A Reinforcement Learning Model of Social Referencing

Hector Jasso [1] (hjasso@ucsd.edu), Jochen Triesch [2,3] (triesch@fias.uni-frankfurt.de)
Gedeon Deák [2] (deak@cogsci.ucsd.edu)

[1] University of California - San Diego
9500 Gilman Drive; La Jolla, CA 92093, USA
[2] Frankfurt Institute for Advanced Studies, J.W. Goethe University, Frankfurt,
Ruth-Moufang-Str. 1,60437 Frankfurt am Main, Germany
[3] Department of Cognitive Science; University of California - San Diego
9500 Gilman Drive, MC 0515; La Jolla, CA 92093-0515, USA

Abstract

We present a novel computational model of social referencing. The model replicates a classic social referencing experiment where an infant is presented with a novel object and has the choice of consulting an adult's informative facial expression before reacting to the object. The infant model learns the value of consulting the adult's facial expression using the temporal difference learning algorithm. The model is used to make hypotheses about the reason for a lack of social referencing found in autistic individuals, based on an aversion to faces. Comparisons are made between this reinforcement learning model and a previous model based on mood contagion.

Introduction

Infants of a certain age, when presented with a novel object such as an unfamiliar toy, will sometimes consult the facial expression of a trusted adult before reacting to the object (refer to Fig. 1). If the adult shows a positive expression such as a smile, the infant will interact with the object (case 1), but if instead the expression is negative as in a fearful face, the infant will avoid the object (case 2). And if the object is not novel, then the infant will not look at the adult before reacting to the object (cases 3 and 4).

This behavior, called *social referencing* [1, 2, 3], is the focus of this paper. While social referencing is also employed in other situations such as reacting to an unknown adult [4] or crossing a visual cliff that might or might not be dangerous [5], we will concentrate on the above example involving reacting to a novel object. The common thread is that the infant is presented with a situation where the best way to react is ambiguous, and at the same time has the opportunity of consulting the emotional expression of an adult to inform its response.

Social referencing is thought to be a key to understanding the origins of emotional expressions. According to some theorists [6], social referencing is part of a larger set of joint attention behaviors, which also include gaze following [7]. These behaviors enable non-linguistic communication between infants and caregivers and, ultimately, serve the development of social understanding: that is, the ability to interpret, predict, and influence other people's behaviors [8, 9].

The only computational model of social referencing developed so far is implemented using Leonardo, a humanoid robot used to explore expressive social interactions [10, 11]. Leonardo has a cognitive-affective architecture where behavior is guided not only by its visual system, but also by its affective appraisals of objects (i.e. its interactions with objects are influenced by how it "feels about them").

In the social referencing scenario presented in [10, 11], Leonardo encounters a novel object, of which it does not yet have an affective appraisal stored in memory. This novelty causes Leonardo's emotion system to evoke a "state of anxiety", which triggers a search for human faces. The human then makes sure that Leonardo attends to her facial expression as well as the novel object. Watching the human's facial expression and hearing her vocalizations causes Leonardo to change its affective state, associating it to the object and storing this association in the object's template, which is committed to memory. This will affect subsequent emotional reactions when the object is either remembered or encountered in the future: Leonardo will avoid objects and show negative affect towards them if the caregiver was expressing negative affect when the object was first encountered, and will interact with objects and show positive affect towards them if the caregiver's expression was positive. Leonardo's reaction to the object can change in the future after interactions with it.

We present in this paper a novel reinforcement learning [12, 13] model of social referencing that learns the ability as it experiences the world in which it acts. Our model performs social referencing without need to model emotions, or through mood contagion, as Leonardo does. This same reward-driven modeling framework has been used to investigate gaze following [14, 15]. Such models not only give explanations of the possible origins of such abilities, but also allow modeling their developmental trajectory, and hold the key to understanding why such abilities can fail to develop [16].

A model of social referencing

In this section we describe our model of social referencing, and how it was used to simulate the experiments described in the previous section.

