HRI 2015 PORTLAND/USA
March 2–5, 2015
10th ACM/IEEE International Conference
on Human-Robot Interaction

Workshop

Cognition: A Bridge between Robotics and Interaction.

Organizers: Alessandra Sciutti, Katrin Lohan and Yukie Nagai

Cognition: A Bridge between Robotics and Interaction.

Abstract

A key feature of humans is the ability to anticipate what other agents are going to do and to plan accordingly a collaborative action. This skill, derived from being able to entertain models of other agents, allows for the compensation for intrinsic delays of human motor control and is a primary support for efficient and fluid interaction. Moreover, the awareness that other humans are cognitive agents who combine sensory perception with internal models of the environment and others, enables easier mutual understanding and coordination [1]. Cognition represents therefore an ideal link between different disciplines, as the field of Robotics and that of Interaction studies performed by neuroscientists and psychologists. From a robotics perspective, the study of cognition is aimed at implementing cognitive architectures leading to efficient interaction with the environment and other agents (e.g., [2,3]). From the perspective of the human disciplines, robots could represent an ideal stimulus to study which are the fundamental robot properties necessary to make it perceived as a cognitive agent, enabling natural human-robot interaction (e.g., [4,5]). Ideally, the implementation of cognitive architectures may raise new interesting questions for psychologists, and the behavioral and neuroscientific results of the human-robot interaction studies could validate or give new inputs for robotics engineers. The aim of this workshop will be to provide a venue for researchers of different disciplines to discuss the possible points of contact and to highlight the issues and the advantages of bridging different fields for the study of cognition for interaction.

References

[1] Lohan, K. S., Rohlfing, K. J., Pitsch, K., Saunders, J., Lehmann, H., Nehaniv, C. L., ... & Wrede, B., 2012, 'Tutor spotter: Proposing a feature set and evaluating it in a robotic system', International Journal of Social Robotics, 4(2), 131- 146

[2] Nagai, Y., Asada, M., & Hosoda, K., 2006, ' Learning for joint attention helped by functional development', Advanced Robotics, vol. 20, no. 10, pp. 1165-1181

[3] Nagai, Y., Kawai, Y., & Asada, M., 2011, 'Emergence of Mirror Neuron System: Immature vision leads to self-other correspondence', in Proceedings of the 1st Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics

[4] Sciutti, A., Patanè, L., Nori, F. & Sandini, G., 2014, 'Understanding object weight from human and humanoid lifting actions', IEEE Transactions on Autonomous Mental Development, vol. 6, no. 2, pp. 80-92 doi: 10.1109/TAMD.2014.2312399

[5] Sciutti, A., Bisio, A., Nori, F., Metta, G., Fadiga, L. & Sandini, G., 2013, 'Robots can be perceived as goal-oriented agents', Interaction Studies, in Broz, Frank, Hagen Lehmann, Bilge Mutlu and Yukiko I. Nakano (eds.), Gaze in human-robot communication. Special Issue of Interaction Studies 14:3 . xv, 179 pp. (pp. 329–350)

The organizers:
Alessandra Sciutti, Katrin Lohan and Yukie Nagai
Website: `http://www.macs.hw.ac.uk/~kl360/HRI2015W`

Cognition: A Bridge between Robotics and Interaction.

Schedule

9:00 – 9:05 *Opening* (A. Sciutti, K. Lohan and Y. Nagai)

9:05 – 9:40 Invited talk: Dr. Greg Trafton

9:40 – 10.00 *Embodiment is a Double-Edged Sword in Human-Robot Interaction: Ascribed vs. Intrinsic Intentionality*
Authors: Tom Ziemke, Serge Thill, David Vernon

10:00 – 10:30 Coffee Break

10:30 – 11:05 Invited talk: Prof. David Vernon

11:05 – 11:25 State Prediction for Development of Helping Behavior in Robots
Authors: Jimmy Baraglia, Yukie Nagai and Minoru Asada

11:25 – 11:45 Social Robots and the Tree of Social Cognition
Author: Bertram F. Malle

11:45 - 13:10 Break + Lunch (12:00 – 13:00)

13:10 – 13:45 Invited talk: Prof. Ayse P. Saygin

13:45 – 14:05 Predictive coding and the Uncanny Valley hypothesis: Evidence from electrical brain activity
Authors: Burcu A. Urgen, Alvin X. Li, Chris Berka, Marta Kutas, Hiroshi Ishiguro and Ayse P. Saygin

14:05– 14:25 The audio-motor feedback: a new rehabilitative aid for the developing blind child.
Authors: Giulia Cappagli, Elena Cocchi, Sara Finocchietti, Gabriel Baud-Bovy, Monica Gori

14:25 – 15:00 Invited talk: Prof. Andrew N. Meltzoff

15:00– 15:30 Coffee Break

15:30 – 15:50 Interaction as a bridge between cognition and robotics
Authors: Serge Thill, Tom Ziemke

15:50 – 16:50 Panel discussion

16:50 – 17:00 *Closing Remarks* (A. Sciutti, K. Lohan and Y. Nagai)

Invited Speakers

Dr. Greg Trafton, Naval Research Laboratory
Dr. Trafton received his degree in cognitive science from Princeton University. His research in Human Robot Interaction involves creating computational cognitive models that perceive and think the way that people do, putting those models on embodied platforms (Mobile/Dexterous/Social robots), and using those models to increase interaction capabilities between robots and people. He also works on predicting and preventing procedural errors and predicting when an operator becomes overloaded when managing multiple UAVs during supervisory control.

Prof. David Vernon, University of Skövde
From 2004 to 2010, David Vernon was a member of the team coordinating the EU-funded RobotCub Integrated Project, the goal of which is to develop an open-source cognitive humanoid robot: the iCub. His focus and specific responsibility was for the cognitive architecture. In late 2006, he returned to the UAE as Professor of Computer Engineering at Etisalat University College (now Kalifa University of Science, Technology, and Research) with the specific brief to help develop the post-graduate degree programmes. He was re-appointed Head of Department in 2009. In 2011 he joined the Institute of Cognitive Systems at the Technical University of Munich as a senior researcher. Currently he is Professor of Informatics at the Informatics Research Centre, University of Skövde since the 1st March 2013. David Vernon is working on a major new project, DREAM, which is funded by the European Commission to deliver the next generation robot-enhanced therapy (RET) for children with autism spectrum disorder (ASD). The goal is to develop clinical interactive capacities for social robots that can operate autonomously for limited periods under the supervision of a psychotherapist. DREAM is a great example of a new breed of robotics - cognitive robotics - and I have the privilege of serving as one of the co-chairs of the new IEEE Robotics and Automation Society Technical Committee on Cognitive Robotics.

Prof. Ayse P. Saygin, University of California
Prof. Ayse P. Saygin directs the Cognitive Neuroscience and Neuropsychology Lab (Saygin Lab) at the University of California, San Diego, where she is an Associate Professor of Cognitive Science and Neurosciences. She received a PhD in Cognitive Science from UC San Diego, followed by a European Commission Marie Curie fellowship at the Institute for Cognitive Neuroscience and Wellcome Trust Centre for Functional Neuroimaging at University College London. She holds an MSc. in Computer Science from Bilkent University, and a BSc. in Mathematics from Middle East Technical University, both in Ankara, Turkey. Dr. Saygin and her lab study human perception and cognition using a range of experimental and computational methods, including psychophysics, EEG, MRI, fMRI, brain stimulation, neuropsychological patient studies, machine learning, and brain-computer interfaces. As an NSF CAREER awardee, Dr. Saygin has built upon her PhD and postdoctoral work to develop a research program exploring the perceptual and neural mechanisms supporting the processing biologically and socially important objects and events such as the body movements and actions of other agents. With additional support from DARPA, Kavli Institute for Mind and Brain, the Qualcomm Institute, and the Hellman Foundation, Saygin lab also aims to inform human-robot interaction by integrating methods and theory from cognitive neuroscience, neuroimaging, human perception, artificial intelligence, computational modeling, social robotics and social cognition.

Prof. Andrew N. Meltzoff, University of Washington
Dr. Andrew N. Meltzoff holds the Job and Gertrud Tamaki Endowed Chair and is the Co-Director of the University of Washington Institute for Learning Brain Sciences. A graduate of Harvard Uni-

versity, with a PhD from Oxford University, he is an internationally renowned expert on infant and child development. His discoveries about infant imitation have revolutionized our understanding of early cognition, personality, and brain development. His research on social-emotional development and children's understanding of other people has helped shape policy and practice. Dr. Meltzoff's 20 years of research on young children has had far-reaching implications for cognitive science, especially for ideas about memory and its development; for brain science, especially for ideas about common coding and shared neural circuits for perception and action; and for early education and parenting, particularly for ideas about the importance of role models, both adults and peers, in child development.

Embodiment is a Double-Edged Sword in Human-Robot Interaction: Ascribed vs. Intrinsic Intentionality

Tom Ziemke
iLab, School of Informatics
University of Skövde, Sweden
& HCS, IDA
Linköping University, Sweden
+46-705-441444
tom.ziemke@his.se

Serge Thill
Interaction Lab
School of Informatics
University of Skövde
54128 Skövde, Sweden
+46-500-448389
serge.thill@his.se

David Vernon
Interaction Lab
School of Informatics
University of Skövde
54128 Skövde, Sweden
+46-500-448392
david.vernon@his.se

ABSTRACT
This very short paper makes a relatively simple point: The human embodied cognitive capacity / tendency to attribute intentionality, goals, etc. to others, and to interpret their behavior in intentional terms, is fundamental to many types of social interaction. There are at least two quite different conceptions of embodied cognition though, underlying much research in cognitive robotics and human-robot interaction, which also differ regarding whether robots (a) could actually have their 'own' intrinsic intentionality, or (b) could only be ascribed/attributed intentionality, similar to the way cartoon characters are. For robotics research as such the distinction might be secondary, and for philosophy of mind the questions might not be resolvable any time soon. For society and the general public, however, the issue potentially has quite significant social and ethical implications – therefore researchers might need to pay more attention to this than they have so far.

Categories and Subject Descriptors
I.2.0 [**Computing Methodologies**]: Artificial Intelligence – *Philosophical foundations.*

General Terms
Human Factors, Theory.

Keywords
Embodied cognition, social interaction, intentionality, autonomy.

1. INTRODUCTION
As Sciutti and colleagues recently pointed out, the "ability to understand others' actions and to attribute them mental states and intentionality is crucial for the development of a theory of mind and of the ability to interact and collaborate" [1]. While this is central to human embodied social interaction, it is clear that the underlying cognitive capacity is not limited to interpreting the behavior of other people. It is well known from the classic studies of Heider and Simmel [2] that humans also tend describe the behavior of simple moving objects (triangles, circles, etc.) in intentional terms. Naturally, this also applies to more complex and human-like objects, such as cartoon characters. When, for example, we see Donald Duck angrily chasing chipmunks Chip and Dale because they are stealing his popcorn, it comes very natural to us to interpret their behavior in intentional terms – as illustrated by the first half of this sentence. At the same time, however, we presumably all understand that Donald, Chip, and Dale are not real and therefore also do not really have intentions.

Not surprisingly, this attribution also extends to different types of technology, in particular more or less autonomous systems. In the case of autonomous vehicles, recent research [3] indicates that anthropomorphism – "a process of inductive inference whereby people attribute to nonhumans distinctively human characteristics, particularly the capacity for rational thought (agency) and conscious feeling (experience)" – increases the trust people have in such systems. Hence, it is very likely that in social interactions with robots, humanoid ones in particular, (a) humans will attribute agency, intentionality, etc. to such artifacts (cf. [1]), and (b) interactions benefit from such attributions.

This, however, also raises the question what exactly is the status of the intentions, goals, etc. that we ascribe to robots? After all, they are real (physical), they are interactive, and in the humanoid case they behave and look human-like to some degree. Does that mean that, like humans, they potentially have their own intrinsic intentionality? Or is their intentionality, as in the case of Donald Duck, necessarily only ascribed?

2. DISCUSSION
Researchers in AI and robotics typically avoid answering this question explicitly and might prefer to dismiss it as a purely philosophical question. Implicitly, however, research in embodied AI and robotics, touches on the issue quite commonly. For example, embodied AI researchers commonly refer to Searle's 1980 *Chinese Room Argument* [4] to illustrate that traditional – *disembodied* – approaches to AI were deeply flawed because they only dealt with the internal manipulation of representations by computer programs. Understanding the details of the argument is not relevant to this short paper, but in a nutshell, Searle's criticism was that "the operation of such a machine is defined solely in terms of computational processes over formally defined elements", and that such "formal properties are not by themselves constitutive of intentionality" [4]. Researchers in embodied AI/robotics commonly argue that these problems of traditional AI can be overcome by *'embodied'* (robotic) approaches to AI, which allow internal representations/mechanisms to be grounded in sensorimotor interactions with the physical and social environment. This, however, completely ignores the fact that already back in 1980, in the original paper, Searle presented – and rejected – what he called the *'robot reply'*, which entailed pretty much exactly what is now called *embodied AI*, i.e. computer programs running on robots that interact with their environment.

What is interesting in this context is that there are quite many researchers who – like Searle – take the Chinese Room Argument to be a valid argument against traditional AI, but at the same time

– unlike Searle – consider the *physical/sensorimotor embodiment* provided by today's robots to be sufficient to overcome the problem. In Harnad's terms, this type of embodied AI has gone from a *computational functionalism* to a *robotic functionalism* [5]. Zlatev, for example, formulated the latter position very explicitly, arguing that there is "no good reason to assume that intentionality is an exclusively biological property (pace e.g. Searle)", and "thus a robot with bodily structures, interaction patterns and development similar to those of human beings … could possibly recapitulate [human] ontogenesis, leading to the emergence of intentionality" [6]. Others, including Searle naturally, do indeed believe that there are good reasons to assume that intentionality is in fact a biological property intrinsic to living bodies [7, 8, 9].

This illustrates Chemero's point that there currently are (at least) two very different positions that are both referred to as *'embodied cognitive science'* [10]. The one that Chemero refers to as *radical embodied cognitive science* is grounded in anti-representationalist and anti-computationalist traditions. The other, more mainstream version, on the other hand, in line with robotic functionalism, is derived from more or less traditional representationalist and computationalist theoretical frameworks.

It is therefore not surprising that researchers in embodied AI and robotics, inspired by theories of embodied cognition, are confused regarding the relevance of 'embodiment', and subsequently find it difficult to answer the question to what degree their robots have, or could have, their own intrinsic intentionality. If you adopt the more mainstream position of robotic functionalism, then, as Zlatev put it, there is "no good reason to assume that intentionality is an exclusively biological property" [6]. If, on the other hand, you adopt a more radical, non-functionalist position, e.g. the enactive view, which has also gained some influence in embodied AI [8, 9, 11, 12], then intrinsic intentionality indeed might very well be "an exclusively biological property" and therefore most probably not replicable in robots with current technology.

If at this point you are about to dismiss (once more) the issue of robot intentionality as a purely philosophical question – which obviously the philosophers cannot answer either – it should be noted that the context of human-robot interaction adds a novel dimension to this old problem and gives new social and ethical relevance. The point is that human interaction with robots is to some degree comparable to interaction with animals: It is certainly the case that scientists, philosophers, and the general public are divided regarding whether or not animals have and experience human-like feelings. However, neither the fact that we cannot conclusively answer that question, nor the fact that we are likely to have different opinions (e.g., vegans vs. vegetarians vs. meat-eaters), change the fact that we need to have legislation, ethical guidelines, personal positions, etc. for how animals are to be treated in our society. Likewise, in the case of robots, whether or not we want to deal with the philosophical problem of robot intentionality, if or when robots become a part of human society, we will need to come to some kind of conclusion anyway.

