Cogn Process (2007) 8:151–158
DOI 10.1007/s10339-007-0168-9

REVIEW

Prediction of intent in robotics and multi-agent systems

Yiannis Demiris

Received: 11 January 2007 / Revised: 13 March 2007 / Accepted: 19 March 2007 / Published online: 4 May 2007

Abstract Moving beyond the stimulus contained in observable agent behaviour, i.e. understanding the underlying intent of the observed agent is of immense interest in a variety of domains that involve collaborative and competitive scenarios, for example assistive robotics, computer games, robot–human interaction, decision support and intelligent tutoring. This review paper examines approaches for performing action recognition and prediction of intent from a multi-disciplinary perspective, in both single robot and multi-agent scenarios, and analyses the underlying challenges, focusing mainly on generative approaches.

Introduction

Designing and implementing algorithms for enabling machines, and in particular robots to recognise the actions of humans is a task that, although challenging, has substantial application potential. Applications for such algorithms include:

- Surveillance: monitoring public areas for automatic recognition of threatening or abusive behaviour; crowd monitoring during evacuation of large buildings.
- Ambient intelligence and assistive devices: monitoring indoor environments and the actions of humans for assisted living. Applications in this area are usually focused on monitoring and assisting disabled or elderly people.
- Entertainment and sports: recognising the actions of humans as an interface to games and virtual environments; better monitoring of athletes' performance.
- Robotics: recognising the actions of humans has novel robot applications such as learning by demonstration and imitation (Schaal 1999; Schaal et al. 2003; Demiris and Hayes 2002; Demiris and Khadhouri 2006) which have the potential to lead to easily programmable robots.

A number of detailed surveys (Aggarwal and Cai 1999; Moeslund and Granum 2000; Moeslund et al. 2006, among others) have already explored how such actions can be captured, analysed and understood; what this paper will concentrate on is the different approaches to move beyond the demonstrated stimulus, and investigate how less tangible aspects of the demonstration, particularly the underlying goals and intentions of the demonstrator, can be inferred. This is a task that is particularly difficult, and might prove to be impossible in certain cases; however, it is worthwhile to pursue because equipping machines with such capabilities will elevate their capacities as effective assistants.

We will first examine some of the definitions related with intention and prediction, and proceed to examine alternative approaches for the prediction of intent; we will subsequently focus our discussion on generative approaches, using the HAMMER architecture (Demiris and Khadhouri 2006) as a representative example. The paper will conclude with a review of the more general and less explored problem of predicting the intention of groups of

Y. Demiris (✉)
Department of Electrical and Electronic Engineering,
Imperial College London, South Kensington Campus,
Exhibition Road, London SW7 2BT, UK
e-mail: y.demiris@imperial.ac.uk
URL: http://www.iis.ee.ic.ac.uk/yiannis/

agents. Intention recognition is studied extensively in different disciplines and it is not possible to do justice to all of them in the space of a short review article. The purpose instead is to serve as an interdisciplinary introduction and demonstrate links between the different approaches, hopefully inspiring further interdisciplinary cooperation in intention recognition.

Background—intentions and goals in humans

People act not only as a response to external or internal stimuli, but also in order to achieve internally or externally posed goals. There has been a lot of theoretical and experimental work in determining the mechanisms involved in these processes, as well as clearly defining the associated terminology (e.g. Bratman 1990; Cohen and Levesque 1990; Tomasello et al. 2005).

Living in societies, humans also direct much of their behaviour in response to their interpretation and prediction of the intentions of others. Humans are quite good at this inference task, starting from a very young age. In an experiment by (Meltzoff 1995, 2007b), 18-month-old children were shown unsuccessful acts involving a demonstrator trying but failing to achieve his goal, i.e. the children did not see a successfully reached end-state. The children, however, did not replicate the unsuccessful surface behaviour of the adult but proceeded to imitate the intended goal, even when it was never shown to them. In adults, neuroscience data have been pointing to specialised human brain mechanisms for perceiving actions and intentions of other humans (for a review, see Blakemore and Decety 2001).

