

Human Behavior Understanding for Robotics

Albert Ali Salah, Javier Ruiz-Del-Solar, Cetin Mericli, Pierre-Yves Oudeyer

► **To cite this version:**

Albert Ali Salah, Javier Ruiz-Del-Solar, Cetin Mericli, Pierre-Yves Oudeyer. Human Behavior Understanding for Robotics. Salah, Albert and Ruiz-del-Solar, Javier and Mericli, Cetin and Oudeyer, P-Y. Human Behavior Understanding, 7559, Springer, pp.1-16, 2012, LNCS, <10.1007/978-3-642-34014-7_1>. <hal-00788617>

HAL Id: hal-00788617

<https://hal.inria.fr/hal-00788617>

Submitted on 14 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Human Behavior Understanding for Robotics

Albert Ali Salah¹, Javier Ruiz-del-Solar², Çetin Meriçli³,
and Pierre-Yves Oudeyer⁴

¹ Boğaziçi University,

Department of Computer Engineering, Istanbul, Turkey
salah@boun.edu.tr

² Universidad de Chile, Department of Electric Engineering
Av. Tupper 2007, Santiago, Chile
jrui@d@ing.uchile.cl

³ Carnegie Mellon University, Computer Science Department
Pittsburgh, PA 15213, USA
cetin@cmu.edu

⁴ FLOWERS Research Team, INRIA Bordeaux Sud-Ouest
33405 Talence, Cedex, France
pierre-yves.oudeyer@inria.fr

Abstract. Human behavior is complex, but structured along individual and social lines. Robotic systems interacting with people in uncontrolled environments need capabilities to correctly interpret, predict and respond to human behaviors. This paper discusses the scientific, technological and application challenges that arise from the mutual interaction of robotics and computational human behavior understanding. We supply a short survey of the area to provide a contextual framework and describe the most recent research in this area.

1 Introduction

Personal robots are predicted to arrive in homes and everyday life in the coming decades, and assist humans physically, socially, and/or cognitively. They are expected to become an integral part of the lives of people with physical or cognitive disabilities, for example, allowing the elderly or the handicapped to maintain a comfortable and autonomous life in their homes for a prolonged period of time. Furthermore, with a drastic paradigm shift in the industrial robotics, robots are also becoming closer to humans in the factories, where we observe a shift towards robots that can be intuitively and dynamically re-programmed by workers, and work jointly with them to achieve manufacturing and maintenance tasks.

Nevertheless, considering that the robots becoming so ubiquitous would result in their operating in uncontrolled environments and interacting with non-expert users, several challenging issues need to be addressed. One of these issues is human behavior understanding: in order to act in a useful, relevant, and socially acceptable manner, robots will need to understand the behavior of humans at various levels of abstractions (ranging from identifying the current action of the

human to identifying goals in the discussion of two humans) and at various time scales (ranging from milliseconds to minutes and days).

A large body of work exists in the field of computational human-behavior understanding, and the International Workshop of Human Behavior Understanding, previously organized with a focus on pattern recognition and ambient intelligence, brings together scientific and technological responses to some of the challenges in this field [56,57]. While some of the proposed methods can be readily re-used for robots, novel scientific and technological challenges arise when one considers achieving human behavior understanding in the context of human-robot interaction:

- First, humans who interact with a social robot behave in ways that differ significantly from natural human-human interaction, and there is an associated new repertoire of behaviors and contextual interpretations. Thus, it is paramount to design techniques that understand human behavior specifically in the **context** of human-robot interaction.
- Second, and in a related manner, interaction with an intelligent system (be it a robot, or any artificial or ambient intelligence system) in the loop can produce dynamical evolution of human behavior, where new semiotic conventions can emerge [53]. New dynamic conventions (for example, through linguistic alignment) can be negotiated between a particular robot and a particular human, and a corresponding **dynamic update** of human behavior understanding is needed.
- Third, what makes robots specific as compared for example to classical intelligent ambient systems is that they typically have a rich repertoire of motor behaviors and actions. To be useful, relevant and socially acceptable, they need to act properly. This implies that techniques for human behavior understanding need to provide **internal representations** that are compatible and reusable by the robot's action system.

A second key challenge is the capability of robots to adapt to and learn from humans. Each human user may typically have its own preferences and habits, which a robot needs to infer. The interaction between learning and human behavior understanding can be expressed in two complementary directions:

- Robots need to be capable of learning dynamically how to interpret, and thus understand human multi-modal behavior. This includes for example learning the meaning of new linguistic constructs used by a human [18], learning to interpret the emotional state of particular users from para-linguistic or non-verbal behavior [34,58,38], characterizing properties of the interaction [44] or learning to guess the intention, and potentially the combinatorial structure of goals [39] of a human based on its overt behavior [1].
- Robots also need to be capable of learning new tasks or refining existing tasks through interaction with humans, for example using imitation learning or learning by demonstration [59,9,4,42]. This heavily involves the capacity for decoding linguistic and non-linguistic cues [34,58,38], feedback and guidance provided by humans, as well as inferring reusable primitives in human

behavior [39]. Thomaz and Breazeal [66] have for example shown that prior studies of how humans use social cues to teach can be transferred into highly useful mechanisms used by a robot to learn from humans. Such a study, related to the problem of how non-expert humans can teach new words to a robot, is presented in this volume [18].