Model details

Learning and experiment-making happen during a series of trials, where a model infant is presented with an object, and has access to the facial expression of an adult (caregiver) (refer to Fig. 2). Trials end when the infant either interacts with the object, or ignores it. In these

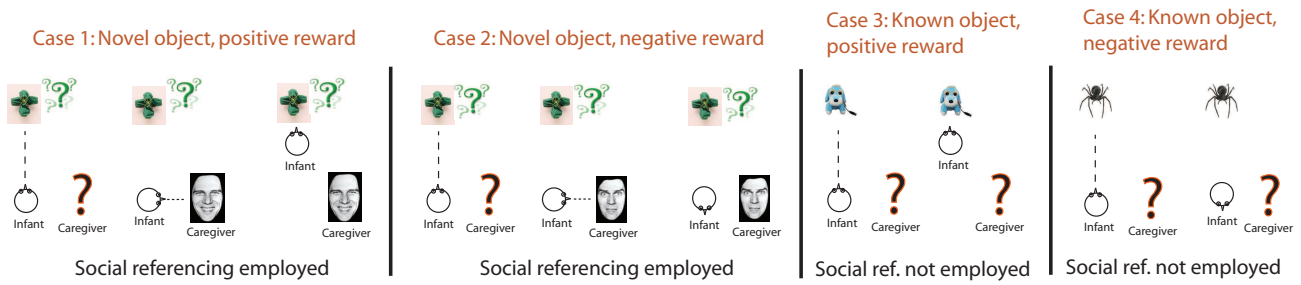

Figure 1: Description of social referencing as performed by older infants (as well as adults). Case 1: Infant is presented with a novel object, which might be nice (rewarding) or not (not rewarding), so it decides to consult the adult first. The adult's smiling face prompts the infant to interact with the object. Case 2: Similar to case 1, but here the adult's fearful face prompts the infant to avoid the object. Case 3: Object is evidently nice (rewarding), so infant interacts with it without consulting the adult. Case 4: Object is evidently not nice (not rewarding), so infant avoids it without consulting the adult.

simulated trials, time is discretized into time steps, each corresponding to about 5 seconds.

The caregiver can be smiling, showing a neutral face, or a fearful face ($cg_{\text{expression}} = smiling/neutral/fearful$). The object presented has an intrinsic reward (obj_{reward}) that can take any real value ($-\infty < obj_{\text{reward}} < \infty$). The object can be a novel object or not ($obj_{\text{novel}} = true/false$). If it is not novel, it will have a remembered intrinsic reward ($obj_{\text{remembered\_reward}}$) that can also take any real value ($-\infty < obj_{\text{remembered\_reward}} < \infty$). These last two variables allow us to model non-perfect memories, where the infant does not necessarily remember correctly about the object's reward.

During any trial step, the infant can be either looking at the object, looking at the caregiver, interacting with the object, or ignoring both the caregiver and the object. (In fig. 1, for example, the three steps in *case 2* correspond to infant looking at object, infant looking at the caregiver, and infant ignoring both the object and the caregiver. The third step in *case 1* shows the infant interacting with the object. Also, note that because activities are exclusive, the infant can look at the caregiver or the object but not both at the same time.) We keep track of whether the infant has looked at the caregiver's expression during the trial or not ($inf_{consulted\_expression} = true/false$), whether the infant has interacted with the object during the trial or not ($inf_{interacted\_with\_object} = true/false$), and whether the infant has ignored the object and the caregiver during the trial ($inf_{ignored} = true/false$).

The state of the world from the infant's perspective (s) is the combination of two elements: infant's knowledge of the object's reward ($s_{knowledge\_object\_reward}$), and infant's knowledge of the caregiver's facial expression during the present trial ($s_{knowledge\_expression}$) ($s = [s_{s_{knowledge\_object\_reward}, s_{knowledge\_expression}}]$).

The first element ($s_{knowledge\_object\_reward}$) is either:

- 'object remembered as rewarding' ($s_{knowledge\_object\_reward} = remembered\_rewarding$),
- 'object remembered as not rewarding' ($s_{knowledge\_object\_reward} = remembered\_not\_rewarding$),
- 'object reward unknown' ($s_{knowledge\_object\_reward} = unknown$).
- 'object rewarding, as manipulated during this trial' ($s_{knowledge\_object\_reward} = rewarding\_this\_trial$).
- 'object not rewarding, as manipulated during this trial' ($s_{knowledge\_object\_reward} = not\_rewarding\_this\_trial$).
- 'object ignored' ($s_{knowledge\_object\_reward} = ignored$).

