Int J Soc Robot (2012) 4:5–14
DOI 10.1007/s12369-011-0127-6

Perception and Action as Two Sides of the Same Coin. A Review of the Importance of Action-Perception Links in Humans for Social Robot Design and Research

Agnieszka Wykowska · Anna Schubö

Accepted: 16 November 2011 / Published online: 9 December 2011

Abstract This paper focuses on the topic of human cognitive architecture in the context of links between action and perception. Results from behavioural studies, neuroimaging, human electrophysiology as well as single-cell studies in monkeys are discussed. These data as well as theoretical background are brought forward to argue that a close connection between action and perception should be considered in designs of artificial systems. Examples of such systems are described and the application of those approaches to robotics is stressed.

Keywords Human action planning · Action-perception links · Perceptual selection · Social robotics

1 Introduction

When designing an artificial robotic system, which is supposed to act and interact with its environment, one needs to equip it with some sort of perceptual module that connects with its motor module in order to provide information for efficient behaviour with respect to the incoming input. However, the perception-action connection, in order to be efficient should most probably not be unidirectional—at least this is what the human cognitive architecture shows. In the following, based on empirical evidence from human cognitive neuroscience, we will argue that action and perception system need to be tightly coupled in a bidirectional way, in order to provide an efficient architecture for interacting with the environment. The views presented in this paper might constitute an example of a framework that belongs to the embodied cognition approach in the cognitive sciences (for a review see [1]). In contrast to traditional accounts in cognitive sciences, embodied cognition views the cognitive capabilities of humans not in isolation from action, environment, and the body, but rather as embedded in the interaction between the three. This might imply that human's inner representations do not consist in abstract symbols and operations thereupon, but rather on simulated actions and interactions with the environment. A prototypical example of such an embodied representational architecture is the work of Lynn Andrea Stein [2] implemented in a robotic system of Toto and MetaToto—designed by Maja Matarić [3]. Those systems, equipped with sensors to detect walls, corridors and obstacles, develop inner representations of the environment that are "action-related". That is, instead of coding landmarks in an abstract way, these systems code them in terms of the robots' movement and perceptual input. Such "embodied" approach might prove to be an extremely efficient way of designing artificial systems, given that growing body of evidence supports the idea of a close coupling between perception, action and environment in humans.

2 Ideomotor Perspectives on Action Control in Humans

In order to efficiently interact with the environment, humans must have developed adequate mechanisms for action planning. According to ideomotor perspectives, e.g., [4–6],

A. Wykowska (✉)
Department of Psychology, Ludwig-Maximilian University, Munich, Germany
e-mail: agnieszka.wykowska@psy.lmu.de

A. Schubö
Department of Psychology, Philipps University, Marburg, Germany
e-mail: anna.schuboe@staff.uni-marburg.de

humans plan their actions according to action goals. Action goals, in turn, are formed by anticipated sensory consequences of the intended actions. Greenwald [4] as well as Hommel et al. [5] suggest that representation of action goals in terms of anticipated sensory consequences is possible thanks to life-long experience with various actions having not only proximal effects (re-afferent consequences of particular movements) but also distal effects, i.e., sensory consequences of changes in the surrounding that are related to the performed action (e.g., if one strikes a ball with a baseball bat, one will see the ball moving and also hear a sound of the ball hitting the bat). To account for the ideas of action representation in the form of anticipated sensory consequences, Harless [7] postulated that acquiring voluntary action control mechanisms takes place in two stages. In the first stage, actions are carried out randomly; the effects of those actions are perceived and memorized. Subsequently, as the links between given action types and their sensory consequences become available, they can be used to intentionally plan actions to produce desired effects. Such a view on development of voluntary action control fits to everyday observations of how babies first randomly move and only subsequently learn to control their actions by anticipating sensory consequences thereof. When the links between actions and their perceptual consequences are acquired and action effects can be anticipated, the human brain is able to evaluate if the actual consequences of particular actions match the anticipated/desired ones (for a similar account, see forward models, e.g., [8]).

Greenwald [9] has been one of the first to provide empirical evidence for the ideomotor principles. Results of his studies showed, for example, that verbal responses were faster to auditory stimuli as compared to visual stimuli—presumably because speaking usually produces auditory sensations, and hence humans link verbal behaviour with auditory sensory effects. Elsner and Hommel [10] designed experiments based on Harless' [7] idea of the two-stage model of action control with action effects acquisition in the first stage and anticipation of the acquired effect in the second stage. In their studies, participants first participated in the acquisition part of the experiment, in which they were asked to respond freely to a stimulus with either a left or right key. Upon the response, a sound was presented that differed depending on the key (e.g., right key-high tone, left key-low tone). The sounds were completely task-irrelevant and the participants were asked to ignore them. In the second (test) phase, participants were asked to perform a forced choice task (response assignment was fixed) to an auditory stimulus (e.g., high tone-left key; low tone-right key). Half of the participants' response assignment was compatible with the previous response-sound event (same sounds were associated with the same responses). The other half of participants had a reversed response assignment. The results showed a significantly better performance for the non-reversal group of participants indicating that the sensory consequences of particular responses have been acquired leading to anticipatory mechanisms when producing the given actions. In another experiment, when given a free choice to select a response in the test phase, participants selected a response with the compatible key significantly more often than with the incompatible key (compatibility was defined as same stimulus-response pairs as in the acquisition phase). These experiments show that humans learn the sensory effects of actions they produce. Representation of those effects is presumably subsequently used to guide efficient action planning.

