European Review, Vol. 15, No. 2, 223–240 (2007) © Academia Europæa, Printed in the United Kingdom

Mind-reading as Control Theory

PETER GÄRDENFORS

Lund University Cognitive Science, Kungshuset, S-222 22 Lund, Sweden.
E-mail: Peter.Gardenfors@lucs.lu.se

In contrast to other animals, humans are good at mind-reading in the sense that they can represent the contents of the minds of others. In this article, the competence for inter-subjectivity is divided into representing the emotions, the attention, the intentions and the beliefs and knowledge of others. Recent attempts to exploit control theory for modelling various cognitive functions are discussed and it is outlined how this modelling approach can be combined with the analysis of inter-subjectivity.

Metaphors for the architecture of the brain

The functioning of the brain has always been a topic of speculation. Aristotle's view was that the brain is a secondary organ that serves as a *cooling system* for the heart. During the 1950s, in connection with the birth of cognitive science, the guiding metaphor became that the brain is a *computer*. This metaphor was the starting point of much research in artificial intelligence, psychology and linguistics. For example, memory research looked for the areas of the brain where certain memories were 'stored', in analogy with the memory cells of the computer.

Another metaphor, that has never really generated a research paradigm, was that the brain functions like a *hologram*.[1] In the 1970s, it became popular to model different cognitive functions as parallel distributed processes in artificial neural networks. An appropriate metaphor for this is that the brain was seen as an *anthill* with a large number of simple processes that jointly resulted in emergent phenomena.[2] From a medical angle, a metaphor proposed by Gerald Edelman is that the brain functions as an *immune system*.[3]

Recently, yet another metaphor has generated a new wave of modelling: the brain as a *control system*. Concepts from control theory, such as feedback loops, oscillators and emulators, have been proposed as new ways of understanding various cognitive functions.[4–7]

In this article, the focus will be on how models from control theory can throw new light on how we think about the thoughts of others. The term *inter-subjectivity* will be used broadly as the sharing and understanding of others' mentality. This is the everyday meaning of reading others' minds. Different levels of inter-subjectivity will be presented and the question is then to what extent control theory can be useful in understanding this kind of mind-reading.

Sensations, perceptions and imaginations

In order to understand how most higher forms of thinking function, we must recall that animals and humans *represent* the world around them and its possibilities in different ways. The different elements in consciousness will be divided into:[8]

- *Sensations* that are the immediate sensory impressions.
- *Perceptions* that are interpreted sensory impressions.
- *Imaginations* (or images) that are not directly governed by sensory impressions.

This is the evolutionary order in which the different functions appear. Even simple organisms have sensations. More advanced animals have perceptions that consist of interpretations of sensations. In particular, perceptions can provide knowledge about what is going on in the animal's immediate surroundings. It is probably only mammals and birds that have imaginations: they can direct their thinking towards something that is not immediately present in the surrounding environment.

On the first level, consciousness contains sensations. Our subjective world of experiences is full of them: tastes, smells, colours, itches, pains, sensations of cold, sounds, etc (what philosophers of mind call *qualia*). They provide an awareness of the world. Nicholas Humphrey argues that the biological role of sensations is to tell about what is *happening right now* to the organism.[9]

However, an organism can not only find out what is happening to its own body if it also receives signals about *what is going on out there in the world*; it then has better opportunities to foresee the future and thus survive in an inhospitable world. This is the role of the perceptions. First and foremost, perceptions provide information about the spatial structure of the world and the objects that are in it. Humphrey says that sensations give us *egocentric* information while perceptions can also contribute *allocentric* (outside the self) information.[9]

The brain is full of mechanisms that contribute new information. In particular, there are many well-studied examples of the visual process. When we see an object, we sense that it has contours, for example. But if we examine the influx of light that hits the retina, we find nothing that corresponds to such contours – they are part of the information that the visual process *constructs*.

Phenomena such as this are very common and well-known from psychophysics. They show that we have a plethora of mechanisms that complement the signals provided by our senses. These mechanisms create the *representations* with which thinking works, since what we experience is not only that which is presented by our sensory receptors but also that which is recreated, that is represented. The filled-in sensations are what are called here perceptions.

Our senses do not give us information about everything that exists in the world. The evolutionary point is that the richer representations help us *predict* what the world will look like. They create the future in advance. In the following section, a control-theory model of this form of prediction will be presented.

There is nothing that requires that there be any feedback from the senses to let the filling-in mechanisms start working. Actually, it is only a little extra step for the evolutionary process to let the mechanisms work without any signals being sent to the body at all. This is what happens when the organism creates for itself an image of what will happen if a certain action is undertaken. Animals' brains, especially humans', utilize *imaginations* of this kind – things that we can move around in our heads before trying to move them in reality.

