WILEY
Publishers Since 1807

THE ATRIUM, SOUTHERN GATE, CHICHESTER, WEST SUSSEX P019 8SQ

*****IMMEDIATE RESPONSE REQUIRED*****

Your article may be published online via Wiley's EarlyView® service (http://www.interscience.wiley.com/) shortly after receipt of corrections. EarlyView® is Wiley's online publication of individual articles in full-text HTML and/or pdf format before release of the compiled print issue of the journal. Articles posted online in EarlyView® are peer-reviewed, copy-edited, author-corrected, and fully citable via the article DOI (for further information, visit www.doi.org). EarlyView® means you benefit from the best of two worlds - fast online availability as well as traditional, issue-based archiving.

Please follow these instructions to avoid delay of publication

□ **READ PROOFS CAREFULLY**

- This will be your only chance to review these proofs**. Please note that once your corrected article is posted online, it is considered legally published, and cannot be removed from the Web site for further corrections.**
- Please note that the volume and page numbers shown on the proofs are for position only.

□ **ANSWER ALL QUERIES ON PROOFS** (Queries for you to answer are attached as the last page of your proof.)

- List all corrections and send back via e-mail to the production contact as detailed in the covering e-mail, or mark all corrections directly on the proofs and send the scanned copy via e-mail. Please do not send corrections by fax or in the post.

□ **CHECK FIGURES AND TABLES CAREFULLY**

- Check size, numbering, and orientation of figures.
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. These images will appear at higher resolution and sharpness in the printed article.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

□ **COMPLETE CTA (if you have not already signed one)**

- Please send a scanned copy with your proofs and post your completed original form to the address detailed in the covering e-mail. **We cannot publish your paper until we receive the original signed form.**

□ **OFFPRINTS**

- 25 complimentary offprints of your article will be dispatched on publication. Please ensure that the correspondence address on your proofs is correct for dispatch of the offprints. If your delivery address has changed, please inform the production contact to the journal - details in the covering e-mail. Please allow six weeks for delivery.

Additional reprint and journal issue purchases

- Additional **paper reprints** (minimum quantity 100 copies) are available on publication to contributors. Quotations may be requested from mailto:author_reprints@wiley.co.uk. Orders for additional paper reprints may be placed in advance in order to ensure that they are fulfilled in a timely manner on publication of the article in question. Please note that offprints and reprints will be dispatched under separate cover.
- PDF files of individual articles may be purchased for personal use for $25 via Wiley's Pay-Per-View service (see http://www3.interscience.wiley.com/aboutus/ppv-articleselect.html).
- Please note that regardless of the form in which they are acquired, reprints should not be resold, nor further disseminated in electronic or print form, nor deployed in part or in whole in any marketing, promotional or educational contexts without further discussion with Wiley. Permissions requests should be directed to mailto:permreq@wiley.co.uk
- Lead authors are cordially invited to remind their co-authors that the reprint opportunities detailed above are also available to them.
- If you wish to purchase print copies of the issue in which your article appears, please contact our Journals Fulfilment Department mailto:cs-journals@wiley.co.uk when you receive your complimentary offprints or when your article is published online in an issue. Please quote the Volume/Issue in which your article appears.

ICD : 541
pp.1–18 (col.fig.: NIL)
PROD.TYPE: COM
3B2
ED: VIJAYA
PAGN: KG.USHAMANI SCAN: XXX

Infant and Child Development
Inf. Child Dev. **15**: 000–000 (2006)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/icd.541

Shared Challenges in Object Perception for Robots and Infants

Paul Fitzpatrick[a,d,*], Amy Needham[b], Lorenzo Natale[a,c,d] and Giorgio Metta[a,d]

[a] *LIRA-Lab, DIST, University of Genova, Genova, Italy*
[b] *Duke University, Durham, NC, USA*
[c] *MIT CSAIL, Cambridge, MA, USA*
[d] *Italian Institute of Technology, Genova, Italy*

Robots and humans receive partial, fragmentary hints about the world's state through their respective sensors. These hints—tiny patches of light intensity, frequency components of sound, etc.—are far removed from the world of objects which we feel and perceive so effortlessly around us. The study of infant development and the construction of robots are both deeply concerned with how this apparent gap between the world and our experience of it is bridged. In this paper, we focus on some fundamental problems in perception which have attracted the attention of researchers in both robotics and infant development. Our goal was to identify points of contact already existing between the two fields, and also important questions identified in one field that could fruitfully be addressed in the other. We start with the problem of object segregation: how do infants and robots determine visually where one object ends and another begins? For object segregation, both the fields have examined the idea of using 'key events' where perception is in some way simplified and the infant or robot acquires knowledge that can be exploited at other times. We propose that the identification of the key events themselves constitutes a point of contact between the fields. Although the specific algorithms used in robots do not necessarily map directly to infant strategies, the overall 'algorithmic skeleton' formed by the set of algorithms needed to identify and exploit key events may in fact form the basis for mutual dialogue. We then look more broadly at the role of embodiment in humans and robots, and see the opportunities it affords for development. Copyright © 2007 John Wiley & Sons, Ltd.

Key words: infant development; robotics; object segregation; intermodal integration; embodiment; active perception

*Correspondence to: Paul Fitzpatrick, LIRA-Lab, DIST, University of Genova, Viale F. Causa 13, 16145 Genova, Italy. E-mail: paulfitz@liralab.it

INTRODUCTION

Imagine if your body's sensory experience were presented to you as column after column of numbers. One number might represent the amount of light hitting a particular photoreceptor, another might be related to the pressure on a tiny patch of skin. Imagine further that you can only control your body by putting numbers in a spreadsheet, with different numbers controlling different muscles and organs in different ways.

This is how a robot experiences the world. It is also a (crude) *model* of how humans experience the world. Of course, our sensing and actuation are not encoded as numbers in the same sense, but aspects of the world and our bodies are transformed to and from internal signals that, in themselves, bear no trace of the signals' origin. For example, a neuron firing selectively to a red stimulus is not itself necessarily red. In telepresence applications (Steuer, 1992), this model becomes literal, with an interface of numbers lying between the human operator and a remote environment. Understanding how to build robots requires understanding in detail how it is possible to sense and respond to the world, in terms of an interface of numbers representing sensor readings and actuator settings rather than symbolic descriptions of the world.