The value 'object remembered as rewarding' corresponds to a known object ($obj_{\text{novel}} = false$) the infant has not yet interacted with or ignored during this trial ($inf_{interacted\_with\_object} = false$ and $inf_{ignored} = false$), with the object's remembered reward greater than zero ($obj_{\text{remembered\_reward}} > 0$). The value 'object remembered as not rewarding' corresponds to a known object the infant has not yet interacted with or ignored during this trial, with the object's remembered reward less than or equal to zero ($obj_{\text{remembered\_reward}} \leq 0$). The value 'object reward unknown' corresponds to a novel object ($obj_{\text{novel}} = true$) the infant has not yet interacted with or ignored during this trial. The value 'object rewarding, as manipulated during this trial' corresponds to an object with reward greater than zero ($obj_{\text{reward}} > 0$) that the infant has interacted with during this trial ($inf_{interacted\_with\_object} = true$). The value 'object not rewarding, as manipulated during this trial' corresponds to an object with reward less than or equal to zero ($obj_{\text{reward}} \leq 0$) that the infant has interacted with during this trial. Finally, it will be 'object ignored' if the infant has ignored the object during this trial ($inf_{ignored} = true$). [1]

The second element ($s_{knowledge\_expression}$) is either:

- 'facial expression unknown' ($s_{knowledge\_expression} = unknown$),

[1] Because of the way trials are structured (see below), the infant can not have interacted with an object and ignored it on the same trial.

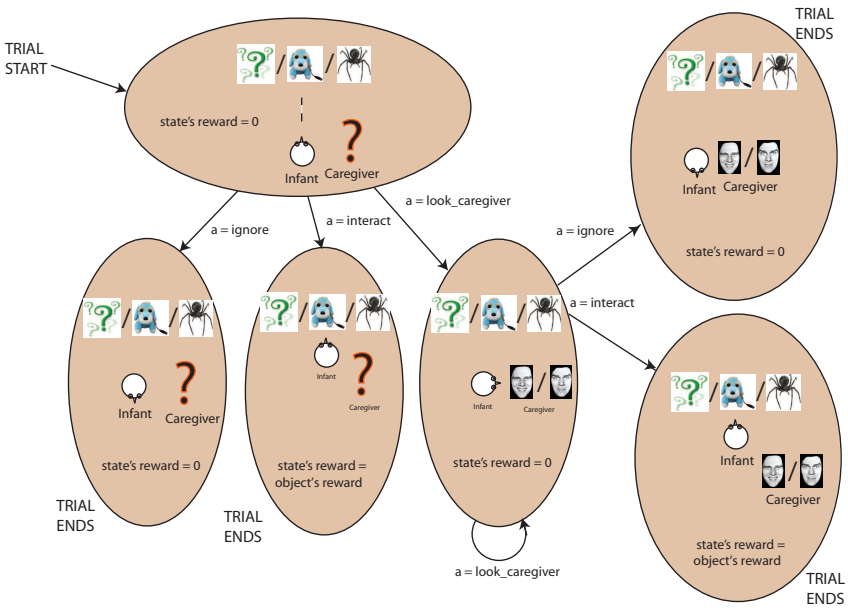

Figure 2: Modeling social referencing. Top left node corresponds to the trial's start. The two lower left nodes correspond to infant not doing social referencing, and the three lower right nodes correspond to infant doing social referencing. Refer to text for details.

- 'facial expression smiling' ($s_{knowledge\_expression} = smiling$),
- 'facial expression neutral' ($s_{knowledge\_expression} = neutral$),
- 'facial expression fearful' ($s_{knowledge\_expression} = fearful$)

The value 'facial expression unknown' corresponds to the infant not having consulted the caregiver during the trial ($inf_{\text{consulted\_expression}} = false$). Otherwise, the value will match the caregiver's expression ($s_{knowledge\_expression} \longleftarrow cg_{\text{expression}}$).