3. SUMMARY AND CONCLUSION

To summarize, the proverbial double-edged sword mentioned in the title is this: The human cognitive capacity/tendency to interpret behavior as intentional is central to embodied social interactions and to our way of interpreting both the animate and the inanimate world. For human-robot social interaction, this capacity is likely to be very useful because it tends to "fill in the gaps" and make interactions more natural and trustworthy. However, the tendency to attribute human-like mental states also comes with the tendency to view things as more human-like than they maybe really are. For cartoon characters this is obvious, and most people have no problems at all to understand that Donald chases Chip and Dale because he is angry and wants his popcorn back, and at the same time understand that neither Donald nor the chipmunks are real, and therefore none of them eats or wants popcorn anyway. For robots, this is much less obvious, because unlike cartoon characters, they are real, physical, 'embodied', etc., they are physically and socially interactive, and in the humanoid case they often also look and behave human-like to some degree.

The question therefore is if their apparent intentionality is only ascribed by the observer, as in the case of cartoon characters, or in fact is genuine and intrinsic to those robots themselves. Your answer to the question is likely to depend on your conception of embodied cognition – and the body underlying it. If you adopt the current mainstream position of robotic functionalism, according to which intentionality arises from the physical body's sensorimotor interaction with the environment, then intrinsic intentionality in robots is at least possible. If, on the other hand, you adopt the view that embodied cognition is ultimately grounded in the living body, then the intentionality of at least current-technology robots is necessarily only ascribed. Which of these positions we, as a society, adopt in the future is likely to have significant social and ethical consequences for the way we deal with robots.

4. ACKNOWLEDGMENTS

Supported by the European Commission, FP7 project 611391, DREAM (*Development of robot-enhanced therapy for children with autism spectrum disorders*), and the Knowledge Foundation, SIDUS project AIR/TINA (*Action and intention recognition in human interaction with autonomous systems*).

5. REFERENCES

[1] Sciutti, A., Bisio, A., Nori, F., Metta, G., Fadiga, L., and Sandini, G. 2014. Robots can be perceived as goal-oriented agents. *Interaction Studies*. 14, 3, 329-350.

[2] Heider, F., Simmel, M., 1944. An experimental study of apparent behavior. *American J. of Psychology* 57, 243–259.

[3] Waytz, A., Heafner, J., & Epley, N. 2014. The mind in the machine. *J. of Exp. Soc. Psychology* 52*,* 113-117.

[4] Searle, J. 1980. Minds, brains, and programs. *Behavioral and brain sciences* 3, 3 , 417-424.

[5] Harnad, S. 1989. Minds, machines and Searle. *Journal of Experimental & Theoretical Artificial Intelligence* 1, 1, 5-25.

[6] Zlatev, J. 2001. The epigenesis of meaning in human beings, and possibly robots, *Minds and Machines* 11, 2, 155-195.

[7] Varela, F. 1997. Patterns of Life: Intertwining Identity and Cognition. *Brain & Cognition* 34, 72-87.

[8] Ziemke, T. 2008. On the role of emotion in biological and robotic autonomy. *BioSystems* 91, 401-408.

[9] Froese, T., and Ziemke, T. 2009. Enactive artificial intelligence. *Artificial Intelligence* 173, 466-500.

[10] Chemero, T. 2009. *Radical embodied cognitive science*. MIT Press, Cambridge, MA.

[11] Vernon, D. 2010. Enaction as a conceptual framework for developmental cognitive robotics. *Paladyn* 1, 2, 89-98.

[12] Vernon, D. 2014. *Artificial cognitive systems: A primer*. MIT Press, Cambridge, MA.

State Prediction for Development of Helping Behavior in Robots

Jimmy Baraglia*
Osaka University, Department of Adaptive Machine System
2-1 Yamadaoka, Suita
Osaka, Japan

Yukie Nagai
Osaka University, Department of Adaptive Machine System
2-1 Yamadaoka, Suita
Osaka, Japan

Minoru Asada
Osaka University, Department of Adaptive Machine System
2-1 Yamadaoka, Suita
Osaka, Japan

ABSTRACT

Robots are less and less programmed to execute a specific behavior, but develop abilities through the interactions with their environment. In our previous studies, we proposed a robotic model for the emergence of helping behavior based on the minimization of the prediction-error. Our hypothesis, different from traditional emotion contagion models, suggests that minimizing the difference (or prediction-error) between the prediction of others' future action and the current observation can motivate infants to help others. Despite promising results, we observed that the prediction of others' actions generated strong perspective differences, which ultimately diminished the helping performance of our robotic system. To solve this issue, we propose to predict the effects of actions instead of predicting the actions per se. Such an ability to predict the environmental state has been observed in young infants and seems promising to improve the performance of our robotic system.

1. INTRODUCTION

Young infants, from the beginning to the middle of their second year of life, are able to altruistically help others with no expectation of future rewards [7, 5, 4]. Traditional approaches suggest that an early form of empathy, or emotional contagion, is the primary behavioral motivation for young infants to act altruistically [7, 2, 3]. Yet, recent experiments tend to show that a more general source of motivation prompts infants to help others achieving their unfulfilled goal [4]. To better understand the origin of altruistic behavior and to program this ability into robots, we developed a hypothesis for the emergence of altruistic behavior in which infants are not motivated to help others based on emotional contagion, but in order to minimize the prediction-error (hereafter PE) between others' predicted future actions and current observations [1]. Although our results gave significant proofs that PE minimization could be used as a behavioral motivation for robots to help others, computing PE based on action prediction could not solve the differences between the own and others' perspective. Therefore, our robotic system failed to reliably achieve the expected helping behavior. To solve this new issue, we must change the way our robot perceives others' actions and the consequences of these actions on the environment. Warneken and Tomasello [7] showed that infants from 14 months of age could help others by handing out an out-of-reach object directly to others, with almost no cases where infants kept the object. This seems to indicate that infants prefer to perform actions that would help achieving others' goals, rather than imitating the predicted actions. Furthermore, other evidences strongly suggest that infants, already from the age of 3 to 5 months, represent actions in terms of goals, which is the relation between actors and objects. [6, 8].

Based on these evidences, it is clear that infants predict the goal of observed actions rather that the actions themselves. Our model then needs to predict the future goal, or targeted state, of an action and to estimate PE when the state is not achieved as predicted. Consequently, PE will be minimized when the goal is reached either by others or by the robot regardless of the mean. The rest of this paper is organized as follows: first, each module of our model is briefly described, then the expected results are presented. Finally a conclusion based on our previous results and literature evidences is given.

2. ROBOTIC MODEL

Our robotic model is a continuation of the work presented by Baraglia et al. [1]. This model consists of five modules and tries to minimize PE by executing actions in the environment to reach a predicted state. The details of each module are presented in the following sections.

2.1 Scene recognition

The scene recognition module recognizes the environment's state including objects and others. An important point here is that others are not differentiated from objects, instead they are detected as parts of the environment. The recognized signals were chosen based the developmental studies previously presented [6, 8].

2.2 Action-state memory

The action-state memory is built as a Markov decision process (hereafter MDP) based on the robot's own experience of executing actions. When an action performed by the robot changes the environment's state, the action and

*email: jimmy.baraglia@ams.eng.osaka-u.ac.jp

Human Robot Interaction 2015, Portland, OR, USA

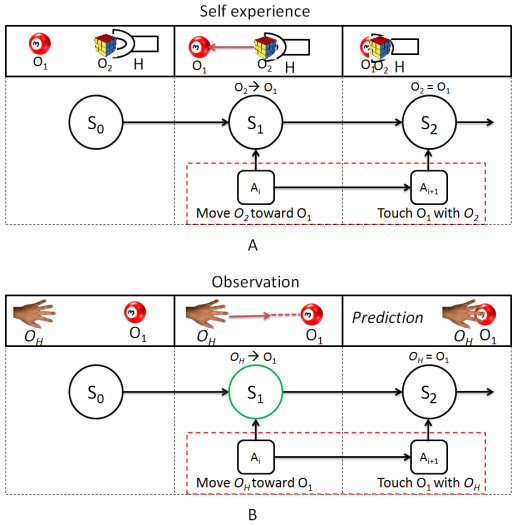


Figure 1: Example of action-state memory. A: the system updates his action-state memory by experiencing the action "Moving an object O_2 toward another object O_1". B. The system generalizes its memory to other objects and recognizes the current state of O_H and O_1, namely S_1 highlighted in green.

the new state are memorized. As we assumed that others are not differentiated from the environment, the system's own experience can be generalized for the recognition of the environment's state. For instance in Fig. 1 A, the robot experienced putting two objects close to each other and can generalize this experience to recognize the state of O_H and O_1 in Fig. 1 B.

2.3 State prediction

The state prediction module estimates the future state based on the current observation and using the action-state memory. The prediction is applied to all the states recognized by the scene recognition module and the targeted goal is predicted as the possible future state with the highest probability. In Fig. 1 B, the recognized state is S_1, thus the predicted state would be the future state with the highest probability, here S_2.

2.4 Estimation of prediction-error

The estimation of prediction-error module estimates PE between the current state of the environment and the future state predicted by the state prediction module. If the predicted state is not achieved within a predicted duration, PE increases accordingly.

2.5 Minimization of prediction-error

The minimization of the prediction-error module tries to minimize PE when its value becomes larger than a predefined threshold. Using the action-state memory and the predicted future state, the system performs an action to minimize PE. For example, in Fig. 1 B, if the predicted state is S_2, the system will perform the action A_i and A_{i+1}, namely "move O_H toward O_1" and "touch O_H with O_1" to reach S_2.

3. EXPECTED RESULTS

Our previous results presented in [1] showed that estimating PE based on the prediction of actions caused strong perspective biases. For instance, if the experimenter was attempting to grasp a ball but failed during the reaching, our robotic model predicted the next action as being "grasping" and performed the same action to minimize PE. This action was successful from the robot perspective, but failed in helping the experimenter and could not replicate the behavior observed in infants. However, if the future state of the environment is predicted instead of the action, we can expect that the minimization of PE will lead to a behavior that would be helpful from the experimenter's perspective. Indeed, when observing others failing to achieve an action, the robot will first recognize the current state of the environment. In a second time, it will predict the future state based on its own experience and finally perform an action that can achieve the predicted state and minimize PE.

4. CONCLUSIONS

To solve the perspective difference, we hypothesized that our system should predict the targeted goal (or state) of an action instead of predicting the future action. By generalizing self experience to the recognition of objects' state in the scene, our robot is then able to minimize PE by performing an action that achieves the predicted state, regardless of the perspective differences. Such an approach is strongly supported by developmental studies and its benefices on the helping performances of our robotic system seem promising. Future experiments will test our assumption and prove whether the state prediction can indeed improve the emergence of altruistic behavior.

5. REFERENCES

[1] J. Baraglia, Y. Nagai, and M. Asada. Prediction error minimization for emergence of altruistic behavior. *4th International Conference on Development and Learning and on Epigenetic Robotics*, pages 281–286, Oct. 2014.

[2] F. B. M. de Waal. Putting the altruism back into altruism: the evolution of empathy. *Annual review of psychology*, 59:279–300, Jan. 2008.

[3] J. Decety and M. Svetlova. Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental cognitive neuroscience*, 2(1):1–24, Jan. 2012.

[4] B. Kenward and G. Gredebäck. Infants help a non-human agent. *PloS one*, 8(9):e75130, Jan. 2013.

[5] H. Over and M. Carpenter. Eighteen-month-old infants show increased helping following priming with affiliation: Research report. *Psychological Science*, 20(10):1189–1193, Oct. 2009.

[6] J. a. Sommerville, A. L. Woodward, and A. Needham. Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96(1):B1–11, May 2005.

[7] F. Warneken and M. Tomasello. Helping and cooperation at 14 months of age. *Infancy*, 11(3):271–294, 2007.

[8] A. L. Woodward. Infants' grasp of others' intentions. *Current directions in psychological science*, 18(1):53–57, Feb. 2009.

Social Robots and the Tree of Social Cognition

Bertram F. Malle
Brown University
Department of Cognitive, Linguistic, and Psychological Sciences
Providence, RI 02906
bfmalle@brown.edu

ABSTRACT
I depict many of the elements of human social cognition within one hierarchical structure—the tree of social cognition—and examine the implications of this structure for human responses to social robots and the proper design of such robots.

Categories and Subject Descriptors
I.2.9 [**Artificial Intelligence**] Robotics
J.4 [**Social And Behavioral Sciences**]

Keywords
Human-Robot Interaction; Social Psychology; Cognitive Robotics

1. INTRODUCTION
In response to intense demands of social life, human beings have evolved a number of capacities that allow them to make sense of other agents—to interpret, explain, and predict their behavior, share their experiences, and coordinate interactions with them. These capacities include simple processes such as gaze following or mimicry; complex processes such as imaginative simulation and mental state inference; and abstract concepts such as *intentionality* and *belief*. These capacities are typically subsumed under the label *social cognition,* but they belong together not because they form a "module" or can somehow be found in the same brain area; rather, what unites them is their responsiveness to a social environment of other intentional agents—who are minded, intelligent, and themselves engaging in social cognition.

I attempt here to integrate many of these social-cognitive capacities into one structural representation. I propose that the structure of social cognition is hierarchical in several ways: lower-order capacities (LC) are often requirements for higher-order capacities (HC); LC have weaker processing demands and can often operate continuously and unconsciously; LC develop earlier in life and are likely to have evolved earlier in human history. The evidence for these claims is distributed over a large literature, exemplified by [1], [2] on developmental orderings, [3], [4] on evolutionary orderings, and [5], [6] on processing orderings.

After introducing the structure as a whole I will comment briefly on a few key relationships among its elements. Then I turn to implications of this integrative view of social cognition for the design and deployment of "social robots"—robots that are meant to interact with, live with, and most likely grow up with humans.

2. THE TREE OF SOCIAL COGNITION
Figure 1 displays many of the capacities of social cognition arranged in an approximate hierarchy, starting on the bottom with the fundamental identification of agents in the environment and building from simpler processes to the most complex processes. (For a glossary and detailed discussion of many of these components, see [7].) In addition, the figure displays a few essential activities that human beings perform in social life but that are not strictly part of social cognition; they are more general phenomena that are *enabled* by the availability of many social-cognitive capacities.

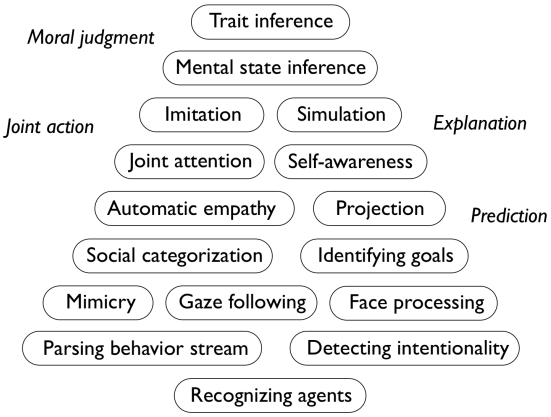


Figure 1. The tree of social cognition, with additional further enabled functions (e.g., moral judgment, joint action)

Concepts. A number of the depicted processes lead, over time, to abstractions—concepts such as *agent*, *intentionality*, *desire,* or *belief*. With experience, these concepts are triggered by a large range of stimuli and guide further processing—for example, if the *agent* concept is activated (by some triggering feature in the perceived entity), behavior parsing is initiated; when *intentionality* is activated (by some triggering feature in the parsed behavior), then goal identification, simulation, and mental state inferences are likely to be initiated as well. Moreover, these concepts stand in specific logical relations to one another that form expectations about what can actually be observed. For example, *intentionality* $\supset$ *belief* & *desire*: an intentional action implies that the agent had some desire and belief (though inferential processes have to determine what those states are).