3 Common Code for Perception an Action

If humans plan their actions through representing anticipated sensory consequences of the intended actions, then the human brain must have developed a mechanism for efficient "communication" between action and perception domains. Therefore, authors favouring the ideomotor principles postulate that action and perception share a common representational code, e.g., [5, 6]. The format of the code presumably consists in a distributed network of features [5] that are bound together across domains to form events. That is, if hitting the ball with a bat is always associated with a given sound and a given movement of the ball, the action of striking the ball with the bat and the perception of sound and motion will probably tend to being bound together to one event.

Furthermore, according to Hommel [11], action and perception codes are functionally equivalent. This presupposes that perception is an active process: perceiving means active acquisition of information through selective mechanisms, eye-, head- and other body movements. On the other hand, action would not be successful or efficient without perceptual feedback. Therefore, action and perception serve the same purpose: efficient interaction with the environment. Importantly, however, the common coding theories refer to representations of actions in terms of their distal effects but not proximal sensory effects. That is, common codes between perception and action do not refer to such characteristics like neuronal pattern of muscular activity while performing a movement. This is a crucial claim, as it would be rather implausible that neural activity related to a stimulation of retina by a given visual input is equivalent to muscular activity related to performing a grasping movement.

Evidence for bindings between perceptual and action features comes from, e.g., studies reporting partial-repetition costs [12] or inverse-compatibility effects [13]. Partial repetition effects have been observed in studies with paradigms consisting of two tasks: simple reaction[1] to Stimulus 1 (S1) onset and binary forced-choice reaction to Stimulus 2 (S2). In this way, stimulus and response characteristics could converge or vary orthogonally. Results demonstrated that repetition of stimuli and responses as well as change of both yield better performance than partial repetition, i.e., same response with alteration of stimulus or same stimulus with a different response. These effects suggested that encountering a particular event (S1) and issuing a relative response (R1) were bound together within an event file. The second encounter of the same stimulus (S2) activated R1 that was bound to it in the respective event file. However, if S2 required a different response (R2), then the event file needed to be updated which was time-consuming and elicited costs relative to the situation when the same stimulus-response binding could have been used to complete the task.

Similarly, inverse-compatibility effects have been observed by Müsseler and Hommel [13]. In these studies, action planning impaired perceptual processes when their activated codes overlapped: perception of a visual stimulus was impaired if it was presented while a particular action was being prepared, given that the stimulus shared certain characteristics with that action. That is, if participants were preparing, for example, a right-key response, and a right-pointing arrow was briefly presented (subsequently masked), accuracy of detecting the right arrow was lower as compared to detection of the left arrow.

Analogous results have been found by Zwickel et al. [14]. The authors asked participants of their experiments to produce simple movements along a linear trajectory in a given direction while observing a visual stimulus also moving along a linear trajectory. Subsequently, participants were asked to compare trajectory of a test stimulus against the reference stimulus (the stimulus presented concurrently with the produced movement). The results showed that the estimated trajectory was repulsed relative to the produced movement's trajectory. This showed a form of contrast/interference effects arguing in favor of shared representational medium for action and perception.

Neuroimaging techniques allowed for pinpointing brain areas that are active while certain stimuli are perceived. As it turned out, e.g., [15], not only perceptual areas but also motor areas of the brain can be automatically activated when a predictable sequence of stimuli is perceived. That is, in an fMRI study [15], participants observed sequences of either visually presented disks of various sizes or sounds of various pitches. Participants' task was to judge whether certain elements of the sequence matched the sequence "rule": e.g., if the sequence consisted in increasing size, the final three elements were also increasing in size. Activation of the premotor areas of the brain has been observed. These areas are typically described in terms of control of motor behavior and selection of appropriate movements for voluntary action. In the described study, regions that are usually involved in hand movements were activated when sequences of visually presented disks were of different sizes. On the other hand, premotor areas involved in articulation were activated when judgments were concerned with auditory stimuli of different pitch. These results show that processing perceptual features automatically activates actions that might be related with those features, and therefore, speaks in favour of strong action-perception links. Moreover, these studies stressed the importance of human ability to anticipate certain steps in a predictable sequence, which might be the base for action understanding, and therefore, might involve action-related areas of the brain.

Based on the results of these studies, a clear picture regarding human cognitive architecture emerges: action and perception domains seem to be strongly coupled and often co-active. Such coupling presumably allows efficient action planning which consists in anticipation of perceptual action effects. A common representational code is the most parsimonious solution to efficient communication between action and perception domains.

4 The Concept of Intentional Weighting

Bidirectional links between action and perception imply that action planning can influence and bias perception as much as perceptual mechanisms might influence action plans. This has indeed been one of the main postulates of the selection-for-action view [16], the premotor theory of selection [17] or of authors arguing in favour of a close coupling between selection for perception and selection for action [18]. Authors supporting the idea of a common code, e.g., [5] have postulated the concept of "intentional weighting" that should operate on perceptual processing in a similar way as other types of weighting mechanisms, see e.g., [19, 20]. That is, neural activity related to given characteristics of the environment might be modulated by top-down control to fit a given task, e.g., [21, 22] or action [5], see Fig. 1. Hommel [11] postulates that action control consist of offline action planning and online action adjustment. The first specifies invariant features of action representations while the latter fills in still open parameters of a given action plan in an online manner. In order for the system to have efficient access to relevant information that needs to be fed into the adjustment

---

[1] A task requiring a simple reaction means that participants are asked to respond to a presented stimulus as soon as they detect it, and the response is given with only one possible key whereas a binary forced-choice reaction means that participants are presented with two types of stimuli and are asked to issue one of two possible responses. Each of the stimuli is assigned to a pre-determined response (e.g., left orientation of a bar-stimulus = left key; right orientation = right key)