At each trial step, the infant chooses an action a, which can be either:

- 'look at caregiver' ($a = look\_caregiver$)
- 'interact with object' ($a = interact$)
- 'ignore object and caregiver' ($a = ignore$)

With the action 'look at the caregiver' the infant looks at the caregiver's facial expression (and sets $inf_{\text{consulted\_expression}}$ to $true$). With the action 'interact with object' the infant interacts with the object (and sets $inf_{\text{interacted\_with\_object}}$ to $true$). With the action 'ignore object and caregiver' the infant ignores both the object and the caregiver (and sets inf_{ignored} to $true$).

Reinforcement learning model

The state of the world from the infant's perspective (s) serves as input to an actor-critic reinforcement learning algorithm [12, 13] that drives actions (a).

The *critic* holds an estimate of the value of the current state of the world v_s as an array of 6x4 real values (i.e. it holds a value for each possible state s of the world). v_s is updated as:

$$v_s(t+1) = v_s(t) + \epsilon\delta(t)$$

where ϵ is a parameter defining the learning rate ($\epsilon > 0$), and $\delta(t)$ specifies the temporal difference error defined as

$$\delta(t) = r(t) + v_s(t+1) - v_s(t)$$

$r(t)$ being the reward obtained after taking the action, and $v_s(t+1)$ the estimate of the value of the new state after taking the action.

The *actor* specifies the action to be taken. This is chosen probabilistically according to:

$$P[a] = \frac{\exp(\beta m_{s,a})}{\sum_{a'} \exp(\beta m_{s,a'})}$$

($m > 0$) where $m_{s,a}$ is action value parameter associated with taking action a while being in state s, and β is an 'inverse temperature' parameter, which increases exploration versus exploration with a larger value. (Thus, an action a in state s is more likely to be chosen the higher the value of $m_{s,a}$ with respect to the corresponding value for alternate actions.) Once the action a is chosen in state s, $m_{s,a}$ is updated according to:

$$m_{s,a}(t+1) = m_{s,a}(t) + \epsilon\delta(t)$$

where ϵ and $\delta(t)$ are the same as defined above.

Experiments and results

1. Parameter setting Before the first training trial, all values of V and m are initialized to zero to reflect an absence of previous, or "innate" knowledge about social referencing. β is set to 5 and ϵ is set to 0.05 for smooth learning.

The reward scheme is defined as follows: no reward is obtained ($r(t) = 0$) if the infant is looking at the object, looking at the caregiver, or ignoring the object and the caregiver. If the infant interacts with the object, it receives the object's intrinsic reward ($r(t) = obj_{\text{reward}}$).

2. Training setup Training trials are as depicted in Fig. 2. At the start of each training trial, a new object is presented, with an intrinsic reward extracted from a normal distribution with average zero and standard deviation of one. The caregiver's facial expression will match the object's intrinsic reward as follows: If the intrinsic reward is greater than 0.5 ($obj_{\text{reward}} > 0.5$), the caregiver smiles ($cg_{\text{expression}} = smiling$), if it is less than -0.5 ($obj_{\text{reward}} < -0.5$), the caregiver shows a fearful extpression ($cg_{\text{expression}} = fearful$), otherwise ($-0.5 \leq obj_{\text{reward}} \leq 0.5$) it shows a neutral face ($cg_{\text{expression}} = neutral$) (this models a caregiver that is neutral to objects that are not negative or positive enough, as in the case of everyday objects such as pieces of paper or glasses).

The object's intrinsic reward, its remembered reward, and the caregiver's facial expression do not change throughout the trial. The infant starts all trials with no knowledge of the caregiver's facial expression ($inf_{consulted\_expression} = false$).

The infant starts all trials looking at the object. The object is familiar ($obj_{\text{novel}} = false$) 80% of trails, and novel ($obj_{\text{novel}} = true$) the rest (this reflects a relatively stable environment, where most of the objects are known to the infant, and new ones are introduced every once in a while.) If the object is familiar, the infant's remembered reward is set to the object's intrinsic reward ($obj_{\text{remembered\_reward}} \longleftarrow obj_{\text{reward}}$).

The model infant chooses its actions based on the reinforcement learning mechanism described above. Trials end when the infant either interacts with the object ($inf_{\text{interacted\_with\_object}} = true$), or ignores the object and caregiver ($inf_{\text{ignored}} = true$).