Given that human behavior understanding in general needs to be at least partially learnt, and that learning new tasks from humans require human behavior understanding, a long-term challenge for research is to study what mechanisms can allow the joint developmental and potentially simultaneous learning of feed-back/guidance/cueing models and new task models (see for example [35]).

At the same time, robotics offers stimulating opportunities for improving human behavior understanding, and especially to allow a deeper analysis of the semantics and structure of human behavior. Indeed, it is now widely known that the human action system mediates the understanding of other people's actions, in particular through the mirror neurons system [19]. Humans tend to interpret the meaning and the structure of other's behaviors in terms of their own action repertoire, which acts as a strong helping prior for this complex inference problem. Robots are also embodied and have an action repertoire, which can be similarly used to decode and interpret human behavior. For example, in this volume, Schillaci et al. show how generative action forward and inverse models of previously learnt motor primitives can be used to recognize ambiguous human movements, or to infer the target of a movement [61]. Mangin and Oudeyer show how biases on action representations can not only allow to infer the underlying combinatorial structure of complex movements demonstrated by humans, but also can be used to reproduce them [39].

In the next sections, we deal with the major contact points of human behavior understanding and robotics. Section 2 is a brief overview of systems for sensing human behavior, including pervasive systems, action and activity recognition. Section 3 discusses the social and affective aspects of human behavior from a robotics standpoint. Section 4 focuses on human-robot interaction, and Section 5 describes recent issues in imitation and learning from demonstration. Before concluding, we review a few relevant application areas briefly in Section 6 to show the practical implications of this line of research.

2 Sensing Human Behavior

The first task of a robot interacting with humans in uncontrolled environments is to sense the location of the interacting parties, as well as to recognize the relevant actions and activities. Since a lot of information can be gained by analyzing the context of interaction, multiple pattern recognition tasks are overlapped for this challenge.

2.1 Pervasive Systems

Pervasive systems describe a paradigm in which computational elements enhance interaction and intelligence of environments and objects of interaction

in a person's daily life. While many sensors are used to collect data to guide these systems, visual sensors provide perhaps the richest data over short periods [54]. François Brémont describes five levels of computer vision functionality for understanding a scene: those of detection, localization, tracking, recognition and understanding. Especially for localization, vision based sensors provide the highest accuracy for acceptable convenience levels. While recently popularized RGB-D camera technologies provide fast and accurate body tracking, most RGB-D cameras operate in limited ranges, and only under controlled illumination conditions. For mobile robots, the use of these cameras have proven to be very useful, as face-to-face interaction with humans usually occurs over small distances. The depth camera based approaches also seem to help with the high computational demands of the traditional vision-based solutions.

Cameras installed in a smart environment are typically static, configured to cover a maximal area of interest. It is possible to use multiple cameras to deal with problems of occlusion and view angles that may not be adequate at any given situation, but multi-camera systems require more complex algorithms to integrate information coming from different cameras, and are subsequently more difficult to deploy. In [14] a low-cost silhouette-based pose representation is obtained from multiple cameras and fused for action recognition.

It is obvious that installing sensors on a robot is fundamentally different than deploying the sensors on a smart environment. While the former provides a certain flexibility, it is limited by resource constraints of the robot. A promising approach to overcome some of these limits is the combination of sensors in a smart environment with the sensors on the robot. In [23], a Bayesian framework is described where a ceiling mounted camera is used for detection and tracking of people in conjunction with a laser range finder located on a mobile robot.

2.2 Action and Activity Recognition

Understanding human action mostly boils down to finding good representations of the sensed primitives. The chosen representation should be rich enough to differentiate between the action classes targeted by the application, but often it is not chosen to be much richer than that. The reason for this is purely pragmatic; more powerful representations require correspondingly complex training procedures, more training samples for learning, and longer computation time during operation. Consequently, the human body, for instance, is often represented by a graph structure made up of nodes representing landmark points on the body, and edges that connect these nodes in a fixed topology. Refinement on such a representation may be achieved by adding more landmarks (i.e. nodes) to the body parts being modeled.

In approaches where interest points do not necessarily correspond to known landmarks, space-time corners and similar 'salient' points are detected and used for learning spatio-temporal representations of actions [33]. In the present volume, Çeliktutan et al. propose an approach to solve the point set matching problem for establishing the correspondence between an action, represented by

interest points, to a template [13]. In [14], silhouettes are used for action recognition. The action template in this case is a bag of key poses representing the action in a temporal sequence.