How closely does this match the concerns of psychology? In works concerned with modelling phenomena deeply rooted in culture, history, and biology, connections may exist at a rather high-level abstraction—for example, one can investigate theories of how language evolves in a group (Steels, 1997). In works concerned with immediate perception of the environment, we believe that there is a value in forging connections at a detailed level. We expect that there will be commonality between how infants and successful robots operate at the information-processing level, given the common constraints imposed and opportunities afforded by the physical world they share. For example, natural environments are of mixed, inconstant observability—there are properties of the environment that can be perceived easily under some circumstances and with great difficulty (or not at all) under others. This network of opportunities and frustrations should place limits on information processing that applies both to infants and robots with human-like sensors.

In this paper, we focus on early perceptual development in infants. The perceptual judgements infants make may change over time, showing an evolving sensitivity to various cues. This progression may be at least partially due to knowledge gained from experience. We identify opportunities that can be exploited by both infants and robots to perceive properties of their environment that cannot be directly perceived in other circumstances. We review some of what is known about how robots and infants can exploit such opportunities to learn how object properties not directly given in the display correlate with observable properties. The topics we focus on are object segregation and intermodal integration. In the last section, we discuss the role of the body for perception and how this contributes to creating points of contacts between the two fields.

OBJECT SEGREGATION

The world around us has a structure, and to an adult it appears to be made up of more-or-less well-defined objects. Perceiving the world this way sounds trivial,

DOI: 10.1002/icd

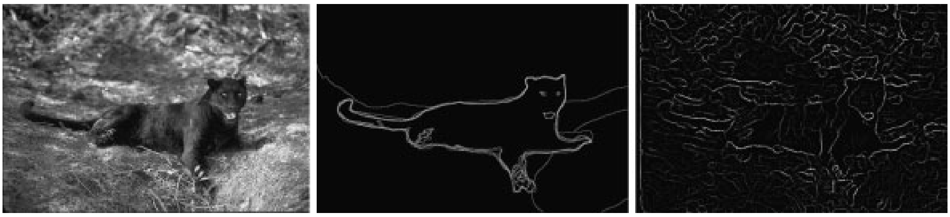

Figure 1. An example from Martin, Fowlkes, and Malik (2004) to highlight the difficulties of bottom-up segmentation. For the image shown on the left, humans see the definite boundaries shown in white in the middle image. The best machine segmentation of a set of algorithms gives the result shown on the right—a mess. This seems a very difficult scene to segment without having some training at least for the specific kinds of materials in the scene.

but from an engineering perspective, it is heart-breakingly complex. As Spelke wrote in 1990:

> ... the ability to organize unexpected, cluttered, and changing arrays into objects is mysterious: so mysterious that no existing mechanical vision system can accomplish this task in any general manner. (Spelke, 1990)

This is still true today. This ability to assign boundaries to objects in visually presented scenes (called 'object segregation' in psychology or 'object segmentation' in engineering) cannot yet be successfully automated for arbitrary object sets in unconstrained environments (see Figure 1). On the engineering part, there has been some *algorithmic* progress; for example, given local measures of similarity between each neighbouring element of a visual scene, a globally appropriate set of boundaries can be inferred in efficient and well-founded ways (see, for example, Felzenszwalb & Huttenlocher, 2004; Shi & Malik, 2000). There is also a growing awareness of the importance of collecting and exploiting empirical *knowledge* about the statistical combinations of materials, shapes, lighting, and viewpoints that actually occur in our world (see, for example, Martin *et al.*, 2004). Of course, such knowledge can only be captured and used effectively because of algorithmic advances in machine learning, but the knowledge itself is not specified by an algorithm. Empirical, non-algorithmic knowledge of this kind now plays a key role in machine perception tasks of all sorts. For example, face detection took a step forward with Viola and Jone (2004); the success of this work was due to both algorithmic innovation and better exploitation of knowledge (features learned from 5000 hand-labelled face examples). Automatic speech recognition is successful largely because of the collection and exploitation of extensive corpuses of clearly labelled phoneme or phoneme-pair examples that cover well the domain of utterances to be recognized. These two examples clarify the ways the 'knowledge' can play a role in machine perception. The bulk of the 'knowledge' used in such systems takes the form of *labelled examples*—examples of input from the sensors (a vector of numbers) and the corresponding desired output interpretation (another vector of numbers). More-or-less general purpose machine-learning algorithms can then approximate the mapping from sensor input to desired output interpretation based on the examples (called the *training set*), and apply that approximation to novel situations (called the *test set*). Generally, this approximation will be very poor unless we transform the sensory input in a manner that highlights properties that the programmer believes may be relevant. This transformation is called *preprocessing and feature selection*. This transformation and a corresponding transformation that applies the results of the learning system back to the original problem together make up a very important part of the full system. This 'infrastructure' is often downplayed or not reported. For this paper, we will group all this infrastructure and call it the *algorithmic skeleton*. This set of carefully interlocking algorithms is designed so that, when fed with appropriate training data, it produces a functional system. Without the algorithmic skeleton, there would be no way to make sense of the training data, and without the data, perception would be crude and uninformed.

The algorithmic skeleton, seen as a set of choices about preprocessing and feature selection, gives a specific bias to the final performance of the interlocking algorithms. With it, the designer guides the learning system towards an interpretation of data likely to be appropriate for the domain in which the system will find itself. Clearly, this is also a crucial point where informed choices can be made starting from infant studies or from neuroscience evidence. These biases and choices are 'knowledge' that is just as important as the data that come from the specific interaction of the learning machine with the environment. An ongoing research goal is to maximize the amount that a system can learn with the minimum of hand-designed bias (Bell & Sejnowski, 1997; Simoncelli & Olshausen, 2001). This generally means adding algorithms to infer extra parameters from data rather than setting them from human judgement. This can seem a little confusing, since in the quest to reduce the need for designer bias, we actually increase designer effort—the designer is now adding complex algorithms rather than picking a few numbers. What is really happening is that bias is not being removed, but rather moved to a higher level of abstraction. This is very valuable because it can greatly increase the number of situations in which a fixed algorithmic skeleton can be successfully applied.

What, then, is a good algorithmic skeleton for object segregation? What set of algorithms, coupled with what kind of training data, would lead to best performance? We review suggestive results in both infant development research and robotics.