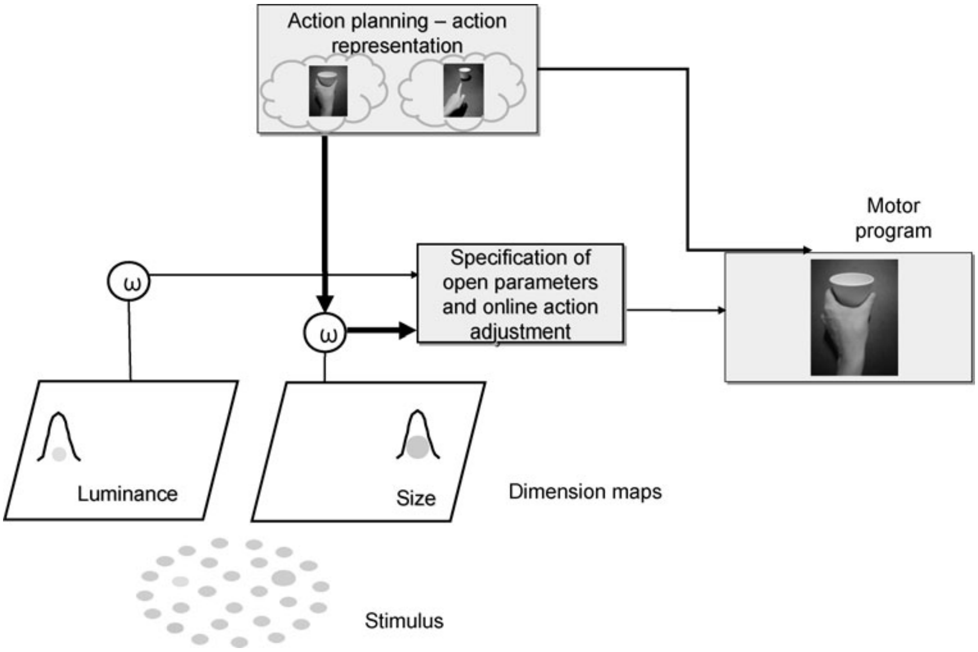

Fig. 1 A schematic illustration of the postulated mechanism that weighs perceptual processing with respect to action planning (based on a model in Hommel [11]). The visual scene is supposedly coded in separate dimension-specific maps, where saliency signals are elicited by the stimuli. Those saliency signals can be directly weighted with respect to particular task-relevant action plans (e.g., grasping or pointing). Such weighted signals are specifications of particular parameters for given action plans that subsequently are fed into a motor program

processes, an intentional weighting mechanism biases perception in such a way that processing action-relevant characteristics is prioritized over other features that the perceptual apparatus processes. That is, if one is planning a grasping action, particular grip aperture needs to be specified (parameter left open for online adjustment). Hence, size of the perceived (and to-be grasped) object needs to be calculated while information concerning for example its color is not relevant. Perceptual system, therefore, processes information about size with priority, allowing for efficient specification of open parameters (e.g., particular grip aperture) during online action adjustment. This mechanism is what Hommel [11] terms *intentional weighting*.

Such biased processing has been observed in a laboratory setup. For example, Craighero et al. [23] demonstrated that preparing for a grasping movement facilitated processing of a visual stimulus when the stimulus shared characteristics of the to-be grasped object (same orientation of the visually presented bar and to-be-grasped object). Analogously, Fagioli et al. [24] observed that participants tuned their perception to a concurrently prepared action. That is, the authors asked participants to prepare either a pointing or a grasping movement. At the same time, participants observed a sequence of stimuli and were asked to detect an oddball item in the sequence—an item that did not fit into a sequence “rule” (e.g., if a “rule” consisted in alternating in size disks: small-large-small, an oddball would occur when two disks of the same size were presented in a row). The oddballs could occur also through repetition of the same location. Hence, the perceptual task was concerned with either size (dimension relevant for grasping) or location (dimension relevant for pointing). As expected, participants exhibited better performance in size oddball detection when they were preparing to grasp, as compared to point, and they were better in location oddball detection when they prepared for pointing, relative to grasping. In a similar line of reasoning, Wykowska et al. [25] designed an experiment in which participants performed two tasks: a perceptual task that consisted in detecting a smaller circle (size target) or a lighter circle (luminance target) among many other circles presented on the computer screen (a typical visual search task), see Fig. 2.

The responses in this task were to be made on a computer mouse with one key being related to the target present trials and another key being related to target absent trials. The second task (a grasping or a pointing movement) was performed on items of especially designed device positioned below the computer screen. This task was executed with a different hand than the one used in the perceptual task. The experiment was designed to create two congruent perception-action pairs (size with grasping and luminance with pointing)—as it was assumed, in line with the ideomotor theories, that through lifelong experience with grasping, humans have learned size to be relevant for grasping whereas luminance (feature allowing fast localization) to be associated with pointing. Importantly, as described above, the two tasks of this paradigm were experimentally entirely unrelated—both motorically and perceptually. Thus, the results showing congruency effects (better performance in search for size when grasping as compared to pointing and better performance in search for luminance when pointing as compared to grasping) indicate that the facilitation of certain feature detection occurs at the representational level in the brain that links action and perception codes—and it is not due to motor priming (different effectors) or percep-

