

Robotic Partners' Bodies and Minds: An Embodied Approach to Fluid Human-Robot Collaboration

Guy Hoffman and Cynthia Breazeal

MIT Media Laboratory
20 Ames Street E15-468
Cambridge, MA 02139
{guy,cynthiab}@media.mit.edu

Abstract

A mounting body of evidence in psychology and neuroscience points towards an *embodied* model of cognition, in which the mechanisms governing perception and action are strongly interconnected, and also play a central role in higher cognitive functions, traditionally modeled as amodal symbol systems.

We argue that robots designed to interact fluidly with humans must adopt a similar approach, and shed traditional distinctions between cognition, perception, and action. In particular, embodiment is crucial to *fluid joint action*, in which the robot's performance must tightly integrate with that of a human counterpart, taking advantage of rapid sub-cognitive processes.

We thus propose a model for embodied robotic cognition that is built upon three propositions: (a) modal, perceptual models of knowledge; (b) integration of perception and action; (c) top-down bias in perceptual processing. We then discuss implications and derivatives of our approach.

"[T]he human being is a unity, an indivisible whole. [...] ideas, emotions and sensations are all indissolubly interwoven. A bodily movement 'is' a thought and a thought expresses itself in corporeal form."

— Augusto Boal (2002)

Introduction

We aim to build robots that work fluidly with humans in a shared location. These robots can be teammates in a joint human-robot team; household, office, or nursing robots assisting everyday people in their tasks; or robotic entertainers performing in conjunction with a human actor. In each of these cases, the robot is expected to maintain continuous interaction with a person, tightly meshing its actions with that of the human.

Drawing from a growing body of knowledge in psychology and neuroscience, we propose an architecture that diverges from much of the traditional work in artificial intelligence in that it emphasizes the *embodied* aspects of cognition (Wilson 2002). It is becoming increasingly clear

that human (as well as animal) perception and action are not mere input and output channels to an abstract symbol processor or rule-generating engine, but that instead thought, memory, concepts, and language are inherently grounded in our physical presence (Pecher & Zwaan 2005; Barsalou 1999; Wilson 2001). The same principles also play a substantial role in social cognition and joint action, when it becomes crucial to coordinate one agent's activity with another's (Sebanz, Bekkering, & Knoblich 2006; Barsalou *et al.* 2003). It is our belief that these lessons can and should be transferred to robotic cognition as well, and in particular if we are to design robots that act in natural and fluid dialog with a human partner.

A central goal of this work is to steer away from the stop-and-go turntaking rigidity present in virtually all human-robot interaction to date. Based on many of the same underlying assumptions of symbolic AI, robotic dialog systems, robotic teammates, and robotic stage actors take little or no advantage of the social and non-verbal behaviors that enable humans to easily display an impressive level of coordination and flexibility. The premise of this work is that a physically grounded cognitive architecture can give rise to appropriate behavior enabling a much more fluid meshing of a robot's actions with those of its human counterpart.

In this paper we propose an architecture built on three interrelated principles backed by psychological and neurological data:

Modal, perceptual models of knowledge Much of AI is focused on amodal theories of knowledge, which assert that information is translated from perceptual stimuli into nonperceptual symbols, later used for information retrieval, decision making, and action production. This view also corresponds to much of the work currently done in robotics, where sensory input is translated into semantic symbols, which are then operated upon for the production of motor control. In contrast, a growing number of psychologists and neuroscientists support a *perceptual* model of cognition, in which perceptual symbols are stored through modus-specific mechanisms, subsequently used by ways of "simulation" or "imagery" (Barsalou 1999; Kosslyn 1995). Perceptual symbols are organized in cross-modal networks of activation which are used to reconstruct and produce knowledge.

Integration of perception and action Evidence also points to a close integration between perceptual mechanisms and action production. A large body of work points to an isomorphic representation between perception and action, leading to mutual and often involuntary influence between the two (Wilson 2001). These mechanisms could also underlie the rapid and effortless adaptation to a partner that is needed to perform a joint task (Sebanz, Bekkering, & Knoblich 2006). For robots that are to work in a similarly fluid manner, it makes sense to break from the traditional modular approach separating perception, action, and higher-level cognition. Instead, intentions, goals, actions, perceptions, concepts, and object representations should interact in a complex network of activation, giving rise to behaviors that are appropriate in timing and context.