3 Social and Affective Signals

Action recognition literature mostly focuses on simple actions, performed by a single actor [48]. A broad class of actions, however, are social in nature, and require either detailed analysis of multiple actors performing in tandem, or the distinction of very fine cues that can easily change the meaning of an action semantically. For instance, it takes a very small cue, like the creasing of the eye corners to change the meaning of a smile. Social signal processing arose from the need of intelligent systems interacting with humans to interpret and reproduce social signals, and to increase the sensitivity of the computer (or of the robot) to the interacting person's emotional and mental state [7,55]. Social signals are communicative or informative signals or cues "that directly or indirectly provide information about 'social facts': social interactions, social emotions, social attitudes, evaluations and stances, social relations, and social identities." [47].

3.1 Multimodal Analysis of Social Signals

Humans convey social information in many different ways. Facial expressions, posture, gait, body and hand gesture, speech, vocal prosody, and nonverbal cues like turn-taking behavior can all contain information relevant for interactions. Not all these signals are consciously or cognitively produced. In the present volume, Vincze et al. discuss problems that arise when people provide a certain information in a vague or approximate way, as well as the case where detectable cognitive qualities are associated with conveying information, like hesitation or hastiness [68]. An important point we made in the Introduction section of this paper is that semiotic conventions need to be established between a robot and a human in communication. While vagueness can arise because of an information gap, it can also be a device to leave open the goals designated in the communicated message. What would, for instance, be the benefit of employing vagueness when communicating with a robot? It can very well be to set up a situation where the robot decides on the correct level of abstraction or a most plausible resolution of the vague reference by examining other information available to it. This is a flexibility people have in human-human communication, and would eventually require in human-robot communication.

In natural interactions, humans also emit signals that have no real counterpart for robots. Research into human behavior understanding creates methods of analyzing these signals, which will open up new response patterns for robotic systems. In [38], an algorithm is described to determine a laughter index from visual input. This research is part of the EU-ICT FET Project ILHAIRE, which is aimed at endowing machines with automated detection, analysis, and synthesis of laughter. The authors use psychophysical descriptions of the laughter process

and propose a set of features including shoulder and body movement energy and periodicity. Obviously, a better understanding of the features that lead to accurate detection of laughter will also help us build systems that can synthesize realistic instances of laughter.

3.2 Perception of Affect

Emotions are important modifiers of human behavior, serving to enrich the response palette, but also allowing faster and contextualized decisions to help the human function better. Part of the importance of emotions also comes from the fact that humans are quite adept at recognizing emotional displays in others, and this forms the backbone of a social existence. In fact, this capability is so strong that humans easily attribute affect even to technological artifacts, as the well known Heider-Simmel study has demonstrated with simple moving geometric shapes [22]. In [58], Hylozoic Soil, a responsive architectural geotextile environment, is used to induce affective responses in viewers. Basic emotions like anger, sadness and happiness can be conveyed with simple movements of these dynamic structures. The authors also establish that there are gender differences in the perception of these affective movements [58]. These studies confirm that social interaction between humans and robots cannot ignore the affective dimension.

Movement is rarely used for automatic affect analysis of humans. In face to face communication, robots can observe the facial expressions of the interacting humans, as well as analyze the voice for affective signals. These are the most typically used modalities for affect analysis. In the present volume, Lim and Okuno show that a robot can also use the gait of a person to determine affective states [34]. In their approach, speed, intensity, irregularity, and extent features are extracted from the gait and speech of persons to determine affective states like happiness, sadness, anger, and fear. The advantage of using gait is that the face may not be available to a robot at all times, and the movement and resolution of the face may make emotion recognition difficult.

Ziemke and Lowe characterize emotion as (a) being closely connected to embodied cognition, (b) grounded in homeostatic bodily regulation, and (c) a powerful and useful organizational principle for modulation of behavioral and cognitive mechanisms [70]. Their focus is on maintaining emotion as an integral part of the internal environment of a robot, and as they admit, the role of emotion in social interactions is not addressed in their work, but they do note that the interplay of internal (i.e. individual) and external (i.e. social) aspects of emotion is still not very well known [3]. Robotic platforms can be excellent experimental tools for probing into these relatively unexplored areas.

4 Human-Robot Interaction

One of the long term ambitious goals of robotics research is to have robots capable of seamlessly integrating themselves in our daily environments. Therefore, recognizing, interpreting, and reasoning about the human behavior is a critical

skill for a robot that co-inhabits the human environments and interacts with humans on a regular basis. Particularly difficult challenges in human behavior understanding from the robotics point of view are the necessity to perform the processing using the limited computational resources on board, and using the sensors that can be mounted on a robotic platform.

4.1 Interacting with Robots

In general, the human-robot interaction (HRI) research can be divided into two main categories:

- **Human-centered HRI** investigates issues like the design and usability of proper interaction interfaces, robot platforms, and behaviors through extensive user studies.
- **Robot-centered HRI** focuses on algorithms, engineering innovations, and other computational approaches that would improve the overall performance of the interaction.

Although there is no clear distinction, the majority of the research on synthesizing behaviors, facial expressions and whole body gestures, and the development of proper interaction media fall into the human-centered HRI branch, especially from the validation point of view, while perceiving and interpreting behaviors, recognizing speech, and interactive learning applications fall into the robot-centered HRI branch.