Hierarchical dependencies. To illustrate the many hierarchical dependencies among social-cognitive capacities, consider the following sample relations (marked with x $\prec$ y for x *precedes* y):

Recognize agent & parse behavior $\prec$ detect intentionality

Detect intentionality & follow gaze $\prec$ identify goal

Parse behavior & process face & mimicry $\prec$ automatic empathy

Detect intentionality & simulation $\prec$ mental state inference

These relations will typically be probabilistic, as one can arrive at, say, a goal representation in numerous ways (e.g., infer, hear, or read about it).

Other resources. Numerous other cognitive resources aid and facilitate the more complex social-cognitive capacities. For example, simulating another's internal state must be enriched by *knowledge structures* in order to lead to an explicit mental state

inference. Some of this knowledge is activated by social categorizations (e.g., male, young) and may come with trait ascriptions that are not the result of bottom-up inferences but of top-down assumptions (Fiske Neuberg). Further, these resources and complex capacities interact to enable general functions such as explanation, prediction, and moral judgment. For example, moral judgment requires behavior parsing and mental state inferences, but it can only operate in light of a *norm system* that the perceiver knows and endorses.

3. IMPLICATIONS FOR SOCIAL ROBOTS

Two sets of implications of this hierarchical structure are worth considering. The first is what this kind of hierarchical system means for *human beings'* perception, judgments, and interaction with artificial agents. The second is what kinds of design implications for *robots* we need to take into account if the goal is to create robots that themselves have social cognitive capacities.

Humans responding to robots. The sketched hierarchical system is highly responsive to initial activations at lower levers, and its operations and processes are likely to cascade upwards once set in motion. Thus, even a more primitive artificial entity may quickly engage the human social-cognitive system. If the agent detector is triggered, for example, the system is likely to search for behavior, parses it, assigns intentionality to some of the parsed units, and tries to infer goals. It take very little (e.g., eyes and a little bit of self-propelled movement) to promptly lead to goal inferences—even when, in reality, the observed entity may have no goals. This over-responsiveness can occur at many levels of the system. If the artificial agent has gaze and facial expressions (e.g., Kismet), for example, projection will invite the social perceiver to assume the agent has interest, emotional reactions, and intentions quite similar to the perceiver's own. If the agent looks female and young (as the new headline-making robot "Pepper"), knowledge and expectations are likely to be triggered about personality or ability traits. And once intentionality, mental states, and traits have been inferred and the robot's behavior violates a norm, moral judgment may well be a natural human response.

The upshot of this brief analysis is that it may take very little to trigger the components of the human social-cognitive system, and because of the system's structure, any one triggered component will engage others as well. A robot that has such power in engaging human social cognition but that does not have all the inferred or expected properties may easily disappoint, mislead, confuse, and ultimately be rejected. The implication for robot design is clear: We need to know which features of robots trigger which social-cognitive processes (and which other components the latter initiate in turn). Robots that have these triggering features either must have the abilities, states, and traits that human partners are prone to infer, or else the robot must communicate its own limitations or its early stage of development, and perhaps invite the human partner to tutor, coach, and better the robot.

Robots' social cognition. The growing sophistication and proliferation of robots in society pose a challenge to science—given that robots of the near future are likely to participate in many aspects of human social life, such as healthcare, education, and law enforcement [8], what kinds of robots can safely take these roles? Among many important requirements for such robots, two closely related ones are autonomy and social cognition. Why do robots need autonomy? Because otherwise they would utterly fail in the demanding tasks of interacting with humans, collaborating with them, and taking care of them. Successful performance in such tasks cannot simply rely on prior programs, because human behavior is too complex, variable, and therefore difficult to predict, and if robots have difficulty predicting their interaction partners' behavior, their own behavior in social interactions cannot be pre-programmed. Instead, the robot must monitor a person's responses to small changes in the situation and in turn flexibly respond to them. This flexibility requires autonomous decision making and, as the input to such decision making, social cognition. Human behavior may be hard to predict but it is a whole lot easier if the perceiver can infer goals, beliefs, emotions, and skills, and if the perceiver knows the social and moral norms that the human agent is obeying—because these norms are good predictors of at least aggregate behavior. So the successful social robot of the future will need to be autonomous and adept at social (and moral) cognition.

How, then, should we build such a robot? Does it need to have the same hierarchical tree of social cognition that humans have? I would argue, yes, in many respects its social cognition must be similar (at least functionally) so that humans can interact with the robot just like they expect to interact with other, real human agents. The human expects interest, trust, politeness, loyalty, and many other social traits from their interaction partners, and robots better deliver on those expectations.

Finally, here is where the two sets of implications converge: If human social cognition weren't so easily triggered, robots could have all kinds of features and properties, even preprogrammed and highly limited ones. But because we can fairly confidently assume that people *do* form strong expectations about their interaction partners, robots must meet these demands. Not only must they be smart, communicative, and helpful, but they need to show interest in "other minds," in the mental states and personality of their human interaction partners.

4. REFERENCES

[1] H. M. Wellman, D. Cross and J. Watson, Meta-analysis of theory-of-mind development: The truth about false belief, *Child Development* **72** (2001), 655–684.

[2] D. Poulin-Dubois, I. Brooker and V. Chow, The developmental origins of naïve psychology in infancy, *Advances in Child Development and Behavior* **37** (2009), 55–104.

[3] D. J. Povinelli and T. M. Preuss, Theory of mind: Evolutionary history of a cognitive specialization, *Trends in Neurosciences* **18** (1995), 418–424.

[4] J. Call and M. Tomasello, Does the chimpanzee have a theory of mind? 30 years later, *Trends in Cognitive Sciences* **12** (2008), 187–192.

[5] B. F. Malle and J. Holbrook, Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality, *Journal of Personality and Social Psychology* **102** (2012), 661–684.

[6] F. Van Overwalle, M. Van Duynslaeger, D. Coomans and B. Timmermans, Spontaneous goal inferences are often inferred faster than spontaneous trait inferences, *Journal of Experimental Social Psychology* **48** (2012), 13–18.

[7] B. F. Malle, The fundamental tools, and possibly universals, of social cognition, in *Handbook of motivation and cognition across cultures*, R. M. Sorrentino and S. Yamaguchi, Eds. New York, NY: Elsevier/Academic Press, 2008, pp. 267–296.

[8] I. R. Nourbakhsh, *Robot futures*. Cambridge, MA: MIT Press, 2013.

Predictive coding and the Uncanny Valley hypothesis: Evidence from electrical brain activity

Burcu A. Urgen
Cognitive Science, UC San Diego
9500 Gilman Drive, La Jolla, CA
92093-0515
+1-858-822-1994
burgen@cogsci.ucsd.edu

Alvin X. Li
Cognitive Science, UC San Diego
9500 Gilman Drive, La Jolla, CA
92093-0515
+1-858-822-1994
axl002@ucsd.edu

Chris Berka
Advanced Brain Monitoring, Inc.
2237 Faraday Ave, Suite 100,
Carlsbad, CA, 92008
+1-760-720-0099
chris@b-alert.com

Marta Kutas
Cognitive Science, UC San Diego
9500 Gilman Drive, La Jolla, CA
92093-0515
+1-858-534-2440
mkutas@ucsd.edu

Hiroshi Ishiguro
System Innovation, Graduate
School of Engineering Science,
Osaka University, Japan
+81-6-6850-6360
ishiguro@sys.es.osaka-u.ac.jp

Ayse P. Saygin
Cognitive Science, UC San Diego
9500 Gilman Drive, La Jolla, CA
92093-0515
+1-858-822-1994
asaygin@cogsci.ucsd.edu

ABSTRACT

The uncanny valley hypothesis suggests that robots that are humanoid in appearance elicit positive and empathetic responses, but that there is a point where the robot design is very close to human, the robot becomes repulsive [1]. A possible mechanism underlying this phenomenon is based on the predictive coding theory of neural computations [2,3]. According to this framework, certain neural systems in the brain can ascribe humanness to a robot that is highly human-like in its appearance, and if the robot's behavior does not match in realism to the appearance, there will be a processing conflict the neural network will need to resolve. Although this hypothesis is consistent with previous results in the field, empirical work directly testing it is lacking. Here we addressed this gap with a cognitive neuroscience study: We recorded electrical brain activity from the human brain using electroencephalography (EEG) as human subjects viewed images and videos of three agents: A female adult (human), a robot agent closely resembling her (android), and the same robot in a more mechanical appearance (robot). The human and robot had *congruent* appearance and movement (human with biological appearance and movement; robot with mechanical appearance and movement), and the android had *incongruent* appearance and movement (biological appearance but mechanical movement). We hypothesized that the android would violate the brain's predictions since it has a biological appearance, but mechanical movement, whereas the other agents would not lead to such a conflict (robot looks mechanical and moves mechanically; human looks biological and moves biologically). We focused on the N400 ERP component derived from the EEG data. Since the N400 has a greater amplitude for anomalies and violations based on preceding context, we hypothesized the amplitude would be significantly greater for the android in the moving condition than the still condition, whereas the moving and still conditions of the robot and human stimuli would not differ. Our results confirmed out hypothesis, indicating that the uncanny valley might at least partially be due to violations of the brain's internal predictions about almost-but-not-quite-human robots. Interdisciplinary studies like this one not only allows us to understand the neural basis of human social cognition but also informs robotics about what kind of robots we should design for successful human-robot interaction.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, human information processing*.

General Terms

Human Factors, Experimentation, Theory

Keywords

Android science, predictive coding, social cognition, action perception, cognitive neuroscience, neuroimaging

1. INTRODUCTION

As humanoid robots become participants in our lives in areas such as education, healthcare, and entertainment, we need to consider an important issue: How should we design artificial agents so that humans socially accept them and can interact with them successfully? An intuitive approach might be to make the robots as humanlike as possible so that they will be more familiar and tap into neural systems for social cognition that are already well-developed in the human brain.

However, increasing humanlikeness does not necessarily result in increasing acceptance [4]. The uncanny valley is a phenomenon that refers to people's response to artificial agents

such as robots and animated characters that possess almost but not exactly human-like characteristics. [1], who introduced the term, proposed that the relationship between humanlikeness and people's response to artificial agents is not a linear one. Instead, increasing humanlikeness would elicits positive responses up to a certain point, whereafter increasing humanlikeness starts to elicit negative responses, which forms a "valley" where the agents are perceived to be creepy, odd, zombielike, or disturbing (Figure 1). Mori also posited that the effects would be more pronounced if the agent were moving rather than stationary.

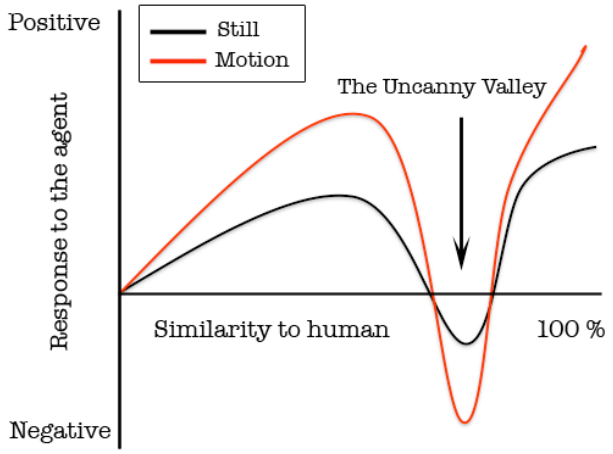


Figure 1. A depiction of Mori's proposal [1], plotting the expected human responses as a function of a robot's humanlikeness. The uncanny valley refers to the sharp dip in the acceptability of the robot as the appearance becomes increasingly humanlike. Note that Mori expected motion would exaggerate the effect and deepen the valley.

As humanoid robots became more feasible to develop in recent years, the uncanny valley became a frequently discussed issue from both a theoretical and a practical viewpoint. [5] contended that the social, cognitive, and neurosciences would be invaluable if we were to understand this intriguing phenomenon. Indeed, empirical studies have recently been exploring the anecdotally well-known, but scientifically uncharted valley. Several subjective rating studies with a range of humanoid stimuli claimed that the uncanny valley might be a legitimate psychological phenomenon: For example, [6,7,8] used computer-animated faces and asked human subjects to rate such dimensions as humanlikeness, eeriness, attractiveness or pleasantness. In a similar fashion, [9] recently used human, robot, and prosthetic hand stimuli and reported eeriness ratings that were broadly compatible with the hypothesis. [10] also reported evidence for the hypothesized valley for very realistic humanoid agents in a study that collected social acceptability ratings with full-body computer animated agents as stimuli. However, their data did not support Mori's proposal that there would be a more pronounced effect with moving stimuli. More broadly, some studies did not reveal evidence for Mori's hypothetical curve with experimental data. [11] varied several motion parameters and explored how they influenced humanlikeness, familiarity, and eeriness ratings of human avatars, and did not find results resembling the hypothesized uncanny valley. The inconsistencies between the studies may be due to different dependent measures that were used in the ratings (e.g., likeability is a complex measure that is correlated with the humanlikeness dimension that is used as the x-axis of the Mori graph [12,13].

In addition to rating studies, researchers have attempted to use less explicit measures such as gaze behavior to characterize the uncanny valley. Using eye-tracking and a parametrically varying set of avatar faces, [14] showed that ambiguous avatar faces (i.e., those that are at the category boundary between human and avatar) required greater depth of processing in the eye and mouth regions compared with unambiguous avatar faces. Similarly, [15] found that by 9-10 months of age, infants looked longer to highly familiar and strange faces compared with morphed faces near the boundary these categories. Furthermore, [16] reported that that monkeys looked longer at faces of monkey-like agents that were either of their own species or unrealistic animations compared with very realistic animations. Although this suggests the uncanny valley has earlier evolutionary origins, [17] used analyses of infants' gaze behavior that early exposure to typical human faces in development constrained uncanny valley-like responses.

Although these and similar recent studies have been a good step to scientifically characterize the uncanny valley, the underlying mechanism remains unclear. Possible mechanisms that have been proposed include threat or disease avoidance, mate selection, and Bayesian estimation or predictive coding hypotheses [6, 18, 19]. The latter hypothesis is linked to a more general description of neural computational properties of the brain [3, 20], and therefore promises a scientifically testable framework. According to predictive coding, the uncanny valley is related to expectation violations in neural computing when the brain encounters almost-but-not-quite-human agents. A growing body of work has linked Mori's hypothetical curve to the processing of conflicting perceptual or cognitive cues, varying whether the stimuli are compatible with the elicited expectations or are in violation of them [13, 14, 19, 21-25].

Behavioral studies alone are insufficient to directly test and identify mechanisms that underlie the uncanny valley, or to distinguish between alternative theories for numerous reasons. First, dependent measures such as subjective ratings require overt responses, whereas the uncanny valley phenomenon might be better studied with covert responses of humans' subjective states that may even occur outside of awareness (cf. [26]). In addition, it is difficult to characterize a complex phenomenon with a single measurement such as pleasantness, familiarity, or eeriness as each can imply different cognitive and emotional states, and captures uncanny valley curve for different robot characteristics [27]. Second, behavioral studies only provide the output of the system, and do not address what kind of information processing underlies the phenomenon. Although methods such as eye-tracking or automatic attention paradigms have advantages over rating studies in this respect, to provide a mechanistic account of the uncanny valley, methods from social and cognitive neuroscience are likely to be more fruitful [16, 28-31].