Top-down bias in perceptual processing Finally, the two principles outlined above, as well as a large body of related experimental data, give rise to the following insight: perceptual processing is not a strictly bottom-up analysis of raw available data, as it is often modeled in robotic systems. Instead, simulations of perceptual processes prime the acquisition of new perceptual data, motor knowledge is used in sensory parsing, and intentions, goals, and expectations all play a role in the ability to parse the world into meaningful objects. This seems to be particularly true for the parsing of human behavior in a goal-oriented manner, a vital component of joint action.

In this paper, we will briefly present evidence for the principles outlined above and discuss their pertinence to fluid joint action. We will describe how we use these principles in a cognitive framework for a robot designed to act jointly with a human counterpart. We then discuss a number of interrelated concepts as they stem from this approach, notably: (a) *attention*, (b) *routine action*, (c) *practice*, (d) *flexible resolution*, (e) *anticipatory action*, and (f) *mutual responsiveness*.

Embodied Cognition: Evidence from Neuropsychology

During the second half of the twentieth century artificial intelligence has not only *drawn* from theories of cognitive psychology, but also shaped notions of amodal, symbolic information processing. According to this view, perceptual input is hierarchically processed in isolated modules, and eventually gives rise to a non-perceptual representation of concepts, which are processed symbolically and may subsequently give rise to physical behaviors.

An increasing body of recent findings challenges this view and suggests instead that concepts and memory are phenomena which are grounded in modal representations utilizing many of the same mechanisms used during the perceptual process. Moreover, action production and perception are not separated by supervisory, symbolic rule operators, but instead intimately linked. Models of action production affect perceptual acquisition and retention, and give rise to a number of intermodal effects.

Perceptual Models of Knowledge

In recent years, many authors in language, psychology, and neuroscience have supported perceptual theories of cognition (Spivey, Richardson, & Gonzalez-Marquez 2005; Lakoff & Johnson 1999; Stanfield & Zwaan 2001). A prominent theory explaining these findings is one of “simulators”, siting long-term memory and recall in the very neural modules that govern perception itself (Barsalou 1999). By this theory, concepts are formed as activation patterns of experiences related to the pattern, stored in associative networks relating them to other modalities, and re-invoked using perceptual simulators of such experiences. This view is supported by a large number inter-modal behavioral influences, as well as by the detection of perceptual neural activation when a subject is engaging in cognitive tasks. Thus, when memory—and even language—is invoked to produce behavior, the underlying perceptual processes elicit many of the same behaviors normally used to regulate perception.

A similar case has been made with regard to hand signals, which are viewed as instrumental to lexical lookup during language generation (Krauss, Chen, & Chawla 1996), and is supported by findings of redundancy in head-movements (McClave 2000) and facial expression (Chovil 1992) during speech generation.

A number of experiments evaluating perceptual tasks have shown that task time is related to the mental simulation of kinematic configuration (Parsons 1994). Similarly, a study shows that children who mispronounce certain letters confuse the same letters in word recall, even when the memory cue is purely visual (Locke & Kutz 1975). In visual processing, imagery is a conscious simulation mechanism (Kosslyn 1995), and is believed by some to use the same neural mechanisms as perception (Kreiman, Koch, & Fried 2000; Kosslyn 1995).

Perception-Action Integration

In parallel to a perception-based theory of cognition lies an understanding that cognitive processes are equally interwoven with motor activity. Evidence in human developmental psychology shows that motor and cognitive development are not parallel but highly interdependent. For example, artificially enhancing 3-month old infant’s grasping abilities (through the wearing of a ‘sticky’ mitten), equated some of their cognitive capabilities to the level of older, already grasping¹, infants (Somerville, Woodward, & Needham 2004).

Additionally, neurological findings indicate that—in some primates—observing an action and performing it causes activation in the same cerebral areas (Gallese *et al.* 1996; Gallese & Goldman 1996). This common coding is thought to play a role in imitation, and the relation of the behavior of others to our own, which is considered a central process in the development of a Theory of Mind (Meltzoff & Moore 1997). For a review of these so-called mirror neurons and the connection between perception and action, as it relates to imitation in robots, see (Matarić 2002).

¹In the physical sense.