There are significant difficulties in perceiving intentions as well as action goals. The main one is the problem of inversion, the fact that an observed action can be the result of more than one intention. Consider the example of someone intentionally pushing you. The immediate goal of the other agent is to displace you from a location, but the underlying intention is not clear until additional information are added into the equation—is the person that pushed me angry at me? Am I in danger in my previous location? The perception of the current context is crucial to correctly infer the intentions of other agents.

The example above highlights the close relation of intentions and action goals, and the difficulty in drawing an exact division line between them**.** The terms goals and intentions are frequently used interchangeably, but in general goals refer to more immediate desirable end-states, whereas frequently intentions have a longer term or higher-level connotation. Tomasello et al. (2005) define intentions as ''a plan of action the organism chooses and commits itself to the pursuit of a goal—an intention thus includes both a means (action plan) as well as a goal'' (p. 676). We will use Tomasello's definition as the working definition for this paper.

Approaches to intention recognition

(Kanno et al. 2003) defines three types of intention recognition: keyhole recognition, intended recognition and obstructed recognition. In the first one, the observed agent is unaware of the observer, and proceeds executing the plan without any special consideration for the observer. In the second type the observed agent is aware of the observer and actively cooperates in the recognition, for example, by ensuring that crucial parts of the demonstration are not obstructed. In the third type, the observed agent is again aware of the observer but is actively trying to disrupt the recognition process and hide its intentions. More challenging issues such as adversarial reasoning and deception (Kott and McEneaney 2006) can also come into play, where the agent will even execute actions that do not correspond to its intentions, in order to deceive or mislead the observer. The latter cases are however beyond the scope of the paper, and we will restrict the discussion to the keyhole and intended types of intention recognition.

Recognizing the goals and intentions of the actions of an agent is essentially a problem of model matching; the observer agent deploys a number of sensors, each reporting its observations about the state of the observed agent at a specified sampling rate. The collected data can be acted upon through two different approaches, *descriptive* versus *generative.*

Within the descriptive approach, patterns are characterised through the extraction of a number of low-level features, and the use of a set of restrictions at the feature level, for example through Markov Random Fields (Isham 1981), or Deformable Models, popular in computer vision-based applications (see Jain et al. 1998 for a review). The observer agent subsequently matches the observed data against pre-existing representations, and depending on what the task is (imitation of observed actions, collaboration etc.), generates the actions corresponding to these representations. Pre-existing representations can have associated data that label these representations with the goals, beliefs and intentions that underlie their execution. This approach corresponds to the ''action–effects associations'' method for intention interpretation in the review of Csibra and Gergely (2007), and to the ''Theory of Event Coding'' approach put forward by Hommel et al. (2001) based on William James' ideomotor principle in which bidirectional action–effects associations are used to predict the goals of an action.

Within the generative approach, a set of latent (hidden) variables is introduced; this set encodes the causes that *can* *produce* the observed data. They represent the intrinsic degrees of freedom underlying the structure of the observations, usually using probability distributions. Using these variables for a recognition task involves modifying the parameters of the generating process until the generated data can be favourably compared against the observed data. Generative models are very popular in the machine learning community, with many variations in existence (e.g. Roweis and Ghahramani 1999; Bishop 2006; Buxton 2003).

The idea that the generative model can be used to explain or predict observed data has been gaining popularity in the robotics community who have been approaching the problem armed with an additional constraint, that of *embodiment*. The internal models here take the form of motor control models capable of driving an embodied system. These internal models exist in various forms, including *forward and inverse models* (explained later), as well as *behaviours* (Arkin 1998), *schemas* (Acosta-Calderon and Hu 2005; Pezzulo and Calvi 2006), varying in whether they act in a feedback (usually behaviours) or feedforward (usually schemas, inverse models) manner. A number of architectures have been proposed, using combinations of these internal models, including HAMMER (Demiris and Hayes 2002; Demiris and Khadhouri 2006), with an emphasis on modelling mirror neurons and robot learning by imitation applications, and MOSAIC (Wolpert et al. 2003) with an emphasis on motor control. HAMMER, in particular, was designed with the aim of using the internal models of robots to produce movement as well as perceive it when produced by others. We will proceed to explaining this in more detail in the next section, as a prototypical example of the prediction through synthesis approach. Alternative approaches also exist, including for example the use of repeated imitation games between agents (Jansen and Belpaeme 2006).