Segregation Skills in Infants

By 4–5 months of age, infants can visually parse simple displays like the one in Figure 2 into units, based on something like a subset of static Gestalt principles—see, for example, Needham (1998, 2000). Initial studies indicated that infants use a collection of features to parse the displays (Needham & Baillargeon, 1997, 1998; Needham, 1998); subsequent studies suggested that object shape is the key feature that young infants use to identify boundaries between adjacent objects (Needham, 1999). Compared with adult judgements, we would expect such strategies to lead to many incorrect parsings, but they will also provide reasonable best guess interpretations of uniform objects in complex displays.

Infants do not come prepared from birth to segregate objects into units that match adult judgement. It appears that infants learn over time how object features can be used to predict object boundaries. More than 20 years ago, Kellman and Spelke (1983) suggested that infants may be born with knowledge about solid, three-dimensional objects and that this knowledge could help them

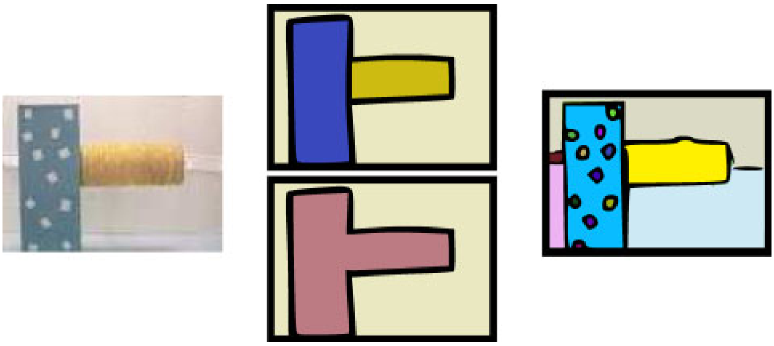

Figure 2. Object segregation is not necessarily well defined. On the left, there is a simple scenario, taken from Needham (2001), showing a rectangle attached to a yellow tube. Two plausible ways to segregate this scene are shown in the middle, depending on whether the tube and rectangle make up a single object. For comparison, automatically acquired boundaries are shown on the right, produced using the algorithm in Felzenszwalb and Huttenlocher (2004). This algorithm does *image segmentation*, seeking to produce regions that correspond to whole objects (such as the yellow tube) or at least to object parts (such all the blue rectangle and all the small white patches on its surface, and various parts of the background). Ideally, regions that extend across object boundaries are avoided. Image segmentation is less ambitious than object segregation, and allows context information to be factored in as a higher level process operating on a region level rather than pixel level.

interpret portions of a moving object as connected to other portions that were moving in unison. This assertion was put to the test by Slater and his colleagues (Slater *et al.*, 1990), a test that resulted in a new conception of the neonate's visual world. Rather than interpreting common motion as a cue to object unity, neonates appeared to interpret the visible portions of a partly occluded object as clearly separate from each other, even when undergoing common motion. This finding was important because it revealed one way in which learning likely changes how infants interpret their visual world.

Although segregating adjacent objects present a very similar kind of perceptual problem ('are these surfaces connected or not'), the critical components of success might be quite different. Early work with adjacent objects indicated that at 3 months of age, infants tend to group all touching surfaces into a single unit (Kestenbaum, Termine, & Spelke, 1987). Subsequent experiments have revealed that soon after this point in development, infants begin to analyse the perceptual differences between adjacent surfaces and segregate surfaces with different features (but not those with similar features) into separate units (Needham, 2000). Although infants can use the boundary seam between two objects as a source of information about the likely separation between them (Kaufman & Needham, 1999), other work comparing boundary-occluded and fully visible versions of the same displays suggests that boundary information is not the only information infants use to parse the objects in a display (Needham, 1998). Still later, 8.5-month-old infants have shown to also use information about specific objects or classes of objects to guide their judgement (Needham, Cantlon, & Holley, 2006).

It might be that extensive amounts of experience are required to 'train up' this system. However, it might also be that infants learn on the basis of relatively few exposures to key events (Baillargeon, 1999). This possibility was investigated

DOI: 10.1002/icd

within the context of object segregation by asking how infants' parsing of a display would be altered by a brief prior exposure to one of the objects in the test display.

In this paradigm, a test display was used that was known to be ambiguous to 4.5-month-old infants. Infants were given a prior experience that could help disambiguate the test display. This prior experience consisted of a brief exposure (visual only) to a portion of the test display. If infants used this prior experience to help them interpret the test display, they should see the display as two separate objects rather than a single aggregate. In that case, they should look reliably longer when the objects moved as a single unit (unexpected) than when they move separately (expected). If, however, the prior experience was ineffective in altering infants' interpretation of the display, their behaviour should be similar to the infants in the initial study with no particular prior experience (Needham & Baillargeon, 1998). In fact, prior experiences with either portion of the test display turned out to be effective in facilitating infants' parsing of the test display.

Segregation Skills in Robots

This idea that *exposure to key events* could influence segregation is intuitive, and evidently operative in infants. Yet, it is not generally studied or used in mechanical systems for object segregation. In this section, we attempt to reformulate robotics work by the authors in these terms. For object segregation in robotics, we will interpret 'key events' as moments in the robot's experience where the true boundary of an object can be reliably inferred. They offer an opportunity to determine *correlates* of the boundary that can be detected outside of the limited context of the key events themselves. Thus, with an appropriate algorithmic skeleton, information learned during key events can be applied more broadly. Key events used by infants include seeing an object in isolation or seeing objects in relative motion, as discussed in segregation skills in infants section. In the authors' work, algorithmic skeletons have been developed for exploiting constrained analogues of these situations.

Natale, Orabona, Metta, and Sandini (2005) used a very simple key event to learn about objects—*holding an object up to the face*. The robot can be handed an object or happen to grasp it, and will then hold it up close to its cameras. This gives a good view of its surface features, allowing the robot to do some learning and later correctly segregate the object out from the background visually even when out of its grasp (see Figure 3). This is similar to an isolated presentation of an object, as in Needham's experiments. In real environments, true isolation is very unlikely, and actively moving an object so that it dominates the scene can be beneficial. Fitzpatrick and Metta (2003) used the 'key event' *hitting an object with the hand/arm*. This is a constrained form of relative object motion. In the real world, all sorts of strange motions happen which can be hard to parse, so it is simpler at least to begin with to focus on situations the robot can initiate and at least partially control. Motion caused by body impact has some technical advantages; the impactor (the arm) is modelled and can be tracked, and since the moment and place of impact can be detected quite precisely, unrelated motion in the scene can be largely filtered out.