The capability to parse the actions and intentions of others is a cornerstone of joint action. Mirror mechanisms have been found to play a role in experiments of joint activity, where it has been suggested that the goals and tasks of a team member are not only represented, but coded in functionally similar ways to one's own (Sebanz, Bekkering, & Knoblich 2006).

Another important outcome of the common mechanism of motor simulators and perception is that there seem to be “privileged loops” between perception and action, which are faster and more robust than supervised activity. Humans can “shadow” speech with little effort or interference from other modalities, and this effect is dampened when even the simplest translation is required (McLeod & Posner 1984).

Much of our critical cognitive capabilities occur in what is sometimes called a “cognitive unconscious” (Lakoff & Johnson 1999), and it seems that these are more prominent in routine activity than those governed by executive control. Routine actions are faster and seem to operate in an unsupervised fashion, leading to a number of action lapses (Cooper & Shallice 2000). From introspection we know that supervisory control is often utilized when the direct pathways fail to achieve the expected results. These unsupervised capabilities seem to be highly subject to practice by repetition, and are central to the coordination of actions when working on a joint task, a fact that can be witnessed whenever we enjoy the performance of a highly trained sports team or a well-rehearsed performance ensemble.

Top-down Perceptual Processing

An important conclusion of the approach laid out herein is that intelligence is neither purely bottom-up nor strictly top-down. Instead, higher-level cognition plays an important role in the mechanisms of lower-level processes such as perception and action, and vice versa. This view is supported by neurological findings, as stated in (Spivey, Richardson, & Gonzalez-Marquez 2005): “the vast and recurrent interconnectedness between anatomically and functionally segregated cortical areas unavoidably compromises any assumptions of information encapsulation, and can even wind up blurring the distinction between feedback and feedforward signals.”

Rather than intelligence and behavior emerging from a hierarchical analysis of perception into higher-level concepts, a continuous stream of information must flow in both directions—from the perception and motor systems to the higher-level concepts, intentions and goals, and back from concepts, intentions, goals, and object features towards the physically grounded aspects of cognition, shaping and biasing those for successful operation.

Experimental data supports this hypothesis, finding perception to be predictive (for a review, see (Wilson & Knoblich 2005)). In vision, information is sent both upstream and downstream, and object priming triggers top-down processing, biasing lower-level mechanisms in sensitivity and criterion. Similarly, visual lip-reading affects the perception of auditory syllables indicating that the sound signal is not processed as a raw unknown piece of data (Mas-

saro & Cohen 1983)². High-level visual processing is also involved in the perception of human figures from point light displays, enabling subjects to identify gender and identity from very sparse visual information (Wilson 2001).

This evidence leads to an integrated view of human cognition, which can collectively be called “embodied”, viewing mental processes not as amodal semantic symbol processors with perceptual inputs and motor outputs, but as integrated psycho-physical systems acting as indivisible wholes.

Our work proposes to take a similar approach to designing a cognitive architecture for robots acting with human counterparts, be it teammates, helpers, or scene partners. It is founded on the hope that grounding cognition in perception and action can hold a key to socially appropriate behavior in a robotic agent, as well as to context and temporally precise human-robot collaboration, enabling hitherto unattained fluidity in this setting.

Architecture

Inspired by the principles laid out above, we are developing a cognitive architecture for robots that follows an embodied view. The following sections describe some of the main components of the proposed architecture.

Perception-Based Memory

Grounded in perceptual symbol based theories of cognition, and in particular that of simulators (Barsalou 1999), we propose that long-term memory and concepts of objects and actions are based on perceptual snapshots and reside in the various perceptual systems rather than in an amodal semantic network (Figure 1).

These percept activation are heavily interconnected both directly and through an associative memory mechanism. Perceptions of different modalities are connected based on concurrency with other perceptual snapshots, as well as concurrency with perceptions generated by simulators during activation.

Bidirectional Percept Trees Incoming perceptions are filtered through percept trees for each modality, organizing them on a gradient of specificity (Blumberg *et al.* 2002). At the same time, predictions stemming from more schematic perceptions bias perceptual processes closer to the sensory level. An example of how higher level parsing is both constructed from and influencing lower level perception is proposed in a separate technical report (Hoffman 2005). In that work, intentions are build up gradually from motor trajectories, actions, and goals. Conversely, probability distributions on intentions bias goal identification, which in turn set the priors for action detection, and subsequently motion estimation.