Neuroscience methods such as neuroimaging have advantages that can help "demystify" the uncanny valley. First of all, there is the potential to provide valid dependent measures that can be used to operationalize the uncanny valley, and to situate it as part of a cognitive domain. Decades of cognitive neuroscience research have informed us about the basic functions of the human brain, perception and social cognition. It would be fruitful to use accumulated knowledge in these areas to inform robotics about how humans respond to and interact with social

agents [5, 30, 31]. Second, neuroimaging does not require overt responses since brain activity can be monitored on an ongoing basis. Last but not least, neuroimaging provides a rich a set of data, which can be more informative than individual behavioral measures. Temporally sensitive methods in particular provide a means to understand the time course of processing in comparison to ratings or reaction times, which only provide the output of the system. Overall, neuroimaging research has the potential to reveal the underlying mechanisms of the uncanny valley phenomenon.

Indeed, there is now growing interest in using neuroimaging in the field. [22] used functional magnetic resonance imaging (fMRI) with a face stimulus set along a human-avatar continuum and found that ambiguous faces at the category boundary of human and avatar are processed differentially than the unambiguous faces within each category. Behavioral studies with animated faces support this category conflict explanation for uncanny valley, which is in line with the prediction error hypothesis [18, 32, 33]. Another example is our previous work [19], which used fMRI and based on the results, proposed predictive coding as a framework for future studies on the uncanny valley and the underlying mechanisms. In this study, brain responses to body movements of agents of varying degrees of humanlikeness with and without conflicting perceptual cues was compared. The agents were a human with biological appearance and biological motion, a very human-like robot (referred to as android) with biological appearance but mechanical motion, and a less human-like robot with mechanical appearance and mechanical motion (Figure 3). Notably, neural activity, especially in the parietal cortex differentiated the android from the other two agents. The data suggested, based on the functional properties of this brain region in the social cognition network, that the uncanny valley might be related to the violation of the brain's internal predictions due to conflicting perceptual cues (appearance and motion). The human and robot agents exhibited congruent appearance and motion profiles (i.e. human looks biological, moves biologically; robot looks mechanical, moves mechanically) whereas the android exhibited incongruent appearance and motion (looks biological but moves mechanically). For the latter agent, the human appearance would elicit predictions that the motion will also be; when the agent instead moves mechanically, the brain network processing the agent would show evidence of processing the violations. The differential activity measured in parietal cortex could reflect this prediction error [2, 19, 34].

Although [19] used neuroimaging to situate the uncanny valley phenomenon in the scientific context of violation of predictions and the predictive coding theory of neural computations, the study was not a priori designed to test this theory. Thus, the proposed framework ideally needs to be further validated with independent experiments. Furthermore, fMRI has methodological limitations, most notably due to its limited temporal resolution.

Electroencephalography (EEG) is an alternative neuroimaging method that allows recording brain activity with electrodes located on the scalp with excellent temporal resolution (Figure 2). Importantly, a specific dependent measure derived from EEG, the N400 event-related potential (ERP) component, is an ideal measure with which to test the prediction error hypothesis or the uncanny valley. N400 is a negative-going ERP, which peaks around 400 ms after stimulus onset, and is maximal in fronto-central regions of the human scalp [35] (Figure 2). Although the N400 is elicited in response to any meaningful stimulus, its amplitude is greater for semantically or contextually anomalous stimuli (i.e., items that violate expectations).

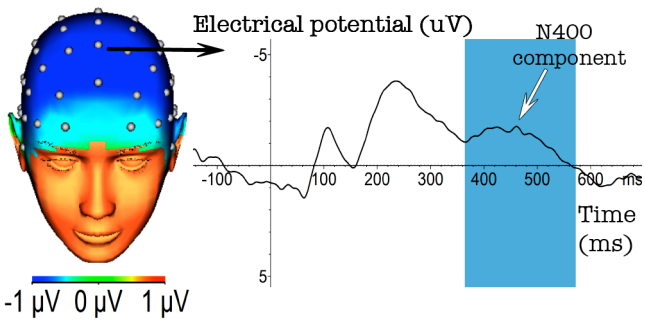


Figure 2. Depiction of the event-related brain potential N400 component in a representative frontal channel (Fz) on the human scalp, which usually peaks in the time interval between 300-600 ms. N400 generally is distributed over the fronto-central channels on the human scalp for non-linguistic visual stimuli [35].

In the present study, we used EEG, and report on analyses of the N400 component to directly test the prediction error hypothesis for the uncanny valley phenomenon. We presented agents of varying humanlikeness in still and moving forms as we recorded EEG (Figure 3). We used the same stimuli as the fMRI study by [19]: a human agent with biological appearance and motion, an android with biological appearance and mechanical motion, and a robot with mechanical appearance and motion (Figure 3). We hypothesized that the android would elicit a greater N400 in the moving condition than the still condition, as its mechanical movement would result in violation of subjects' predictions due to the biological appearance it possesses. The N400 amplitude for the still and moving conditions would not differ for the robot and human agents, who possess appearance-motion congruence. Such a pattern of activity would provide strong evidence for the prediction error hypothesis, and inform us about the timing of the uncanny valley phenomenon.

2. METHODS

2.1 Participants

Twenty right-handed adults (10 females; mean age = 23.8; SD = 4.8) from the student community at University of California, San Diego with normal or corrected-to-normal vision, and no history of neurological disorders participated in the study. Informed consent was obtained in accordance with the university's Human Research Protections Program. Participants were paid $8 per hour or received course credit. One subject's data was excluded due to high noise during EEG recording.

2.2 Stimuli

Stimuli consisted of video clips of actions performed by the humanoid robot Repliee Q2 (in Robotic and Human-like appearance, Figure 3 left and middle images, respectively) and by the human 'master', after whom Repliee Q2 was modeled (Figure 3, right image). We refer to these agents as the Robot, the Android (dressed up robot), and the Human conditions (even though the former two are in fact the same robot).

Biological Motion	No	No	Yes
Biological Appearance	No	Yes	Yes
Congruent Motion and Appearance	Yes	No	Yes

Figure 3. Sample frames depicting the three agents: Robot, Android, Human. Robot and Human had congruent motion and appearance, whereas Android had incongruent motion and appearance. In the present EEG experiment, both the still and moving forms of these agents were used.

Repliee Q2 has 42 degrees of freedom and can make face, head and upper body movements [36]. The robot's movements cannot match the dynamics of biological motion; it is mechanical or "robotic". The same movements were videotaped in two appearance conditions. For the Robot condition, Repliee Q2's surface elements were removed to reveal its wiring, metal arms and joints, etc. The silicone 'skin' on the hands and face and some of the fine hair around the face could not be removed but was covered. The movement kinematics for the Android and Robot conditions was identical, since these conditions comprised the same robot, carrying out the very same movements. For the Human condition, the female adult whose face was molded and used in constructing Repliee Q2 was videotaped performing the same actions. She was asked to watch each of Repliee Q2's actions and perform the same action naturally. All agents were videotaped in the same room with the same background. Video recordings were digitized, converted to grayscale and cropped to 400x400 pixels. Videos were clipped such that the motion of the agent began at the first frame of each video.

2.3 Procedure

Since prior knowledge can affect judgments of artificial agents differentially [37], each participant was given exactly the same introduction to the study and the same exposure to the videos. Before starting EEG recordings, participants were shown each video and told whether each agent was a human or a robot, and the name of the action. Participants went through a practice session before the experiment.

EEG was recorded as participants watched video clips of the three agents performing eight different upper body actions (drinking from a cup, examining an object with hand, handwaving, turning the body, wiping a table, nudging, introducing self, and throwing a piece of paper) (Figure 2A). The videos were presented in two modes that we call *motion alone* and *still-then-motion*. In the *motion-alone* condition, 2-second videos were presented. In the *still-then-motion* condition, the first frame of the video was presented for 600-1000 ms (with a uniform probability jitter), and then the full video was played. The experiment consisted of 15 blocks. In each block, the eight videos of each agent were presented once in the *motion-alone* condition, and once in the *still-then-motion* condition. Stimuli were presented in a pseudo-randomized order ensuring that a video was not repeated on two consecutive trials. Each participant experienced a different pseudo-randomized stimuli sequence.

Stimuli were displayed on a 19" Dell Trinitron CRT monitor at 90 Hz using Psychophysics Toolbox [38, 39]. To prevent an augmented visual evoked potential at the beginning of video onset that might occlude subtle effects between conditions, we displayed a gray screen with a white fixation cross before the start of the video clip or still frame on each trial. Participants were instructed to fixate the fixation cross at the center of the screen for 900-1200 ms (with a uniform probability jitter). A comprehension question was displayed every 6-10 trials, asking participants a true/false question about the action in the just seen video (e.g. Drinking?), after which they responded with a manual key press (Yes/No response).

2.4 EEG Recording and Data Analysis

EEG was recorded at 512 Hz from 64 ActiveTwo Ag/AgCl electrodes (Brain Vision, Inc.) following the International 10/20 system. The electrode-offset level was kept below 25 k-Ohm. Two additional electrodes were placed above and below the right eye to monitor oculomotor activity (1 additional electrode was placed on the forehead as a ground of the eye electrodes). The data were preprocessed with MATLAB and the EEGLAB toolbox [40]. Each participant's data were first high-pass filtered at 1 Hz, low-pass filtered at 50 Hz, and re-referenced to average mastoid electrodes behind the right and left ear. Then the data were epoched ranging from 200 ms preceding video or first frame onset to 700 ms after video onset, and were time-locked to the onset of the video clips (*motion-alone* condition, see Procedures) or the first frame (*still-then-motion* condition, see Procedures) to compare the motion and still forms of the agents (we refer to these as *motion* and *still* conditions from now on). Atypical epochs of electromyographic activity were removed from further analysis by semi-automated epoch rejection procedures (kurtosis and probability-based procedures with standard deviation = 6). After preprocessing, grand average event-related brain potentials (ERP) and scalp topographies were computed and plotted for each condition using Brain Vision Analyzer.

2.5 Statistical Analysis

The time window between 370-600 ms was considered for N400 analysis based on the grand average ERPs across all conditions. The area under curve measure was used to extract the N400 values for each agent under both motion and still condition for each subject in frontal channels (AF3, AFz, AF4, Fz, F1, F2, F3, F4, F5, F6). After preprocessing, data were exported to ERPLAB (http://erpinfo.org/erplab) and area under curve measures were extracted by using this toolbox.

We then applied paired t-tests on the average frontal channel activity to compare the motion and still conditions for each agent (Robot, Android, Human). Since we expected motion condition to be greater than the still condition for Android (and no effect for Human and Robot), our t-tests were one-tailed.

2.6 Localization of the EEG activity

For identifying the neural generators (sources) of the activity during the N400 period, we used the LORETA method [41]. LORETA estimates the distributed neural activity in the cortex based on the scalp measurements of ERP differences. Localization of the EEG activity was as follows: First, we computed the N400 differences between the static and motion conditions of each agent (Robot, Android, Human), and then we took the grand average of the N400 differences. We then applied LORETA to the N400 difference waveform in the time interval between 370-600 ms to estimate the distributed neural activity underlying N400.

3. RESULTS

3.1 N400 component

Our results indicate that observation of all agents elicited an N400 component regardless of the presentation mode (still or motion) in frontal sites (electrodes AF3, AFz, AF4, Fz, F1, F2, F3, F4, F5, F6; Figure 4A shows ERPs on a representative frontal channel Fz). The amplitude of N400 (measured with area under curve between 370-600 ms averaged across all frontal channels) in the still condition was significantly greater than the motion condition for Android ($t(18) = 2.401$, $p<0.05$), whereas still and motion conditions did not differ neither for Robot ($t(18) = 0.388$) nor for Human ($t(18) = -0.346$) (Figure 4A for ERPs, Figure 4B for bar graphs, and Figure 4C for scalp topographies to see the distribution of the effect on the whole scalp).

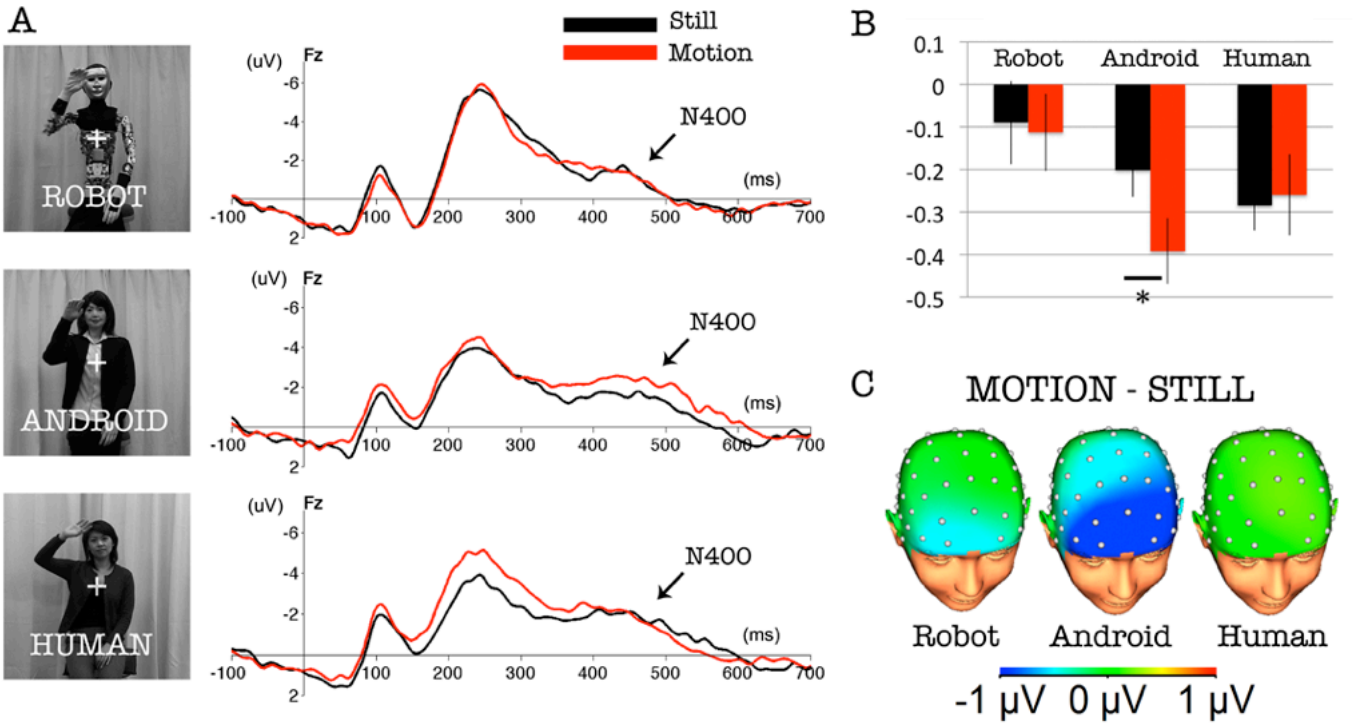


Figure 4. (A) ERP plots of a representative frontal site (Fz) for still and motion conditions for each agent, (B) Bar graphs representing the amplitude (area under curve) for N400 (370-600 ms) for each of the conditions. N400 amplitude was significantly greater for motion than still condition for the Android (* $p < 0.05$), whereas they did not differ for Robot or Human ($p > 0.05$), (C) ERP scalp topographies representing the difference between still and motion conditions for each agent in the time interval of the N400 (370 ms – 600 ms).

3.2 Localization of the EEG activity

Our source analysis with LORETA, which estimates the brain regions that generate the EEG activity, suggests that the generator of the N400 component is a distributed network including the middle and superior temporal areas, temporal-parietal junction, and prefrontal areas (Figure 5, all agents' motion-still differences collapsed). These areas align with the network that has been implicated for N400 with intracranial recordings and MEG in humans [35]. More interestingly, the maximal source density of this network was identified as Brodman area 40 (x = -60, y = -32, z = 29, MNI coordinates) in the inferior parietal lobule (Figure 5), which is the same area that differentiated the agent with appearance-motion incongruence (Android) from the other agents in an independent fMRI study with excellent spatial resolution [19].