The founding father of psychology, William James writes:[10] 'The commonly received idea [about imagination] is that it is only a milder degree of the same process which took place when the thing now imagined was sensibly perceived.' Although James did not have much evidence for this position, there is now quite strong support from brain research that the same mechanisms are involved in controlling the body as in imagining controlling it. The so-called motor imagining ability has been investigated thoroughly by Marc Jeannerod and others.[11,12] It turns out that the same parts of the motor cortex are involved when an action is carried out as when one only imagines the action.

Being able to use imaginations requires that one can *suppress* the sensations one has for the moment, otherwise they will come into conflict with the representation. Arthur Glenberg says that imaginations put reality in quarantine.[13] That places new demands on mental capacities. The suppression of information coming in from reality is probably managed by the frontal lobes of the brain, which are the parts that are in charge of planning and fantasizing and the so-called 'executive functions' of self-control.

Glenberg presents a theory of how memory works in which he distinguishes between 'automatic' and 'effortful' memory.[13] The automatic memory is the one that is used to fill in sensations so that they become perceptions. When you recognise a person in a crowd it is because you blend what you see of the person with your memories. Sometimes there is a mistake – you overlay your memory of one person on top of the sensory impression you receive of another person.

The effortful memory is the one we use when we create images. Images do not appear from nowhere – they build on our previous experiences. What we usually call remembering is only a special kind of image that we think corresponds to something that has actually happened. But memory is also used in fantasies: you cannot imagine a centaur (if you have not seen a picture of one) without memories of horses and people. But images must not be confused with perceptions. An effort is required to ignore sensations. That is why we often close our eyes or look up at the ceiling when we want to remember something or when we fantasize. This effort can be investigated since, for example, it is more difficult to carry out some action at the same time as we are remembering or imagining.

A control-theory approach to perceptions

The collection of perceptions and imaginations corresponds to what is often called the *inner world* of an organism. The existence of an inner world has clear evolutionary advantages as was pointed out by Craik:

> If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which are the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react on a much fuller, safer and more competent manner to the emergencies which face it.[14]

One of the main evolutionary advantages of an inner world is that it frees an animal that is seeking a solution to a problem from dangerous trial-and-error behaviour. Jeannerod says that his general position is that 'actions are driven by an internally represented goal rather than directly by the external world.'[10] By exploiting its inner world, the animal can *simulate* a number of different actions in order to 'see' their consequences and evaluate them.[4,15,16] After these simulations, it can choose the most appropriate action to perform in the outer environment.

It seems that many animal species have inner worlds. For example, the searching behaviour of rats is best explained if it is assumed that they have some form of 'spatial maps' in their heads. Evidence for this, based on their abilities to find optimal paths in mazes, was collected by Edward Tolman in the 1930s.[17] It should be noted that it is not assumed that the animal is *aware* of its inner world, nor of the processes utilizing this construct.

In this section, a way of modelling perceptions that is based on ideas from control theory will be presented. This approach has its root in theories of *motor control*, so it is appropriate to begin by a brief description of how the brain controls an action such as throwing. The throw begins in the cerebrum and is guided by the cerebellum. Naturally, the brain receives feedback from the sensors in the muscles of the hand and arm about the direction they are taking during the throw. This feedback is part of what is called proprioception. In control theory, a feedback loop is described as in Figure 1. In this figure the controller corresponds to the brain, the process to the muscles carrying out the action and the measurement to

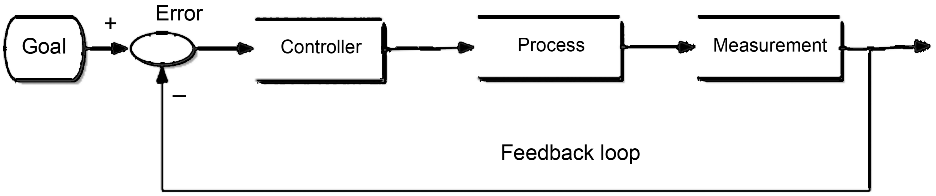

Figure 1. A model of a feedback loop

the signals sent back to the brain (and maybe also the visual perception of the movement).

The feedback signal that comes from the muscles is compared to the goal of the organism and the difference (the error) is then used to adjust the control of the process. The catch is that the signals that come back from the muscles via the nerve fibres are too *slow*. It has been calculated that the loop from the signals that go from the brain to the muscles and back to the brain takes something between 200 and 450 ms.

This is where the big news comes in. A kind of *emulator* has been created in the brain that quickly estimates what the *anticipated* result of the signals to the muscles will be.[4,5,16] (An emulator is a simulator that runs in parallel with the real process.) The signals that leave the motor part of the cortex are sent by the cerebellum to both the arm and the emulator. A calculation is made in the brain of what is about to happen in the arm and the result is sent back to the cerebellum, which adjusts the arm's continuing movement. Richard Grush writes:[4] 'The idea is that in addition to simply engaging with the body and environment, the brain constructs neural circuits that act as models of the body and environment.' The emulators are called *feed-forward models* in control theory since they predict what is going to happen in the real process.