Retention Perceptions are retained in memory, but decay over time. Instead of subscribing to a firm division of short- and long-term memory, the proposed architecture advocates

²As reported in (Wilson 2001).

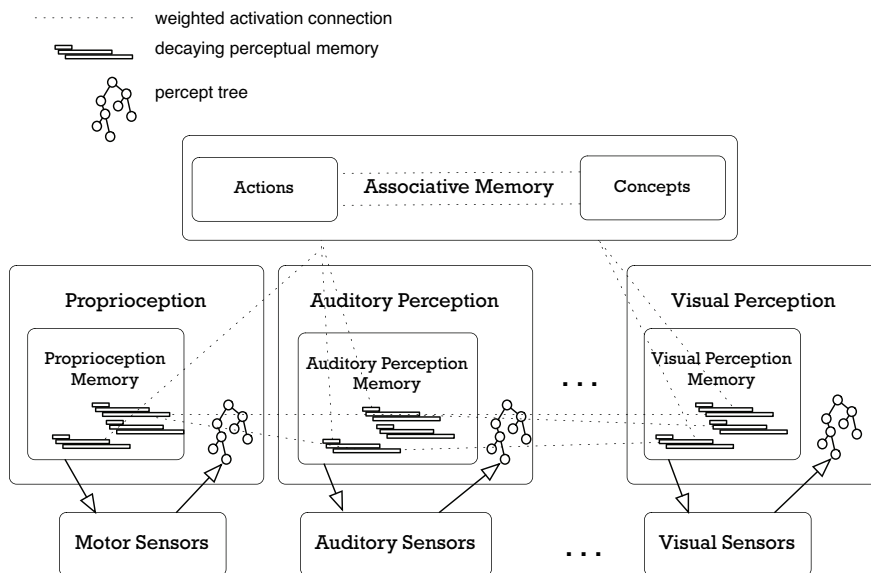


Figure 1: Perception-Based Memory

a gradient information-based decay in which more specific perceptual memory decays faster, and more general perceptual information is retained longer. Decay is governed by the amount of storage needed to keep perceptual memory at various levels of specificity, retaining more specific memories for a shorter period of time, and more schematic perceptions for a longer time span.

Retention is also influenced by the relevance of the perceptual information to the current task and their attentive saliency: perceptual elements that are specifically attended to at time of acquisition will be retained longer than ones that were peripheral at that time. In a similar vein, affective and motivational states of the system can also influence memory retention.³

For example in the visual processing pathway, raw images are retained for a short period of time, while line orientations, colors, blob location, overall rotation, and fully recognized objects remain in memory for a longer time span (Figure 2).

Weighted Connections Memories are organized in activation networks, connecting them to each other, as well as to concepts and actions in associative memory (the dotted lines in Figure 1). However, it is important to note that connections between perceptual, conceptual, and action memory are not binary. Instead, a single perceptual memory—for example the sound of a cat meowing—can be more strongly connected to one memory—like the proprioceptive memory of tugging at a cat’s tail—and more weakly connected to an-

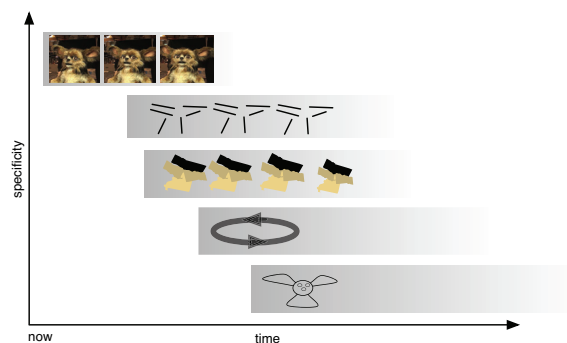


Figure 2: Perceptual memory is filtered and remains in memory for a variable amount of time, based—among others—on the information at each specificity level.

other memory, such as that of one’s childhood living room scent.

Weights of connectivity are learned and refined over time. Their value is influenced by a number of factors, such as frequency of co-occurrence, attention while creating the perceptual memory, and affective states during perception.