A good example of the first approach is [28] in this volume, which reports the use case development for an outdoor robotic tour guide. In this work, abstractions of human behaviors appropriate for robot tour guides were developed. These abstractions form the basis of implemented robotic behaviors, which are then assessed in the real application scenario, where the robot meets visitors in a fairly unconstrained manner.

4.2 Closing the Interaction Loop

In the present volume, Fischer and Saunder investigate how people's initial expectations from an interaction, and their increasing experience and acquaintance with the robot over prolonged interaction sessions affect the way people tend to interact with robots [18]. Speech-based interaction has been heavily studied over the past decade. Grounding spatial commands given using unrestricted natural language for commanding a robot to navigate in the environment and manipulate objects have been studied in [65,24,29].

Humans also use gaze and gestures heavily to narrow down the uncertainties about the context when conversing verbally. Especially, forming joint attention through modeling the gaze of a human can be very useful in human-robot collaboration scenarios or when a human teacher teaches tasks or concepts involving the objects in the environment [69,63]. In [69], object saliency is used in conjunction with head pose estimates to allow a humanoid robot to determine the

visual focus of attention of the interacting human, while in [63] a fixed mapping between head pose directions and gaze target directions was not assumed, and models are investigated that perform a dynamic (temporal) mapping implicitly accounting for varying body/shoulder orientations of a person over time, as well as unsupervised adaptation.

Closing the interaction loop requires robots that behave closer to humans, and have more exploratory behavior than currently allowed for. An important concept related to the exploration capabilities of the robot is the notion of “*Symbiotic Autonomy*”. Accepting the fact that the robot has physical and cognitive limitations, and assuming the robot is also aware of some of its limitations, symbiotic autonomy advocates the benefits of engaging with the humans in the environment in a symbiotic relationship so that the robot does tasks for people, and asks people for help whenever its capabilities fall short of dealing with a certain situation [50]. Human interaction with the objective of asking for help raises new challenges like how and where to find humans who would likely provide help [51,52], and if there are more than one human present in the scene, whom to approach, as well as how to approach. Especially for the latter case, the ability to infer the intent of people as well as their predicted movement trajectories can drastically improve the way the robot interacts with the humans, and hence, the quality of the help it receives.

5 Imitation and Learning from Demonstration

Imitation is a process of paramount importance in both human-human and human-robot interaction. It is used for diverse functions, ranging from interaction regulation and social bonding to learning new knowledge and new competencies from others. In the recent years, imitation has been highly explored in various robotics contexts: its role for natural, intuitive and usable human-robot social interaction [46], robot learning of new tasks from demonstration [6,4], and its origins and functions in the course of epigenesis in developmental robotics [2,27,5]. Imitation learning in particular poses fundamental and challenging scientific problems [45], related to what, when and who to imitate, and it may be achieved at various levels of abstractions. Lopes et al. [36] describe three main levels of abstraction in imitation, which are respectively addressed by three chapters in this book: Mimicking behavior and trajectory-level imitation [44], imitation mediated by the action system and motor primitives [61], and imitation of goals and intentions [39].

The first level of abstraction in imitation is **mimicking**, where imitation consists of directly trying to reproduce the observed movements without an attempt to infer their underlying structure or goals [36]. A large amount of methods and approaches have been developed within this approach, and more particularly in the context of imitation learning where many researchers have studied how machine learning regression techniques could be used to reproduce smooth and generalizing trajectories out of sets of noisy human demonstrations (e.g. [6,11,20,12]). Besides robot learning of new skills by trajectory-level imitation, it

is also highly useful that robots be able to detect when two humans, or a human and a robot, are imitating each other [44]. Michelet et al. propose a combined computer vision and machine learning approach, which targets automatic identification of imitative interaction among humans [44]. A useful feature of this approach for easy deployment in the real-world is that the approach is capable of analyzing fine aspects of movements without the need to identify and track human skeletons.

The second level of abstraction in imitation leverages the **mediation of the action repertoire**, and consists in first interpreting observed behavior in terms of one's own repertoire of motor primitives, which are then re-used to generate the imitation [36]. Such approaches have been gaining popularity in robot learning recently, in particular because such an approach allows to reduce the dimensionality of movement and behavior representations meaningfully, which in turn often allows for better robustness and generalization [41,26]. A central question within these approaches, both in biology and robotics, is to understand how these motor primitives form initially. Some works have explored various learning techniques that allow to automatically infer and learn motion primitives from observation of human behaviors, e.g. [60,32,31,40]. In the present volume, Schillaci et al. study a complementary question [61]: once motor primitives have been learnt - in this case with the help of an annotated database -, how can they be used to recognize human actions and disambiguate potential targets? Using such a motor primitive representation, in the form of paired forward and inverse models, is highly useful, since it can allow direct reproduction of the observed behavior.