The idea that you can view perception as internal simulation, using your action models to predict ongoing demonstration (as in HAMMER) has many links with the simulationist perspective of cognitive functions (Hesslow 2002). Similar ideas to this have been put forward in other research fields, demonstrating the generality of the principle. For example, in the field of intelligent tutoring, John Anderson put forward a technique known as model tracing (Anderson et al. 1990), where a runnable model of the student's cognitive skills in a particular domain is executed and compared with the student's actions. Inserting ''buggy rules'' into the model results in suboptimal performance and errors; if these errors correlate well with the student errors, the rules are taken as a possible explanation of the deficiencies in the student's knowledge, and actions are taken to repair these. In the field of speech perception, Liberman's theory of speech perception (Liberman et al. 1967) employs a similar perception through motor simulation approach; you understand speech through internal generation and reproduction of the acoustic signal. The neuroanatomical basis of this approach and its alternatives are examined in Scott and Johnsrude (2003).

An alternative to goal recognition that has been put forward is also worth noting, that is the ''teleological interpretation of actions''. A comparative review against the other two approaches can be found in Csibra and Gergely (2007), but briefly the approach performs a normative evaluation of observed actions based on the principle of rational actions (Csibra and Gergely 1998), which ''allows for the assessment of the relative efficiency of the action performed to achieve the goal within the situational constraints given'' (Csibra and Gergely 2007, p. 70). The effect of an observed action can be seen as the goal depending on whether the outcome is judged to justify the action in the given context it was observed in.

The generative embodied simulationist approach—the single agent case

We will now use the HAMMER (Hierarchical Attentive Multiple Models for Execution and Recognition) architecture as a representative example of the generative embodied simulationist approach to understanding intentions. We will explain the operation of the architecture by starting from the second half of its acronym (MER—how a model can be used both for execution and recognition of an action) in the next section, and proceed to explain how multiple models can be used concurrently, organised in hierarchies and incorporate attention, in the subsequent sections.

Principles

HAMMER utilises the concepts of inverse and forward models. An inverse model is akin to the concepts of a controller, behaviour, action, or motor plan. The inverse model's function is to receive as input a measurement or estimate of the current state of the system and the desired target goal(s) and outputs the control commands that are needed to achieve or maintain those goal(s). A forward model of a modelled system (akin to the concept of internal predictor) is a function that takes as inputs the current state of the system and a control command to be applied to it and outputs the predicted next state of the controlled system (Miall and Wolpert 1996). It is worthwhile to note that the term *forward models* have also been used in a modified version in different contexts (for a review of different usages, see Karniel 2002).

The building block of HAMMER is an inverse model paired with a forward model (Fig. 1). When HAMMER is asked to rehearse or execute a certain action, the corresponding inverse model module is given information about the current state and, optionally, about the target goal(s). The inverse model then outputs the motor commands that are necessary to achieve or maintain these implicit or explicit target goal(s). The forward model provides an estimate of the upcoming states should these motor commands get executed. This estimate is returned back to the inverse model, allowing it to adjust any parameters of the action (an example of this would be achieving different movement speeds (Demiris and Hayes 2002)). The estimate can also be compared with the target goal to produce a reinforcement signal for the inverse model depending on how much the model's motor commands brought the estimate closer to the target goal. Architectures involving combinations of inverse and forward models (in varying configurations, for example differing in how control is switched between multiple models) are used in motor control (Narendra and Balakrishnan 1997; Wolpert and Kawato 1998) due to their flexible modular structure, and have been advocated for use in imitation and learning (Demiris and Hayes 2002; Demiris and Khadhouri 2006; Schaal 1999; Schaal et al. 2003; Wolpert et al. 2003).

