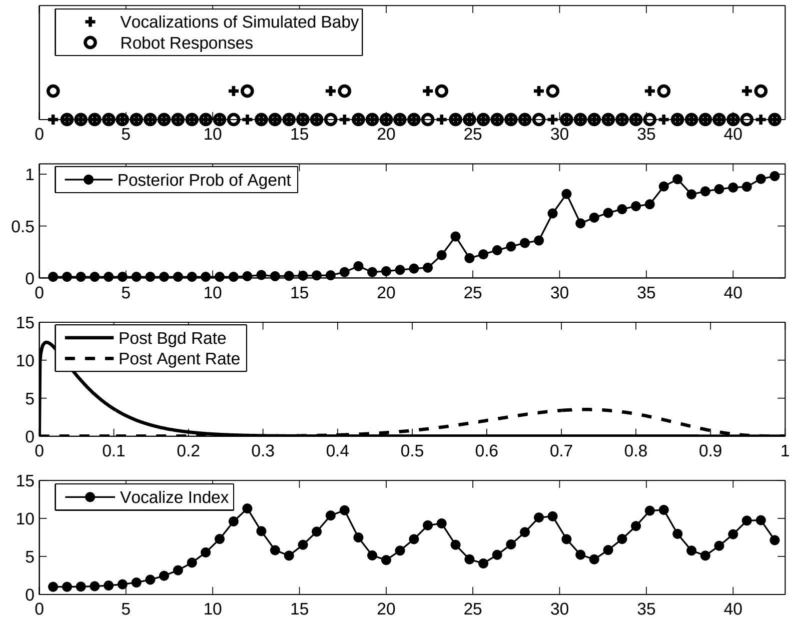

Fig. 7. *The horizontal axis represents time in seconds. From top to bottom: (1) Responses of the Infomax controller (which simulates a baby). Note the environment was silent and the controller was responded to every time it vocalized. (2) Posterior probability for the presence of a responsive agent as a function of time. (3) Posterior distribution for the agent and background rates after 43 seconds. (4) Ratio of the uncertainty about the agent's response rate vs the uncertainty about the background's response rate.*

how humans are "put together" scientists must attempt to put together systems that operate in everyday life, and learn from the problems encountered when doing so. Based on experimentation with infants, John Watson proposed that contingency is a primal source of information in early social interaction but, to our knowledge, no attempt has been made to test whether contingency is indeed a reliable source of information in real life environments.

To investigate this issue I implemented the optimal Infomax controller developed above on RobovieM, a humanoid robot developed at ATR's Intelligent Robotics laboratory. While the robot was not strictly necessary to test the real time controller, it greatly helped improve the quality of the interactions developed between humans and machine thus providing a more realistic way for testing the controller. RobovieM has 22 degrees of freedom (1 dof shoulder, 1 dof waist, 2×4 dof arms, and 2×6 dof legs). It's height is 29 cm, and it weights approximately 1.9 Kg. Control of the corresponding 22 servos is handled by an H8 16MHz CPU and a NiMH 6V battery, 2000mAh. The real time Infomax controller was implemented in Java and run on a host computer, a Mac PowerBook G4 which displayed graphically different states of the controller in real time, e.g., the posterior distribution of the different hidden variables. Communication between the host computer and the controller was handled in a wireless manner using a blue-tooth to serial adapter from Wireless Cables Inc. The current version of the Infomax controller requires a 1 bit sensor and a 1 bit actuator. For sensor we chose to average acoustic energy over 500 msec windows and discretize it using a 1 bit optimal coder. The actuator was a small loudspeaker producing a 200 msec robotic sound. The self-time delay parameters of the controller were chosen by measuring the time delay between issuing a command to produce a sound and receiving feedback from the audio sensor. The agent delay parameters were chosen by asking people to respond to the robot sounds and finding an interval that bracketed 95 % of the vocalizations.

In addition to the robot's vocalizations its posture changed based on the controllers's belief on the presence/absence of a responsive agent: a posture that indicated a high level of attention when the controller believed an agent was present, and a posture that indicated boredom when the it believed an agent was not present.