Simulators and Production A key capability of perceptual memory is the production of new perceptual patterns and concepts. Using simulation, the activation of a perceptual symbol can evoke the construction of both an experienced and a fictional situation. Citing (Barsalou 1999): “Productivity in perceptual symbol systems is approximately the symbol formation process run in reverse. During

³Such an approach could also explain why certain marginal perceptions are retained for a long time if they occurred in a traumatic setting.

symbol formation, large amounts of information are filtered out of perceptual representations to form a schematic representation of a selected aspect. During productivity, small amounts of the information filtered out are added back.” In the proposed system, the production of such memories is made possible by running a constructive process on the temporally information-based filtered perceptive memories described above.

The produced perceptual activations can be used in reasoning as well as guiding motor activity. In addition, according to the principle of top-down processing, they also bias current perception by both affecting the filtering in the percept tree and the parameters of the sensory modules.

Action-Perception Activation Networks

Associative memory not only governs and interconnects perceptual memory, but also establishes weighted connections between concepts and actions (see: Figure 1). Inspired by architectures such as Contention Scheduling (Cooper & Shallice 2000), activities, concepts, and perceptions occur in a perpetually updating activation relationship.

Action-Perception Activation Networks operate as follows: currently acquired perceptions exert a weighted influence on concepts and activities, leading to (a) potential action selection; (b) the simulation of other, related perceptions; and (c) the activation of an attention mechanism which in turn impacts the sensory layer and the perceptual task, aiding in its pertinent operation.

Thus, for example, the presentation of a screwdriver may activate a grasping action, as well as the activity of applying a screwdriver to a screw. This could in turn activate a perceptual simulator guiding the visual search towards a screw-like object, and the motor memory related to a clockwise rotation of the wrist. A key point to notice regarding these networks is that they don’t follow a single line of action production, but instead activate a number of interrelated mechanisms which can — if exceeding a certain threshold — activate a particular motor pattern.

Bregler demonstrated a convincing application of this kind of top-bottom feedback loop in the realm of Machine Vision action recognition (Bregler 1997), and others argued for a symbolic task-level integration of schemas, objects, and resources (Cooper & Shallice 2000). In this work, we aim to span this approach across both action selection and perception.

Intentions, Motivations, and Supervision

The above architecture is predominantly suited for governing automatic or routine activity. While this model can be adequate for well-established behavior, a robot acting jointly with a human must also behave in concordance with internal motivations and intentions, as well as higher-level supervision. This is particularly important since humans naturally assign internal states and intentions to animate and even inanimate objects (Baldwin & Baird 2001; Dennett 1987; Malle, Moses, & Baldwin 2001). Robots acting with people must therefore behave according to internal drives as well as clearly communicate these drives to their human counterpart.

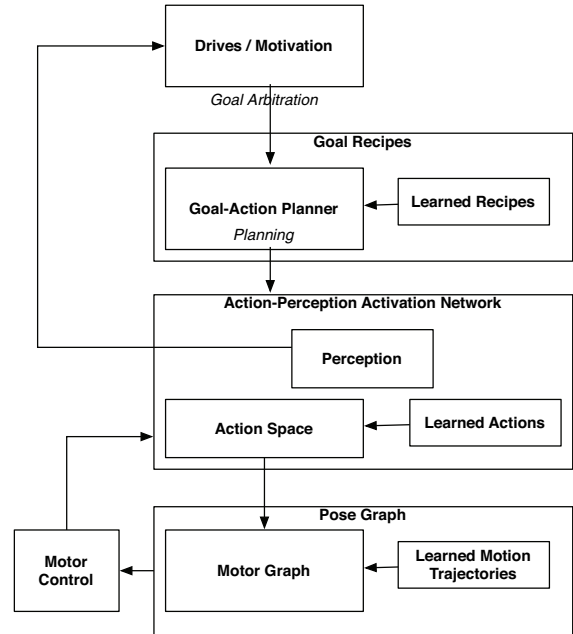


Figure 3: Motivational drives are used to form plans overriding or swaying action selection in the Action-Perception Activation Network.