The *algorithmic skeleton* by Fitzpatrick and Metta (2003) processes views of the arm moving, detects collisions of objects with the arm, and outputs boundary estimates of whatever the arm collides with based on a motion cue. These boundaries, and what they contain, are used as training data for another

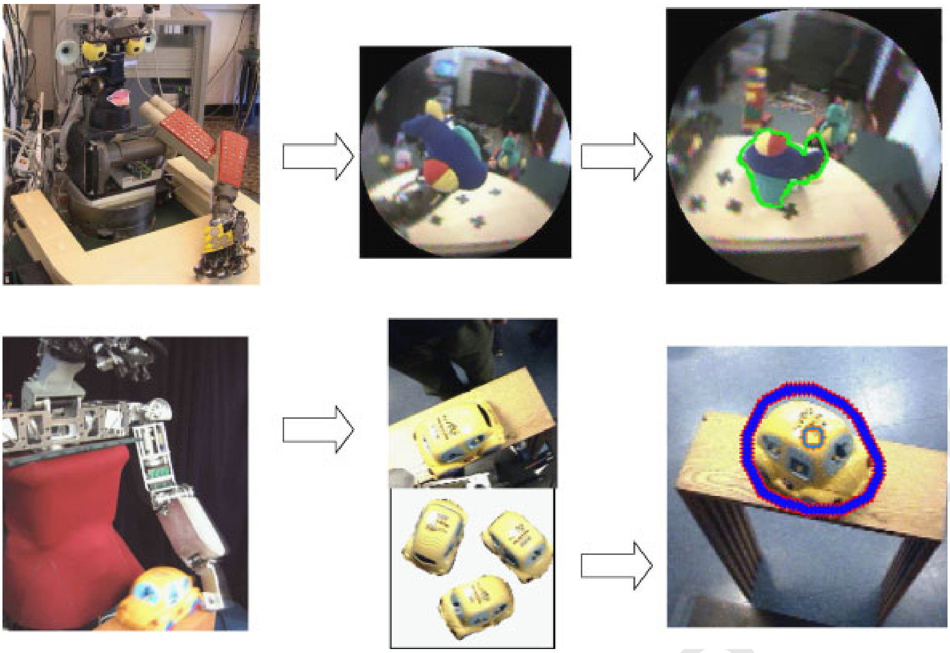

Figure 3. The upper row shows object segregation by the robot 'Babybot' based on prior experience. The robot explores the visual appearances of an object that it has grasped; the information collected in this way is used later on to segment the object (Natale *et al.*, 2005). Left: the robot. Middle: the robot's view when holding up an object. Right: later segmentation of the object. The lower row shows the robot 'Cog' detecting object boundaries experimentally by poking (Fitzpatrick, 2003). During object motion, it finds features of the object that contrast with other objects, and that are stable with respect to certain geometric transformations. These features are then used to jointly detect and segment the object in future views. Left: the robot. Middle: segmentations of a poked object. Right: later segmentation of the object on a similarly coloured table.

algorithm, whose purpose is to estimate boundaries from visual appearance when motion information is not available. See Fitzpatrick (2003) for technical details. As a basic overview, the classes of algorithms involved are as follows:

1. *Behaviour system*: An algorithm that drives the robot's behaviour, so that it is likely to hit things. This specific, rather idiosyncratic goal is chosen in order to enable a broader set of outcomes.
2. *Key event detection*: An algorithm that detects the event of interest, in this case when the arm/hand hits an object.
3. *Training data extraction*: An algorithm that can, within the specific context of the key event, extract boundary information—in this case using object motion caused by hitting.
4. *Machine learning*: An algorithm that uses the training data to identify features that are predictive of boundaries and which can be extracted in other situations outside the key event (for example, edge and color combinations).
5. *Application of learning*: An algorithm that actually uses those features to predict boundaries. This must be integrated with the very first algorithm, to influence

the robot's behaviour in useful ways. In terms of observable behaviour, the robot's ability to attend and fixate specific objects increases, since they become segregated from the background.

This skeleton gives the robot an initial behaviour that changes during learning, once the robot actually starts hitting objects and extracting specific features predictive of the boundaries of specific objects. A set of different algorithms performing analogous roles are given by Natale *et al.* (2005). Fitzpatrick and Metta (2003) used a very specific condition (objects being hit by people or the robot itself) to extract good motion-based object boundaries; surface features of the object could then be used to segregate that object out in static presentations (Fitzpatrick, 2003). Arsenio and Fitzpatrick (2005) used rhythmic motion of objects to segment their boundaries both visually and acoustically. Arsenio and Fitzpatrick (2005) developed a set of techniques for acquiring all sorts of segmentations. Some methods work for small, grasp-size objects, others work for large background objects like walls or tables. At the algorithmic level, the technical concerns are quite diverse, but for a complete system all five points listed above must be addressed. At the skeletal level, the concerns seem quite close in spirit to those of infant perceptual development, apart from differences of terminology caused by the synthetic rather than analytic nature of robotics.

Specificity of Knowledge Gained from Experience

In the robotic-learning examples in the previous section (Fitzpatrick, 2003; Natale *et al.*, 2005), information learned by the robot is intended to be specific to one particular object. The specificity could be varied algorithmically, by adding or removing parts of a feature's 'identity'. Too much specificity, and the feature will not be recognized in another context. Too little, and it will be 'hallucinated' everywhere. We return now to Needham's experiments, which probed the question of generalization in the same experimental scenario described in segregation skills in infants section. When changes were introduced between the objects seen during familiarization and that seen as part of the test display, an unexpected pattern emerged. Nearly, any change in the object's features introduced between familiarization and test prevented infants from benefiting from this prior experience. So, even when infants saw a blue box with yellow squares prior to testing, and the box used in testing had white squares but was otherwise identical, they did not apply this prior experience to the parsing of the test display. However, infants did benefit from the prior exposure when the change was not in the features of the object but rather in its orientation (Needham, 2001). A change in the orientation of the box from horizontally to vertically oriented led to the facilitation in parsing seen in some prior experiments. Thus, infants even as young as 4.5–5 months of age know that to probe whether they have seen an object before, they must attend to the object's features rather than its spatial orientation (Needham, 2001).