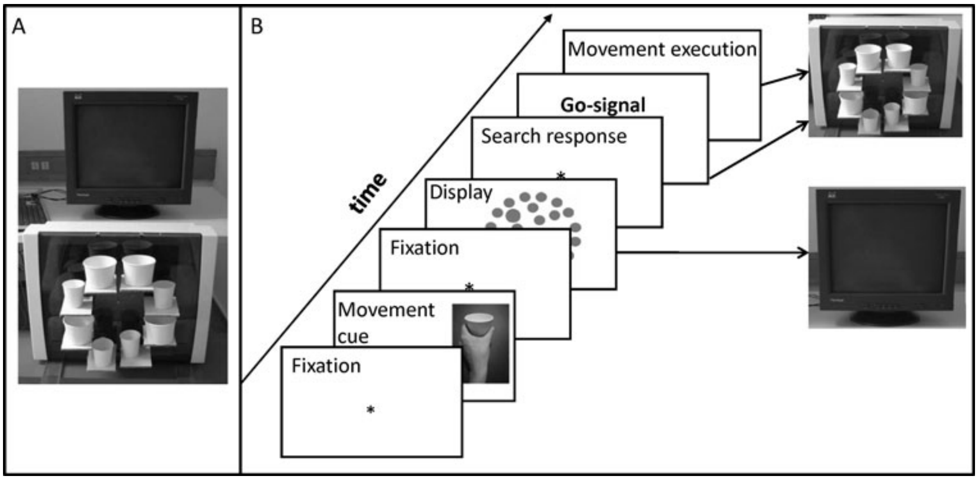

Fig. 2 An illustration of a paradigm similar to the one used in Wykowska et al. [25]. The experimental setup (**A**) and the trials sequence (**B**). The participants were first asked to fixate their eyes on the fixation star. Then, they observed a movement cue (a picture with a grasping or pointing movement) and they were asked to prepare the movement but not execute it at that point. Then, they were presented with a visual search display on the computer screen, in which they were to detect the presence of a circle differing from others in size or luminance. The participants were asked to press one mouse key for target presence and another for target absence. Subsequent to this response, one of the elements on the movement execution device (MED) lit up and served as a go- signal for the movement execution. Participants grasped or pointed to the element that lit up and this was the end of a trial

tual priming (different objects in the visual search and in the motor task). In a subsequent series of studies, Wykowska et al. [26] showed that these results occur even when perceptual similarity between the items (and their arrangement) on the computer screen and the device on which the movements were executed is reduced to minimum. Hence, the congruency effects seem not to be due to overlap of perceptual characteristics between the setup of the perceptual task and the movement task. These results suggest that perception can be biased with respect to action planning in a similar way as it can be biased by other top-down factors, e.g., [19, 20].

5 Perceiving the World in Action-Related Manner: The Concept of Affordances

The common coding perspectives share certain aspects of their account with the ecological approach of Gibson, e.g., [27]. The ecological approach also stresses the idea of active perception and is strongly related to the concept of affordances. Affordances are characteristics of the environment or specifications thereof in terms of possible interaction with it. For example, a cup would have a (learned) characteristic of “something to drink from”. A chair would have a characteristic of “sittability”. As such, affordances constitute close links between action and perception. In other words, perception of an object carries information not only about such characteristics like “whiteness” or “roundness” but also about “graspability”, and as such, activates the action system automatically.

Evidence for such affordances-based link between perception and action has been described by, for example, Humphreys and Riddoch [28], where the authors reported a case study of a patient with a neglect syndrome after damage in his right brain hemisphere. This patient was unable to direct attention to the neglected (left) side, and in effect, had difficulties in detecting objects presented in that visual hemifield, when the objects were defined by their name or other characteristics, such as colour. Remarkably, however, when the objects were defined through their affordances, i.e., “something to drink from”, for example, the patient’s performance in detecting the objects in the neglected side improved. Similarly, Tucker and Ellis [29] reported results supporting the idea of automatic processing of action-related characteristics of perceived objects. The authors conducted an experiment with healthy participants. In their study, participants were asked to sort presented objects into two categories: “natural” or “manufactured”. Responses were to be made either by a precision grip or a power grip, dependent on the category. Interestingly, the objects varied in size, although the size dimension was completely task-irrelevant and orthogonal to the categories that the objects were supposed to be sorted into. Results showed that the type of responses biased (implicit) processing of the size dimension: reaction times in the categorization task were faster when the size of presented object was congruent with the response type (smaller object + precision grip, and larger object + power grip). The incongruent pairs (small size + power grip; large size + precision grip) yielded worse performance.

Furthermore, a growing body of evidence provided by neuroimaging studies speaks in favor of action-related way of perceiving the world, in the form of affordances. For example, Grèzes and Decety [30] in a study using PET methodology showed that brain areas related to motor representation were automatically activated when participants only watched objects that have certain action affordances. Similar results have been obtained by Grafton et al. [31] who recorded neural activity in premotor cortex in the area related to hand/arm movements when participants only passively viewed tools of common use (such as scissors, hammer, etc.). Also single-cell studies on monkeys reported that neurons in certain motor-related areas of the brain are selective for object affordances, e.g., [32]. Therefore, all these results speak in favour of the idea that learned associations between certain objects and actions they afford might result in an automatic activation of motor-related areas while such objects are perceived.