3. Testing procedure After initialization, as well as after every 1,000 training trials, the infant is tested 1,000 times. During testing, all learning is disabled (i.e. no updates of V and m are made, in order to eliminate any bias that multiple testing might cause). The following four different setups were tested (refer to Fig. 1):

- **case 1: novel object, positive reward**: A novel object is presented ($obj_{\text{novel}} = true$). The object is rewarding enough that the caregiver shows a smiling face ($obj_{\text{reward}} = +1$, $cg_{\text{expression}} = smiling$)).
- **case 2: novel object, negative reward**: A novel object is presented ($obj_{\text{novel}} = true$). The object is non-rewarding enough that the caregiver shows a fearful face ($obj_{\text{reward}} = -1$, $cg_{\text{expression}} = fearful$)).
- **case 3: familiar object, positive reward**: A known object is presented ($obj_{\text{novel}} = false$). The object is rewarding enough that the caregiver shows a smiling face ($obj_{\text{reward}} = +1$, $cg_{\text{expression}} = smiling$)).
- **case 4: familiar object, negative reward**: A known object is presented ($obj_{\text{novel}} = false$). The object is non-rewarding enough that the caregiver shows a fearful face ($obj_{\text{reward}} = -1$, $cg_{\text{expression}} = fearful$)).

The values of -1 and +1 result in the caregiver showing the facial expression we need for the test: any value under -0.5 can be used for the "negative reward" cases, and any value over 0.5 can be used for the "positive reward" cases.

Results

Fig. 3 shows the results. After about 2,000 training trials, in cases 1 and 2 the infant consults the adult before making a choice of how to react to the novel object: In case 1, the infant interacts with the object after seeing that the caregiver was smiling, while in case 2, it prefers to ignore the object after seeing that the caregiver was showing a fearful face. In cases 3 and 4, after training, the infant either interacts with the known object if it is known to be rewarding (case 3), or ignores it if it is known to be not rewarding (case 4). This is done without looking at the caregiver before.

Cases 1 and 2 show that the infant learns to do social referencing when confronted with a novel object, correctly interpreting the caregiver's facial expression. Cases 3 and 4 show that the infant does not perform social referencing when it is not needed (i.e. when the object is not novel), reacting correctly to the object according to its knowledge of the object's reward. Overall, then, the model infant has correctly learned to do social referencing.

Note that the infant does not learn to consult the caregiver on every trial. In fact such behavior would not lead to the highest possible sum of discounted rewards because of the reduction of future rewards with the discount factor γ. This reward discounting leads to a preference for immediate rewards over future rewards. Since consulting the caregiver postpones the reward for interacting with the object, the model learns to avoid consulting the caregiver when the object is known to be positive (or negative). Only in the case of an ambiguous object the knowledge gained from looking at the caregiver will outweigh the cost of delaying the interaction with the object.

In a previous gaze following model [17, 14], delays or failures to develop the ability resulted from introducing a negative reward for looking at the caregiver's face (i.e. simulating an aversion to faces), based on experimental evidence with autistic individuals [18]. We can simulate autistic behavior in our model by assigning a negative

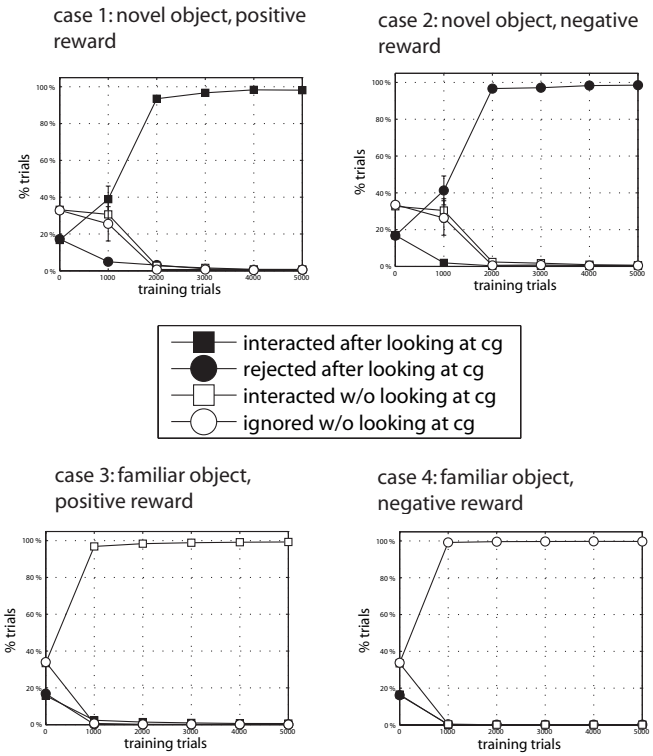

Figure 3: Experiment results. Refer to text, and to Fig. 1. Error bars indicate standard errors after 10 repetitions.