Of course, the emulator should be updated with what has happened to the real process, even if this must occur with some time delay. In control theory, the update is often generated with the aid of *Kalman filters*. The mathematical details will not be presented here. A qualitative description of its function is that the Kalman filter compares its expectation of what the signal should be to what it actually is, and on the basis of the mismatch adjusts its estimate of what state the real process is in. The adjustment is called the Kalman gain. The gain can be weighted in different ways so that when the sensors are expected to be unreliable, the value from the process emulator is given more weight than the sensor signal. In contrast, when the emulator is less precise but sensor information is good, the value provided by the emulator is given less weight than the signal.

In motor control, the calculation loop in the emulator is faster than the loop that goes via the body's muscles. In Figure 2, this is the loop going back from the

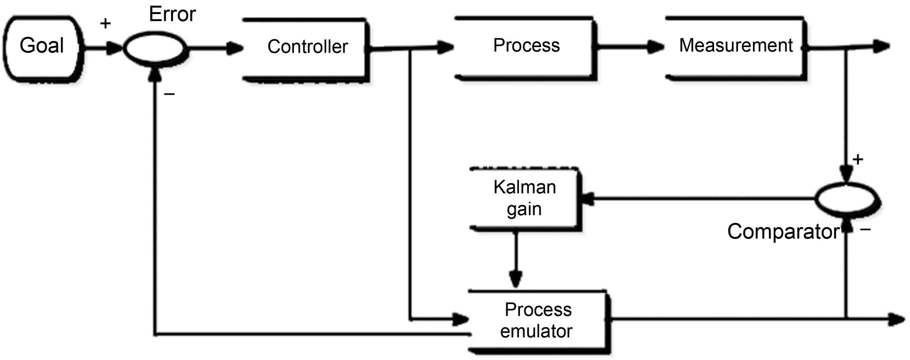

Figure 2. An emulator running in parallel with a control process

process emulator to be compared with the desired value. It has been estimated that the brain can actually correct a signal it has sent to the muscles within 70 ms. That is significantly faster than the 200–450 ms it takes for the signal to reach the muscle and return. So the emulator manages to adjust the arm's movements during a throw much more quickly than can be accomplished with muscular feedback. The faster you want to throw a projectile, the less time you have to steer your arm. Without such an emulator we would never be able to solve the motor control problem involved in the art of aiming. However, we still do not know very much about how the emulator is implemented in the brain.

A general theory must also account for how an emulator can *learn* to control a system. Supposedly, it slowly adjusts its filter settings (and set of variables) on the basis of some form of reward or punishment feedback from the process to be controlled. This would be analogous to how artificial neuron networks learn. Such a form of learning may pick up higher-order correlations between input and output. These correlations may be expressed by the 'hidden' variables of the emulator (see below).

The central hypothesis is that perceptions are generated by emulators of the type described here. Even if it cannot be justified for all of the processes generating perceptions, the assumption is that there are similar complementing mechanisms for many other types of sensations. Perceptions are sensations that are *reinforced* with emulations.

One important property of an emulator is that it does not need to rely exclusively on the signals coming from sense organs: it can also *add on* new types of information that can be useful in simulating. It does not matter much if this information has no direct counterpart in the surrounding world as long as the simulations produce the right result, that is, lead to appropriate actions. In this way the emulator's output will be richer than the influx of sensations.[4,8,16,18]

In particular, the emulators can become more efficient by exploiting 'hidden variables'.

The information provided by these variables is what generates the difference between sensations and perceptions. For example, when an organism observes a moving object, its sensations consist only of the positions of the object, while the *forces* that influence the movement of the object are not sensed. However, if the organism has been able to extract 'force' as a hidden variable and relates this to the sensations via something like Newton's second law, then the system would be able to make more efficient and general, if not more accurate, predictions.

It can be argued that much of our reasoning about *causes* depends on this kind of addition.[8] Interestingly enough, there is a marked difference between humans and other animals in their capacities for causal reasoning. It seems that humans are considerably more skilled at perceiving physical causality. In particular, recent experiments indicate that monkeys and apes have great difficulties understanding the causes of physical events and of mental events.[8,19] Michael Tomasello provides the following explanation for why monkeys and apes cannot understand causal mechanisms and intentionality in others: 'It is just that they do not view the world in terms of intermediate and often hidden "forces," the underlying causes and intentional/mental states, that are so important in human thinking.'[20]

In contrast, even very small human children show strong signs of interpreting the world with the aid of hidden forces and other causal variables. Alison Gopnik claims that 'other animals primarily understand causality in terms of the effects of their own actions on the world. In contrast, human beings combine that understanding with a view that equates the causal power of their own actions and those of objects independent of them.'[21]

My interpretation of this crucial difference between humans and other apes is that the inner emulators we use to understand the physical world are much better at using hidden variables to make predictions about the future. The causes are not part of our sensations, but the emulators that generate our perceptions fill them in. In this way we *perceive* the causes. These causal fillings produce in humans' inner worlds a much more fully formed toolbox for reasoning about the future than exists in other apes.