These results also support two additional conclusions. First, infants' object representations include detailed information about the object's features. Because infants' application of their prior experience to the parsing of the test display was so dependent on something close to an exact match between the features, one must conclude that a highly detailed representation is formed on the initial

DOI: 10.1002/icd

exposure and maintained during the inter-trial-interval. Because these features are remembered and used in the absence of the initial item and in the presence of a different item, this is strong evidence for infants' representational abilities. Secondly, 4.5-month-old infants are conservative generalizers—they do not extend information from one object to another very readily. But would they extend information from a *group* of objects to a new object that is a member of that group?

Generalization of Knowledge Gained from Experience

This question was investigated by Needham, Dueker, and Lockhead (2005) in a study using the same test display and a similar procedure as by Needham (2001). Infants were given prior experiences with collections of objects, no one of which was an effective cue to the composition of the test display when seen prior to testing. A set of three similar objects seen simultaneously prior to test did facilitate 4.5-month-old infants segregation of the test display. But no subset of these three objects seen prior to testing facilitated infants' segregation of the test display. Also, not just any three objects functioned in this way—sets that had no variation within them or that were too different from the relevant test item provided no facilitation. Thus, experience with multiple objects that are varied but that are similar to the target item is important to infants' transfer of their experience to the target display. This finding with artificial objects was tested in a more natural setting by investigating infants' parsing of a test display consisting of a novel key ring (Needham *et al.*, 2006). According to a strict application of organizational principles using object features, the display should be seen as composed of (at least) two separate objects—the keys on one side of the screen and the separate ring on the other side. However, to the extent that infants recognize the display as a member of a familiar category—key rings—they should group the keys and ring into a single unit that should move as a whole. The findings indicate that by 8.5 months of age, infants parse the display into a single unit, expecting the keys and ring to move together. Younger infants do not see the display as a single unit, and instead parse the keys and ring into separate units. Infants of both ages interpreted an altered display, in which the identifiable portions of the key ring were hidden by patterned covers, as composed of two separate units. Together, these findings provide evidence that the studies of controlled prior exposure described in the previous section are consistent with the process as it occurs under natural circumstances. Infants' ordinary experiences present them with multiple similar exemplars of key rings, and these exposures build a representation that can then be applied to novel (and yet similar) instances of the key ring category, altering the interpretation that would come from feature-based principles alone. Supporting a differentiation view of the development of generalization, Bahrick's findings suggest that young (i.e. 2-month-old) infants are more likely to generalize farther from the specific experiences they received than infants just a few months older (Bahrick, 2002). This finding suggests that experience might serve to initially narrow and then extend the range of stimuli over which young children will generalize.

These results from infant development suggest a path for robotics to follow. There is currently no developmental robotics work to point to on generalization of object categories, despite its importance. Robotics work in this area could potentially aid infant psychologists since there is a strong theoretical framework in machine learning for issues of generalization.

Intermodal Integration

We have talked about 'key events' in which object boundaries are easier to perceive. In general, the ease with which any particular object property can be estimated varies from situation to situation. Robots and infants can exploit the easy times to learn statistical correlates that are available in less clear-cut situations. For example, *cross-modal* signals are a particularly rich source of correlates and have been investigated in robotics and machine perception. Most events have components that are accessible through different senses: a bouncing ball can be seen as well as heard; the position of the observer's own hand can be seen and felt as it moves through the visual field. Although these perceptual experiences are clearly separate from each other, composing separate 'channels', we also recognize meaningful correspondences between the input from these channels. How these channels are related in humans is not entirely clear. Different approaches to the development of intermodal perception posit that infants' sensory experiences are (a) unified at birth and must be differentiated from each other over development, or (b) separated at birth and must be linked through repeated pairings. Although the time frame over which either of these processes would occur has not been well defined, research findings do suggest that intermodal correspondences are detected early in development.

On what basis do infants detect these correspondences? Some of the earliest work on this topic revealed that even newborn infants look for the source of a sound (Butterworth & Castillo, 1976) and by 4 months of age have specific expectations about what they should see when they find the source of the sound (Spelke, 1976). More recent investigations of infants' auditory–visual correspondences have identified important roles for synchrony and other amodal properties of objects—properties that can be detected across multiple perceptual modalities. An impact (e.g. a ball bouncing) provides amodal information because the sound of the ball hitting the surface is coincident with a sudden change in the direction of the ball's path of motion. Some researchers have argued that detection and use of amodal object properties serve to bootstrap the use of more idiosyncratic properties (e.g. the kind of sound made by an object when it hits a surface). Bahrick and Lickliter have shown that babies (and bobwhite quail) learn better and faster from multimodal stimulation (see their Intermodal Redundancy Hypothesis, Bahrick & Lickliter, 2000).

In robotics, amodal properties such as *location* have been used—for example, sound localization can aid visual detection of a talking person. *Timing* has also been used. Prince and Hollich (2005) developed specific models of audio-visual synchrony detection and evaluated compatibility with infant performance. Arsenio and Fitzpatrick (2005) exploited the specific timing cue of *regular repetition* to form correspondences across sensor modalities. From an engineering perspective, the redundant information supplied by repetition makes this form of timing information easier to detect reliably than synchrony of a single event in the presence of background activity. However, in the model of Lewkowicz (2000), the ordering during infant development is opposite. A more complete understanding of the practical benefits of different types of intermodal regularity for robots and infants is a clear and important point of contact between the respective fields.

THE ROLE OF EMBODIMENT

The study of perception in biological systems cannot neglect the role of the body and its morphology in the generation of the sensory information reaching the

DOI: 10.1002/icd

brain. One of the big steps forward in neurophysiology during the last 20 years in understanding brain function is the realization that the brain controls actions rather than movements. That is, the most basic unit of control is not the activation of a specific muscle but rather an action unit that includes a goal, a motive for acting, specific modes of perception tailored to this goal, and the recombination of functional modules and synergies of muscles to attain the goal (Hoftsen, 2004). This shift in perspective is supported by evidence accumulated through the study of the motor system in animals and humans: for a comprehensive treatment, see, for example, (Rizzolatti & Craighero, 2004; Rizzolatti & Gentilucci, 1988).

A modern view of biological motor control considers multiple controllers that are *goal* specific (rather than effector specific) and multiple homunculi and somatotopies that expand into multiple controllers for these goals. This particular type of generalization is, for example, crystal clear in one of the premotor areas that is correlated to the act of grasping. This area, called F5 (frontal area 5), contains neurons that are used for grasping with the left hand, the right hand or even with the mouth (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996).