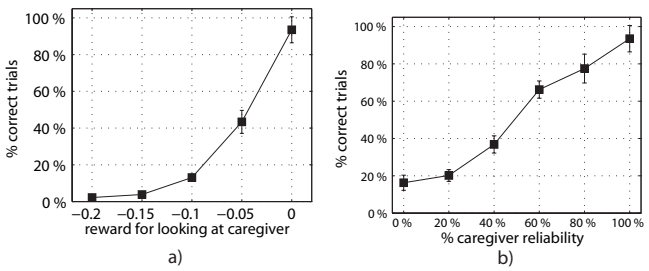

Figure 4: a) Effect of introducing an aversion to looking at the caregiver: percentage correct social referencing after 2,000 trials for case 1, for different rewards for looking at caregiver (similar results are obtained for the analogous case 2). b) Effect of an unreliable caregiver: percentage correct social referencing after 2,000 trials for case 1, for different caregiver reliabilities (similar results are obtained for the analogous case 2). Error bars indicate standard errors after 10 repetitions.

reward for looking at the caregiver (the basic setup of our model does not include any rewards, positive or negative, for looking at the caregiver): Fig. 4 a) shows how the percentage of correct social referencing after 2,000 trials for case 1 (i.e. infant interacting with the object after consulting the caregiver) decreases as a negative reward is introduced for looking at the caregiver, with not social referencing if the reward is -0.2.

Similarly, the model predicts that an unreliable caregiver (i.e. one showing a random facial expression for some of the training trials) leads to a slower learning of social referencing (see Fig. 4 b)).

The model also predicts that an unstable environment will lead to a faster development of social referencing, and slower learning with a less stable environment: When the percentage of training trials with non-novel objects is increased from 80% to 95%, correct social referencing in case 1 takes 5,000 trials to reach 80% while it normally takes less than 2,000 trials, as shown in Fig. 3 (with similar results for case 2). And when when the percentage of training trials with non-novel objects is decreased to 50% familiarity, social referencing learning is speeded up, with an average of about 90% correct social referencing after 1,000 learning trials, as opposed to the typical 40% shown in Fig. 3.

The model also predicts that if the infant is not presented with new objects, then social referencing will not emerge: if no new objects are presented during training, then correct social referencing in case 1 (see above) remains at 20% even after 5,000 training trials, which is the baseline (i.e. this is the value that the infant tests to immediately after initialization (0 training trials), as shown in Fig. 3).

The model also predicts that if the infant has a poor memory, incorrectly recalling objects as either rewarding or not (specifically, when a familiar object is presented but we want to simulate incorrect recall for that trial, the remembered reward for the object is extracted from a normal probability distribution with average zero and standard deviation 1, without affecting obj_{reward} nor the resulting caregiver's expression), then the development of social referencing is not affected (i.e. plots for cases 1 and 2 do not change significantly with only 50% correct recall). And with 0% correct recall the model infant will perform social referencing even when presented with a familiar object (cases 3 and 4).

It takes about 2,000 trials for the model to learn social referencing (in Fig. 3, this is about 95% correct social referencing response in cases 1 and 2, and close to 100% for cases 3 and 4). It seems reasonable that during the 6 months that infants take to learn social referencing they are exposed to that amount of such experiences.

Discussion

We have presented a model of social learning based on reinforcement learning [12, 13]. The model shows how social referencing can develop as the infant learns the value of consulting the caregiver's facial expression before acting on a novel object[2].