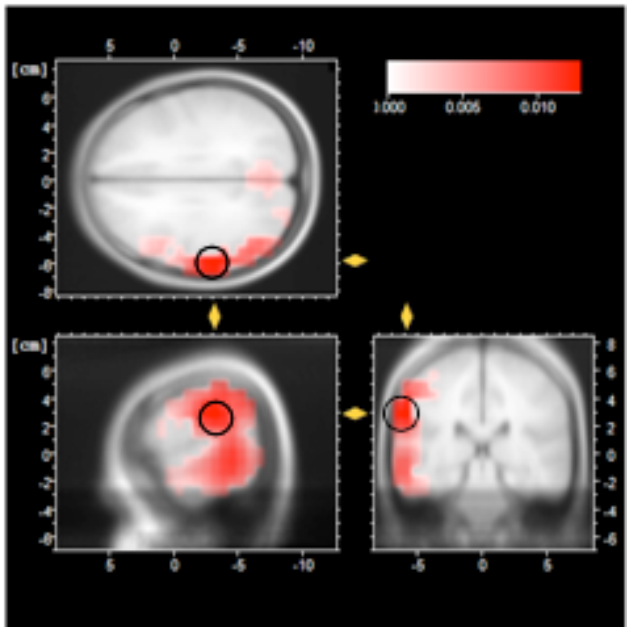


Figure 5. Distributed brain network that is computed as the most probable generator of the N400 component of interest: middle and superior temporal areas, temporal-parietal junction, and prefrontal cortical areas. The maximal data point was found to lie in the inferior parietal lobule (marked with the circles), highly consistent with the prior fMRI study [19].

4. DISCUSSION

In the present study, we tested the predictive coding as a potential underlying mechanism of the uncanny valley phenomenon. To this end, we used an established method from cognitive neuroscience, namely EEG, specifically focusing on the event related brain potential N400, which has been reliably associated with violation of predictions. Our stimuli consisted of three agents that had different levels of humanlikeness in appearance and motion dimensions: a human agent with biological appearance and motion, an android with biological appearance and mechanical motion, and a robot with mechanical appearance and motion. In this design, human and robot agents exhibited congruence in appearance and motion dimensions, whereas android agent had incongruity in appearance and motion. The agents were presented both still and moving, with the hypothesis that the android condition would elicit an N400 differential (motion-still difference) due to its incongruent appearance and motion, whereas human and robot agents would not, as they had congruent appearance and motion. Our results confirmed these predictions: the moving android elicited a greater N400 component than the still android, whereas no difference was found for the moving and still presentations of the human and robot agents. Thus, these results provides support for the hypothesis that uncanny valley might involve the violation of the brain's predictions.

Our study demonstrates the benefit of using neural dependent measures in testing hypotheses about uncanny valley, whose underlying mechanism has remained unknown. Previous research mainly has focused on behavioral ratings in studying the uncanny valley. While these efforts have been a good step to operationalize the uncanny valley, they fall short for a number of reasons. For one thing, these studies generally ask for an explicit (or conscious) response in a certain dimension such as humanlikeness, eeriness, or familiarity. However, explicit measures might be too restrictive and might not be sufficient to characterize the reaction of the human subjects for uncanny stimuli. Neuroimaging has the advantage to measure human responses implicitly without asking for a specific response. In the present study, N400 was used as such an implicit measure. For another thing, behavioral measures provide only the output of the system (one data point), which is not very informative about the processing stages. Neuroimaging provides a rich set of data, and especially the temporally sensitive methods such as EEG allows one to monitor the information processing during stimulus presentation. In addition, well-established dependent measures, such as N400 as used in this study, help one to situate the uncanny valley in a well-studied cognitive domain.

The use of event-related brain potentials in the present study is complementary to our previous neuroimaging study of action perception that used fMRI with the same stimuli. [19] found differential activity in parietal cortex for the android compared with the human and robot, which was interpreted as supporting evidence for the hypothesis that uncanny valley might be due to the incongruity of appearance and motion in the action processing network. The N400 effect for the android in the present study corroborates this interpretation. Using EEG has allowed us to link the uncanny valley phenomenon to cognitive processing using the well-established dependent measure N400.

The current study has broader implications for future research on characterizing the uncanny valley. First of all, the appearance-motion incongruence presented in the current study is one specific violation of one's predictions. In fact, [21] has shown that conflicting visual and auditory cues (appearance and voice) increase eeriness ratings in evaluating human and robot agents. In another study, [23] showed that incongruent appearance and touch (non-humanlike appearance and humanlike touch) resulted in fear from the robot in human subjects. Based on their exploratory rating study with a number of robot videos, [13] suggest that uncanny valley effect might be seen for agents that have mismatches in a variety of dimensions. On the other hand, robots differ from humans in a variety of ways, not only in their physical properties such as appearance, motion, and voice, but also in the way they accomplish tasks. In fact, [25] showed that children of 2-3-year-old showed different behavioral patterns based on the congruity of the robot appearance and contingency of its behavior upon their own behavior. Similarly, it has been suggested that congruity of facial expressions and actions determines uncanny valley

responses [24]. Thus, creating broader range of violations of one's predictions could allow us to understand the sensitivity of humans to deviation from human dimensions.

In addition, individual differences are a recent highlighted aspect of uncanny valley [42, 43] suggesting that people may show different patterns of reactions to humanoid robots, and this could well be studied with neural dependent measures. The role of experience and learning can also be studied, by testing people who have been exposed to robots or animated characters compared to those who have not. [44]. Previous research suggests that culture (e.g., western vs. eastern) as well as context are important factors in perceiving and interacting with humanoid robots [45]. Thus, these different groups' expectations from robots might differ considerably from each other, resulting in different brain activity patterns in general.

In conclusion, our study demonstrates that studying the uncanny valley with neuroscience methods can help us not only understand the underlying mechanisms but it can also inform human robot interaction. Furthermore, the uncanny valley serves as a window into human social cognition. Studying human brain responses when viewing robots has allowed us to study the functional properties of the neural systems that underlie agent and action processing, which are the building blocks for social interaction [5, 19, 29-31]. On the other hand, our study also gives insights about the design parameters of robots that will interact with humans. Based on the current data and emerging underlying mechanisms of the uncanny valley, we suggest the human brain's expectations from a human-like agent should be considered in the design process for successful human-robot interaction. More broadly, we demonstrate that interdisciplinary work can not only improve our understanding of human-robot interaction, but also make individual contributions to both neuroscience and robotics.

5. ACKNOWLEDGMENTS

This research was supported by NSF (CAREER BCS1151805), DARPA, Kavli Institure for Brain and Mind and the Qualcomm Institure (Calit2). We thank Intelligent Robotics Laboratory at Osaka University for the preparation of the stimuli, Akila Kadambi, Wayne Khoe, Edward Nguyen and Markus Plank for assistance in data collection and analysis.

6. REFERENCES

[1] Mori, M. 1970. The uncanny valley. *Energy,* 7(4): 33-35.

[2] Kilner, J.M., Friston, K.J., Frith, C.D. 2007a Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, 8:159-166.

[3] Friston, K. 2010 The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11:127-138.

[4] Pollick, F.E. 2010. In search of the uncanny valley. *UCMedia 2009, LNICST 40*, 69-78.

[5] MacDorman, K.F. and Ishiguro, H. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies,* 7(3): 297-337.

[6] MacDorman, K.F., Green, R.D., Ho, C. and Koch, C.T. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior,* 25(3): 695-710.

[7] Flach, L. M., Dill., V. and Lywkawka, C. 2012. *Evaluation of the uncanny valley in CG characters,* XI SBGames, Brazil.

[8] Seyama, J. and Nagayama, R. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments,* 16: 337-351.

[9] Poliakoff, E., Beach, N., Best, R., Howard, T., and Gowen, E. 2013. Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception,* 42: 998-1000.

[10] Piwek, L., McKay, L.S., and Pollick, F.E. 2014. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition,* 130(3): 271-277.

[11] Thompson, J.C., Trafton, J.G. and McKnight, P. 2011. The perception of humanness from the movements of synthetic agents. *Perception,* 40: 695–705.

[12] Bartneck, C., Kanda, T., Ishiguro, H. and Hagita, N. 2009. My Robotic Doppelganger - A Critical Look at the Uncanny Valley Theory. *18th IEEE International Symposium on Robot and Human Interactive Communication*, RO-MAN2009, Toyama, IEEE.

[13] Ho, C. and MacDorman, K.F. 2008. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, Amsterdam, The Netherlands, ACM.

[14] Cheetham, M., Pavlovic, I., Jordan, N., Suter, P. and Jancke, L. 2013. Category processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Frontiers in Psychology.* 4.

[15] Matsuda, Y. T., Okamoto, Y., Ida, M., Okanoya, K., Myowa-Yamakoshi, M. 2012. Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biologial Letters* 8(5): 725-728.

[16] Steckenfinger, S.A. and Ghazanfar, A.A. 2009. Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America* 106(43): 18362-18366.

[17] Lewkowicz, D. J. and Ghazanfar, A.A. 2012. The development of the uncanny valley in infants. *Developmental Psychobiology,* 54(2): 124-132.

[18] Moore, R.K. 2012. A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports*, 2: 864.

[19] Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J. and Frith, C. 2012. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience,* 7(4): 413-422.

[20] Rao, R.P. and Ballard, D.H. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87.

[21] Mitchell, W.J., Szerszen, K.A., Lu, A.S., Schermerhorn, P.W., Scheutz, M. and MacDorman, K.F. 2011. A mismatch in the human realism of face and voice produces an uncanny valley. *Iperception,* 2(1): 10-12.

[22] Cheetham, M., Suter, P. and Jancke, L. 2011. The human likeness dimension of the "Uncanny Valley Hypothesis":

Behavioral and functional MRI findings. *Frontiers in Human Neuroscience,* 5: 126.

[23] Nie, J., Park, M., Marin, A.L., Sundar, S.S. 2012. Can you hold my hand? Physical warmth in human-robot interaction. *Human-Robot Interaction*. Boston, Massachusetts, USA.

[24] Tinwell, A., Nabi, D.A. and Chalton, J.P. 2013. Perception of psychopathy and the Uncanny Valley in virtual characters. *Computers in Human Behavior,* 29(4): 1617-1625.

[25] Yamamoto, K., Tanaka, S., Kobayashi, H., Kozima, H. and Hashiya, K. 2009. A non-humanoid robot in the "uncanny valley": Experimental analysis of the reaction to behavioral contingency in 2-3 year old children. *Plos One*, 4(9).

[26] Li, A.X., Florendo, M., Miller, L.E., Ishiguro, H., Saygin, A.P. 2015. Robot form and motion influences social attention. *10th ACM/IEEE International Conference on Human-Robot Interaction*, Portland, USA.

[27] Rosenthal-von der Pütten, A.M. and Krämer, N.C. 2014. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior,* 36: 422-439.

[28] Cheetham, M. and Jancke, L. 2013. Perceptual and category processing of the Uncanny Valley hypothesis' dimension of human likeness: some methodological issues. *Journal of Visualized Experiments,* (76).

[29] Urgen, B.A., Plank, M., Ishiguro, H., Poizner, H. and Saygin, A.P. 2013. EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics,* 7: 19.

[30] Saygin, A. P., Chaminade, T., Urgen, B.A. and Ishiguro, H. 2011. Cognitive neuroscience and robotics: A mutually beneficial joining of forces. *Robotics: Systems and Science,* Los Angeles, CA.

[31] Saygin, A.P. 2012. What can the Brain Tell us about Interactions with Artificial Agents and Vice Versa? *Workshop on Teleoperated Androids, 34th Annual Conference of the Cognitive Science Society*, Sapporo, Japan.

[32] Burleigh, T.J., Schoenherr, J.R. and Lacroix, G.L. 2013. Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior,* 29(3): 759-771.

[33] Yamada, Y., Kawabe, T. and Ihaya, K. 2013. Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese Psychological Research*, 55(1), 20-32.

[34] Kilner, J.M., Friston, K.J. and Frith, C.D. (2007b). The

mirror-neuron system: a Bayesian perspective.

Neuroreport,18:619-623.

[35] Kutas, M. and Federmeier, K.D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Reviews of Psychology,* 62: 621-647.

[36] Ishiguro, H. 2006. Android science: conscious and subconscious recognition. *Connection Science,* 18(4): 319-332.

[37] Saygin, A.P. and Cicekli, I. 2002. Pragmatics in human-computer conversations. *Journal of Pragmatics,* 34(3): 227-258.

[38] Brainard, D. H. 1997. The Psychophysics Toolbox. *Spatial Vision,* 10(4): 433-436.

[39] Pelli, D. G. 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision,* 10(4): 437-442.

[40] Delorme, A. and Makeig, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods,* 134(1): 9-21.

[41] Pascual-Marqui, R.D., Michel, C.M. and Lehmann, D. 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology, 18(1), 49-65.

[42] Walters, M.L., Syrdal, D.S., Dautenhahn, K., te Boekhorst, R. and Koay, K.L. 2008. Avoiding the uncanny valley: Robot apparence, personality, and consistency of behavior in an attention seeking home-scenario for a robot companion. *Autonomous Robots*, 24(2), 159-178.

[43] Macdorman, K.F. and Entezari, S.O. (in press). Individual differences predict sensitivity to the uncanny valley. *Interraction Studies*.

[44] Chen, H., Russell, R., Nakayama, K. and Livingstone, M. 2011. Crossing the 'uncanny valley': adaptation to cartoon faces can influence perception of human faces. *Perception,* 39(3): 378-386.

[45] Oyedele, A., Hong, A. and Minor, M.S. 2007. Contextual factors in the appearance of consumer robots: exploratory assessment of perceived anxiety toward humanlike consumer robots. *CyberPsychology and Behavior,* 10(5): 624-632.

The audio-motor feedback: a new rehabilitative aid for the developing blind child.

Giulia Cappagli[1], Elena Cocchi[2], Sara Finocchietti[1], Gabriel Baud-Bovy[1], Monica Gori[1]
[1] Robotics, Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, via Morego 30, Genoa, Italy
[2] Istituto David Chiossone, Corso Italia 3, Genoa, Italy
+39 338 1796578
giulia.cappagli@iit.it

ABSTRACT

Early onset of blindness adversely affects psychomotor, emotional and social development [1], that mostly depend on spatial cognition. Some studies suggest that the lack of vision could potentially explain this delayed or weakened development since vision is the most accurate and robust sense to encode spatial information [2,3,4]. We recently found that blind people are severely impaired when asked to judge the orientation in the haptic modality and bisect intervals in the auditory modality [5,6]. These results confirm that vision is essential in building up spatial representations that might be essential to navigate in the environment and make interactive contacts with the others [6].
Here we report for the first time also a substantial spatial impairment in proprioceptive reproduction and audio distance evaluation in early blind children and adults. Interestingly, the deficit is not present in a small group of adults with acquired visual disability. Our results support the idea that in absence of vision the audio and proprioceptive spatial representations may be delayed or drastically weakened due to the lack of visual calibration over the auditory and haptic modalities during the critical period of development.

Recent findings suggest that the acquisition of spatial capabilities is driven by the reciprocal influence between visual perception and execution of movements [7]. We recently found that multisensory integration develops late (around 8-10 years of age) and that the absence of one sensory signal might impact on the development of perceptual skills in another sensory modalities. These results suggest that the absence of multisensory integration between vision and motion might cause perceptual disabilities in spatial perception.
We aim to rehabilitate the sense of space in visually impaired children by strengthening the natural sensory-motor association of the intact senses. To do this, we developed a new rehabilitative device (ABBI, Audio Bracelet for Blind Interactions) meant to restore the sense of space in blind children. ABBI will provide an audio feedback about body movements that might help the blind child to understand and internalize the spatial structure around his own body. This approach is innovative, because unlike most existing sensory-substitution devices introduced in late childhood or adulthood, it does not require learning new "languages", and can be applied in the first years of life.