It thus makes sense to embellish the perception-action subsystem with supervisory intention- and motivation-based control that affects the autonomic processing scheme outlined above (Figure 3). At the base of this supervisory system lie core drives, such as hunger, boredom, attention-seeking, and domain-specific drives, modeled as scalar fluents, which the agent seeks to maintain at an optimal level (similar to (Breazeal 2002)). If any of those fluents falls above or below the defined range, they trigger a goal request. The number and identity of the agent’s motivational drives are fixed. Goals are represented as a vector indexed to the various motivational drives, and encoding the effect of achieving a goal on each of the drives. This representation can be used by the Goal Arbitration Module to decide what goal to pursue. The relative deviance of the motivational drive from its optimal value in combination with the expected effect can be used to compute a utility function for selecting each goal. The most useful goal according to this function is thus selected.

Once a goal is selected, a planner is used to construct a plan that satisfies the goal. The planner uses recipes for achieving a goal in a STRIPS-like precondition/action/effect representation. If the planner has a recipe that achieves a certain goal (i.e. a hierarchical set of actions from its action space with the overall effect of satisfying the goal), it will attempt to use this recipe, while monitoring the perceptual input to evaluate goal satisfaction (in a similar manner as described in (Hoffman & Breazeal 2004)). Some recipes may be innate, although none need to

be. Given a set of actions and a set of goals, the planner can build up new recipes by searching the action space using, for example, regression planning (Russell & Norvig 2002).

The action space is modeled as a set of STRIPS-like precondition/action/effect actions, and their translation into paths into motor graph(s). The action space-motor graph relationship is modeled as described in (Blumberg *et al.* 2002).

Actions that are selected based on the above approach influence, and in some cases override, the automatic action selection mechanisms in the action-perception activation network. Once an action is activated in such a way, it also triggers simulators and guides perception as described above.

This motivation-goal-action model holds an additional benefit for joint action, as it can be used as a basis for an experienced-based intention reading framework, as we described in a separate technical report (Hoffman 2005).

Joint Action Derivatives

The approach presented in this paper aims to allow for a number of important cognitive mechanisms that underly fluid human-robot co-acting. We plan to investigate these derivatives under the proposed approach:

Attention The generic “perception problem” is quite intractable. Humans (and probably other animals) use attention systems to filter incoming data, and attention has also proved useful for robotic agents (Breazeal & Scassellati 1999). In addition, mutual attention is a key factor in the successful behavior of a human-robot team. In the framework advocated herein, attention emerges from the interaction between actions, concepts, and perceptual processes, and in a top-down fashion biases sensation and perception. Attention also determines learning in the system, by affecting the weighted ties between elements in perceptual memory as well as between perceptual memory and concepts.

Routine Action Performing a well-known task is likely to involve non-conscious and sub-planning action selection mechanisms. In humans, routine action performance is often found to be faster and more robust than conscious or planned activity. Teams and ensembles working together increasingly rely on routine action for action coordination, and these become more emphasized when tight meshing of physical activity is required. We believe that true fluidity in human-robot collaboration can only be achieved through reliance on these “privileged loops” between perception and action.

Practice In direct relation to routine activity, *practice* is an important facet of joint action, and of any kind of performance art. The mechanisms underlying repeated practice of a task, and in particular joint practice to achieve fluid collaboration are not well understood. We believe that the above architecture allows for a novel investigation into concepts of practice and refinement of routine joint activities.

Flexible Resolution Cognition often appears to conduct itself on a ‘need-to-know’ or ‘need-to-do’ basis. It can be

argued that planning and adjustment of trained concepts and action schemas only occurs when well-practiced routine activation models fail. Similarly, planning, while informed by lower-level perceptual phenomena, does not need to delve into percept-level resolution unless more information than available is needed. Perception and action can rely on existing patterns and routine simulations unless there is a significant mismatch between the expected and the observed. An embodied approach spanning both low-level processing and high-level goal-directed action could allow for the flexibility of cognition needed for human-robot joint action.

Anticipatory Action Among other factors, successful coordinated action has been linked to the formation of expectations of each partner’s actions by the other (Flanagan & Johansson 2003), and the subsequent acting on these expectations (Knoblich & Jordan 2003). We have presented initial work aimed at understanding possible mechanisms to achieve this behavior in a collaborative robot (Hoffman & Breazeal 2006). A further research goal of this work to evaluate whether more fluid anticipatory action can emerge from an approach that uses perception-action activation as its underlying principle.