6 The Mirror Neuron System, Motor- and Perceptual Resonance

The discovery of mirror neurons (cf. [33]) clearly tagged a common neural mechanism for action and perception domains. Mirror neurons have traditionally been observed in motor areas of monkey brains. Those neurons are related to planning a particular action but they also get activated when an animal only observes others performing that action type. Interestingly, those neurons are active also when a meaning of an action can be inferred from sounds [34] or other hints [35]. Therefore, these results have been brought forward as argument in favour of the idea that the mirror neuron system has a functional role of action understanding [36]. Subsequent to the discovery of the mirror neuron system in monkeys, a large body of evidence has been collected for the analogue of a mirror neuron system in humans. EEG studies have shown that a certain EEG frequency (the mü-rhythm) is related not only to active movements but also to action observation, e.g., [37]. Finally, many brain imaging studies (PET, fMRI) revealed activation of motor areas when participants observed actions performed by others, e.g., [38], see [36] for a review. Certain authors have postulated that the mirror neuron system is responsible not only for action understanding but also for imitative learning, e.g., [39] and maybe even a base for communication and language acquisition, e.g. [40]. Some authors challenge the view that the mirror neuron system is specialized in action understanding and/or action imitation stressing the point that those areas are also involved in predictions of subsequent steps of regular sequences of stimuli, e.g., [15]. As such, action understanding might only fall within the broad category of anticipatory processes involved in known or predictable sequences (typical familiar actions usually consist in predictable sequences of movements). In either case, the mirror neuron system is a prominent example of close action-perception links which only shows how much readiness for action is involved in perception of the external world.

The literature examples related to the mirror neuron system described so far focus on the so-called "motor resonance" (see [36, 41] for reviews), i.e., the idea that observing an action activates similar brain mechanisms and structures as actually performing the action. Some authors, e.g., [41], stress the importance of the link in the other direction as well: the idea of *perceptual* resonance. According to this idea, action observation not only activates the observer's motor programs, but it also influences observer's perception. Evidence for perceptual resonance has been brought forward by, for example Hamilton et al. [42]. In these studies, participants were asked to lift or hold boxes of various weights. Simultaneously, they observed films of other people lifting or carrying similarly-looking boxes. Interestingly, participants under- or over-estimated the weight of the boxes carried by the other agents dependent on the weight of their own boxes. Along similar lines, Repp and Knoblich [43] showed that perception of ambiguous auditory stimuli depend on the produced actions.

The idea of perceptual resonance is in line with the previously described concept of intentional weighting, with the latter being actually a candidate mechanism for the phenomena described as perceptual resonance.

7 Application for Social Robotics

All the above considerations, theories and empirical evidence have been brought forward to argue in favour of a close coupling between action and selection mechanisms in perception. This has been described in order to emphasize that design of artificial systems needs to take into account such neural architecture and functionality of the human brain. This is important for a few reasons: one is that it might be beneficial to draw from evolutionary solutions when designing robots. Current artificial robotic systems have limited computing power and therefore, it should definitely be advantageous to implement certain selection mechanisms especially in the case of fast real-time systems that produce very rich and fast data streams. When deciding what sort of selection mechanism should be implemented, "copying" from evolution seems to be the most efficient way to solve the problem. Such an approach to design of robots that are to assist humans in the kitchen has been described in [44]. Another important argument for designing artificial systems based on similar mechanisms as those that operate in the human brain is the need for behaviour similar to humans for the sake of safe and efficient human-robot interaction. If robots are to accompany humans, they should produce behaviour predictable for humans, and this might be possible only when similar computing architecture is employed. Moreover, also robots need to be equipped with systems that understand the meaning of human actions. Here, the evolutionary solution might prove particularly useful. This direction has been taken in the work of [45] in which the researchers designed a model of functioning of the mirror neuron system that has been implemented in a humanoid robot which acquired knowledge about objects in terms of their affordances, became capable of action recognition, and finally managed to mimic actions performed by a human on various objects.

The influential work of Matarić and colleagues has followed similar direction. For example, Billard and Matarić [46] reported results of a validation study of their biologically-inspired connectionist model that imitated human arm movements. The model consisted of a hierarchy of neural networks that represented the functionality of brain areas involved in motor control. For example, the "lowest level" network represented the functionality of spinal cord, another one represented cerebellum, two networks simulated the role of M1 (primary motor cortex—area involved in control of voluntary movement execution) and premotor (PM) cortex (area involved in selection of movements; also the area where mirror neurons have been found), and finally, there was a network that represented visual perception at the level of temporal cortex. Most importantly, the network representing the workings of the PM cortex translated and linked visual representations of observed actions to motor representations. More specifically, nodes of PM transferred the activity of the nodes in the "visual" network into activity of M1 nodes representing motor commands of the corresponding movements. Therefore, neurons in that network responded to both visual information and the corresponding motor commands. Learning of movements consisted of storing the movement sequences registered by the "visual" network and mapping them onto the motor commands of the M1 network. As such, the PM network simulated the functionality of the mirror-neuron system. This model was implemented on a biomechanical simulation of a human avatar, which managed to reproduce human movements with high level of accuracy. In [47], Matarić reported and summarized work devoted to the design of biologically-inspired imitation systems. The author described the main components of their models, i.e., the "visual and attention" system that detected and selected biological motion from a stream of visual input by focusing on effectors or tools; the system that mapped visual input onto motor programs (mirror-neuron system functionality); the motor control system consisting of motor primitives; and finally, a classification-based learning mechanism that learned "from a match between the observed and executable movements" and continuously expanded movement repertoire. These models have been implemented on artificial humanoid simulation avatars or robotic systems with high degree of success.