Keywords

Visual disability, Blindness, Spatial perception, Multisensory integration, Development, Sensory Substitution Devices

1. INTRODUCTION

Spatial cognition is essential in everyday life for numerous human activities, as it entails the ability to understand and internalize the representation of the structure, entities and relations of space with respect to one's own body [8]. There is a general consensus on the crucial role of visual experience in guiding the maturation of space cognition in the brain. Vision takes advantages respect to other senses in encoding spatial information because it ensures the simultaneous perception of multiple stimuli in the environment [2,3] despite the apparent motion of the array on the retina during locomotion and enables us to extract more invariant spatial properties from the surrounding layout [4]. Indeed, data from sensorial conflict situations [9-11] show that spatial auditory and tactile perception are strongly biased by simultaneously presented visual information, suggesting that sighted people tend to organize spatial information according to a visual frame of reference. These results suggest that vision typically provides the most accurate and reliable information about the spatial properties of the external world, and therefore dominates spatial perception.

In people with visual disabilities, the absence of vision might cause substantial impairments in spatial cognition that are related to psychomotor, emotional and social competencies. The current literature provides few clues about how to reconcile the hypothesis that visual perception is essential to build up spatial representations in the other sensory modalities [6,12] with findings showing that blind people compensate for the lack of vision by strengthening the others senses [13]. The cross-sensory calibration hypothesis proposed by Gori et al. [6] states that during the early development vision calibrates the other senses to process specific aspects of spatial information for which it is the most robust sense. As a consequence, blind people would be impaired in those specific aspects of spatial cognition

In the past years we have investigated how different senses are integrated during development, and how an impairment of one modality, such as in blindness, can impact on other modalities. The ultimate goal of our research is to exploit this knowledge to understand the brain and to create new rehabilitation programs and devices to increment sensory-motor abilities of children with sensory disabilities. We have recently conducted behavioral and rehabilitative studies on the development of spatial cognition and mobility skills in sighted and blind children and adults.

Behavioral studies.
In the past years we observed that unlike adults, children of less than eight years of age do not integrate visual and haptic spatial information, with one or the other modality dominating totally [13]. This result suggests that during the early years of development, children use the cross-sensory information to

calibrate the senses to physical reality: the more robust sense calibrates the other. An important question is what happens when the calibrating sense is impaired or absent, as is the case for children born without sight.
We run some behavioral studies to assess how the loss of vision impacts on the spatial sense of blind people. We found that both early children and adults are impaired in auditory spatial tests, like bisection and distance perception, and haptic spatial tests, like proprioceptive localization of arm and body in the space. The ability to localize sound sources in an environment is critically impaired in childhood but improves with ages. Moreover, the performance of the late blinds in both auditory and haptic tests is similar to the performance of the sighted individuals, showing that even a small period of visual experience allows the creation of a reliable spatial representation of the world.

Rehabilitative studies.
Since our studies reveal the presence of spatial impairments in blind people, our main goal is to successfully rehabilitate their sense of space. ABBI (Audio Bracelet for Blind Interaction) is the main device to achieve this goal [Fig.1]. This bracelet produces a sound when the acceleration of the arm movement exceed a fixed threshold: it gives precise information about when and how the movement is occurring, producing an auditory feedback about body movement similar to the one provided by the visual modality for the sighted person. This would help creating a spatial representation of the surrounding environment. Since the ABBI system can be used at an early age and in diverse contexts, it would represent an innovative and powerful rehabilitative tool for visually impaired people.
A preliminary study with blindfolded adults showed that a short training session with the ABBI bracelet might improve the pointing accuracy in a sound localization task. We run a 3-months rehabilitative protocol with ABBI with 20 early blind and low-vision children. They used ABBI for 1 hour per week in a controlled environment with expert rehabilitators and 1 hour per week of free play at home with parents. We found general spatial improvements in mobility skills such as walk velocity, auditory skills such as distance perception and proprioceptive skills such as sense of body location in the environment.

2. DISCUSSION

A veridical representation of space is fundamental for social interactions. The early absence of vision might adversely affect the full development of spatial cognition, leading to social impairments for the blind population. We think that social competence can improve by allowing blind people to understand and internalize an accurate representation of the surrounding space. We propose a new rehabilitative device (ABBI) that can help the young blind child to build up the spatial auditory maps necessary to navigate in the environment and make interactive contacts with the others.

3. ACKNOWLEDGMENTS

The research presented here has been supported by the European ABBI project (FP7-ICT-2013-10-611452).

3. FIGURES/CAPTIONS

Figure 1. On the right, the ABBI (Audio Bracelet for Blind Interactions) device on the wrist. On the left, the smartphone on which is installed the app used to control ABBI functions.

REFERENCES

[1] Gilbert, C., and Awan, H. (2003). Blindness in children. BMJ 327, 760-761.

[2] Foulke, E. (1982). Perception, cognition, and mobility of blind pedestrians. . In M. Potegal (Ed.), Spatial orientation: Development and physiological foundations. New York: Academic Press., 55-76.

[3] Millar, S. (1981). Crossmodal and intersensory perception and the blind. In Intersensory perception and sensory integration (pp. 281-314): Springer.

[4] Thinus-Blanc, C., & Gaunet, F. (1997). Representation of space in blind persons: vision as a spatial sense? Psychological bulletin, 121(1), 20.

[5] Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. Curr Biol, 20(3), 223-225.

[6] Gori, M., Sandini, G., Martinoli, C., & Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. Brain, 137(Pt 1), 288-293.

[7] Bremner, A. J., Holmes, N. P., & Spence, C. (2008). Infants lost in (peripersonal) space? Trends in cognitive sciences, 12(8), 298-305.

[8] Hart, R. A., & Moore, G. T. (1973). The development of spatial cognition: A review: AldineTransaction.

[9] Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Current biology, 14(3), 257-262.

[10] Anderson, P. W., & Zahorik, P. (2014). Auditory and visual distance estimation. Paper presented at the Proceedings of Meetings on Acoustics.

[11] Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. Perception & Psychophysics, 6(4), 203-205

[12] Eimer, M. (2004). Multisensory integration: how visual experience shapes spatial perception. Current biology, 14(3), R115-R117.

[13] Gori, M., Del Viva, M., Sandini, G., and Burr, D.C. (2008). Young children do not integrate visual and haptic form information. Current biology : CB 18, 694-698.

Interaction as a bridge between cognition and robotics

Serge Thill
Interaction Lab, School of Informatics
University of Skövde, Sweden
serge.thill@his.se

Tom Ziemke
Interaction Lab, School of Informatics
University of Skövde, Sweden
and
Human-Centered Systems
Department of Computer & Information Science
Linköping University, Sweden
tom.ziemke@his.se

ABSTRACT

The triplet formed by studies of cognition, interaction, and robotics offers a number of opportunities for symbiotic relationships and mutual benefits. One such avenue is explored by the workshop's main theme in which cognition is seen as a bridge between interaction and robotics. Exploring ideas along that direction leads, as also discussed here, amongst others to the question of how theory-of-mind mechanisms might facilitate interaction between humans and robots.

A complementary view that we explore more fully here sees interaction as the bridge that connects robotics to relevant research on cognition. We follow recent trends in social cognition that go beyond studying social interaction as the outcome of the individuals' cognitive processes by seeing it as a constitutive and enabling element of social cognition. Here, we discuss this idea and show that it leads, amongst others, to the question of how interaction can be a constitutive element of a robot's cognitive architecture. It also leads to pointers towards research in the cognitive sciences that is beneficial to robotics but goes beyond cognitive architectures themselves. We show that considering the degree to which the robot is perceived by its end user as a tool and/or social partner points, for instance, to distributed and/or social cognition approaches for methodologies to evaluate human-robot interaction.

Categories and Subject Descriptors

I.2.9 [**Computing Methodologies**]: AI—*Robotics*

Keywords

Human-robot interaction evaluation, Distributed cognition, Social cognition, Social interaction

1. INTRODUCTION

Cognitive science and robotics research have long found common ground, whether this is to demonstrate the principles behind theories of cognition using, for example, robots or to improve artificial agents based on insights gained about human cognitive abilities. It is, however, noteworthy, that this common ground does not necessarily consider the *interaction* between robots and humans – most progress has indeed only necessitated one direction: human cognition inspiring robot cognitive architectures [45] or, alternatively, robotic models illustrating insights that could be relevant to the study of cognition [48].

Simultaneously, robots and other types of artificial agents are increasingly playing important roles in our society. Examples include research on robots for use in medical or therapeutic contexts [32, 43, 30, 3, 40], game-playing robots [1], but also the increasingly intelligent, adaptive and decision-making cars in use today [17, 39].

As such agents become more ubiquitous, there is thus a need to extend the common ground covered by robotics and the cognitive sciences into territories that concern such interactions. The 2015 HRI conference workshop tackling these issues head-on is titled "Cognition: a bridge between robotics and interaction". This cognition-centric view places an emphasis on cognitive mechanisms such as Theory of Mind (ToM): one agent's ability to create an internal model of another agent and use that to predict that agent's behaviour also improves the ability to interact with that agent [12]. Understanding these mechanisms – including, for instance, what social signals human pick up on, or how robot analogues of human ToM mechanisms might be constructed – leads to better interaction between humans and robots: it becomes the bridge between robots as such and interaction.

Here, as the paper title suggests, we explore the consequences of viewing this characterisation from a different angle in which interaction is the bridge between cognition and robotics. We do this to highlight aspects of cognitive science that are highly relevant to (social or sociable) robots that interact with humans and that complement the necessary focus on cognitive architectures. The purpose is therefore not to disagree with the view of cognition as a bridge between robotics and interaction, but rather to extend it with the complementary insights that interaction as a bridge between cognition and robotics leads to.

2. COGNITIVE SCIENCE AND ROBOTICS

The idea that robotics research might help to further research in the cognitive sciences (and vice versa) has been around for a while, fuelled in particular by the developing prominence of embodied theories of cognition [44, 6, 49, 52] on one side and increasingly well-engineered and cheap(er) robots on the other. In this section, we briefly discuss three traditional approaches to research that explore this symbiotic potential. This will serve as a background against

which to discuss the added contributions of a focus on the interactive aspects.

2.1 Proof of concept

The first approach demonstrates principles of cognitive science (usually embodied or situated flavours thereof) using robots (see [48] for a brief review of relevant work). The general theme is that robots, through their design, display specific behaviours that are not themselves explicitly represented within the system, thus illustrating that "embodiment and embedding can therefore *replace* internal algorithms and lead to stable, functional behaviour" ([48], p. 4, emphasis in original). A similar example is that of morphological computation [28, 27]; the idea that computations can be offloaded into a suitably designed morphology. Well-designed legs on a quadruped robot can for example lead to an appropriate quadruped walking gait without the need for complicated control mechanisms [28].

As such, the purpose of these models is first and foremost to *illustrate by example* concepts that would otherwise be difficult to verify in a living organism. There is a benefit to robotics because these illustrations tend to be viable implementations of behaviours that might be useful for robotic applications too – such as pointers on how to simplify locomotory control as in the example above. The relevance to the study of cognition, on the other hand, is weaker: it is possible that predictions generated by such a mechanism turn out to match the biological counterpart (as in the case of Webb's cricket robots [47], see the discussion in [48]) but this is not a requirement since the original purpose is typically the demonstration of the concept (as in the case of robots that show tidying behaviour even though their underlying controllers do not explicitly implement any such behaviour [21], again discussed in [48]).

2.2 Embodying models of human cognition

The second approach attempts to more directly study human cognition using artificial agents. The motivation follows more or less as a consequence of accepting embodied or grounded theories of cognition according to which the body plays a fundamental, non-abstractable role in cognition. The Chinese Room argument [34], or the symbol grounding problem [16] are frequently cited in this context and the conclusion drawn tends to be that a cognitive model must be instantiated in a physical agent (how else could the role of the body otherwise be represented?). It is, however, worth noting that the mere provision of a robot instantiation does not by itself overcome the problems described, for instance, by the Chinese room argument: indeed, the "robot reply", in which a robot body is used to provide a sensorimotor apparatus in which to "embody" the computational model has already been considered and rejected by Searle in his original paper [34] (for a fuller discussion, see [55]).

Another challenge that these robot-reliant ways of studying human cognition face is simply that a robot body is not like a human body, even if it is described as "humanoid" [54]. Embodied accounts (irrespective of the particular theoretic flavour) ascribe a role to the body (and/or environment) that is fundamental in shaping cognition and cannot be abstracted away; yet robotic implementations often begin with a sensory apparatus that is radically different from the human senses and by necessity includes several simplifications and abstractions. Vision, for example, is often simplified, for instance by using brightly coloured and easily discriminable objects [22]. Although an advantage of robotic models is that they force integration from sensory perception to motor action [23, 26], this integration is not as forceful as it seems.

It is of course true of all models that they must contain abstractions and simplifications (otherwise they would not be a model). It has famously been said that "all models are wrong; the practical question is how wrong do they have to be to not be useful" [4]. When the model is not just of the cognitive process (because it is, in this view, meaningless to talk of "just" the cognitive process), but also of the body, and therefore all sensorimotor aspects, as well as the environment (whether this is because the agent is simulated or put into a purpose-engineered artificial situation), one has to exercise extra care when discussing the relevance of insights from such embodied models to human cognition [41].

However, this is not to say that such models have no utility beyond illustration of concepts (in which case, we should group them under the *proof-of-concept* approach discussed previously). For instance, any cognitive process that requires interaction with the environment needs to be modelled in a manner in which such interactions are possible. Even strongly abstracted sensorimotor mechanisms can provide insights into minimal requirements for the cognitive process of interest [26].

2.3 Cognitive science for the benefit of robots

The previous two approaches were examples of research whose aim is first and foremost a contribution to the study of cognitive mechanisms. By virtue of necessitating a robotic implementation, there is also a benefit to the field of robotics since, as previously argued, the algorithms and controllers that are developed may find new approaches or solutions to problems and challenges in robotics.

At the same time, there is an approach to research at the intersection between the study of cognition and robotics that aims first and foremost to benefit robotics research: knowledge and results from the cognitive sciences can be used to create "better" (defined, for instance, as an increased ability to cope with uncertainty or unpredicted events) robots. ToM mechanisms are an important example of cognitive mechanisms that have been used to this effect (see [12] for a discussion of the two main flavours of ToM – *theory theory* and *simulation theory* – in the context of social robots). Indeed, to interact proficiently with humans, such robots may simply require at least a rudimentary ToM; an internal model that can be used to estimate mental states of humans, in particular their intentions, expectations and predicted reactions to actions by the agent [31, 40, 41]. A second example is given in [18] (as cited in [50]) - here the insight that anticipation and perceptual simulation are important, for humans, in the perception of conspecifics and joint action are used to design a robot that can interact fluently with human partners. Finally, see [11] for an early review of a large number of socially interactive robots and the design principles and inspirations behind them.

3. INTERACTION AT THE CENTRE

The previous section has illustrated a number of active research areas that explore the symbiotic relationship between research in the cognitive sciences and robotics[45, 33]. It is readily apparent that interaction does not necessarily need to be considered in these areas – it is naturally not excluded: the ToM mechanisms discussed in section 2.3 are a prime example of a benefit that the study of cognition brings to robotics whereas research on human interpretation of robot movements leads to what aspects of robot motion may involve mechanisms thought to underlie, for instance, social interaction [13].

