Memories and Dreams of Social Interaction

Goren Gordon
Personal Robotics Group
MIT Media Lab
20 Ames Street E15-468
Cambridge, MA 02139
goren@gorengordon.com

Cythina Breazeal
Personal Robotics Group
MIT Media Lab
20 Ames Street E15-468
Cambridge, MA 02139
cynthiab@media.mit.edu

ABSTRACT
Social interaction has a long lasting effect on all participants. However, most HRI research has been focused on the interaction itself and not on the memories it leaves on both human and robot subjects. On the other hand, neuroscience has progressed in the study of memories formation, mostly from a bottom-up approach, i.e. mechanistic, molecular and cellular perspectives. We propose a new view on HRI, in which the focus is placed on the lasting effect of the interaction, namely, the memories of the interaction itself. An even more challenging goal is to probe the remnants of the social interaction that surface during dreams. Robots can serve a dual goal in this investigation. First, the interaction with a robot is, still, a novel and intensive situation that can be manipulated so as to stick to memories and surface in the dreams of the human subjects. Second, computational models of memories formation and dreams' roles in them can be implemented on the robots and compared to human reports. Here we propose a simple experimental paradigm to probe human subject memory formation and retention and a novelty-based computational model of memories and dreams to be implemented on the robot.

Keywords
Memory formation, Memory retention, Dreams

1. INTRODUCTION
The study of social interaction, be it with robots or humans, has mainly focused on the interaction itself, whereas its long lasting effect on the participants has received less attention. While HRI has served as a field of learning and teaching, the formation of memories of the interaction itself, e.g. its initiation, its peak, in the human subject has been largely neglected. Since interacting with a robot is still a novel situation for most people (not including HRI researchers), what subjects remember and dream of after such an encounter remains largely a mystery. Furthermore, for a robot to behave in a more human-like manner, the same memories of previous interactions should also be computationally modeled and implemented [11, 5]. Perhaps the robot can dream of this interaction?

Neuroscience and psychology has a long history of research of both the formation of memories and dreams. From a bottom-up approach, much has been learned in previous years on the mechanisms of short- and long-term memory, from the molecular [9], cellular [10] and systems [3] levels. A top-down approach has been investigated for more than a century in psychology [12, 4], whereas the study of dreams has regained new favor [1, 7]. Many insights from these studies can be applied to HRI-induced memory formation.

From a computational perspective, several neuro-inspired models have been suggested [8]. However, in the HRI field, most of the attention of computational models has been on either the behavioral side of the robot or the learning it performs on specific tasks. The episodic memories formed due to the interaction itself has been largely neglected [11, 5]. As for dreams, they have been suggested to serve as a "batch-learning" phase, where recorded data is re-iterated for the learning process.

Human-robot interaction can serve as a unique test-bed for the study of social interaction memory formation from both a psychological, as well as a computational side. For the former, we propose an experimental paradigm that enables the probing of human subject memories, as well as dream recall, following an HRI. We also propose a novelty-based computational model that allows comparison of robot memory formation as well as the coveted "robot-dreams'.

2. MEMORY-INDUCING HRI
The experimental paradigm is based on the observation that surprising, exciting and novel situations induce long-lasting memories and often reappear in dreams. It thus has a mandatory requirement that all subjects are naive with regard not only to the experimental paradigm, but also to the robot itself, whatever it may be. The experiment is composed of three stages, namely, introduction, task and farewell. The first stage is composed of the introduction of the human subject to the robot, which should be an exciting scenario, since interacting with a robot is, still, a rather uncommon situation. The introduction should "overwhelm" the subject with the sophistication of the robot, e.g. high-level cognitive reasoning and animated behavior, so as to induced a memorable event.

The second stage is composed of a rather tedious task that the human subject and robot perform together, e.g. place-and-drill or block-building. The task should be a simple one so as not to tax the robot's as well as the subject's mental capacities. During this stage, though, a surprising event should be deliberately induced by the robot, e.g. a deliberate mistake in performance or an unrelated animated behavior, in order to induce a second memorable event.

The last stage is composed of concluding remarks/gestures by the robot and a farewell gesture from the human subject. A human-like behavior, again exemplifying the state-of-the-art robot sophistication should be performed.