The third level of abstraction in imitation is **goal imitation** [36]. Here, the imitator tries to infer the intention, or the goal of the observed behavior, and then tries to reproduce this goal, possibly with different means (for example with a different motor policy). Mathematically, this amounts to learning the hidden cost function that the observed behavior may try to maximize, and then using this cost function to define a surrogate optimization/learning problem which the imitator has to solve. From a theoretical point of view, these approaches have been studied in two fields, optimal feedback control [49] and inverse reinforcement learning (IRL) [1], respectively. In recent years, they have been applied to imitation learning in robotics, where they have been shown to be powerful for generalization and robust to environment change at the same time [1,67]. For example, Abbeel and Ng[1] showed how autonomous helicopters could learn to achieve acrobatic flights better than professional human demonstrators using this approach. Lopes et al. [37] showed how active learning techniques could be used to increase the efficiency of IRL. In the present volume, Mangin and Oudeyer explore a frontier of these approaches [39]: how can a robot learn the combinatorial structure of the hidden goals underlying demonstrated behaviors? Previous approaches assumed that observed behavior corresponds to a single hidden cost function/goal. On the contrary, Mangin and Oudeyer consider the case when the demonstrator has a repertoire of hidden goals and only produces behaviors which concurrently target several goals. The proposed approach relies

on establishing a bridge between inverse feedback control techniques and dictionary learning techniques [25]. Like [61], [39] infers a motor representation of observed behaviors of a demonstrator, which allows the learning system to both recognize and reproduce behavior with adequate generalization.

6 Applications

In the future, we envision robots that not only assist humans in domestic environments, but also interact with them in public spaces and factories. These robotic applications will require proper social interactions to be maintained between robots and humans. We will move from the current situation in which robots carry out certain tasks without any interaction with humans (e.g. automotive factory) to situations in which humans and robots will co-work, and subsequently robots becoming co-workers and co-inhabitants [21], carrying out tasks of increasing complexity that require understanding the behavior of humans. Even the simple task of cooperatively carrying a table by a human and a robot [64], requires the synchronization of the individual's movements, which can only be achieved if the other's behavior is correctly analyzed in real-time with the correct resolution and level of abstraction.

In this section, we give some application examples to make the requirements more concrete.

6.1 Socially Assistive Robotics

One of the envisioned applications of robotics is assisting specific human populations, such as children, elderly people, and patients. These are tasks that require specific expertise in relatively restricted domains, embodiment, and most importantly, a social aspect that makes robots preferable to automated systems that are less suited to display and interpret social and affective signals. Socially assistive robotics defines the robot's goal to be the creation of "close and effective interaction with a human user for the purpose of giving assistance and achieving measurable progress in convalescence, rehabilitation, learning, etc." [17]. Some related applications are robots as exercise coaches, evaluating the moves of the interacting humans [16], and guiding robots providing context-dependent information to people [28].

In these applications human behavior understanding will be crucial for interpreting the human needs and requirements, but also for understanding the mood and for taking actions to manage it appropriately. Understanding human moods and needs requires having some basic functionality. One of them is understanding the visual focus of attention of humans while interacting with robots. This is addressed in this volume [63].

Applications in which robots interact with humans have increased largely in the last years thanks to the development of the Microsoft Kinect sensor. The sensor's ability of obtain 3D-images at a low cost, and the availability of libraries with functionalities such as human body segmentation, have boosted

the development of HRI applications. Most of these applications are related with entertainment, although they can be expanded to education. However, the use of infrared lighting makes the sensor being surface dependent (e.g. on black surfaces the reflection is very limited), and not appropriate to be used outdoors. Therefore, complementary sensors need to be used in applications with those constraints.

6.2 Playful Interactions

The work of Cynthia Breazeal and others has established that people interacting with robots will treat the robot as a social entity [8,10]. Consequently, robots have the potential to be much more than elaborate toys in children's games. In social games of children, interactions are not pre-determined, but emerge through mutual interaction. The ideal game partner is thus one that adapts to a game scenario, and one that can assume one of many different roles, each as coherent as possible in the social and affective displays that belong to the designated role. The contribution of human behavior understanding to this kind of a scenario would be the detailed analysis of gaming roles to create the coherent role models, as well as real-time observation of the playing partners to determine which mode should be selected and put into action.

A less ambitious, but worthy goal is to use robots as mediators in playful interactions. A very important research direction is for instance the work with autistic children, who may shun social contact in the form provided by their peers, but may come to like what a social robot has to offer. An example is the work of Michaud and Théberge-Turmel, who used robust robotic toys in play experiments with children to obtain promising results [43]. The AURORA project is an important initiative in this area with the aim of to encouraging autistic children "to become engaged in a variety of different interactions important to human social behavior" [15]. Another good example of toy robots for interacting with children is the Keepon robot, which is capable of conveying limited emotion and attention, promoting social playful interaction [30].

The basic idea that underlies these applications is that play is a fundamental activity in learning social interactions. While human behavior understanding has been used in gaming scenarios in the design phase, to specify interaction scenarios, real-time behavior analysis is only recently being integrated into games [62].