Levels of inter-subjectivity

Apart from physical forces, the emulators help us perceive the *mental forces* – the emotions, beliefs, desires and intentions – that govern the behaviour of others. Perceiving these forces forms the basis for inter-subjectivity. In this context, inter-subjectivity means *the sharing and understanding of others' mentality.* The term 'mentality' is taken here to involve not only beliefs and other proposition-like entities, but all sorts of forms of consciousness such as emotions, desires, attentional foci and intentions. In the philosophical debate, inter-subjectivity is commonly called having a 'theory of mind'.[8,20,22] (This term should be avoided since it often presumes that one can understand the beliefs of others, something which, on the account presented here, is but one aspect of inter-subjectivity).

The question of whether an animal or a child exhibits inter-subjectivity does not have a simple yes or no answer. In this article, different *levels of inter-subjectivity* will be distinguished in order to understand the cognitive capacities of animals and children at different ages. Implicit in the argument is also the assumption that these levels correspond to different stages in the evolution of animal and human cognition.

A necessary prerequisite for an animal (or a human) to entertain any form of inter-subjectivity is that it has an *inner world*. The crucial issue is whether an individual has any representation of other individuals' inner worlds. To achieve sufficient precision in the analysis of the cognitive abilities of animals and children, the theory of mind will be split into the following four competencies.[8]

(1) Representing the *emotions* of others. At this level one can, for example, understand that someone else is in pain. This is what is usually meant by *empathy*. Even though one can understand others' emotions, it does not mean that one understands what they want or believe.
(2) Representing the *attention* of others. This means that one can understand, for example, what someone else is looking at. However, this ability does not presuppose any conception of other parts of their inner world.
(3) Representing the *intentions* of others. This capacity means, above all, being able to understand the objective that may lie behind another individual's behaviour.
(4) Representing the *beliefs* and *knowledge* of others. This ability involves, among other things, understanding that others don't know the same things as you do.

These four levels grade in a new way what is involved in inter-subjectivity. For comparison, Tomasello distinguishes between three levels of a child's development: (1) seeing others as agents; (2) seeing others as intentional agents; and (3) seeing others as mental agents.[20] His second and third levels correspond roughly to my capacities 3 and 4. Joëlle Proust presents a longer list of capacities that are part of a theory of others' minds, but she does not sort them into levels or discuss how they are related.[28] The account that comes closest is Daniel Stern's theory of infant development where he discusses inter-subjective behaviour as inter-affective, inter-attentional, or inter-intentional.[24] What is missing from his list is that sharing of beliefs is also a form of inter-subjectivity – which may be the most important form for adults.

Representing the emotions of others

Bodily expressions of emotions have a communicative purpose. The expressions are most obvious among social animals. The evolutionary explanation for this seems to be that a capacity for *empathy* leads to greater solidarity within the group. This reduces the risk of violent conflicts in which individuals thoughtlessly injure one another. The resulting solidarity increases the chances of survival for the individuals in the group.

Stephanie Preston and Frans de Waal have argued that as a mechanism involving the linkage of perception and action, a basic form of empathy is available to most, if not all, mammalian species.[25] Defining empathy as 'any process where the attended perception of the object's state generates a state in the subject that is more applicable to the object's state or situation than to the subject's own prior state or situation,' they see a clear evolutionary motivation for its emergence in the ability to recognize and understand the behaviour of con-specifics. It is characteristic that facial signals, such as the famous 'play-face' expression of chimpanzees and gorillas (quite likely an evolutionary precursor to the human smile), typically carry emotional rather than referential meaning.

There is a more advanced form of representing the emotions of others that Preston and de Waal call *cognitive empathy*.[25] This requires a differentiation between subject and object where 'the subject is thought to use perspective-taking processes to imagine or project into the place of the object.' Cognitive empathy develops later in children, at about the same time that they learn to represent the belief and knowledge of others.

Representing the attention of others

Humans are very good at following the *direction* of other people's gazes. Even very young children can understand where other people are looking. In a series of experiments, mothers were asked to sit opposite their infants and look in different directions (sometimes the mother did this while turning her head, sometimes not) and then the observers followed where the children aimed their gaze.[26] At the age of just six months a child can follow its mother's gaze if she turns her head at the same time, and at the age of 12 months it can manage this even if she just moves her eyes. Such young children, however, can only follow their mother's gaze if it falls within their own field of vision. It is not until the age of 18 months that that they can turn to find a point behind them that the mother is looking at. This task is more difficult since it requires that the child can understand that space also continues beyond their field of vision. Following Humphrey, one can say that the child expands its perception from an egocentric to an *allocentric* space.[9]