3.1 Interaction as a constitutive component of cognition

When interaction is considered in robotics research, however, it is often understood as two agents[1], one human, one artificial, each with their own cognitive apparatus, using that apparatus to engage in interaction with the other agent. This both reflects traditional views in social cognition (which are mainly interested in the individual's internal mechanisms underlying interaction) and features the same pitfall: not explicitly recognising that the interaction itself is fundamental, and *part of the overall cognitive process* as opposed to merely the result thereof [10]. In other words, the interactive setting does not merely play a contextual role for an individual's cognitive mechanisms but also takes on enabling and constitutive roles [10]. Just as a cognitive architecture in which the body does not play a fundamental, irremovable and irreplaceable part of the cognitive process is not an embodied architecture [55, 56], a cognitive architecture in which interaction is merely a contextual aspect lacks something.

This is the first core insight we gain from a focus on interaction: as the field of social cognition is moving away from an individualistic view of interaction, robotic cognitive architectures need to consider the implications of an enabling, constitutive role of interaction with other agents in their overall functionality (see also *e.g* [9] for a similar argument). For example, robots are often built for specific purposes – their desired behaviour is therefore given by that application. Yet, to deal with uncertainty and unforeseen events, it is not desirable to specify all behaviours axiomatically at design time – rather, the ability for appropriate behaviour to emerge from the robot's experience is needed. In this context, it can be shown that casting the objective function modulating such emergence in terms of *interaction* may lead to desirable, yet not unnecessarily constrained behaviour [41].

3.2 Evaluating human-robot interaction (HRI)

Robots (and other artificial agents), as discussed before, can in almost all cases be expected to interact with humans to some degree. There is therefore also a need to evaluate these robots *in terms of their interaction with humans*. There are no "simple" metrics to this end since successful performance, by definition, depends on the human/artificial agent system as a whole.

In other words, one cannot consider the robot's performance in isolation; its success is a function of how well the agent/human system functions (see [2, 10, 41] for related arguments). In some applications, for instance robot-enhanced therapy (RET) for children with autism spectrum disorder (ASD) [32, 40], the ideal measure (*e.g.* long-lasting benefits) is also simply unavailable since it can only be meaningfully be sampled after years if not decades after the artificial agent is deployed. Other scenarios might be entirely open-ended and without any direct task to be achieved, yet the need to evaluate the robot remains. Further, although it is possible to achieve some form of evaluation by asking persons who interacted with artificial agents to fill out questionnaires and similar (see also [39] for an example in which just that has been done), such options are typically not available if the persons interacting with the artificial agent are in fact children [2]. More generally speaking, these methods usually require the subjects to have a substantial degree of insight into their own cognitive processes.

[1]The present argument easily extends to multi-agent systems, but two are sufficient for illustrative purposes

How to characterise and evaluate interaction has long been a topic in HRI (see for instance the extensive survey and introduction to the topics in [14]). An immediate realisation in such efforts is that there is no "one size fits all" solution; robots can interact with humans in a number of ways that then define and shape what one expects from such interaction. This then leads to a number of proposals for dimensions along which to rate the precise nature of the interaction at hand. The ubiquitous example is that of autonomy: in 1978, Sheridan and Verplank proposed a 10-step scale describing degrees of automation, ranging from machines that are entirely remote-controlled to machines that ignore human beings altogether [36]. Since then, there have been numerous discussions of the scale in particular and the concept of autonomy in HRI in general (*e.g.* [14, 51, 38, 42]). It is for instance repeatedly argued that "human-robot interaction cannot be studied without consideration of a robot's degree of autonomy" [42] (p. 14).

It is therefore worth emphasising that autonomy is a particularly difficult term that can mean very different things to different people [45, 53]. In HRI, for instance, the take on autonomy is often task-oriented – referring, for example (as in Sheridan and Verplank's scale), to the degree to which the human has to assist the machine in accomplishing a given task [51], thus measuring the degree of automation. Cognitive scientists, on the other hand, might consider autonomy more in terms of self-sufficiency, or behaviour that is *not* determined entirely by external events but shaped by internal goals of an agent [35].

This highlights an important point pertinent to the possible benefits between the study of cognition and robotics: it needs to be kept in mind that autonomy is an overloaded term (as are others) when researchers from different disciplines meet. In [45], for instance, no less than 19 different takes on autonomy are discussed, a list that is by no means complete. Although we cannot possibly do the concept justice here (and instead point to [45], Ch. 4), the relevant insight is that, when the study of cognition and robotics meet, it is critical to be clear about the terms one uses; a symbiotic relationship depends on a common understanding of such concepts.

When autonomy refers to the degree of automation, it is a dimension in which social interactions occupy the middle of the range (since there is no meaningful interaction in the fully automated case and merely tele-operating a robot does not constitute social interaction with another cognitive agent). Likewise, other metrics that fundamentally seek to evaluate HRI performance in terms of task performance (*e.g.* robot efficiency and effectiveness in the task and human situation awareness [38]) do not assess the social interaction itself. Metrics that do would need to measure, it has been suggested [38], interaction characteristics, persuasiveness, trust, engagement and compliance, but the exact methodologies for that remain unclear.

3.3 Interaction-focussed HRI evaluation

It has been suggested [42] that we may not actually want to interact with robots in precisely the same way as we interact with other humans. Whether or not one reserves the term "social interaction" for human-human interaction or opens it up to human-robot interaction is a different debate and does not *per se* invalidate the idea of evaluating HRI as a type of interaction that can be usefully characterised by metrics similar to those used for human-human interaction.

It does, however, lead to the interesting question of how robots (and other artificial agents) are *perceived* – it is for instance known that,

for some robots and actions at least, the human mirror system is activated when observing robot actions [13] that can then be interpreted as being goal-directed [33], which does point towards the likelihood that interacting with humans and robots – when they are perceived as having some agency at least – may not be entirely different.

The interesting question therefore is to what degree robots are actually perceived as agents by the people they interact with. With that characterisation, we can then return to the central theme of the paper and discuss methodologies in the cognitive sciences that may be useful for characterising human-robot interaction based on how the robot is perceived.

In the context of increasingly automated vehicles, it has been suggested that a useful way to characterise human-vehicle interaction is by establishing the degree to which the vehicle is perceived as a *tool*, used in navigation tasks, as opposed to an *intelligent agent*, with whom the driver collaborates in solving the task [39]. Here, we explore a similar characterisation for robots and artificial agents in general. In particular, we illustrate in the next two subsections that they can be understood *by their end users* as, to varying degrees, both tools and social partners.

Such characterisations have been used in the past: the "robot role" (ranging from tool/machine to companion/partner) is, for example, one of the suggested dimensions for determining the requirements on a robot's social skills [8]. Here, however, we use this dimension to identify theories of cognitive science useful in evaluating human-robot interaction. It is difficult to find such theories in the traditional overlap between cognitive science and robotics discussed in section 2: the first two approaches, proof-of-concept and embodying models, mainly use artificial agents for theoretical insights that could include interaction between agents (see *e.g.* robot language games [37]), but do not have to. When cognitive models are primarily used as an inspiration for better robots, validation is given by an adequate implementation of the targeted cognitive ability.

3.3.1 *Artificial agents are tools*

Artificial agents are usually created for a purpose - this can be academic (*e.g.* as demonstrators of cognitive theories or as tools for studying cognition as discussed above) or with a practical application in mind (*e.g.* for use in elderly care, therapy, navigation of dangerous or inhospitable terrain and so on). They exist, therefore, to assist humans in achieving certain goals (even if they are designed as autonomous agents). Artefacts used by people in addition to their own body to achieve a certain purpose are, by definition, tools.

3.3.2 *Artificial agents are social partners*

Although artificial agents are, as argued above, always created for a purpose, significant research efforts [45] are dedicated to creating agents with interesting cognitive abilities (whether it is to showcase models of these abilities or more directly to allow the agents to tackle more complicated and less trivial tasks). It is therefore clear that artificial agents can be seen as more than tools: indeed, they can be social partners with whom we interact, *collaborating* in solving the task for which they were created.

This highlights (again) that the artificial agent should not be seen by itself but rather as *interacting* with humans. [2] for instance argues that technical challenges in cHRI (HRI in which the humans are specifically children) may be overcome if we see the cognition

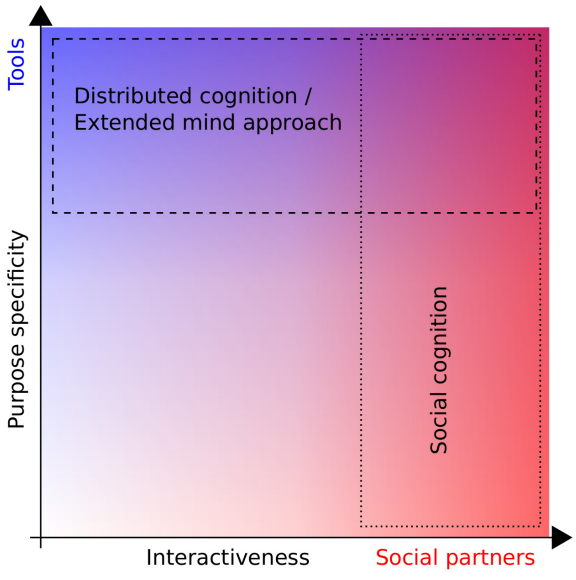


Figure 1: Diagram positioning artificial agents in function of how their interactiveness and purpose specificity are perceived by the end users. Boxes inside the diagram indicate the cognitive science research strands one should primarily consider when evaluating artificial agents in that area of the spectrum.

of a human/artificial agent ensemble as the product of their interaction. A critical point these authors make is that one agent (*e.g.* the human) can cover for potential failings of the other (*e.g.* the robot), which in itself illustrates that one cannot evaluate the robot by itself (see the credit assignment problem).

Viewing artificial agents as social partners also has consequences for how one expects humans to interact with them. For instance, humans tend to modulate their behaviour based on their beliefs about, amongst others, the cognitive abilities of the agent they interact with [5]. This has been shown to extend to robots [46, 20]. Furthermore, our recent research indicates that this extends even to cars [39].

3.3.3 *Artificial agents are tools and social partners*

It is worth emphasising that the two views of artificial agents, as sketched out above, are not mutually exclusive. In other words, they do not form two ends on a scale as in previous examples of similar scales [9]. Rather, artificial agents can be, to varying degrees, both:

- If a robot is built for very specific purposes, it is a tool created to achieve that purpose. But not all robots are created for such specific purposes: another scale used by Dautenhahn [8] considers "robot functionalities" which can range from clearly defined to open and adaptive. In a similar vein, we use *purpose specificity* as a dimension along which to measure whether or not an artificial agent may be perceived as a tool.
- Similarly, the degree to which a robot can be considered a social partner depends on the degree to which it is seen as interacting with its end user (by the end user). Again, in

Dautenhahn's set of scales used to determine social requirements, the analogue is the "contact with humans" dimension. Here, we refer to this dimension as *interactiveness* of a robot to more explicitly capture the fact the contact involves interaction.

These two dimensions - purpose specificity and interactiveness create a 2D map of artificial agents (as sketched in Fig. 1). We resist the temptation to populate the sketch with placements of example robots and other artificial agents. To give but two examples:

- It can be argued that cars would typically score high on purpose specificity (they are built purposefully for navigating from A to B) and low on interactiveness (since – until recently – they do not interact with the drive beyond providing information about their internal state). There are, however, significant technological developments [17, 39] that will increase the interactiveness of cars. In the near future, cars might thus move further to the right in Fig. 1.
- Therapeutic robots, for instance as used for ASD therapy [32] are naturally highly interactive but their purpose is more open-ended, [8], reducing their purpose specificity (especially as perceived by the child). One could conceivably expect to place them around the middle of the right side of the graph.

3.3.4 Cognitive theories for HRI

With this in mind, we can now consider cognitive theories that have traditionally dealt with human interaction with tools and social partners. First, robots that score highly on the purpose specificity scale more or less directly speak to *extended and distributed views of cognition* [19, 7].

From the extended mind view [7], we can take the position that the artificial agent becomes just such an extension of the mind. The cognitive process according to which the human uses the artificial agent to achieve a certain purpose cannot be defined within the human alone; the artefact at a minimum becomes a resource (of what type depends on the agent).

From distributed cognition [19], we similarly get the perspective that cognition should be understood in terms of the interaction with the material and social world. The paradigm additionally comes with a large set of tools for analysing such interactions, most dominantly ethnography (see [24] for an extensive review of these aspects of distributed cognition, including criticisms and rebuttals). Distributed cognition has also already found applications in HCI, for instance as a method "with which to understand the underlying mechanisms of the relationships between humans and computer" [24] (p. 63). For instance, a distributed cognition-inspired methodology for studying the interaction between humans and machines in a maritime control room has been developed [25]. While it may of course be too bold to refer to such a control room as an artificial agent, the example illustrates that it is possible to take the basic ideas from distributed cognition into a more formalised approach to studying the interaction between man and machine.

When robots score high on the interactiveness scale, meanwhile, it is possible (and necessary) to go beyond distributed cognition and explicitly treat the interaction as social. Consequently, this points to insights from social cognition. Here, social interaction can, for instance, be defined as "two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics *acquire an autonomy of their own*" [10] (p. 441, emphasis added). A highly interactive robot would necessarily possess some autonomy in the same sense (and notably not necessarily the sense usually given to autonomy in HRI, see the previous discussion in section 3.2); it is therefore clear that any take on this agent that ignores the interactive aspect will fail to adequately take into account this coupling.

A comprehensive review of methods that are useful in studying social interaction can also be found in [10]. These include conversation and gesture analysis, with the particular insight that Motion Energy Analysis [15] could predict subjective assessments of a therapy session's quality based on bodily coordination between patient and therapist [29] (as cited in [10]). Work such as this provides a clear entry point by which one could possibly evaluate therapeutic robots, addressing for instance the concerns of [2] that one cannot easily make children fill out a questionnaire. Even though putting the focus of social cognition on embodied social interaction is, as noted at the beginning, a relatively recent trend [10], it is clear that the field is developing a range of techniques that are useful for evaluating the quality of this interaction. These techniques may well find further applications in the study of the interaction between humans and robots.

4. CONCLUSIONS

We have highlighted the importance of interaction in the ⟨*cognition, robotics, interaction*⟩ triplet. This perspective has enabled us to illustrate that interaction is not just contextual, but rather an enabling and constitutive component of social cognition [10]. Although cognition can, as also illustrated here, rightfully be seen as a bridge between robotics and interaction, the latter also functions as a bridge between robotics and cognition; in particular enabling robotics research to develop cognitive architectures in which the interaction likewise plays a constitutive, enabling component (as opposed to being the outcome).

The perspective has also enabled us to consider the roles that robots play when interacting with humans. We have argued that the degree to which the robot is perceived as fulfilling a specific purpose as well as the degree to which it is perceived as interacting with humans – in both cases as seen from the end user – are useful dimensions to consider in this respect. In particular, the relative degree to which robots score on these dimensions form a guide to theories in cognitive science that can be useful to understand and evaluate the interaction between the human and the robot.

Given that robots and other artificial agents (we have mentioned cars in particular) are increasingly entering into our daily lives, such evaluations become increasingly important. It of course remains to be seen to what extent exactly one can translate the methodologies, explanatory tools and techniques from distributed and social cognition onto the study of artificial agents. Here, our purpose has been to highlight that the relevance of cognitive research for robotics goes beyond inspiration for better cognitive architectures as such to include the study of how the human-robot system as a whole functions. Such a perspective has relevance in many application areas. In robots used for therapy, for instance RET aimed at children with ASD, the child-robot system is more than just the sum of a child and a robot - a relationship between the two exists that cannot be abstracted away [9] and that has implications both for the design of the cognitive architecture of the robot [41] and, as argued here, for the evaluation of the robot.