Similarly, Breazeal and colleagues [48] provided an example of a computational model inspired by the functionality of mirror neuron system as well as infant development and the idea of the Theory of Mind. Their robotic implementation of the model, Leonardo, is well capable of imitating facial expressions thanks to the architecture consisting of a perception system that extracts relevant features from the sensory data, for example, faces or its components such as eyes or lips, an action system that selects the appropriate behavior, and a motor system based on movement primitives. Importantly, Leonardo learns through imitation and interaction in a two-step manner. In the first phase, the human participant imitates Leonardo's motor babbling—allowing Leonardo to map perceived human expressions onto its own intermodal representational space. The intermodal representations thus consist in learned correspondences between the model's own expressions and those of the human imitator. In the second phase of imitative learning, once the robot has learned the representations of perceived facial expressions in intermodal space, it tries to imitate the human and optimizes imitation through successive approximations. Finally, the authors describe work in progress and future directions that would aim at equipping the robot with emotion and attention systems. These would allow for joint attention and emotional appraisal of novel situations, which in turn, are necessary for social referencing. Such skill is extremely important in social behavior, especially when it comes to understanding affective messages of others and emotional communication in general.

The above-described research provides prominent examples of how designing an artificial system based on results from human psychology and/or human/monkey neurophysiology might allow for establishing action understanding necessary for efficient human-robot interaction. In [49], Matarić describes how and why biologically-inspired architectures implemented in robotic platforms are important not only for theoretical, but also for practical reasons. Matarić argues that humans have a tendency to "socially engage with, and attribute life-like properties to machines, especially embodied ones such as robots, which exhibit sufficient (often quite simple) attributes of biological-like movement or appearance" (p. 82). This, in turn, leads to the possibility of designing artificial systems whose role is to provide care in hospitals, education and rehabilitation centers, programs for children with special needs and elderly people. Apart from the above arguments stressing the importance of implementing similar mechanisms in artificial systems to those that have developed in human or animal brains, drawing inspiration from nature might also be simply useful: for example, if humans have developed a fast pragmatic route of action-related selection [28] and if human perceptual mechanisms are biased towards action planning [5, 25], then designing systems that make use of such action-perception links might allow for solutions that are better tuned to their needs.

The work of Cabibihan and colleagues [50] speaks in favour of this line of argumentation. The authors described a study in which they investigated a computer-based communication with social cues more rich than usual. That is, in contrast to standard internet-based communication such as Skype (Skype Technologies, Luxembourg), Cabibihan and colleagues developed a mobile robot interface capable of pointing gestures. The experimenter could communicate with the participants either using a standard static mode, i.e., through the notebook monitor positioned on top of the robot platform, or applying the dynamic mode, i.e., using the robot's arms for pointing. The experimenter described to the participants a layout of items in commonly familiar places such as kitchen or living room. In the static condition, the experimenter used terms such as "left/right side of" or "behind". In the dynamic condition, they used such words as "there" or "in that corner" and pointed to the respective directions with the robot arms. Subsequent to the communication session, participants were asked to recall positions of the items. Results showed that participants performed significantly better in the dynamic condition as compared to the static condition. The authors conclude that this might be thanks to that the dynamic condition involved a platform that was socially richer. Based on the argumentation put forward in this paper, an alternative interpretation might emerge: in the dynamic condition, due to interaction with the mobile platform and the pointing gestures, participants might have employed not the symbolic representation, but rather the action system in information encoding. Activating the action system, in turn, might have facilitated memory processes due to the action-perception links present in human cognitive architecture. In any case, this study is a prominent example showing that systems that make use of action-perception links might be more intuitive for humans and their usability might be simply better-tuned to the human needs.

One line of research that also has direct consequences for robotics is grounded in the aforementioned idea of perceptual resonance [41]. Several authors have shown that action-related bias of perception (perceptual resonance) might not only be related to the actions concurrently observed but also, and importantly, to motor constraints of the observers. That is, for example it has been shown [51] that participants perceived the velocity of a dot presented as moving along an elliptical path according to the constraints of biological motion. Another study [52] has revealed that participants best estimated the trajectory and success of dart throws when they observed their own throws as compared to other people's throws. Importantly, in this study, participants did not have access to the information of whose throws they were observing and what were the landing points of the darts. Similarly, Calvo-Merino et al. [53] have found that motor expertise affects perception. That is, motor areas of the brain were activated higher when experts observed other agents performing the actions of their expertise area, i.e., for example ballet dancers observing other people dancing ballet as compared to capoeira. Most importantly, for the purposes of human-robot interaction research, such factors like observers' gender has been found to affect brain activity during observation of other agents' actions [54]. In this study, male and female ballet dancers observed other men and women performing the dance. It was assumed that the aspect of familiarity with the movements has been controlled, as both genders practiced together. However, higher activity in motor-related brain areas was found when the observers watched agents of their own gender performing the dance as compared to the other gender members. These results have been put forward to argue that humans activate their own motor repertoire when observing other people act. A striking piece of evidence for such a thesis is the case of two deafferented patients (patients who lost the sense of proprioception and touch, and hence sensory feedback), who were unable to interpret movement kinematics of other people performing a lifting-box action, that normally would be used as cues for estimates of object weight [55]. In the paper reviewing the studies that support the idea of perceptual resonance [41], Schütz-Bosbach and Prinz, based on the evidence described above, speculate that "we can only perceive and understand in others what we can do ourselves" (p. 354).

If that were to be the case, it would have dramatic consequences for the human-robot interaction enterprise, as it would mean that in order to make humans capable of intuitively understanding and predicting robot movement sequences and intentions, robots would have to have almost identical motoric make-up as humans. Otherwise, perception of robot actions might seem imperfect, which might lead to the problem of uncanny valley, the concept introduced by Masahiro Mori [56].