The next conceptual step in changing our view of the control of movement was made by the discovery of sensory neurons (e.g. visual) in this same premotor cortex, area F5. As far as objects are concerned, it is now well established that the premotor cortex responds both to the sight of objects (visual response), and to a grasping action directed at the same object (motoric response) (Gallese *et al.*, 1996). The two representations—motoric and visual—not only coexist in the same brain areas but also they coexist in the same population of neurons.

Similar responses have been found in the parietal cortex. This forms such a conspicuously bi-directional connection with the premotor cortex that is useful to speak of the fronto-parietal system. Parietal neurons have been found to respond to geometric global object features (e.g. their orientation in 3D) which seem in fact well tuned to the control of action. But the fronto-parietal circuitry is also active when an intended movement does not become an actual one. The natural question to be posed is that what is the purpose of this activation: potential motor action or true object recognition? Multisensory neurons are testimonies of how much action and perception, body and brain are deeply intertwined in shaping each other during development and throughout adulthood.

Active Perception and the Body in Infants

Through the body, the brain performs actions to explore the environment and collect information about its properties and rules. Early in development, exploration of the world occurs through the eyes, hands, and mouth. Infants' earliest competence for exploration is with the eyes—they engage in active visual exploration of the world around them from the first moments following birth (Haith, 1980; Salapatek, 1968). With age and experience, their scanning of visual displays becomes more comprehensive and focused on meaningful features. More recent work has shown that infants' scanning patterns constrain their learning (Johnson & Johnson, 2000; Johnson, Slemmer, & Amso, 2004).

Over the first few months of life, infants gain more control over their limbs and develop a sense of themselves as agents in the world as they make the transition into reaching (Rochat & Striano, 2000; Thelen *et al.*, 1993; White, Castle, & Held, 1964). They often engage in prolonged periods of visual attention to their own hands (White *et al.*, 1964). Interestingly, monkeys deprived of early visual access

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/icd

to one of their arms engaged in intense scrutiny of the arm once they were allowed an unobstructed view of it (Held & Bauer, 1967; see also White, 1971, for related findings with human infants). These results suggest that infants' learning about objects and their own action skills may benefit in very specific ways from their own actions on the world. They exploit the capabilities of their bodies early on to scan objects visually and to explore them with their eyes, mouth, and hands (Rochat, 1983, 1989; Ruff, 1984).

The use of hands for object exploration has received additional attention. In their experiments with human adults, Lederman and Klazky (1987) have identified a set of stereotyped hand movements (*exploratory procedures*) used when haptically exploring objects to determine properties like weight, shape, texture, and temperature. Lederman and Klatzky showed that to each property can be associated a preferential exploratory procedure which is, if not required, at least best suited for its identification.

These observations support the theory that motor development and the body play an important role in perceptual development in infancy (Bushnell & Boudreau, 1993). Proper control of at least the head, the arm, and the hand is required before infants can reliably and repetitively engage in interaction with objects. During the first months of life, the inability of infants to perform skilful movements with the hand would prevent them from haptically exploring the environment and perceive properties of objects like weight, volume, hardness, and shape. But, even more surprisingly, motor development could affect the developmental course of object visual perception (like three-dimensional shape). Further support to this theory comes from the recent experiment by Needham and colleagues (Needham, Barret, & Peterman, 2002), where the ability of pre-reaching infants to grasp objects was artificially anticipated by means of mittens with palms covered with velcro that stuck to some toys prepared by the experimenters. The results showed that those infants whose grasping ability had been enhanced by the glove were more interested in objects than a reference group of the same age that developed 'normally'. This suggests that, although artificial, the boost in motor development produced by the glove anticipated the infants' interest towards objects.

Exploiting actions for learning and perception requires the ability to match actions with the agents that caused it. The sense of agency (Jeannerod, 2002) gives humans a sense of ownership of their actions and implies the existence of an internal representation of the body. Although some sort of self-recognition is already present at birth, at least in the form of a simple hand–eye coordination (Meer, Weel, & Lee, 1995), it is during the first months of development that infants learn to recognize their body as a separate entity acting in the world (Rochat, & Striano, 2000). It is believed that to develop this ability infants exploit correlations across different sensorial channels (combined double touch/correlation between proprioception and vision).

Active Perception and the Body in Robots

In robotics, we have the possibility to study the link between action and perception, and its implications on the realization of artificial systems. Robots, like infants, can exploit the physical interaction with the environment to enrich and control their sensorial experience. However, these abilities do not come for free. Very much like an infant, the robot must first learn to identify and control its body, so that the interaction with the environment is meaningful and, at least to a certain extent, safe. Indeed, motor control is challenging especially when it involves the physical interaction between the robot and the world.

Inspired by the developmental psychology literature, roboticists have begun to investigate the problem of self-recognition in robotics (Gold & Scassellati, 2005; Metta & Fitzpatrick, 2003; Natale *et al.*, 2005; Yoshikawa, Hosoda, & Asada, 2003). Although different in several respects, in each of these efforts the robot looks for intermodal similarities and invariances to identify its body from the rest of the world. In the work of Yoshikawa (Yoshikawa *et al.*, 2003), the rationale is that for any given posture the body of the robot is invariant with respect to the rest of the world. The correlation between visual information and proprioceptive feedback is learned by a neural network that is trained to predict the position of the arms in the visual field. Gold and Scassellati (2005) approached the self-recognition problem by exploiting knowledge of the time elapsing between the actions of the robot and the associated sensorial feedback. In the work of Metta and Fitzpatrick (2003) and Natale *et al.* (2005), actions are instead used for generating visual motion with a known pattern. Similarities in the proprioceptive and visual flow are searched to visually identify the hand of the robot. Periodicity in this case enhances and simplifies the identification. The robot learns a multimodal representation of its hand that allows a robust identification in the visual field.