Mutual Responsiveness Finally, a central goal of this work is to achieve appropriate mutual responsiveness between a robot and a human. Mutual responsiveness has been strongly tied to joint action (Bratman 1992). By reacting to external stimuli in a combined autonomic and supervised fashion, we hope to be able to demonstrate truly fluid mutual responsiveness as a result of the work proposed herein.

Discussion

Surprisingly, while the existence and complexity of joint action has been acknowledged for decades, and its behavioral operation has been investigated extensively, the neuro-cognitive mechanisms underlying shared activity have only received sparse attention, predominantly over the last few years (for a review, see (Sebanz, Bekkering, & Knoblich 2006)). Much of the work in this field exploring the cognitive activity of two humans sharing a task builds on *embodied* structures of cognition, and emphasizes the non-verbal aspects of collaboration. It is our belief that this approach is the most promising for robots working together with people, as well.

Notions of embodiment are not new in robotics research. It has been fifteen years since Brooks postulated that the “fundamental decomposition of the intelligent system is not into independent information processing units which must interface with each other via representations. Instead, the intelligent system is decomposed into independent and parallel activity producers which all interface directly to the world through perception and action, rather than interface to each other particularly much.” (Brooks 1991) In this context, Brooks advocated incremental design of simple reactive robots. Today, it seems that embodied approaches are still mostly confined to simple robots interacting with an inanimate environment.

Designers of robots interacting with humans, however, have taken little note of theories of embodied cognition, with most systems translating perception into amodal symbols for further processing. While active perception has been investigated (Aloimonos 1993; Breazeal & Scassellati 1999), and top-down influences in object recognition have been explored on low-level vision tasks (Bregler 1997; Hamdan, Heitz, & Thoraval 1999), a full human-robot interaction system taking advantage of the ideas brought forth in this paper has yet to be implemented.

Conclusion

To date, most robots interacting with humans still operate in a non-fluid command-and-response manner. In addition, most are based on an amodal model of cognition, employing relatively little embodied intelligence. In the meantime, neuroscience and psychology have gathered impressive evidence for a view of cognition that is based on our physical experience, showing that concept representation is perception- and action-based; that perception and action are coupled not only through supervisory control; and that intelligence is top-down as much as it is bottom-up.

We believe that these mechanisms are at the core of joint action. To build robots that display a continuous fluid meshing of their actions with those of a human, designers of robots for human interaction need therefore take these insights to heart. We have presented a framework that implements the above ideas and discussed possible behavioral derivatives crucial to fluid joint activity emerging from the proposed architecture.