Chimpanzees are also good at understanding where others are looking. Daniel Povinelli and Timothy Eddy tested this by putting a human opposite a chimpanzee with a transparent screen between them.[19] The apes had no trouble in looking at the spot to which the experimenter was directing his gaze, even if it was behind the chimpanzee. In a variant of the experiment, a screen with an opaque lower half was set up between the human and the chimpanzee. When the experimenter now looked at a point on the opaque part of the screen, the ape would lean forward to try to see what was on the other side. This shows that it is not just the direction of the gaze that the apes follow, but also that they understand that the gaze is directed towards a certain point in the surroundings. Apes *see that others see* the object. This is called second-order attention: I notice that you notice. Apart from chimpanzees, the capacity to follow gazes is well attested for several species of primates, for dogs and for goats.[27]

A more sophisticated form of attention is to succeed in drawing *joint attention* to an object. If I see that you are looking at an object and you see that I see the same object, we have joint attention. This presupposes that both you and I have second-order attention.

Representing the intentions of others

Humans have a powerful inclination to look for *causes* in the world. By reasoning about causes and effects, we become better at predicting the future. The ability to distinguish phenomena caused by other agents (animals or humans) is fundamental to being able to understand intentionality in other individuals' behaviour. When we see that something is caused by an agent, our cognitive system *presumes* that there is some purpose for the act, in other words, that it is intentional. Humans thus find it very easy to create a representation of the objective of an action – we see other people's behaviour as being *goal-directed.* Mostly, of course, it is true that other people's actions are intentional, but sometimes it happens that people do things unintentionally.

It is important to notice that, even though one can interpret someone else's behaviour as goal-directed, this does not necessarily mean that one has any conception of the other's beliefs. It is sufficient that one creates for oneself a representation of the *goal* of the action. This representation exists in the one who interprets the behaviour, and it is far from certain that it agrees with the objective of the person who actually performs the act. The ability to see objectives thus requires a less advanced inner world than the ability to have representations of others' beliefs.

Tomasello and his colleagues have discussed the extent to which infants and apes can represent the intentions of others.[28] One type of experiment involves an adult either holding out a toy in a teasing fashion, or holding it out but dropping it (seemingly) accidentally. Nine-month-old infants, but not 6-month-old infants, showed more impatience when the adult was teasing them than when he was simply dropping the toy, indicating that the infant could understand the difference in intention.

In a similar experiment in chimpanzees, a human began giving food to an ape through a hole in a transparent wall, sometimes refusing to give it to the ape and sometimes attempting to give it to the ape but failing. The result was that chimpanzees gestured more and left the area earlier when the human was unwilling than when he was unable. In the latter case they tended to wait patiently throughout the unsuccessful attempts. The interpretation is that the chimpanzees understood the behaviour of the human in the 'unable condition' as an intention to give them food.

This experiment indicates that chimpanzees have some representation of the intentions of others. A limitation is that the experiment only concerns a basic test of whether an action is intentional or not and it does not show that the apes can represent the *contents* of a specific intention. However, it seems difficult to devise an adequate methodology for testing this, so we will have to wait for the development of new experiments, before it can be judged to what extent other species understand the intentions of others.

Representing the beliefs and knowledge of others

Experiments have been performed to test whether apes and monkeys can represent the beliefs of others. Most such experiments focus on whether they understand that 'seeing is knowing'. So far the outcomes are negative and therefore there is nothing to suggest that primates have representations of others' beliefs and knowledge.

It is easier to test whether young children can understand that 'seeing is knowing', since one can communicate with them through language from a fairly early age. In one experiment, children aged between three and five were asked whether another person knew what was in a box. Some of the people had looked inside the box while the others had not. In the same way, some of the children were allowed to see what was in the box while the others were not. None of the children had any problem saying whether a particular person had looked in the box or not. But when the question concerned whether the person knew what was in the box, the three-year-olds consistently replied that the person knew if they themselves had seen what was in the box, and that the person did not know if they had not seen what was in the box. Whether the person had looked or not made no difference. The four- and five-year-olds, on the other hand, were able to connect 'seeing' with 'knowing'. This is one of several experiments that suggest that a representation of others' beliefs and knowledge develops in humans at the age of about four.

A common type of test of children's understanding of other people's beliefs concerns whether they can understand that someone else has a *false* belief about what the world is like.[29–31] The best known of these experiments was about a Smarties tube. The test was carried out on children aged between three and five. The children were first shown the tube and then asked what they thought was in it. All the children replied 'Smarties' (or 'sweets'). When the tube was opened it was found to contain a pencil. The tube was then closed. The children were then asked what a friend, who has not yet seen what is in the tube, will say that it contains. The three-year-olds generally answer 'pencil' whereas most of the older children say 'Smarties'. The older children understand that the friend *does not have the same knowledge* as they do. They thus realize that the friend has a false belief about what is in the tube. This is a clear example that they represent the belief of others. The younger children, on the other hand, do not appear to be able to distinguish between their own beliefs and other people's.[31] Variations of the false belief tasks have also been performed with chimpanzees, but so far there is no indication that they understand the beliefs of others.