The HAMMER architecture uses an inverse–forward model coupling in a dual role: either for executing an action, or for perceiving the same action when performed by a demonstrator. When HAMMER operates in action perception mode, it can determine whether a visually perceived demonstrated action matches a particular inverse–forward model coupling (Fig. 2), by feeding the demonstrator's current state as perceived by the imitator to the inverse model. The inverse model generates the motor commands that it would output *if it was in that state and was executing the particular action*. In a sense, the imitator processes the actions by analogy with the self—''what would I do if I were in the demonstrator's shoes?''

In the perception or planning modes, the motor commands are inhibited from being sent to the motor system. The forward model outputs an estimated next state, which is a prediction of what the demonstrator's next state will be. This predicted state is compared with the demonstrator's actual state at the next time step. As seen in Fig. 2 and the text that follows, this comparison results in an error signal that can be used to increase or decrease the behaviour's confidence value, which is an indicator of how closely the demonstrated action matches a particular imitator's action.

An interesting point that arises here is how to learn these models; interested readers are referred to Dearden and Demiris (2005) for some initial work on a developmental approach on how this can be achieved in robots. In these experiments, the robot associated self-generated actions with the feedback they produce once executed (including learning the feedback delays in the motor system).

So far we have described how the 'MER' (Models for Execution and Recognition) part of HAMMER operates. It remains to be seen why the 'HAM' (Hierarchical Attentive Multiple) part is important, starting from the multiplicity aspect and continuing with the Hierarchies and Attention in the next section.

HAMMER consists of multiple pairs of inverse and forward models that operate in parallel (Demiris and Hayes 2002). As the demonstrator agent executes a particular action, and there are multiple models (possibilities) that can explain the ongoing demonstration, we feed the perceived states into all of the imitator's available inverse models. This will result into the generation of multiple motor commands (representing the multiple hypotheses as to what action is being demonstrated) that are sent to the forward models. The forward models generate predictions about the demonstrator's next state as described earlier and these are compared with the actual demonstrator's state at the next time step. The error signal resulting from this comparison affects the confidence values of the inverse models. At the end of the demonstration (or earlier if required) the inverse model with the highest confidence value, i.e. the one that is the closest match to the demonstrator's action is selected and is offered as an estimate of the intention. Demiris and Hayes (2002) have described the relation of this process to a biological counterpart, the mirror system (Gallese et al. 1996), offering a number of

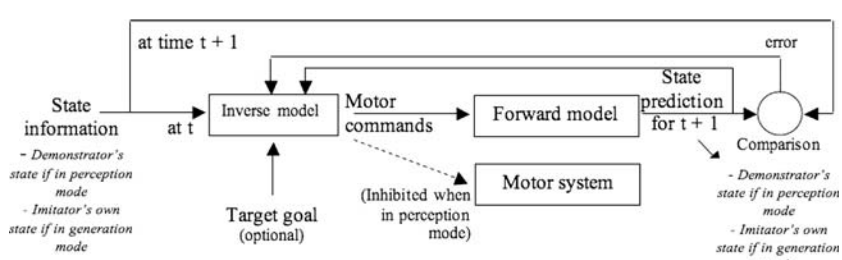

Fig. 1 HAMMER's basic building block, an inverse model paired with a forward model (from Demiris and Hayes 2002, Demiris and Johnson 2003). The target goal (or intention) is marked optional since it might already be implicit in the functionality of the inverse model

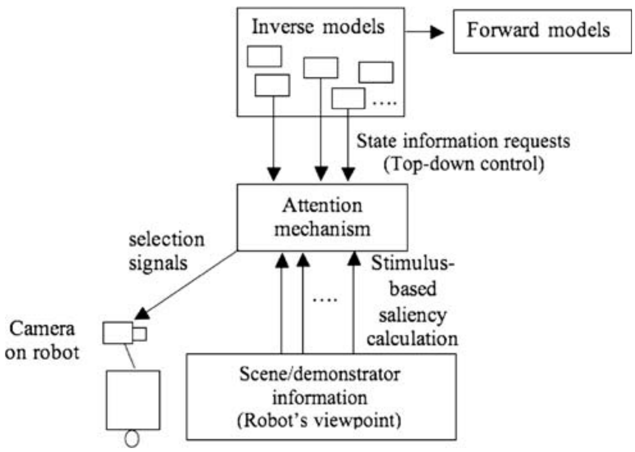

Fig. 2 Inverse models submit requests to the attention mechanism, exerting top-down control