7 Conclusions

In this introductory paper of the 3rd International Workshop on Human Behavior Understanding, our primary aim was to articulate the points of contact between robotics and human behavior understanding. It is clear that progress in the latter will have direct bearing on the design and implementation of robots that have social skills and interact with humans in more natural ways. The proper approach to do this is not mere imitation of the human behavior, but goes through a deeper understanding of the abstract processes leading to particular behavior and ways of interaction, so as to let the counterparts emerge in

the interacting robots. Obviously, a lot of basic skills must be in place before this can be achieved.

The second important point is what robotics has to offer to human behavior understanding, especially in terms of new scientific questions it poses. Since robots need to act in an embodied manner, it is essential that human-behavior understanding capabilities provided to/learned by robots are adapted to allow leveraging this understanding (e.g. the representations) to act appropriately. Purely functional representations may not be sufficient, and robotics is an excellent testbed for this; if the correct abstraction is not achieved, transferring behavior patterns to the robot will not be successful.

A final point is that the presence of the robots causes changes in the behavior of humans. It is important to understand what kind of new social situations are created by putting robots with social capabilities, and social roles, in a natural environment. As the skill palette of robots grows, and they start reading and responding to social and affective displays of humans, these mutual relationships will be increasingly complex, and will require more thorough analysis.

Acknowledgments. This work is supported by INRIA project PAL, Boğaziçi University project BAP-6531, EUCogIII, and ERC EXPLORERS 240007.

References

1. Abbeel, P., Ng, A.Y.: Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the 21st International Conference on Machine Learning (ICML 2004), pp. 1–8 (2004)
2. Andry, P., Gaussier, P., Nadel, J., Hirsbrunner, B.: From sensori-motor development to low-level imitation. *Adaptive Behavior* 12, 117–138 (2004)
3. Arbib, M.A., Fellous, J.M.: Emotions: from brain to robot. *Trends in Cognitive Sciences* 8(12), 554–561 (2004)
4. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57(5), 469–483 (2009)
5. Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., Yoshida, C.: Cognitive developmental robotics: A survey. *IEEE Trans. Autonomous Mental Development* 1(1) (2009)
6. Billard, A., Calinon, S., Dillmann, R., Schaal, S.: Survey: Robot programming by demonstration. In: *Handbook of Robotics*, ch. 59 (2008)
7. Breazeal, C.: Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59(1-2), 119–155 (2003)
8. Breazeal, C.: Toward sociable robots. *Robotics and Autonomous Systems* 42(3), 167–175 (2003)
9. Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., Blumberg, B.: Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life* 11(1-2), 31–62 (2005)
10. Brooks, A.G., Gray, J., Hoffman, G., Lockerd, A., Lee, H., Breazeal, C.: Robot’s play: interactive games with sociable machines. *Computers in Entertainment (CIE)* 2(3), 1–10 (2004)

11. Calinon, S., Guenter, F., Billard, A.: On learning, representing and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man and Cybernetics, Part B* 37(2), 286–298 (2007)
12. Cederborg, T., Li, M., Baranes, A., Oudeyer, P.-Y.: Incremental local inline gaussian mixture regression for imitation learning of multiple tasks. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan (2010)
13. Çeliktutan, O., Wolf, C., Sankur, B., Lombardi, E.: Real-Time Exact Graph Matching with Application in Human Action Recognition. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 17–28. Springer, Heidelberg (2012)
14. Chaaraoui, A.A., Climent-Pérez, P., Flórez-Revuelta, F.: An Efficient Approach for Multi-view Human Action Recognition Based on Bag-of-Key-Poses. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 29–40. Springer, Heidelberg (2012)
15. Dautenhahn, K., Werry, I.: Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition* 12(1), 1–35 (2004)
16. Fasola, J., Matarić, M.J.: Robot exercise instructor: A socially assistive robot system to monitor and encourage physical exercise for the elderly. In: *19th IEEE International Symposium in Robot and Human Interactive Communication*, Viareggio, Italy, pp. 416–421 (September 2010)
17. Feil-Seifer, D., Matarić, M.J.: Defining socially assistive robotics. In: *9th International Conference on Rehabilitation Robotics*, ICORR 2005, pp. 465–468. IEEE (2005)
18. Fischer, K., Saunders, J.: Between Initial Expectations and Acquaintance: Interacting with a Developing Robot. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 125–133. Springer, Heidelberg (2012)
19. Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V.: Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience* 19(11), 1803–1814 (2007)
20. Grollman, D.H., Jenkins, O.C.: Sparse incremental learning for interactive robot control policy estimation. In: *International Conference on Robotics and Automation (ICRA 2008)*, pp. 3315–3320 (May 2008)
21. Guizzo, E., Deyle, T.: Robotics trends for 2012 (the future is robots). *IEEE Robot. Automat. Mag.* 19(1), 119–123 (2012)
22. Heider, F., Simmel, M.: An experimental study of apparent behavior. *The American Journal of Psychology* 57(2), 243–259 (1944)
23. Hu, N., Englebienne, G., Kröse, B.: Bayesian Fusion of Ceiling Mounted Camera and Laser Range Finder on a Mobile Robot for People Detection and Localization. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 41–51. Springer, Heidelberg (2012)
24. Huang, A.S., Tellex, S., Bachrach, A., Kollar, T., Roy, D., Roy, N.: Natural language command of an autonomous micro-air vehicle. In: *Int. Conf. on Intelligent Robots and Systems* (IROS), Taipei, Taiwan (October 2010)
25. Jenatton, R., Mairal, J., Obozinski, G., Bach, F.: Proximal Methods for Hierarchical Sparse Coding. *Journal of Machine Learning Research* 12, 2297–2334 (2011), <http://hal.inria.fr/inria-00516723>
26. Jenkins, O.C., Matarić, M.J., Weber, S.: Primitive-based movement classification for humanoid imitation. In: *IEEE International Conference on Humanoid Robots, Humanoids 2000* (2000)