On the control theory of inter-subjectivity

Most of the modelling involving control theory that was outlined earlier concerned motor control. However, there have been some recent attempts to extend this kind of modelling to social cognition and inter-subjectivity. The analogy with motor control is clearly stated by Daniel Wolpert, Kenji Doya and Mitsuo Kawato:

> Motor control is [...] concerned with inputs and outputs from a controlled object (e.g. the arm) that is part of our own body. When interacting with another person we can think of an analogous social interaction loop in which the controlled object is the other person rather than part of our own body [...]. Again, our motor commands cause muscle contractions and these lead to motor consequences which generate communicative signals, such as speech or gestures. When perceived by another person these can have influences on their hidden (mental) state, which constitutes the set of parameters that determine their behaviour. We can regard the other person as having a state in the same way that our own body has a state. If we know the state of someone else and have a model of their behaviour, we should be able to predict their response to a given input that we or the environment provides. Given the other person's state, the motor command we have generated, and the context provided by the environment, the other person will generate motor commands causing consequences. We can perceive these

> consequences and these can be used to determine our next motor command, thereby closing a social interaction loop.[7]

On the basis of this general analogy, the task is now to model the different levels of inter-subjectivity. First, consider empathy; that is, representing the emotions of others. For this form of inter-subjectivity, it seems that the machinery of control theory will not be required, since many emotions are so clearly correlated with more or less fixed facial and bodily expressions. A simple form of associative learning, connecting the expressions of others with your own emotions, could therefore be sufficient to explain this form of inter-subjectivity. It is even possible that some aspects of this connection are genetically grounded. In support of this position, Jeannerod writes: 'Human facial expressions are visual patterns shaped by evolution and [...] are likely to be directly understandable by human beings without an obvious need for simulation.'[12]

Secondly, consider representing the attention of others. From a modelling point of view, the important fact to note is that this involves a *shift of perspective*.[32] For example, when a child follows the gaze of somebody that focuses on a point outside its own visual field, it has to form an allocentric representation of the space. In general, this can be modelled via a *coordinate transformation* together with an unlimited extension of the space. A general and flexible solution is to *add* an allocentric representation of space to the egocentric one provided by the visual system. Furthermore, such an allocentric representation could be generated (and exploited) by other sensory systems involved in spatial representation, that is, touch and auditory information. Thus, the most efficient solution would be to have a *cross-modal* allocentric representation of space. As we saw earlier, such an allocentric space for representing the attention of others is not functioning at birth, but develops over the first 18 months.

Modelling this in control theory would involve a system with two emulators:[4] one for egocentric (visual) space and one for allocentric cross-modal space (see Figure 3). The output from the two emulators can be combined with each other to improve the updating of both emulators. One important advantage of a cross-modal emulator is that it can be used to represent states that are not directly available to the egocentric process emulator. For example, it can maintain the position of an object as the organism turns around and temporarily loses it out of sight. This capacity is central for the phenomenon of *object permanence* studied by Piaget.

Furthermore, a cross-modal emulator can be used with more than one modality specific emulator, for example an emulator for space based on hearing. On this possibility, Grush writes:

> In such a case the [cross]modal emulator would be subject to two different 'measurements', a visual measurement, yielding an expectation of what should

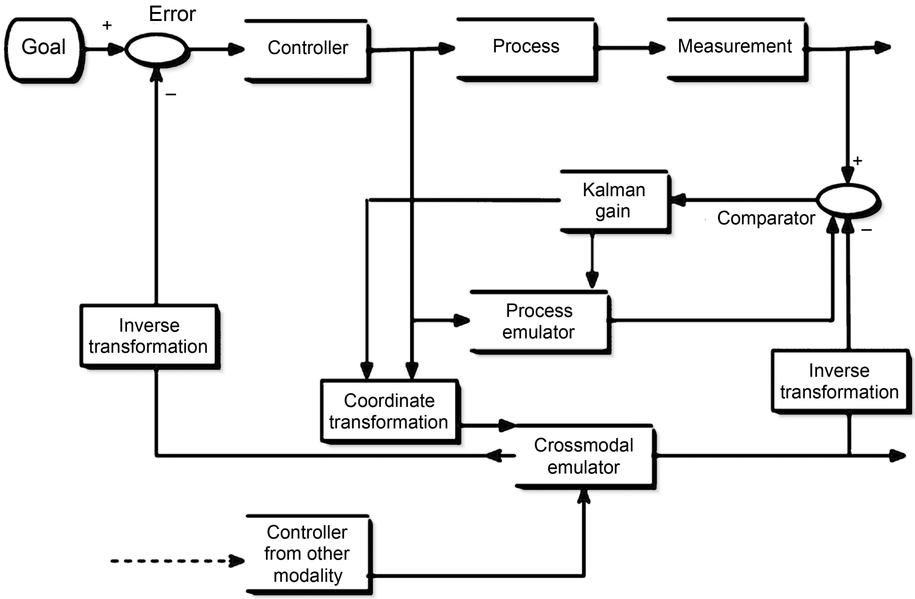

Figure 3. A combination of a modal and a cross-modal emulator

> be *seen* given the current estimate of the state of the environment, and an auditory measurement yielding an expectation of what should be *heard* given the current estimate of the state of the environment. And the [cross]modal emulator would be updated by both sensory residuals, resulting in a state estimate that effectively integrates information from all modalities as well as *a priori* estimates of the state of the environment.[4]