explanations and testable predictions (Demiris and Hayes 2002; Demiris and Simmons 2006), for example, a predicted dependency of the firing rate of the macaque monkey mirror neurons to the velocity profile of the demonstrated act.

Attention, hierarchies and perspective taking

Attention

The multiple models formulation, as stated so far, assumes that the complete state information will be available for and fed to all the available inverse models. Since each of the inverse models requires a subset of the global state information (for example, one might only need the arm position of the demonstrator rather than full body state information), we can optimise this process by allowing each inverse model to request a subset of the information from an attention mechanism, *thus exerting a top-down control on the attention mechanism*. Since HAMMER is inspired by the ''simulation theory of mind'' point of view for action perception, it asserts that, for a given behaviour, the information that it will try to extract during the demonstration is the state of the variables it would control if it was executing this behaviour (Demiris and Khadhouri 2006). Apart from improving on the resource requirements of the architecture above, this novel approach provides a principled way for supplying top-down signals to attention. *The saliency of each request can then be a function of the confidence that each inverse model possesses*, removing the need for ad-hoc ways for computing the saliency of top-down requests. Top-down control can then be integrated with saliency information from the stimuli itself, allowing a control decision to be made as to where to focus the observer's attention. An overall diagram of this is shown in Fig. 2).

Strategies for selecting among the different requests can include ''equal time sharing'', or ''highest priority first'', or other suitable resource scheduling algorithms (Demiris and Khadhouri 2006).

Although this architecture is based on a principled approach on how the observer's internal models and prior knowledge influence what parts of the stimulus will be attended to, the relation to biological (for example, Flanagan and Johansson 2003) and developmental data requires further exploration.

Hierarchical organisation of the inverse and forward models

How are human action models organised? Recent evidence on how infants encode goals suggests hierarchical representations (Bekkering et al. 2000; Gleissner et al. 2000; Wohlschlager et al. 2003), and recent brain imaging data have also begun to shed light into these hierarchical representations in adults (Hamilton and Grafton 2007). In robots, hierarchical formulations have been proposed and used (Demiris and Johnson, 2003; Tani and Nolfi 1999), but their relation to biological data has not been explored (but see Byrne and Russon 1998; Demiris and Simmons 2006).

The important issues to consider in hierarchical organisations are the nature of abstraction or generalisation (if any) that we achieve by moving into higher levels of the hierarchy, i.e. how inverse and forward models can be put together to form ''higher models''. In the 'subsumption architecture' (Brooks 1986) for example, higher levels provide the gating for the lower levels but do not provide any generalisation. In Demiris and Johnson (2003) inverse models are formed by allowing lower level models to be placed in parallel or in sequence based on whether there are overlapping degrees of freedom between the body structures that the inverse models control. HMOSAIC (Wolpert et al. 2003) proposes a three-level hierarchy with the low-level dynamics at the lower level, sequences of elements at the middle level, and symbolic representations of tasks at the higher level. Despite these first attempts, further theoretical advancements will be required, in order to be able to merge the prediction of proximal motor intentions that architectures, such as HAMMER and MOSAIC, can provide and higher ''theory of mind'' type of tasks that a more general simulation theory of mind would require.