5. ACKNOWLEDGEMENTS

This work has been supported by DREAM (www.dream2020.eu), funded by the European Commission (FP7-ICT, project number 611391), and TINA/AIR, funded by KK-SIDUS, Sweden.

6. REFERENCES

[1] P. Baxter, J. de Greeff, and T. Belpaeme. Do children behave differently with a social robot if with peers? In *International Conference on Social Robotics (ICSR 2013)*, October 2013.

[2] T. Belpaeme, P. Baxter, J. de Greeff, J. Kennedy, R. Read, R. Looije, M. Neerincx, I. Baroni, and M. Zelati. Child-robot interaction: Perspectives and challenges. In G. Herrmann, M. Pearson, A. Lenz, P. Bremner, A. Spiers, and U. Leonards, editors, *Social Robotics, volume 8239 of Lecture Notes in Computer Science*, volume 8239, pages 452–459. Springer International Publishing, 2013.

[3] O. A. Blanson Henkemans, B. P. Bierman, J. Janssen, M. A. Neerincx, R. Looije, H. van der Bosch, and J. A. van der Giessen. Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient Education and Counseling*, (PEC-4519):8, 2013.

[4] G. E. P. Box and N. R. Draper. *Empirical Model Building and Response Surfaces*. John Wiley & Sons, New York, NY, 1987.

[5] H. P. Branigan, M. J. Pickering, J. Pearson, J. F. McLean, and A. Brown. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121(1):41 – 57, 2011.

[6] A. Clark. *Being there: Putting brain, body, and world together again*. MIT press, Cambridge, MA, 1997.

[7] A. Clark and D. J. Chalmers. The extended mind. *Analysis*, 58:7–19, 1998.

[8] K. Dautenhahn. Roles and functions of robots in human society: implications from research in autism therapy. *Robotica*, 21:443–452, 8 2003.

[9] K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1480):679–704, 2007.

[10] H. De Jaegher, E. Di Paolo, and S. Gallagher. Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10):441 – 447, 2010.

[11] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3âĂŞ4):143 – 166, 2003. Socially Interactive Robots.

[12] S. Gallagher. Social cognition and social robots. *Pragmatics & Cognition*, 15(3):435–453, 2007.

[13] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers. The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, 35(4):1674 – 1684, 2007.

[14] M. A. Goodrich and A. C. Schultz. Human-robot interaction: A survey. *Found. Trends Hum.-Comput. Interact.*, 1(3):203–275, Jan. 2007.

[15] K. Grammer, M. Honda, A. Juette, and A. Schmitt. Fuzziness of nonverbal courtship communication unblurred by motion energy detection. *Journal of Personality and Social Psychology*, 77(3):487–508, 1999.

[16] S. Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346, 1990.

[17] A. Heide and K. Henning. The "cognitive car": A roadmap for research issues in the automotive sector. *Annual Reviews in Control*, 30(2):197 – 203, 2006.

[18] G. Hoffman and C. Breazeal. Effects of anticipatory perceptual simulation on practiced human-robot tasks. *Autonomous Robots*, 28(4):403–423, 2010.

[19] E. Hutchins. *Cognition in the Wild*. MIT Press, Cambridge, MA, 1995.

[20] S. Kopp. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, 52(6):587–597, 2010.

[21] M. Maris and R. Boeckhorst. Exploiting physical constraints: heap formation through behavioral error in a group of robots. In *Intelligent Robots and Systems '96, IROS 96, Proceedings of the 1996 IEEE/RSJ International Conference on*, volume 3, pages 1655–1660 vol.3, 1996.

[22] A. F. Morse, T. Belpaeme, A. Cangelosi, and L. B. Smith. Thinking with your body: Modelling spatial biases in categorization using a real humanoid robot. In S. Ohlsson and R. Catrambone, editors, *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pages 1362–1367, Austin, TX, 2010. Cognitive Science Society.

[23] A. F. Morse, C. Herrera, R. Clowes, A. Montebelli, and T. Ziemke. The role of robotic modelling in cognitive science. *New Ideas in Psychology*, 29(3):312–324, 2011.

[24] M. Nilsson. *Capturing semi-automated decision making: the methodology of CASADEMA*. PhD thesis, Örebro University, 2010.

[25] M. Nilsson, J. van Laere, T. Susi, and T. Ziemke. Information fusion in practice: A distributed cognition perspective on the active role of users. *Information Fusion*, 13(1):60 – 78, 2012.

[26] G. Pezzulo, L. W. Barsalou, A. Cangelosi, M. H. Fischer, K. McRae, and M. J. Spivey. The mechanics of embodiment: a dialog on embodiment and computational modeling. *Frontiers in Psychology*, 2(5), 2011.

[27] R. Pfeifer, J. Bongard, and S. Grand. *How the body shapes the way we think: a new view of intelligence*. MIT press, Cambridge, MA, 2007.

[28] R. Pfeifer and F. Iida. Morphological computation: Connecting body, brain and environment. *Japanese Scientific Monthly*, 2005.

[29] F. Ramseyer and W. Tschacher. Synchrony: A core concept for a constructivist approach to psychotherapy. *Constructivism in the Human Sciences*, 11(1):150–171, 2006.

[30] B. Robins, K. Dautenhahn, R. Boekhorst, and A. Billard. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2):105–120, 2005.

[31] B. Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24, 2002.

[32] B. Scassellati, H. Admoni, and M. Matarić. Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14:275–294, 2012.

[33] A. Sciutti, A. Bisio, F. Nori, G. Metta, L. Fadiga, and G. Sandini. Robots can be perceived as goal-oriented agents. *Interaction Studies*, 14(3):329–350, 2013.

[34] J. R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–424, 9 1980.

[35] A. Seth. Measuring autonomy and emergence via Granger

causality. *Artificial Life*, 16(2):179–196, April 2010.

[36] T. B. Sheridan and W. L. Verplank. Human and computer control of undersea teleoperators. Technical report, MIT Man-Machine Systems Laboratory, 1978.

[37] L. Steels. Evolving grounded communication for robots. *Trends in cognitive sciences*, 7(7):308 – 312, 2003.

[38] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 33–40, New York, NY, USA, 2006. ACM.

[39] S. Thill, P. E. Hemeren, and M. Nilsson. The apparent intelligence of a system as a factor in situation awareness. In *Proceedings of the 4th IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 52 – 58, 2014.

[40] S. Thill, C. Pop, T. Belpaeme, T. Ziemke, and B. Vanderborght. Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook. *Paladyn*, 3(4):209–217, 2012.

[41] S. Thill and D. Vernon. How to design emergent models of cognition for application-driven artificial agents. In *Proceedings of the 14th Neural Computation and Psychology Workshop (NCPW14)*, submitted.

[42] S. Thrun. Toward a framework for human-robot interaction. *Human–Computer Interaction*, 19(1-2):9–24, 2004.

[43] B. Vanderborght, R. E. Simut, J. Saldien, C. A. Pop, A. S. Rusu, S. Pintea, D. Lefeber, and D. David. Social stories for autistic children told by the huggable robot Probo. In *Cognitive Neuroscience Robotics workshop IROS*, pages 1–6, 2011.

[44] F. J. Varela, E. Rosch, and E. Thompson. *The embodied mind: Cognitive science and human experience*. MIT press, Cambridge, Ma, 1992.

[45] D. Vernon. *Artificial Cognitive Systems: A primer*. MIT Press, Cambridge, MA, 2014.

[46] A.-L. Vollmer, B. Wrede, K. J. Rohlfing, and A. Cangelosi. Do beliefs about a robot's capabilities influence alignment to its actions? In *Development and Learning and Epigenetic Robotics (ICDL), 2013 IEEE Third Joint International Conference on*, pages 1–6, 2013.

[47] B. Webb. Using robots to model animals: a cricket test. *Robotics and Autonomous Systems*, 16(117–134), 1995.

[48] A. D. Wilson and S. Golonka. Embodied cognition is not what you think it is. *Frontiers in Psychology*, 4(58), 2013.

[49] M. Wilson. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636, 2002.

[50] M. Wilson and G. Knoblich. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3):460–473, 2005.

[51] H. Yanco and J. Drury. Classifying human-robot interaction: an updated taxonomy. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 3, pages 2841–2846 vol.3, Oct 2004.

[52] T. Ziemke. WhatâĂŹs that thing called embodiment. In *Proceedings of the 25th Annual meeting of the Cognitive Science Society*, pages 1305–1310, 2003.

[53] T. Ziemke. On the role of emotion in biological and robotic autonomy. *Biosystems*, 91(2):401 – 408, 2008. Modelling Autonomy Modelling Autonomy.

[54] T. Ziemke and J. Lindbolm. Some methodological issues in android science. *Interaction Studies*, 7(4):339–342, 2006.

[55] T. Ziemke and S. Thill. Robots are not embodied! conceptions of embodiment and their implications for social human-robot interaction. In *Proceedings of Robo-Philosophy 2014: Sociable robots and the future of social relations*, pages 49– 53. IOS Press BV, 2014.

[56] T. Ziemke, S. Thill, and D. Vernon. Embodiment is a double-edged sword in human-robot interaction: Ascribed vs. intrinsic intentionality. In *Cognition: a bridge between robotics and interaction. Workshop at HRI2015*, 2015.

Acknowledgements

We would like to thank the support given by our departments (RBCS at IIT, Emergent Robotics Lab at Osaka University and MACS at Heriot-Watt University). Furthermore, we like to thank for the financial support of the HRI consortium, the Osaka University, the Italian Institute of Technology, the Heriot-Watt University, EU CODEFROR project: FP7-PIRSES-2013-612555 and the Institute for Academic Initiatives (2012-2015).
Gratefully, Yukie Nagai, Alessandra Sciutti and Katrin S. Lohan

Institute for Academic Initiatives

The Osaka University Institute for Academic Initiatives (IAI) was set up in order to promote, under the leadership of President Hirano, interdisciplinary cross-boundary education and research. Each of Osaka Universitys schools conducts education and research in professional fields; however, modern society faces many challenges that require creative approaches, approaches that require scholarship from more than one field. Thus, the IAI was set up for the purpose of promoting such cross-border, medium- and long-term learning and research, strategies for a future viewed as a whole.

The Institute for Academic Initiatives (IAI) was set up in order to promote interdisciplinary education and research under the leadership of President Toshio HIRANO. Excellence in education and research has always been pursued at Osaka University's schools. However, modern sociery faces many challenges that require creative approaches, approaches that call upon scholarship from more than one field, one s@ecoa;tu. Thus, the IAI was set up for the purpose of promoting such cross-border, medium-and long-term interdisciplinary learning and research.

CODEFROR

COgnitive DEvelopment for Friendly ROBots and Rehabilitation

FP7-PEOPLE-2013-IRSES
MARIE CURIE ACTIONS: International Research Staff Exchange Scheme
Starting date: 1st February 2014
Duration: 4 years

Social interaction is a bidirectional process based on a shared representation of actions and on mutual understanding and its study will help discovering how infants develop the understanding of actions, intentions and emotions to progressively improve their social behaviours. In addition, implementing models derived from human studies on robots provides a constructive approach to investigate cognitive developments and could benefit both robotics (better robots) and neuroscience, providing a test-bed for the proposed theories.

Objective of the joint exchange project
Investigate aspects of **human cognitive development** with the double goal of :

- developing **robots able to interact** with humans in a friendly way
- designing and testing protocols and devices for **sensory and motor rehabilitation** of disabled children.

Methodology:
Combination of **science driven investigation** of human cognitive development and **engineering based implementation** of devices and protocols. This multidisciplinary program calls for a wide range of expertise both in terms of scientific communities (developmental psychology, robotics, sensory and motor rehabilitation), and in relation to engineering implementation (robots as well rehabilitation devices) and social exploitation (sensory and motor rehabilitation).

The exchange program proposed has the goal of joining the forces and expertises of the participating partners and of helping the formation and establishment of an international community of young researchers that shall effectively bridge the involved groups and their expertise in order to be effective in the long term.

Exploring the development of Human Cognitive Functions

TOPICS:

- The development of multimodal integration
- The development of social skills

OBJECTIVES:
To assess the dynamics of the development of cross-sensory calibration and the transition from basic social behaviours at birth to more complex social interaction abilities.

Sensory, Motor and Cognitive Rehabilitation

TOPICS:

- Sensory rehabilitation
- Motor rehabilitation and multimodal stimulation
- Cognitive rehabilitation

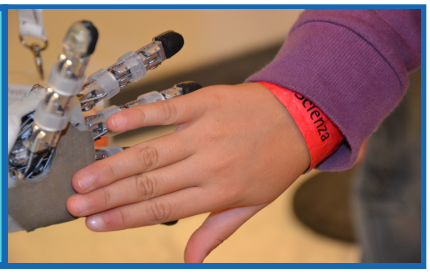
OBJECTIVES:
To design new training protocols and devices to be used in the early years of life to stimulate sensory, motor and social development in children affected by disability, with particular reference to cerebral palsy.

Human-Human(oid) friendly Interaction

TOPICS:

- Basic social skills to bootstrap social interaction
- Social cues for human-robot interaction
- Cultural differences in human-robot natural interaction

OBJECTIVES:
To define the taxonomy of robot behaviour parameters needed to allow for natural interaction and to test the validity of the behavioural models derived from developmental studies .



ISTITUTO ITALIANO DI TECNOLOGIA

Universität Bielefeld

東京大学 THE UNIVERSITY OF TOKYO

大阪大学 OSAKA UNIVERSITY

Cognitive development and architectures for cognitive robotics

Symposium at the EuroAsianPacific Joint Conference on Cognitive Science
Turin, September (between 25th and 27th)

A key feature of humans is the ability to entertain models of other agents, to anticipate what they are going to do and to plan accordingly a collaborative action. Analogously the focus of cognitive robotics is on predictive capabilities: being able to view the world from someone else's perspective, a cognitive robot can anticipate that person's intended actions and needs. Hence, a fundamental aspect of cognition, both natural and artificial, is about anticipating the need for action and developing the capacity to predict the outcome of those actions. But how does this capability develop in humans and how can it be developed in robots?

The goal of this symposium is to address these questions by investigating aspects of cognitive development through the development of cognitive robots. The discussion will focus on what is a cognitive architecture, on how predictive learning could lead to social cognition, and how bio-inspired cognitive architectures in robotics could prove fundamental for (physical) interaction. The session will start with an overview on artificial cognitive architectures by Professor David Vernon (University of Skövde), followed by talks on different aspects of cognitive robotics with a focus on learning and development by selected speakers as Professor Yukie Nagai (Osaka University) and Professor Jochen Steil (Bielefeld University). Professor Giulio Sandini (Istituto Italiano di Tecnologia) and Professor Minoru Asada (Osaka University) will chair the session providing an introduction and a link between the different perspectives of developmental cognitive robotics and discussing its relevance for a multidisciplinary understanding of cognition.

This symposium is part of the CODEFROR project (COgnitive Development for Friendly RObots and Rehabilitation, https://www.codefror.eu/), which aims at joining the forces and expertise of the participating partners (Italian Institute of Technology, Bielefeld University, Osaka University and Tokyo University) to help the establishment of an international community of researchers that shall effectively bridge the expertise of the different disciplines as robotics and cognitive sciences in the investigation of cognitive development.

Program at a glance:

- **Giulio Sandini** – *Introduction*
- **David Vernon** – *"Artificial Cognitive Systems"*
- **Jochen Steil** – *"Biomorphic control as key for cognitive soft robotics"*
- **Yukie Nagai** – *"Predictive learning as a key for cognitive development: new Insights from developmental robotics"*
- **Minoru Asada** – *Conclusion*

Chairs: Alessandra Sciutti, Tomoyuki Yamamoto, Minoru Asada and Giulio Sandini