The human subjects are then either taken to a sleeping facility, if possible, or asked to write down their dreams in the following several days after this interaction. Furthermore, a day, a week and a month following this interaction, the subjects are asked for their recount of their interaction with the robot, so as to probe their memory formation. The goal of this experiment is to probe which aspect of the HRI induces long-lasting memories and surfaces in dreams, e.g. the physical aspect of the robot, its behavior, the interaction itself, a specific period during the interaction. This can then serve as the basis for a computational model for the robot.

3. NOVELTY-BASED COMPUTATIONAL MODEL

From the robot perspective, we propose a novelty-based computational model of memory formation. It is based on the same observation that surprising and novel events are integrated more into long-term memory. In order to mathematically describe a novel event, we propose that the robot attempts to learn to predict the dynamics of the social interaction, e.g. a forward model of the interaction. Based on its sensors input, in whatever level of abstraction, e.g. pixel-wise, facial expression, the model attempts to predict the next state based on the robot actions.

Starting with a reasonable, hard-coded prior, each new encounter results in surprise or prediction errors [2, 6] that can then serve as a measure of novelty. We propose that the peaks of novelty dynamics, i.e. the state-action pairs that instigate the highest prediction errors or information gain, are stored in memory. Thus, for example, one prediction of such a model is that the introduction stage results in many such stored memories, whereas the tedious task stage does not.

"Robot dreams" can be implemented by re-playing those sparsely stored memory events in a randomized manner, so as to continue and train the forward model of the social interaction. Thus, if at least one sensor has a visual component, one can project the robot dreams onto a screen. Obviously, several such interactions with different human subjects instigates various memorable events for the robot, and all of these can coincide in these dreams.

4. CONCLUSIONS

We propose that HRI serves as a test-bed for the neuroscience-inspired research of memory formation and dreams. While the human and robot perspectives of such an interaction is very different, hopefully the same mechanisms can serve both biological and artificial agents. One such example is a novelty-based model that can serve both, given that the human subjects are naive to the robot exemplar, as well as the task at hand. Hence, both participants experience the interaction as a novel event and their memories can be examined and compared.

Taking an extra step, an HRI resurfacing in dreams can also be studied within the same experimental paradigm. A computational model of memories resurfacing can thus serve as the coveted "robot-dreams" within the context of human-robot social interaction.

5. ACKNOWLEDGMENTS

G.G. was supported by the Fulbright commission for Israel, the United States-Israel Educational Foundation.

6. REFERENCES

[1] L. De Gennaro, C. Marzano, C. Cipolli, and M. Ferrara. How we remember the stuff that dreams are made of: Neurobiological approaches to the brain mechanisms of dream recall. *Behavioural Brain Research*, 226(2):592–596, 2012.

[2] G. Gordon and E. Ahissar. Hierarchical curiosity loops and active sensing. *Neural Netw*, 32:119–29, 2012.

[3] M. J. Jutras and E. A. Buffalo. Synchronous neural activity and memory formation. *Current Opinion in Neurobiology*, 20(2):150–155, 2010.

[4] D. Kahneman. *Attention and Effort*. Prentice-Hall Inc., 1973.

[5] Z. Kasap and N. Magnenat-Thalmann. Towards episodic memory-based long-term affective interaction with a human-like robot. In *RO-MAN, 2010 IEEE*, pages 452–457.

[6] D. Y. Little and F. T. Sommer. Learning and exploration in action-perception loops. *Front Neural Circuits*, 7:37, 2013.

[7] B. Rasch and J. Born. About sleep's role in memory. *Physiological Reviews*, 93(2):681–766, 2013.

[8] E. T. Rolls. A computational theory of episodic memory formation in the hippocampus. *Behavioural Brain Research*, 215(2):180–196, 2010.

[9] R. Shema, T. C. Sacktor, and Y. Dudai. Rapid erasure of long-term memory associations in the cortex by an inhibitor of pkmzeta. *Science*, 317(5840):951–953, 2007.

[10] N. Vitureira and Y. Goda. The interplay between hebbian and homeostatic synaptic plasticity. *The Journal of Cell Biology*, 203(2):175–186, 2013.

[11] H. Wan Ching, K. Dautenhahn, L. Mei Yii, P. A. Vargas, R. Aylett, and S. Enz. An initial memory model for virtual and robot companions supporting migration and long-term interaction. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 277–284.

[12] J. T. Wixted. The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55(1):235–269, 2004.