One important point of this model is that the allocentric space needed to understand the attention of others is something that is *constructed* in the brain by the cross-modal emulator. As Grush writes: 'Space is a theoretical posit of the nervous system, made in order to render intelligible the multitude of interdependencies between the many motor pathways going out, and the many forms of sensory information coming in. Space is not spoon-fed to the cognizer, but is an achievement.'[16]

Thirdly, consider representing the intentions of others. On this level, the mapping between the mental variables, in this case the intentions, and their behavioural correlates, becomes more complicated. When it comes to simple motor actions, such as drinking from a glass, my movements may give sufficient clues to the underlying intention,[7] but for more advanced intentions, such as buying a car, my movements are far from sufficient to infer my intention. This makes modelling in terms of control theory more difficult. However, there are some recent attempts that focus on representing the intentions of motor actions.[7]

As an addition to having a cross-modal emulator representing allocentric space, it would be useful to have a cross-modal emulator representing *forces*. Even though our cognitive system may not be built precisely for Newtonian mechanics, our brains are likely to be constructed for extracting the forces that lie behind different kinds of movements and action (see below). In accordance with this, the hypothesis is that the fundamental cognitive representation of an action consists of the pattern of forces that generates it.[33] However, it should be emphasized that the 'forces' represented by the brain are psychological constructs and not the scientific dimension introduced by Newton. Again, the forces would appear as 'hidden' variables generated by the proposed cross-modal emulator.

In support of the representational hypothesis, the experiments by Gunnar Johansson and his followers suggest that the kinematics of a movement contains sufficient information for identifying the underlying *dynamic force patterns*.[34] Sverker Runesson claims that we can directly *perceive* the forces that control different kinds of motion.[35] This is reasonable, if 'perceiving' is interpreted as including the hidden forces added by the proposed type of emulator. However, when it comes to more advanced actions than motor control, physical forces may not be enough to represent intentions, but proper 'mental' forces must be added by the emulators. Wolpert *et al.*[7] present some ideas about how this can be achieved in a control theory model. It remains to be seen how far this kind of modelling can be extended outside the realm of motor control.

Finally, consider the problem of representing the beliefs and knowledge of others. Here, the situation becomes very complex and it is far from clear how this can be handled with the aid of control theory. Again, Wolpert *et al.*[7] provide an analysis of how different *contexts* will affect motor control. If you are about to lift a teapot, your motor control will depend on whether you believe the teapot to be full or empty. However, if one wants to take the beliefs of others into account when deciding what to do, the modelling problems become severe and the corresponding control system will involve several layers of emulators that can provide even more kinds of hidden variables representing the beliefs and knowledge of others.

It is clear that when modelling the higher levels of inter-subjectivity with the aid of control theory, methodological issues become pressing. The number of variables and the level of complexity will be so high that it will become difficult to evaluate a model. Wolpert *et al.* present some of the problems:

> When considering the social interaction loop and regarding another person as the controlled object, we encounter similar, but usually more severe, problems. First, the time delays between our action and the consequences on our own body are of the order of hundreds of milliseconds, whereas with other people the consequences can be of the order of seconds to minutes or even days. Moreover, the response of a person to our actions is not easily predicted. There is usually

> a complex, noisy and nonlinear relationship between our actions and the consequences. In a similar way to the nonlinearity of the arm, knowing how someone will respond to two separate actions we perform does not allow us to predict accurately.[7]

In spite of these cautious remarks, looking upon representing the beliefs of others as a problem of control theory may generate new hypotheses that, via feed-forward models of the kind presented here, can lead to testable predictions.

Conclusion

We read the minds of others because we want to influence them by communication, collaboration or deception. In this article it has been outlined how feed-forward models from control theory can be used to analyse different kinds of mind-reading. Mind-reading – here called inter-subjectivity – has been divided into representing the emotions, the attention, the intentions and the beliefs and knowledge of others. Empathy – representing the emotions of others – arguably can be handled as an effect of associations, which would not require the apparatus of control theory. Representing the attention of others involves primarily shifting one's perspective on the world from an egocentric space to an allocentric one. For this shift and for the guidance of attention, the models from control theory are well suited. When it comes to representing intentions of others and their beliefs and knowledge, the feed-forward models will have to posit a range of 'hidden variables,' representing intentions, beliefs and knowledge, in order to emulate the relevant forms of inter-subjective processes. Some attempts to model this can be found in the literature, but for this programme to succeed it requires much more collaboration between philosophers, cognitive scientists and engineers.