7.1 Non-Stationary Environments

In the model presented here the states of the agent and background, which are represented by the variables R and H are random but stationary. For realistic implementations we need for H and R to be able to change over time. Unfortunately in such case computation of the optimal controller can be shown to be intractable. We approximate the situation by assuming that past observations become irrelevant as a function of time in an exponential manner. Under this approximation we simply collect an exponentially smoothed running average of O_t, Q_t and apply the standard controller on these running averages. The time constant for the exponential smoother was 30 seconds, reflecting the idea that one should not expect for the situation to be stationary beyond 30 seconds.

7.2 Qualitative Evaluation

Our goal was to explore whether or not contingency can be used as a reliable source of information. The answer is a resounding "yes". It is reliable and it has very low requirements in terms of computation and bandwidth. Lacking quantitative evaluations we will present a qualitative evaluation based on our experience demonstrating the system at public gatherings. In standard office environments, with relatively high levels of noise, the controller decides in a few trials whether or not a responsive agent is present. Particularly effective are transition points in which agents switch from talking to the robot to talking to somebody else. The robot detects quite reliably this fact within a few seconds and with a minimum computational cost. I have demonstrated the system at 4 scientific talks, and at two conferences: ICDL04 and NIPS04. Demonstrations at talks, which generally have relatively low noise levels work well. At ICDL04 the poster room was quite noisy and it took a bit longer for the controller to make reliable decisions. Overall the the level of performance was remarkable considering the difficulty of the situation. At NIPS04 the conditions were extremely noisy. Talking loud in many cases was not sufficient to understand each other. Under these conditions for the controller to work reliably humans had to talk loud and stay close to the robot.

7.3 Limitations and Extensions

The main practical limitations of the current system are caused by the simplicity of the social agent's model. In particular the current system describes agents as passive responders but not as autonomous initiators of actions with communicative intent. Extensions of the model to take care of this problem are relatively straightforward. However, rather than handcrafting better models of social agency, it would be useful to learn such models from data. This is part of the "development" problem presented in Section 4.1 which I hope to tackle in future work.

In addition to scheduling his responses in an optimal manner, Baby-9 showed a progressive increment in the tone and affective quality of his responses over the 43 seconds of interaction simulated in this paper. It is possible to model this expression by linking the tone to the changes in belief about the presence of a social agent. While this modification is effective in improving the interaction between the robot model and humans, it did not emerge from the current model in a principled manner, the way turn-taking emerged, for example.

While Baby-9 behaved in an optimal manner with respect of learning whether or not a novel social agent was responsive, most of the infants in the experiment did not do so. The subjective feeling one gets when watching these infants is that they are initially apprehensive of the situation.

8 Conclusions: Do neurons ask questions?

I presented a general approach to the organization of behavior based on the idea that organisms are fundamentally goal driven and that they schedule their behaviors in a manner that optimizes the gathering of information related to their goals. The approach is not cognitivist, in that it is better described as a continuous "dance" of actions and reactions with the world, rather than a turn-taking inferential process. The approach however is intelligent in that it produces behaviors that optimize the long-term gathering of information. The approach does not fit the mold of standard reinforcement learning approaches either. Classical and instrumental learning models emphasize the role of external stimulus as reinforcers of behaviors (food, water, disconforting air-puffs, and mild electric shocks being the most typical ones). In Infomax control however, stimulus and responses do not have intrinsic value. One can think of Infomax as a self-supervised form of control in which the organism itself assigns reinforcement value to stimuli and responses in a dynamic manner. No external reinforcer is required. Instead, Infomax controllers modify their internal states to better explain the available data and produce actions that are expected to provide highly informative data.

In this paper I used the ideas of Infomax control to understand the detection of social contingency in 10 month old infants. Interestingly when the controller makes a response it follows it by a period of silence, as if waiting for the outcome of a question. This "turn-taking" behavior was not built onto the system. Instead it emerged from the requirement to maximize information gain given the time delays and levels of uncertainty typical of social interactions. The results suggest that, in spite of lacking language, some infants may be actively asking questions to humans and to other aspects of the environment, scheduling their actions in a manner that maximizes the expected information return. This is something all parents know at an intuitive level but that is hard to prove formally. The model presented here represents a first step to address this type of questions. The approach works well in practice when applied in robots that need to operate in real time in everyday life situations. This provides credibility to the idea that contingency is a useful and computationally inexpensive source of information and gives credibility to the idea that the infant brain is likely to use contingency for defining and detecting con-specifics.