References

- Aloimonos, Y., ed. 1993. *Active Perception*. Lawrence Erlbaum Associates, Inc.
- Baldwin, D., and Baird, J. 2001. Discerning intentions in dynamic human action. *Trends in Cognitive Sciences* 5(4):171–178.
- Barsalou, L. W.; Niedenthal, P. M.; Barbey, A.; and Ruppert, J. 2003. Social embodiment. *The Psychology of Learning and Motivation* 43:43–92.
- Barsalou, L. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–660.
- Blumberg, B.; Downie, M.; Ivanov, Y.; Berlin, M.; Johnson, M. P.; and Tomlinson, B. 2002. Integrated learning for interactive synthetic characters. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 417–426. ACM Press.
- Boal, A. 2002. *Games for Actors and Non-Actors*. Routledge, 2nd edition.
- Bratman, M. 1992. Shared cooperative activity. *The Philosophical Review* 101(2):327–341.
- Breazeal, C., and Scassellati, B. 1999. A context-dependent attention system for a social robot. In *IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1146–1153. Morgan Kaufmann Publishers Inc.
- Breazeal, C. 2002. *Designing Sociable Robots*. MIT Press.
- Bregler, C. 1997. Learning and recognizing human dynamics in video sequences. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, 568. IEEE Computer Society.
- Brooks, R. A. 1991. Intelligence without representation. *Artificial Intelligence* 47:139–159.
- Chovil, N. 1992. Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction* 25:163–194.
- Cooper, R., and Shallice, T. 2000. Contention scheduling and the control of routing activities. *Cognitive Neuropsychology* 17:297–338.
- Dennett, D. C. 1987. Three kinds of intentional psychology. In *The Intentional Stance*. Cambridge, MA: MIT Press. chapter 3.
- Flanagan, J. R., and Johansson, R. S. 2003. Action plans used in action observation. *Nature* 424(6950):769–771.
- Gallese, V., and Goldman, A. 1996. Mirror neurons and the simulation theory of mind-reading. *Brain* 2(12):493–501.
- Gallese, V.; Fadiga, L.; Fogassi, L.; and Rizzolatti, G. 1996. Action recognition in the premotor cortex. *Brain* 119:593–609.
- Hamdan, R.; Heitz, F.; and Thoraval, L. 1999. Gesture localization and recognition using probabilistic visual learning. In *Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition (CVPR '99)*, 2098–2103.
- Hoffman, G., and Breazeal, C. 2004. Collaboration in human-robot teams. In *Proc. of the AIAA 1st Intelligent Systems Technical Conference*. Chicago, IL, USA: AIAA.
- Hoffman, G., and Breazeal, C. 2006. What lies ahead? expectation management in human-robot collaboration. In *Working notes of the AAAI Spring Symposium*.
- Hoffman, G. 2005. An experience-based framework for intention-reading. Technical report, MIT Media Laboratory, Cambridge, MA, USA.
- Knoblich, G., and Jordan, J. S. 2003. Action coordination in groups and individuals: learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(5):1006–1016.
- Kosslyn, S. 1995. Mental imagery. In Kosslyn, S., and D.N.Osherson., eds., *Invitation to Cognitive Science: Visual Cognition*, volume 2. Cambridge, MA: MIT Press, 2nd edition. chapter 7, 276–296.
- Krauss, R. M.; Chen, Y.; and Chawla, P. 1996. Non-verbal behavior and nonverbal communication: What do conversational hand gestures tell us? In Zanna, M., ed., *Advances in experimental social psychology*. Tampa: Academic Press. 389–450.
- Kreiman, G.; Koch, C.; and Fried, I. 2000. Imagery neurons in the human brain. *Nature* 408:357–361.
- Lakoff, G., and Johnson, M. 1999. *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. New York: Basic Books.

- Locke, J. L., and Kutz, K. J. 1975. Memory for speech and speech for memory. *Journal of Speech and Hearing Research* 18:176–191.
- Malle, B.; Moses, L.; and Baldwin, D., eds. 2001. *Intentions and Intentionality*. MIT Press.
- Massaro, D., and Cohen, M. M. 1983. Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 9(5):753–771.
- Matarić, M. J. 2002. Sensory-motor primitives as a basis for imitation: linking perception to action and biology to robotics. In Dautenhahn, K., and Nehaniv, C. L., eds., *Imitation in animals and artifacts*. MIT Press. chapter 15, 391–422.
- McClave, E. 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics* 32:855–878.
- McLeod, P., and Posner, M. I. 1984. Privileged loops from percept to act. In Bouma, H., and Bouwhuis, D. G., eds., *Attention and performance*, volume 10. Hillsdale, NJ: Erlbaum. 55–66.
- Meltzoff, A. N., and Moore, M. K. 1997. Explaining facial imitation: a theoretical model. *Early Development and Parenting* 6:179–192.
- Parsons, L. M. 1994. Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance* 20(4):709–730.
- Pecher, D., and Zwaan, R. A., eds. 2005. *Grounding cognition: the role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge Univ. Press.
- Russell, S., and Norvig, P. 2002. *Artificial Intelligence: A Modern Approach (2nd edition)*. Prentice Hall.
- Sebanz, N.; Bekkering, H.; and Knoblich, G. 2006. Joining action: bodies and minds moving together. *Trend in Cognitive Sciences* 10(2):70–76.
- Somerville, J. A.; Woodward, A. L.; and Needham, A. 2004. Action experience alters 3-month-old infants' perception of others actions. *Cognition*.
- Spivey, M. J.; Richardson, D. C.; and Gonzalez-Marquez, M. 2005. On the perceptual-motor and image-schematic infrastructure of language. In Pecher, D., and Zwaan, R. A., eds., *Grounding cognition: the role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge Univ. Press.
- Stanfield, R., and Zwaan, R. 2001. The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science* 12:153–156.
- Wilson, M., and Knoblich, G. 2005. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin* 131:460–473.
- Wilson, M. 2001. Perceiving imitable stimuli: consequences of isomorphism between input and output. *Psychological Bulletin* 127(4):543–553.
- Wilson, M. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review* 9(4):625–636.