Perspective taking

The simulationist approach to understanding intentions requires the observer to take the perspective of the demonstrator, i.e. to ''step into the demonstrator's shoes''. Useful information on how such mechanisms can be implemented is available from developmental work on gaze following, which can be viewed as the lowest end of perspective taking. Work by Brooks and Meltzoff (2002, 2005) has shown that one-year-old infants can follow the gaze of adults and realise that it is not a meaningless movement but is directed at an object. The evidence points to a use of first-person experience (our own internal models) to make third-person attributions; for example, Meltzoff (2007a) and Meltzoff and Brooks (2004) have shown that once infants had experience with blindfolds, the interpretation of others who wear blindfolds also changes. Although various algorithmic solutions to perspective taking have been proposed (Johnson and Demiris 2005; Breazeal et al. 2006; Trafton et al. 2005), higher levels of perspective taking, like the ones discussed in this paper, including beliefs, desires and intentions remain difficult challenges in robotics. In Johnson and Demiris (2005) perceptual perspective taking allowed an observer robot to ''place itself in the demonstrator robot's perceptual shoes'' and engage the inverse models that were compatible with the demonstrator's viewpoint rather than its own viewpoint. Although there is still a lot of work to be done in robotics on this aspect, research on the development of perspective taking and its roots in gaze following (Meltzoff 2005, 2007a) as well as relevant neuroscience data for the adult cases (e.g. Jackson et al. 2006) can provide robotics researchers with useful information regarding potential implementation approaches.

The multi-agent case

Intention recognition and prediction is of importance also for applications involving groups of agents, particularly in adversarial scenarios such as competitive sports (Beetz et al. 2005) and military simulations (Tambe 1996). It is also of use in cooperative situations where the behaviour of an agent is dependent on its partner's or team's behaviour (Grosz and Hunsberger 2006; Kanno et al. 2003). The multiplicity of agents involved complicates intention recognition in two important ways:

To predict the intention of the group it is not sufficient to track and predict the actions of individual agents in the group. It is necessary to attempt to infer the joint intention or shared plan of the agents as a group. This is not simply the sum of the intentions of the individual agents, but needs to be found within the agents' ''shared cooperative activity'' (Bratman 1992), which Bratman defined as a combination of mutual responsiveness, commitment to the joint activity and commitment to mutual support. (Tambe 1996) presented a system, RESCteam, which constructs explicit teams models and tracks them at the team level; as a result, they avoid the execution of a large number of individual agent models.

In addition to recognising activity, it is crucial to recognise an agent's identity, and its position in the social structure i.e. in what sub-team does it belong to, and what is its role (Sonenberg and Tidhar 1999). Given that an agent within a team can assume more than one role, the action it is performing can be interpreted in different ways depending on the role it is believed to have.

Methods that attempt to simultaneously identify subgroups as well as recognise their behaviour have begun to appear (Devaney and Ram 1998; Sukthankar and Sycara 2006), but their source of information are spatiotemporal traces of the agents, which convey little information, making the problem particularly hard. The observer does not affect these traces, but remains a passive observer. Mutual support, one of the key aspects of shared cooperative activity (Bratman 1992), might be particularly important here because mutual support might necessitate intention updating depending on the performance of subgroups; the change of activity to a set of agents based on an action we caused on another set of agents might reveal important information regarding the correlation of the activities and roles of the two sets, and give clues as to their joint intention.

Conclusions

We reviewed the different approaches to action recognition and prediction of intent, distinguishing between descriptive and generative approaches, and surveying the generative architectures available, using HAMMER as the main example. Prediction of intent remains a challenging task, with advancements needed at all levels, both theoretical, as well as technological, particularly if the application involves groups of agents. Solutions, as in the past in active learning and active vision, might be found in the active involvement of the observer while the operation is unfolding, so that the intricate correlations between activities of multiple agents can be revealed.

Acknowledgments I would like to thank Simon Butler, Bálint Takács, Ant Dearden and Tom Carlson, as well as the anonymous reviewers for their comments.