27. Kaplan, F., Oudeyer, P.-Y.: The progress-drive hypothesis: an interpretation of early imitation. In: Dautenhahn, K., Nehaniv, C. (eds.) *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge University Press (2007)
28. Karreman, D.E., Evers, V., van Dijk, E.M.A.G.: Contextual Analysis of Human Non-verbal Guide Behaviors to Inform the Development of FROG, the Fun Robotic Outdoor Guide. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) *HBU 2012. LNCS*, vol. 7559, pp. 113–124. Springer, Heidelberg (2012)
29. Kollar, T., Tellex, S., Roy, D., Roy, N.: Toward understanding natural language directions. In: *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, pp. 259–266. IEEE Press, Piscataway (2010), <http://dl.acm.org/citation.cfm?id=1734454.1734553>
30. Kozima, H., Michalowski, M.P., Nakagawa, C.: Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics* 1(1), 3–18 (2009)
31. Krüger, V., Herzog, D., Baby, S., Ude, A., Kragic, D.: Learning actions from observations. *IEEE Robot. Automat. Mag.* 17(2), 30–43 (2010)
32. Kulić, D., Nakamura, Y.: Incremental Learning of Full Body Motion Primitives. In: Sigaud, O., Peters, J. (eds.) *From Motor Learning to Interaction Learning in Robots. SCI*, vol. 264, pp. 383–406. Springer, Heidelberg (2010)
33. Laptev, I.: On space-time interest points. *International Journal of Computer Vision* 64(2), 107–123 (2005)
34. Lim, A., Okuno, H.G.: Using Speech Data to Recognize Emotion in Human Gait. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) *HBU 2012. LNCS*, vol. 7559, pp. 52–64. Springer, Heidelberg (2012)
35. Lopes, M., Oudeyer, P.-Y.: Active learning and intrinsically motivated exploration in robots: Advances and challenges (guest editorial). *IEEE Transactions on Autonomous Mental Development* 2(2), 65–69 (2010)
36. Lopes, M., Melo, F., Montesano, L., Santos-Victor, J.: Abstraction Levels for Robotic Imitation: Overview and Computational Approaches. In: Sigaud, O., Peters, J. (eds.) *From Motor Learning to Interaction Learning in Robots. SCI*, vol. 264, pp. 313–355. Springer, Heidelberg (2010)
37. Lopes, M., Melo, F., Montesano, L.: Active Learning for Reward Estimation in Inverse Reinforcement Learning. In: Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds.) *ECML PKDD 2009, Part II. LNCS*, vol. 5782, pp. 31–46. Springer, Heidelberg (2009)
38. Mancini, M., Varni, G., Glowinski, D., Volpe, G.: Computing and Evaluating the Body Laughter Index. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) *HBU 2012. LNCS*, vol. 7559, pp. 90–98. Springer, Heidelberg (2012)
39. Mangin, O., Oudeyer, P.-Y.: Learning the Combinatorial Structure of Demonstrated Behaviors with Inverse Feedback Control. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) *HBU 2012. LNCS*, vol. 7559, pp. 135–147. Springer, Heidelberg (2012)
40. Mangin, O., Oudeyer, P.-Y.: Learning to recognize parallel combinations of human motion primitives with linguistic descriptions using non-negative matrix factorization. To Appear in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2012)
41. Mataríć, M.J.: Sensory-motor primitives as a basis for learning by imitation: linking perception to action and biology to robotics. In: *Imitation in Animals and Artifacts*. MIT Press (2002)