There is an important difference between the model presented here and Leonardo: our model learns the ability through interactions with others, while Leonardo's social referencing capabilities are better understood as an innate ability. This, because Leonardo's behavior of looking at other people's faces when in an "anxious

[2]A dynamic programming (DP) approach would require a model of the environment, including state transition probabilities and expected rewards for different actions [12]. Since we do not assume such knowledge in the infant model, DP is not an option for our model.

state" is pre-programmed. It is our belief that a model that can learn an ability through interactions with others is a more solid base for bootstraping further abilities. This is the heart of the developmental approach to learning cognitive abilities [16]. For this reason, our developmental approach can give simple yet powerful explanations of different developmental trajectories, as in social referencing failing to develop because of an aversion to faces.

Another important difference between our model and Leonardo is that the adult's emotional expression is simply treated as a signal: the infant model does not need to model (i.e. somehow represent) emotions in order to correctly deal with the novel object. This is not to say that emotions do not play an important role in the development of social referencing in human infants. Instead, our model points out a generic mechanism by which social referencing could in principle be learnt in the absence of any emotions or mood contagion, as long as the different emotional expressions of others can be recognized.

Our model learns through a series of trials, with well-defined structure. In the real world, however, these trials happen within a rich and complex infant-caregiver interaction, and the infant has to identify when the caregiver's facial expression is 'about' the object (i.e. when is the adult doing social referencing, as opposed to just looking around). Such situations should be considered when building a version that will work in a more complex environment.

Finally, this model can also be seen as implementing a form of top-down attention [19], where the infant's attention is guided by stimuli other than the saliency of the visual input. In this case, the introduction of a novel object directs the infant's visual attention to the caregiver in order to acquire new knowledge about the environment without a direct/immediate benefit (there is no reward for looking at the caregiver per se).

References

[1] Campos, J., & Stenberg, C. (1981). Perception, appraisal, and emotion: The onset of social referencing. In M. Lamb & L. Sherrod (Eds.), *Infant social cognition*. Erlbaum, Hillsdale, NJ.

[2] Feinman, S. (1982). Social referencing in infancy. *Merrill-Palmer Quarterly*, *28*, 445-470.

[3] Hornik, R., Risenhoover, N., & Gunnar, M. (1987). The effects of maternal positive, neutral and negative affective communications on infant responses to new toys. *Child Development*, *58*, 937-944.

[4] Zarbatany, L., & Lamb, M. (1985). Social referencing as a function of information source: Mothers versus strangers. *Infant Behavior and Development*, textot8, 25-33.

[5] Sorce, J., Emde, R., Campos, J., & Klinnert, M. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*, 195-200.

[6] Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.

[7] Moore, C. & Dunham, P. J. (1995). *Joint Attention: Its Origins and Role in Development*. Erlbaum, Hillsdale, NJ.

[8] Barresi, J., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, *19*, 107-154.

[9] Moore, C. (2006). *The Development of Commonsense Psychology*. Erlbaum, Hillsdale, NJ.

[10] Thomaz, A. L., Berlin, M., & Breazeal, C. (2005) Robot science meets social science: an embodied computational model of social referencing, *Cog Sci 20005 Workshop Toward Social Mechanisms Android Science*. Trento, Italy.

[11] Thomaz, A. L., Berlin, M., & Breazeal, C. (2005), An embodied computational model of social referencing. *Proceedings of Fourteenth IEEE Workshop on Robot and Human Interactive Communication (Ro-Man05)*. Nashville, TN.

[12] Sutton, R. S., & Barto, A. G. (1998) *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

[13] Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA, USA: MIT Press.

[14] Jasso, H. *A reinforcement learning model of gaze following*, Unpublished Ph.D. dissertation. University of California, San Diego (2007)

[15] Triesch, J., Jasso, H., & Deák, G. O. (2007). Emergence of mirror neurons in a model of gaze following. *Adaptive Behavior*, *15*, 149-165.

[16] Elman, J. L. Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996) *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press, Cambridge, MA.

[17] Triesch, J., Teuscher, C., Deák, G., Carlson, E. (2006). Gaze following: why (not) learn it? *Developmental Science*, *9*, 125-147.

[18] Dawson, G., Meltzoff, A. N., Osterling, J. Rinaldi, J., & Brown, E. Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, *28*, 479-485.

[19] Zelinksy, G., Zhang, W., Yu, B., Chen, S., & Samaras, D. (2005). The role of top-down and bottom-up processes in guiding eye movements during visual search. *Neural Information Processing Systems (NIPS) 2005*, Vancouver, Canada.