References

1. K. H. Pribram (1991) *Brain and Perception: Holonomy and Structure in Figural Processing* (Hillsdale, NJ: Lawrence Erlbaum Associates).
2. D. Hofstadter (1979) *Gödel, Escher, Bach: an Eternal Golden Braid* (New York: Basic Books).
3. G. M. Edelman (1987) *Neural Darwinism: The Theory of Neuronal Group Selection* (New York: Basic Books).
4. R. Grush (2004) The emulator theory of representation: motor control, imagery and perception. *Behavioral and Brain Sciences*, **27**, 377–442.
5. M. Kawato (1999) Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, **9**, 718–727.
6. W. Singer (1999) The observer in the brain, in: A. Riegler, M. Peschl, and A. von Stein (Eds) *Understanding Representation in the Cognitive Sciences* (New York: Kluwer Academic/Plenum Publishers), pp. 253–256.

7. D. Wolpert, K. Doya and M. Kawato (2003) A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London*, **358**, 593–602.
8. P. Gärdenfors (2003) *How Homo Became Sapiens* (Oxford: Oxford University Press).
9. N. K. Humphrey (1993) *A History of the Mind* (London: Vintage Books).
10. W. James (1890) *The Principles of Psychology* (New York: Henry Holt).
11. M. Jeannerod (1994) The representing brain, neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, **17**, 187–202.
12. M. Jeannerod (2006) *Motor Cognition* (Oxford: Oxford University Press).
13. A. M. Glenberg (1997) What memory is for. *Behavioral and Brain Sciences*, **20**, 1–19.
14. K. Craik (1943) *The Nature of Explanation* (Cambridge: Cambridge University Press).
15. L. W. Barsalou (1999) Perceptual symbol systems. *Behavioral and Brain Sciences*, **22**, 577–609.
16. R. Grush (1998) Wahrnehmung, Vorstellung und die sensomotorische Schleife', in: F. Esken and H.-D. Heckmann (Eds) *Bewußtsein und Repräsentation* (Paderborn: Verlag Ferdinand Schoeningh).
17. E. C. Tolman (1948) Cognitive maps in rats and men. *Psychological Review*, **55**, 189–208.
18. G. Hesslow (2002) Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, **6**(6), 242–247.
19. D. Povinelli (2000) *Folk Physics for Apes: The Chimpanzee's Theory of How the World Works* (Oxford: Oxford University Press).
20. M. Tomasello (1999) *The Cultural Origins of Human Cognition* (Cambridge, MA: Harvard University Press).
21. A. Gopnik (1998) Explanation as orgasm, *Minds and Machines*, **8**, 101–118.
22. D. Premack and G. Woodruff (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, **4**, 515–526.
23. J. Proust (1998) Can nonhuman primates read minds?, *Rapport no. 9807* (Paris: CREA).
24. D. N. Stern (1985) *The Interpersonal World of the Infant* (New York: Basic Books).
25. S. Preston and F. de Waal (2002) Empathy: Its ultimate and proximal bases. *Behavioral and Brain Sciences*, **25**, 1–72.
26. G. Butterworth and N. L. M. Jarret (1991) What minds share in common is space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, **9**, 55–72.
27. J. Kaminski, J. Riedel, J. Call and M. Tomasello (2005) Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object task. *Animal Behaviour*, **69**, 11–18.
28. M. Tomasello, M. Carpenter, J. Call, T. Behne and H. Moll (2005) Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, **28**, 675–691.

29. A. Gopnik and J. W. Astington (1988) Children's understanding of representational change, and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, **59**, 26–37.
30. P. Mitchell (1997) *Introduction to Theory of Mind: Children, Autism and Apes* (London: Arnold).
31. J. Perner, S. Leekam and H. Wimmer (1987) Three-year-old's difficulty with false belief: the case for a conceptual deficit. *British Journal of Developmental Psychology*, **5**, 125–137.
32. E. Oztop, D. Wolpert and M. Kawato (2005) Mental state inference using visual control parameters. *Cognitive Brain Research*, **22**, 129–151.
33. P. Gärdenfors (to appear) Representing actions and functional properties in conceptual spaces, in: T. Ziemke and J. Zlatev (Eds) *Body, Language and Mind* (Amsterdam: Elsevier).
34. G. Johansson (1973) Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, **14**, 201–211.
35. S. Runesson (1994) Perception of biological motion: the KSD-principle and the implications of a distal versus proximal approach, in: G. Jansson, S.-S. Bergström and W. Epstein (Eds) *Perceiving Events and Objects*. (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 383–405.

About the author

Peter Gärdenfors is Professor of Cognitive Science at Lund University. His main research interests are concept formation and the evolution of cognition. His published books include *Knowledge in Flux* (1988), *Conceptual Spaces* (2000), *How Homo Became Sapiens* (2003) and *The Dynamics of Thought* (2005).