42. Meriçli, Ç., Veloso, M., Akın, H.L.: Improving biped walk stability with complementary corrective demonstration. *Autonomous Robots* 32(4), 419–432 (2012), <http://dx.doi.org/10.1007/s10514-012-9284-1>
43. Michaud, F., Théberge-Turmel, C.: Mobile robotic toys and autism. *Socially Intelligent Agents*, 125–132 (2002)
44. Michelet, S., Karp, K., Delaherche, E., Achard, C., Chetouani, M.: Automatic Imitation Assessment in Interaction. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 161–173. Springer, Heidelberg (2012)
45. Nehaniv, C.: Nine billion correspondence problems. In: Dautenhahn, K., Nehaniv, C. (eds.) *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge University Press (2007)
46. Nehaniv, C.L., Dautenhahn, K. (eds.): *Imitation and social learning in robots, humans, and animals: behavioural, social and communicative dimensions*. Cambridge University Press (2004)
47. Poggi, I., D’Errico, F.: Social signals: a framework in terms of goals and beliefs. *Cognitive Processing* (2012)
48. Poppe, R.: A survey on vision-based human action recognition. *Image and Vision Computing* 28(6), 976–990 (2010)
49. Ratliff, N., Bagnell, J., Zinkevich, M.: Maximum margin planning. In: Proc. 23rd Int. Conf. Machine Learning, pp. 729–736 (2006)
50. Rosenthal, S., Biswas, J., Veloso, M.: An effective personal mobile robot agent through symbiotic human-robot interaction. In: *International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, vol. 1, pp. 915–922 (May 2010)
51. Rosenthal, S., Veloso, M.M., Dey, A.K.: Acquiring accurate human responses to robots’ questions. I. *J. Social Robotics* 4(2), 117–129 (2012)
52. Rosenthal, S., Veloso, M.M., Dey, A.K.: Is someone in this office available to help me? - proactively seeking help from spatially-situated humans. *Journal of Intelligent and Robotic Systems* 66(1-2), 205–221 (2012)
53. Salah, A., Schouten, B.: Semiosis and the relevance of context for the AmI environment. In: Proc. European Conf. on Computing and Philosophy (2009)
54. Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A.: Computer vision for ambient intelligence. *Journal of Ambient Intelligence and Smart Environments* 3(3), 187–191 (2011)
55. Salah, A.A., Pantic, M., Vinciarelli, A.: Recent developments in social signal processing. In: *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 380–385. IEEE (2011)
56. Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A.: Challenges of Human Behavior Understanding. In: Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A. (eds.) HBU 2010. LNCS, vol. 6219, pp. 1–12. Springer, Heidelberg (2010)
57. Salah, A.A., Lepri, B., Pianesi, F., Pentland, A.: Human Behavior Understanding for Inducing Behavioral Change: Application Perspectives. In: Salah, A.A., Lepri, B. (eds.) HBU 2011. LNCS, vol. 7065, pp. 1–15. Springer, Heidelberg (2011)
58. Samadani, A.-A., Gorbet, R., Kulić, D.: Gender Differences in the Perception of Affective Movements. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 65–76. Springer, Heidelberg (2012)
59. Schaal, S., Ijspeert, A., Billard, A.: Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358(1431), 537–547 (2003)

60. Schaal, S., Peters, J., Nakanishi, J., Ijspeert, A.: Learning movement primitives. In: International Symposium on Robotics Research, ISRR 2003 (2003)
61. Schillaci, G., Lara, B., Hafner, V.: Internal Simulations for Behaviour Selection and Recognition. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 148–160. Springer, Heidelberg (2012)
62. Schouten, B.A.M., Tieben, R., van de Ven, A., Schouten, D.W.: Human Behavior Analysis in Ambient Gaming and Playful Interaction. In: Salah, A.A., Gevers, T. (eds.) Computer Analysis of Human Behavior, pp. 387–403. Springer-Verlag London Limited (2011)
63. Sheikhi, S., Odobez, J.-M.: Recognizing the Visual Focus of Attention for Human Robot Interaction. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 99–112. Springer, Heidelberg (2012)
64. Stuckler, J., Holz, D., Behnke, S.: Robocup@home: Demonstrating everyday manipulation skills in robocup@home. *IEEE Robotics Automation Magazine* 19(2), 34–42 (2012)
65. Tellex, S., Kollar, T., Dickerson, S., Walter, M.R., Banerjee, A.G., Teller, S., Roy, N.: Understanding natural language commands for robotic navigation and mobile manipulation. In: Proceedings of the National Conference on Artificial Intelligence (AAAI) (August 2011)
66. Thomaz, A.L., Breazeal, C.: Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence Journal* 172, 716–737 (2008)
67. Verma, D., Rao, R.: Goal-based imitation as probabilistic inference over graphical models. In: Advances in NIPS 18 (2006)
68. Vincze, L., Poggi, I., D’Errico, F.: Vagueness and Dreams. Analysis of Body Signals in Vague Dream Telling. In: Salah, A.A., Ruiz-del Solar, J., Meriçli, Ç., Oudeyer, P.-Y. (eds.) HBU 2012. LNCS, vol. 7559, pp. 77–89. Springer, Heidelberg (2012)
69. Yücel, Z., Salah, A.A., Meriçli, Ç., Meriçli, T.: Joint visual attention modeling for naturally interacting robotic agents. In: 24th International Symposium on Computer and Information Sciences, ISCIS 2009 (2009)
70. Ziemke, T., Lowe, R.: On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cognitive Computation* 1(1), 104–117 (2009)