In our experience with robots we identified three scenarios in which the body proved to be useful in solving perceptual tasks:

1. *Direct exploration*: The body in this case is the interface to extract information about the objects. For example, in the work of Natale, Metta, and Sandini (2004) haptic information was employed to distinguish objects with different shapes, a task that would be much more difficult if performed visually. In the work of Torres-Jara, Natale, and Fitzpatrick (2005), the robot learned to recognize a few objects by using the sound they generate upon contact with the fingers.
2. *Controlled exploration*: Use the body to perform actions to simplify perception. The robot can deliberately generate redundant information by performing periodic actions in the environment. The robot can also initiate actions and wait for the appearance of consequences (Fitzpatrick & Metta, 2003).
3. *The body as a reference frame*: During action the hand is the place where important events are most likely to occur. The ability to direct the attention of the robot towards the hand is particularly helpful during learning; (Natale *et al.*, 2005) showed how this ability allows the robot to learn a visual model of the objects it manages to grasp by simply inspecting the hand when touch is detected on the palm (see Figure 3). In similar situations, the same behaviour could allow the robot to direct the gaze to the hand if something unexpected touches it. Eye–hand coordination seems thus important to establish a link between different sensory channels like touch and vision.

Specificity and Generalization of Knowledge Gained through the Motor System

Study of the motor system has shown the specificity of the coding of object and action information in the brain. For example, Gallese *et al.* (1996) have shown that neurons in the premotor area F5 respond to the execution of specific actions, for example grasping—and not just any grasp, but specific grasps such as pinch grasp rather than power grasp. At the same time, F5 neurons also generalize and many of them are independent of the effector being employed, e.g. left versus right hand. For visuo-motor neurons, specificity and generalization are sometimes complementary, with visual responses typically being broader (less specific) than motoric ones. A category of visuo-motor neuron (mirror neurons) is also related to the recognition of observed actions and similar considerations of specificity versus generalization apply.

It is striking how the brain neatly balances between specificity (allowing recognition and execution of the intended action) and generalization (to the degree of making the effector unimportant). Another way of looking at these results is to say that the goal is important (thus specificity of the action) but not the means by which it is achieved (left versus right hand) (Rizzolatti & Craighero, 2004).

Robotics had adopted the idea of the active recruitment of the motor system for the construction of perceptual abilities even before the discovery of mirror neurons. For example, the Active Vision paradigm in computer vision (Blake & Yuille, 1992) proposed that the movement of sensors could aid the perceptual system by extracting information directly in relation to the goal of the observer. Similarly, in the field of speech processing, Liberman as early as 1967 (Liberman, Copper, Shankweiler, & Studdert-Kennedy, 1967) suggested that speech production and perception are served by a common pathway and by common mechanisms. While at that time, Liberman's ideas were merely conjectures, now they can be defended with scientific argument because of the advancement of the understanding of the physiology of the motor system (Rizzolatti & Arbib, 1998). More recently, Hinton and Nair (2006) proposed a remarkably similar approach for the recognition of handwritten digits and commented on a possible parallel with speech.

More specifically, many authors have either explicitly modelled mirror neurons or approximately borrowed the general idea of common processing modules shared by action production and action understanding (see, for example: Demeris & Johnson, 2003; Fagg & Arbib, 1998; Miall, 2003; Oztop, Kawato, & Arbib, 2006). With respect to generalization, the authors were able to show how inferring the motor representation before classification can improve performance. In a set of experiments (Metta, Sandini, Natale, Craighero, & Fadiga, 2006), human grasping actions including visual and motor data were analysed with machine-learning methods. It was possible to show that the performance of a visual classifier is improved by mapping visual information into a motoric representation as a preprocessing stage. In particular, both the complexity of the classifier is lower and generalization to novel grasp views is improved. These results can both as supporting on one hand the role of embodiment, and on the other, highlighting the benefit to robotics of learning about the acquisition (development) of certain motor skills in humans (grasping in this case).

CONCLUSIONS

In the field of humanoid robotics, researchers have a special respect and admiration for the abilities of infants. They watch their newborn children with particular interest, and their spouses have to constantly be alert for the tell-tale signs of them running an *ad hoc* experiment. It can be depressing to compare the outcome of a five-year, multi-million-euro/dollar/yen project with what an infant can do after four months. Infants are so clearly doing what we want robots to do; is there any way to learn from research on infant development?

Conversely, can infant development research be illuminated by the struggles faced in robotics? Clearly, both domains struggle with questions of origins of abilities and constraints on learning—if we can discover these constraints in the human, perhaps it could facilitate success in the robot. Similarly, facets of what is learned in robotics can guide infant researchers to look for previously unsuspected difficulties that infants might experience.

Is there a way to create a model of development that applies both to infants and robots? Evolution may have selected for propensities in the basic cognitive system of the human infant that could be beneficial for the humanoid robot as well. Considering ways in which the human infant and humanoid robot could learn within the context of a highly structured natural environment, it seems possible that similar sensory constraints and opportunities will mould both the unfolding of an infant's sensitivities to different cues, and the organization of the set of algorithms used by robots to achieve sophisticated perception. So, at least at the level of identifying 'key events' and mutually reinforcing cues, a shared model is possible. Of course, there is a lot that would not fit in the model, and this is as it should be. It would be solely concerned with the class of functional, information-driven constraints. We have not in this paper developed such a model; that would be premature. We have identified some points of connection that could grow into much more. We hope that the paper will serve as a one more link in the growing contact between the fields.

ACKNOWLEDGEMENTS

We are grateful to the editor and to the anonymous reviewers for their help in shaping this paper. The authors at LIRA-Lab were partially funded by EU projects RobotCub (IST-2004-004370) and CONTACT (NEST-5010).

REFERENCES

Arsenio, A. M., & Fitzpatrick, P. M. (2005). Exploiting amodal cues for robot perception. *International Journal of Humanoid Robotics*, *2*(2), 125–143.

Bahrick, L. E. (2002). Generalization of learning in three-month-old infants on the basis of amodal relations. *Child Development*, *73*, 667–681.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190–201.

Baillargeon, R. (1999). Young infants' expectations about hidden objects: A reply to three challenges. *Developmental Science*, *2*(2), 115–132.

Bell, A. J., & Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters. *Vision Research*, *37*(23), 3327–3338.

Blake, A., & Yuille, A. (Eds.). (1992). *Active vision*. Cambridge, MA: MIT Press.

Bushnell, E., & Boudreau, J. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*(4), 1005–1021.

Butterworth, G., & Castillo, M. (1976). Coordination of auditory and visual space in newborn human infants. *Perception*, *5*(2), 155–160.

Demiris, Y., & Johnson, M. H. (2003). Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning. *Connection Science*, *15*(4), 231–243.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interaction in primate control of grasping. *Neural Networks*, *11*(7–8), 1277–1303.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, *59*(2), 167–181.