This issue has been addressed in [57]. In this work, the authors tested, in an implicit manner, the way a humanoid robot would be perceived by human observers. They based the design of their paradigm on [58]. That is, human participants were asked to perform certain diagonal arm movements while observing either a robot, or another human performing similar movements—either congruent or incongruent with the actually performed movement. In line with the previously reported results [58], congruency effects were expected: larger variance of executed movements in the incongruent condition, as compared to the congruent case. In [58], the congruency effects were observed only when humans, but not robots, were involved in concurrent action production while in [57], congruency/interference effects were observed for both the robot and the human conditions. The difference between the design in [58] and [57] was that in the latter, a humanoid robot was used while in the former, an industrial robot bearing little similarity to humans was employed. These findings speak in favor of the idea that perceptual (and motor) resonance can be induced only by agents that bear close anatomic similarities to the human observers, which imply similar constraints on the motor system. In short, this piece of evidence is a clear example of how robotics can benefit from the findings of human neuroscience.

8 Concluding Remarks

This paper argues that the close coupling between action and perception in the human cognitive system has implications for designs of artificial robotic systems that are to interact with humans. Humans learn by imitation, simulate actions when perceiving others performing them, and bias perception towards action goals. Equipping artificial robotic systems with similar mechanisms should facilitate human-robot interaction, and will allow humans to easily infer action goals from observed robot behavior. Hence, drawing inspiration from naturally developed mechanisms in humans might prove useful when human and robots should share common social contexts.

Acknowledgements Work on this paper has been supported by the Deutsche Forschungsgemeinschaft DFG (SCHU 1330/5-1) and by the Excellence Cluster "Cognition for Technical Systems" (CoTeSys, Project #433 to AS).

References

1. Clark A (1999) An embodied cognitive science? Trends Cogn Sci 3:345–351
2. Stein L (1994) Imagination and situated cognition. J Exp Theor Artif Intell 6:393–407
3. Matarić M (1992) Integration of representation into goal-driven behavior based robots. IEEE J Robot Autom 8:304–312
4. Greenwald A (1970) Sensory feedback mechanisms in performance control: with special reference to the ideomotor mechanism. Psychol Rev 77:73–99
5. Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The Theory of Event Coding (TEC): a framework for perception and action planning. Behav Brain Sci 24:849–937
6. Prinz W (1997) Perception and action planning. Eur J Cogn Psychol 9:129–154
7. Harless E (1861) Der Apparat des Willens [The apparatus of will]. Z Philos Philos Krit 38:50–73
8. Wolpert DM, Ghahramani Z (2000) Computational principles of movement neuroscience. Nat Neurosci 3:1212–1217
9. Greenwald AG (1970) A choice reaction time test of ideomotor theory. J Exp Psychol 86:20–26
10. Elsner B, Hommel B (2001) Effect anticipation and action control. J Exp Psychol Hum Percept Perform 27:229–240
11. Hommel B (2010) Grounding attention in action control: the intentional control of selection. In: Bruya BJ (ed) Effortless attention: a new perspective in the cognitive science of attention and action. MIT Press, Cambridge, pp 121–140
12. Hommel B (1998) Event files: evidence for automatic integration of stimulus-response episodes. Vis Cogn 5:183–216
13. Müsseler J, Hommel B (1997) Blindness to response-compatible stimuli. J Exp Psychol Hum Percept Perform 23:861–872
14. Zwickel J, Grosjean J, Prinz W (2008) A contrast effect between the concurrent production and perception of movement directions. Vis Cogn 26:953–978
15. Schubotz RI, von Cramon DY (2002) Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. NeuroImage 15:787–796
16. Allport A (1987) Selection for action: some behavioral and neurophysiological considerations of attention and action. In: Heuer H, Sanders AF (eds) Perspectives on perception and action. Erlbaum, Hillsdale, pp 395–419
17. Rizzolatti G, Riggio L, Sheliga BM (1994) Space and selective attention. In: Umiltà C, Moscovitch M (eds) Attention and performance, XV. Conscious and nonconscious information processing. MIT Press, Cambridge, pp 231–265
18. Deubel H, Schneider WX (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. Vis Res 36:1827–1837
19. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222
20. Wolfe JM (1994) Guided Search 2.0: a revised model of visual search. Psychon Bull Rev 1:202–238
21. Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782–784
22. Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanism subserve attention in macaque areas V2 and V4. J Neurosci 19:1736–1753
23. Craighero L, Fadiga L, Rizzolatti G, Umiltà CA (1999) Action for perception: a motor-visual attentional effect. J Exp Psychol Hum Percept Perform 25:1673–1692
24. Fagioli S, Hommel B, Schubotz RI (2007) Intentional control of attention: action planning primes action related stimulus dimensions. Psychol Res 71:22–29
25. Wykowska A, Schubö A, Hommel B (2009) How you move is what you see: action planning biases selection in visual search. J Exp Psychol Hum Percept Perform 3:1755–1769
26. Wykowska A, Hommel B, Schubö A (2011) Action-induced effects on perception depend neither on element-level nor on set-level similarity between stimulus and response sets. Atten Percept Psychophys 73:1034–1041
27. Gibson EJ (1977) The theory of affordances. In: Shaw RE, Bransford J (eds) Perceiving, acting and knowing. Erlbaum, Hillsdale, pp 127–143
28. Humphreys GW, Riddoch MJ (2001) Detection by action: neuropsychological evidence for action-defined templates in search. Nat Neurosci 4:84–89
29. Tucker R, Ellis M (2001) The potentiation of grasp types during visual object categorization. Vis Cogn 8:769–800
30. Grèzes J, Decety J (2002) Does visual perception of object afford action? Evidence from a neuroimaging study. Neuropsychologia 40:212–222
31. Grafton ST, Fadiga L, Arbib MA, Rizzolatti G (1997) Premotor cortex activation during observation and naming of familiar tools. NeuroImage 6:231–236
32. Murata A, Fadiga L, Fogassi L, Gallese V, Raos V, Rizzolatti G (1997) Object representation in the ventral premotor cortex (area F5) of the monkey. J Neurophysiol 78:2226–2230
33. Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609
34. Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846–848
35. Umiltà MA, Kohler E, Gallese V, Fogassi L, Fadiga L et al (2001) "I know what you are doing": a neurophysiological study. Neuron 32:91–101

36. Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192
37. Cochin S, Barthelemy C, Lejeune B, Roux S, Martineau J (1998) Perception of motion and EEG activity in human adults. Electroencephalogr Clin Neurophysiol 107:287–295
38. Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L et al (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur J Neurosci 13:400–404
39. Arbib MA (2002) Beyond the mirror system: imitation and evolution of language. In: Dautenhan K, Nehaniv C (eds) Imitation in animals and artifacts. MIT Press, Cambridge, pp 229–280
40. Meister IG, Boroojerdi B, Foltys H, Sparing R, Huber W, Topper R (2003) Motor cortex hand area and speech: implications for the development of language. Neuropsychologia 41:401–406
41. Schütz-Bosbach S, Prinz W (2007) Perceptual resonance: action-induced modulation of perception. Trends Cogn Sci 11:349–355
42. Hamilton A, Wolpert D, Frith U (2004) Your own action influences how you perceive another person's action. Curr Biol 14:493–498
43. Repp BH, Knoblich G (2007) Action can affect auditory perception. Psychol Sci 18:6–7
44. Wykowska A, Maldonado A, Beetz M, Schubö A (2011) How humans optimize their interaction with the environment: the impact of action context on human perception. Int J Soc Robot 3:223–231
45. Metta G, Sandini G, Natale L, Craighero L, Fadiga L (2006) Understanding mirror-neurons. A bio-robotic approach. Interact Stud 7:197–231
46. Billard A, Matarić MJ (2001) Learning human arm movements by imitation: evaluation of a biologically inspired connectionist architecture. Robot Auton Syst 37:145–160
47. Matarić MJ (2002) Sensory-motor primitives as a basis for imitation: linking perception to action and biology to robotics. In: Dautenhan K, Nehaniv C (eds) Imitation in animals and artifacts. MIT Press, Cambridge, pp 392–422
48. Breazeal C, Buchsbaum D, Gray J, Gatenby D, Blumberg B (2005) Learning from and about others: towards using imitation to bootstrap the social understanding of others by robots. Artif Life 11:1–32
49. Coradeschi S, Ishiguro H, Asada M, Shapiro SC, Thielscher M, Breazeal C, Matarić MJ, Ishida H (2006) Human-inspired robots. IEEE Intell Syst 21(4):74–85
50. Cabibihan JJ, So WC, Nazar M, Ge SS (2009) Pointing gestures for a robot mediated communication interface. In: Xie M et al (eds) ICIRA 2009. LNAI, vol 5928, pp 67–77
51. Viviani P, Stucchi N (1992) Biological movements look uniform: evidence of motor-perceptual interactions. J Exp Psychol Hum Percept Perform 18:603–623
52. Knoblich G, Flach R (2001) Predicting the effects of actions: interactions of perception and action. Psychol Sci 12:467–472
53. Calvo-Merino B, Glaser DE, Grèzes J, Passingham RE, Haggard P (2005) Action observation and acquired motor skills: an fMRI study with expert dancers. Cereb Cortex 15:1243–1249
54. Calvo-Merino B, Glaser DE, Grèzes J, Passingham RE, Haggard P (2006) Seeing or doing? Influence of visual and motor familiarity in action observation. Curr Biol 16:1–6
55. Bosbach S, Cole J, Prinz W, Knoblich G (2005) Inferring another's expectation from action: the role of peripheral sensation. Nat Neurosci 8:1295–1297
56. Mori M (1970) Bukimi no tani. The uncanny valley. Energy 7:33–35. (Originally in Japanese, trans. MacDorman KF, Minato T)
57. Otzop E, Franklin DW, Chaminade T, Cheng G (2005) Human-humanoid interactions: is humanoid robot perceived as human? Int J Humanoid Robot 2:537–559
58. Kilner JM, Paulignan Y, Blakemore SJ (2003) An interference effect of observed biological movement on action. Curr Biol 13:522–525

Agnieszka Wykowska works as a postdoc researcher at the Department of Psychology, Ludwig-Maximilian University (LMU), Munich, Germany. She received her M.Sc. in Neuro-Cognitive Psychology from the LMU in 2006 and a M.A. in Philosophy from the Jagiellonian University, Krakow, Poland in 2001. She completed her PhD in 2008 with honors in the Department of Psychology, LMU. Since then, she has been involved in the German Cluster of Excellence CoTeSys (Cognition for Technical Systems). Her research interests are human visual attention, interactions between action and perception, as well as human-robot interaction. In her research, she uses EEG/ERP methodology combined with behavioural measures.

Anna Schubö is a Professor of Experimental Psychology at the Department of Psychology at Philipps-University in Marburg, Germany. She received her Diploma in Psychology from the Department of Psychology, Ludwig-Maximilian University (LMU) and her PhD from the Max-Planck Institute for Psychological Research in Munich. After a Postdoc in Erlangen, she worked as senior lecturer at LMU and as principal investigator in the Cluster of Excellence "Cognition for Technical Systems" (CoTeSys) on cognitive capabilities in human and robot action. Her research focuses on cognitive processes such as attention and action planning in humans.