Due to the mathematical grounding of Infomax control on probability and control theory it can be extended to other situations in a principled manner. One could, for example, extend the current analysis to rats, neurons, or even molecules. Current Infomax models of neural activity cast neurons as passive information relays, i.e., the role of neural responses is to transmit as much information as possible about the information they receive. Infomax control suggests the intriguing possibility that neurons may also "ask questions", i.e., that their spikes may be designed to gather information about other neurons, not just to transmit information to other neurons.

I intend for this document to illustrate a general approach to the study of behavior inspired on David Marr, the founder of computational neuroscience [6, 4]. The goal is to avoid the scholastic debates that have characterized the cognitivist era and instead to focus on understanding the problems that organisms need to solve when operating in the real world.

9 Acknowledgements

This research was sponsored by a UC Discovery Grant, and a grant from National Association for Autism Research. The optimal controller was computed using a Cluster funded by NSF grant IIS0223052. The real time robot implementation of the model was conducted by Javier Movellan at ATR's Intelligent Robotics and Communication Laboratory, as part of an "Innovational Interaction Media," project with funding from the National Institute of Information and Communications Technology (NICT) of Japan.

References

[1] T. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[2] R. E. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, New Jersey, 1957.

[3] Antonio R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain.* Avon Books, NY, 1994.

[4] S. Edleman and L. M. Vaina. David marr. *International Encyclopedia of the Social and Behavioral Sciences*, 2001.

[5] R. Linsker. Self-organization in a perceptual network. *Computer*, 21: 105–117, 1988.

[6] David Marr. Vision, 1982.

[7] J. R. Movellan. Optimal control of dynamic Bayes nets. Technical report, Kolmogorov Tutorials, UCSD, 2005.

[8] J. R. Movellan and J. S. Watson. The development of gaze following as a Bayesian systems identification problem. In *Proceedings of the International Conference on Development and Learning (ICDL02)*. IEEE, 2002.

[9] J. D. Nelson and J. R. Movellan. Active inference in concept induction. In T. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, number 13, pages 45–51. MIT Press, Cambridge, Massachusetts, 2001.

[10] J. D. Nelson, J. B. Tenenbaum, and J. R. Movellan. Active inference in concept learning. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 692–697. LEA, Edinburgh, Scotland, 2001.

[11] R. W. Picard. *Affective Computing.* MIT Press, Cambridge, MA, 1997.

[12] Movellan J. R. and J. S. Watson. Perception of directional attention. In *Infant Behavior and Development: Abstracts of the 6th International Conference on Infant Studies*, NJ, 1987. Ablex.

[13] J. S. Watson. Smiling, cooing and "the game". *Merrill-Palmer Quarterly*, 18:323–339, 1972.

[14] J. S. Watson. The perception of contingency as a determinant of social responsiveness. In E. B. Thoman, editor, *Origins of the Infant's Social Responsiveness*, pages 33–64. LEA, New York, 1979.

10 Appendix I: Summary of the Model

Parameters:

$$\Delta t \in \Re, \quad \text{Sampling period in seconds} \tag{44}$$

$$\pi \in [0,1], \quad \text{Prior on } H \tag{45}$$

$$0 \le \tau_1^s \le \tau_2^s, \quad \text{Delay parameters for self-feedback loop} \tag{46}$$

$$\tau_2^s < \tau_1^a \le \tau_2^a, \quad \text{Delay parameters for social agents} \tag{47}$$