Fitzpatrick P. (2003). Object lesson: Discovering and learning to recognize objects. *Proceedings of the 3rd international IEEE/RAS conference on humanoid robots*, Karlsruhe, Germany.

Fitzpatrick, P., & Metta, G. (2003). Grounding vision through experimental manipulation. *Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences*, *361*(1811), 2165–2185.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gold, K., & Scassellati, B. (2005). Learning about the self and others through contingency. *Developmental robotics AAAI spring symposium*, Stanford, CA.

Haith, M. M. (1980). *Rules that babies look by: The organization of newborn visual activity.* Potomoc, MD: Erlbaum Associates.

Held, R., & Bauer, J. A. (1967). Visually guided reaching in infant monkeys after restricted rearing. *Science*, *155*, 718–720.

Hinton, G. E., & Nair, V. (2006). Inferring motor programs from images of handwritten digits. *Advances in neural information processing systems* (Vol. 18). Cambridge, MA: MIT Press.

Hofsten, C. von. (2004). An action perspective on motor development. *Trends in Cognitive Sciences*, *8*(6), 266–272.

Jeannerod, M. (2002). The mechanism of self-recognition in humans. *Behavioural Brain Research*, *142*, 1–15.

Johnson, S. P., & Johnson, K. L. (2000). Early perception–action coupling: Eye movements and the development of object perception. *Infant Behavior and Development*, *23*, 461–483.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*, 185–201.

Kaufman, J., & Needham, A. (1999). The role of shape and boundary seam in 4-month-old infants' object segregation. Submitted.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.

Kestenbaum, R., Termine, N., & Spelke, E. S. (1987). Perception of objects and object boundaries by three-month-old infants. *British Journal of Developmental Psychology*, *5*, 367–383.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*(3), 342–368.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Martin, D., Fowlkes, C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(5), 530–549.

Meer, A. van der, Weel, F. van der, & Lee, D. (1995). The functional significance of arm movements in neonates. *Science*, *267*, 693–695.

Metta, G., & Fitzpatrick, P. (2003). Early integration of vision and manipulation. *Adaptive Behavior*, *11*(2), 109–128.

Metta, G., Sandini, G., Natale, L., Craighero, L., & Fadiga, L. (2006). Understanding mirror neurons: A bio-robotic approach. *Interaction Studies*, *7*(2), 197–232.

Miall, R. C. (2003). Connecting mirror neurons and forward models. *NeuroReport*, *14*(17), 2135–2137.

Natale, L., Metta, G., & Sandini, G. (2004). Learning haptic representation of objects. *International conference on intelligent manipulation and grasping*, Genoa, Italy.

Natale, L., Orabona, F., Metta, G., & Sandini, G. (2005). Exploring the world through grasping: A developmental approach. *Proceedings of the 6th cira symposium*, Espoo, Finland.

Needham, A. (1998). Infants' use of featural information in the segregation of stationary objects. *Infant Behavior and Development*, *21*, 47–76.

DOI: 10.1002/icd

Needham, A. (1999). The role of shape in 4-month-old infants' segregation of adjacent objects. *Infant Behavior and Development*, *22*, 161–178.

Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, *1*, 131–156.

Needham, A. (2001). Object recognition and object segregation in 4.5-month-old infants. *Journal of Experimental Child Psychology*, *78*(1), 3–22.

Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition*, *62*, 121–149.

Needham, A., & Baillargeon, R. (1998). Effects of prior experience in 4.5-month-old infants' object segregation. *Infant Behavior and Development*, *21*, 1–24.

Needham, A., Barret, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior and Development*, *25*, 279–295.

Needham, A., Cantlon, J. F., & Holley, S. M. O. (2006). Infants' use of category knowledge and object attributes when segregating objects at 8.5 months of age. *Cognitive Psychology*, in press.

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition*, *94*(3), 215–240.

Oztop, E., Kawato, M., & Arbib, M. A. (2006). Mirror neurons and limitation: A computationally guided review. *Neural Networks*, *19*, 254–271.

Prince, C. G., & Hollich, G. J. (2005). Synching models with infants: A perceptual-level model of infant audio–visual synchrony detection. *Journal of Cognitive Systems Research*, *6*, 205–228.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192.

Rizzolatti, G., & Gentilucci, M. (1988). *Motor and visual-motor functions of the premotor cortex*. Chichester: Wiley.

Rochat, P. (1983). Oral touch in young infants: Response to variations of nipple characteristics in the first months of life. *International Journal of Behavioral Development*, *6*, 123–133.

Rochat, P. (1989). Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology*, *25*, 871–884.

Rochat, P., & Striano, T. (2000). Perceived self in infancy. *Infant Behavior and Development*, *23*, 513–530.

Ruff, H. A. (1984). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology*, *20*, 9–20.

Salapatek, P. (1968). Visual scanning of geometric figures by the human newborn. *Journal of Comparative and Physiological Psychology*, *66*, 247–258.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905.

Simoncelli, E., & Olshausen, B. (2001). Natural images statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, *13*(1), 33–49.

Spelke, E. S. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*, 553–560.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, *1*(1), 1–34.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J. P., Schneider, K., & Zernicke, R. F. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child Development*, *64*(4), 1058–1098.

Torres-Jara, E., Natale, L., & Fitzpatrick, P. (2005). Tapping into touch. *Fifth international workshop on epigenetic robotics* (*forthcoming*). Nara, Japan: Lund University Cognitive Studies.

Viola, P., & Jones, M. (2004). Robust real-time object detection. *International Journal of Computer Vision*, *57*(2), 137–154.

White, B. L. (1971). *Human infants: Experience and psychological development*. Englewood Cliffs, NJ: Prentice-Hall.

White, B. L., Castle, P., & Held, R. (1964). Observations on the development of visually-directed reaching. *Child Development*, *35*, 349–364.

Yoshikawa, Y., Hosoda, K., & Asada, M. (2003). Does the invariance in multi-modalities represent the body scheme?—a case study with vision and proprioception. *Second intelligent symposium on adaptive motion of animals and machines*, Kyoto, Japan.

John Wiley & Sons Ltd
The Atrium, Southern Gate, Chichester West, Sussex PO19 8SQ

Author Queries For ICD 541

While preparing this paper/manuscript for typesetting, the following queries have arisen

Query No.	Proof Page/line no	Details required	Authors response
1	References	Please update the references: Kaufman and Needham (1999) and Needham et al. (2006).	