References

Acosta-Calderon CA, Hu H (2005) Robot imitation: body schema and body percept. J Appl Bionics Biomech 2(3–4):131–148. ISSN 1176–2322

Aggarwal A, Cai Q (1999) Human motion analysis: a review. Comput Vis Image Underst 73(3):428–440

Anderson JR, Boyle CF, Corbett AT, Lewis MW (1990) Cognitive modelling and intelligent tutoring. Artif Intell 42:7–49
Arkin RC (1998) Behavior based robotics. MIT Press, Cambridge
Beetz M, Kirchlechner B, Lames M (2005) Computerised real-time analysis of football games. Pervasive Comput 4(3):33–39
Bekkering H, Wohlschläger A, Gattis M (2000) Imitation of gestures in children is goal-directed. Q J Exp Psychol 53A:153–164
Bishop C (2006) Pattern recognition and machine learning. Springer, Heidelberg
Blakemore S-J, Decety J (2001) From the perception of action to the understanding of intention. Nat Rev Neurosci 2:561–567
Bratman ME (1990) What is Intention? In: Cohen PR, Morgan JL, Pollack ME (eds) Intentions in communication, MIT Press, Cambridge, pp 15–32
Bratman ME (1992) Shared cooperative activity. Philos Rev 101(2):327–341
Breazeal C, Berlin M, Brooks A, Gray J, Thomaz A (2006) Using perspective taking to learn from ambiguous demonstrations. Rob Auton Syst 54(5):385–393
Brooks (1986) A robust layered control system for a mobile robot. IEEE J Robot Autom RA-2, pp 14–23
Brooks R, Meltzoff AN (2002) The importance of eyes: how infants interpret adult looking behavior. Dev Psychol 38:958–966
Brooks R, Meltzoff AN (2005) The development of gaze following and its relation to language. Dev Sci 8:535–543
Buxton H (2003) Learning and understanding dynamic scene activity: a review. Image Vis Comput 21:125–136
Byrne RW, Russon AE (1998) Learning by imitation: a hierarchical approach. Behav Brain Sci 21(5):667–684
Cohen PR, Levesque HJ (1990) Intention is choice with commitment. Artif Intell 42:213–261
Csibra G, Gergely G (1998) The teleological origins of mentalistic action explanations: a developmental hypothesis. Dev Sci 1:255–259
Csibra G, Gergely G (2007) 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. Acta Psychologica 124:60–78
Dearden A, Demiris Y (2005) Learning forward models for robots. In: Proceedings of the international joint conference on artificial intelligence (IJCAI), Edinburgh, pp 1440–1445
Demiris Y, Hayes G (2002) Imitation as a dual-route process featuring predictive and learning components: a biologically plausible computational model. In: Dautenhahn K, Nehaniv CL (eds) Imitation in animals and artifacts. MIT Press, Cambridge, pp 327–361
Demiris Y, Johnson M (2003) Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning. Connect Sci 15:231–243
Demiris Y, Khadhouri B (2006) Hierarchical attentive multiple models for execution and recognition of actions. Rob Auton Syst 54:361–369
Demiris Y, Simmons G (2006) Perceiving the unusual: temporal properties of hierarchical motor representations for action perception. Neural Netw 19:272–284
Devaney M, Ram A (1998) Needles in a haystack: plan recognition in large spatial domains involving multiple agents. In: Proceedings of the 15th national conference on artificial intelligence, AAAI-98, pp 942–947
Flanagan JR, Johansson RS (2003) Action plans used in action observation. Nature 424:769–771
Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609
Gleissner B, Meltzoff AN, Bekkering H (2000) Children's coding of human action: cognitive factors influencing imitation in 3-year-olds. Dev Sci 3:405–414
Grosz BJ, Hunsberger L (2006) The dynamics of intention in collaborative activity. Cogn Syst Res 7:259–272
Hamilton A, Grafton ST (2007) The motor hierarchy: from kinematics to goals and intentions. In: Haggard P, Rossetti Y, Kawato M (eds) Sensorimotor foundations of higher cognition, attention and performance XXII, chap 18 (in press)
Hesslow G (2002) Conscious thought as simulation of behaviour and perception. Trends Cogn Sci 6(6):242–247
Hommel B, Musseler 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
Isham V (1981) An introduction to spatial point processes and markov random fields. Int Stat Rev 49:21–43
Jackson PL, Meltzoff AN, Decety J (2006) Neural circuits involved in imitation and perspective-taking. Neuroimage 31:429–439
Jain AK, Zhong Y, Dubuisson-Jolly M-P (1998) Deformable template models: a review. Signal Processing 71:109–129
Jansen B, Belpaeme T (2006) A computational model of intention reading in imitation. Rob Auton Syst 54(5):394–402
Johnson MR, Demiris Y (2005) Perceptual perspective taking and action recognition. Int J Adv Rob Syst 2:301–308
Kanno T, Nakata K, Furuta K (2003) A method for team intention inference. Int J Hum Comput Stud 58:393–413
Karniel A (2002) Three creatures named forward model. Neural Netw 15:305–307
Kott A, McEneaney WM (eds) (2006) Adversarial reasoning: computational approaches to reading the opponent's mind. Chapman & Hall/CRC, London
Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. Psychol Rev 74:431–361
Meltzoff AN (1995) Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. Dev Psychol 31:838–850
Meltzoff AN (2005) Imitation and other minds: the ''Like Me'' hypothesis. In: Hurley S, Chater N (eds) Perspectives on imitation: from neuroscience to social science. MIT Press, Cambridge, vol 2, pp 55–77
Meltzoff AN (2007a) 'Like me': a foundation for social cognition. Dev Sci 10:126–134
Meltzoff AN (2007b) The 'like me' framework for recognizing and becoming an intentional agent. Acta Psychologica 124:26–43
Meltzoff AN, Brooks R (2004) Developmental changes in social cognition with an eye towards gaze following. In: Carpenter M, Tomasello M (eds) Action-based measures of infants' understanding of others' intentions and attention. Symposium conducted at the Biennial meeting of the International Conference on Infant Studies, Chicago
Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural Netw 9:1265–1279
Moeslund TB, Granum E (2000) A survey of computer vision-based human motion capture. Comput Vis Image Underst 81(3):231–268
Moeslund TB, Hilton A, Kruger V (2006) A survey of advances in vision-based human motion capture and analysis. Comput Vis Image Underst 104:90–126
Narendra KS, Balakrishnan J (1997) Adaptive control using multiple models. IEEE Trans Autom Control 42(2):171–187
Pezzulo G, Calvi G (2006) A schema based model of the praying mantis. From animals to animats. In: Proceedings of the 9th international conference on simulation of adaptive behaviour. Springer LNAI, vol 4095, pp 211–223
Roweis S, Ghahramani Z (1999) A unifying review of linear Gaussian models. Neural Comput 11(2):305–345
Schaal S (1999) Is Imitation learning the route to humanoid robots?. Trends Cogn Sci 3:233–242