Static Random Variables:

$$H \sim \text{Bernoulli}(\pi), \quad \text{Presence/Absence of Responsive Agent} \tag{48}$$

$$R_1 \sim \text{Beta}(1,1), \quad \text{Self Activity Rate} \tag{49}$$

$$R_2 \sim \text{Beta}(1,1), \quad \text{Agent Activity Rate} \tag{50}$$

$$R_3, \quad \text{Background Activity Rate:} \tag{51}$$

$$R_3 \sim \text{Beta}(1,1), \;\; \text{If } H = 1 \tag{52}$$

$$R_3 = R_2, \;\; \text{If } H = 0 \tag{53}$$

Stochastic Processes:
The following processes are defined for $t = 1, 2, \cdots$

$$\text{Timer:} \quad Z_t = \begin{cases} \tau_2^a + 1, & \text{for } t = 0 \\ h(Z_{t-1}, U_t), & \text{else} \end{cases} \tag{54}$$

$$\text{Indicators:} \quad I_t = (I_{1,t}, I_{2,t}, I_{3,t})^T \in \{0,1\}^3 \tag{55}$$

$$\text{Indicator of Self Period:} \quad I_{1,t} = \begin{cases} 1 & \text{if } Z_t \in [\tau_1^s, \tau_2^s] \\ 0 & \text{else} \end{cases} \tag{56}$$

$$\text{Indicator of Agent Period:} \quad I_{2,t} = \begin{cases} 1 & \text{if } Z_t \in [\tau_1^a, \tau_2^a] \\ 0 & \text{else} \end{cases} \tag{57}$$

$$\text{Indicator of Background Period:} \quad I_{3,t} = (1 - I_{1,t})(1 - I_{2,t}) \tag{58}$$

$$\text{Drivers:} \quad D_t = (D_{1,t}, D_{2,t}, D_{3,t})^T \in \{0,1\}^3 \tag{59}$$

$$\text{Self Driver:} \quad D_{1,t} \sim \text{Poison}(R_1) \tag{60}$$

$$\text{Agent Driver:} \quad D_{2,t} \sim \text{Poison}(R_2) \tag{61}$$

$$\text{Background Driver:} \quad D_{3,t} \sim \text{Poison}(R_3) \tag{62}$$

$$\text{Robot Sensor:} \quad Y_t = I_t \cdot D_t \tag{63}$$

$$\text{Robot Actuator:} \quad U_t \tag{64}$$

$$\text{Sensor Activity Counters:} \quad O_{1,t} = \sum_{s=1}^{t} I_{i,t} Y_t, \text{ for } i = 1,2,3 \tag{65}$$

$$\text{Sensor Inactivity Counters:} \quad Q_{1,t} = \sum_{s=1}^{t} I_{i,t}(1 - Y_t), \text{ for } i = 1,2,3 \tag{66}$$

Probabilistic constraints:

Figure 4 display Markovian constraints in the joint distribution of the different variables involved in the model.

11 Appendix II: Definitions

- Beta Variables:

$$R \sim \text{Beta}(\beta_1, \beta_2) \tag{67}$$

$$p(r) = \text{Beta}(r, \beta_1, \beta_2) = \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} (r)^{\beta_1 - 1}(1 - r)^{\beta_2 - 1} \tag{68}$$

$$\text{E}(R) = \frac{\beta_1}{\beta_1 + \beta_2} \tag{69}$$

$$\text{Var}(R) = \frac{\beta_1 \beta_2}{(\beta_1 + \beta - 2)^2(\beta_1 + \beta_2 + 1)} \tag{70}$$

- Gamma Function:

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du, \quad \text{for } x > 0 \tag{71}$$

The gamma function has the following properties

$$\Gamma(x + 1) = x\Gamma(x) \tag{72}$$

$$\Gamma(x) = (n - 1)!, \quad \text{for } n = 1, 2 \cdots \tag{73}$$

- Logistic Function:

$$\text{logistic}(x) = \frac{1}{1 + e^{-x}} \tag{74}$$

- Entropy:

$$\mathcal{H}(Y) = -\int p(y) \log p(y)\, dy \tag{75}$$

- Conditional Entropy:

$$\mathcal{H}(Y \mid X) = -\int p(x, y) \log p(y \mid x)\, dxdy \tag{76}$$

- Mutual Information:

$$\mathcal{I}(X, Y) = \mathcal{H}(X) - \mathcal{H}(X \mid Y) = \mathcal{H}(Y) - \mathcal{H}(Y \mid X) \tag{77}$$