Schaal S, Ijspeert A, Billard A (2003) Computational approaches to motor learning by imitation. Philos Trans R Soc Lond B Biol Sci 358:537–547

Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organisation of speech perception. Trends Neurosci 26(2):100–107

Sonenberg L, Tidhar G (1999) Observations on team-oriented mental state recognition. In: Proceedings of the IJCAI-1999 workshop on team modelling and plan recognition

Sukthankar G, Sycara K (2006) Simultaneous team assignment and behavior recognition from spatio-temporal agent traces. In: Proceedings of 21st national conference on artificial intelligence (AAAI-06)

Tambe M (1996) Tracking dynamic team activity. In: Proceedings of the national conference on artificial intelligence (AAAI)

Tani J, Nolfi S (1999) Learning to perceive the world as articulated: an approach for hierarchical learning in sensory motor systems. Neural Netw 12:1131–1141

Tomasello M, Carpenter M, Call K, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. Behav Brain Sci 28:675–735

Trafton J, Cassimatis N, Bugajska M, Brock D, Mintz F, Schultz A (2005) Enabling effective human–robot interaction using perspective taking in robots. IEEE Trans Syst Man Cybern A Syst Hum 35(4):460–470

Wohlschlager A, Gattis M, Bekkering H (2003) Action generation and action perception in imitation: an instance of the ideomotor principle. Philos Trans R Soc Lond B Biol Sci 358:501–515

Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Netw 11:1317–1329

Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci 358:593–602