

Trying to Grasp a Sketch of a Brain for Grasping

Helge Ritter, Robert Haschke and Jochen J. Steil
Cognition and Robotics Laboratory (CoR-Lab) &
Cognitive Interaction Technology Institute (CITEC)
Bielefeld University

Abstract: Brain-like behavior is intimately connected with the ability to actively manage a rich set of interactions with the environment. Originating with very simple movements in homogeneous domains, the gradual evolution of movement sophistication endowed animals with an increasing ability to control their environment, ultimately advancing from the physical into the mental object domain with the advent of language-based communication and thinking. Appearing at the high complexity end of the physical movement evolution ladder, the ability of dextrous manipulation seems in the role of a “transition technology”, leading from movement control into the mental capabilities of language use and thinking. We therefore argue that manual actions and their replication in robots are positioned as a “Rosetta stone” for understanding cognition. Using the example of grasping, we contrast the “clockwork building style” of traditional engineering with more holistic, biologically inspired solutions for grasp synthesis and discuss the potential of the research field of “Manual Intelligence” and its speculative connections with language for making progress towards robots with more brain-like behavior.

Introduction

We are witnessing unprecedented increases in the storage capacity and computing power of man made chips. The raw power of these devices begins to approach the estimated storage and computing capacities of small brains. As a result, our excuses for not being able to realize brain-like behavior in robots due to a lack of adequate processing power will soon have lost their factual basis.

Looking at real brains for realizing brain-like behavior in artificial systems confronts us with many potential levels of analysis: at a micro-level, we can study the morphology of neurons, their interconnections, their activity patterns. On longer time scales, we observe adaptive changes in neural response properties which appear to be correlated with adaptive changes at the behavioral level. From a more abstract dynamics perspective, we have to understand what keeps a highly non-linear and recurrent system stable while admitting a remarkable degree of adaptivity at the same time. Changing to a functional viewpoint, we finally may ask how all these phenomena are connected with the abilities of a brain. To what extent can we decompose the brain into functional modules that can be assigned subtasks? And is it realistic

to assume that any such subtasks can be characterized in terms of familiar functional concepts? To what extent are paradigms such as computation, information, goal, optimization, homeostasis, and other favorites of familiar disciplines useful and adequate, and where are new paradigms needed? And, finally, can our hypothesis about brain phenomena at the neurophysiological and functional levels lead to quantitative models that can replicate the observed phenomena in some reasonably “deep” sense, for instance, leading to testable predictions, or allowing to replicate cognitive abilities in technical systems such as robots?

In view of this daunting pile of questions, it may be helpful to step back for a moment and reconsider what might have been the original “driving force” for endowing organisms with nervous systems and with brains. Put succinctly, “why should a body hire a brain?”.

Brains seem always intimately connected with action and change: as soon as evolution started to endow organisms with modestly rich capabilities of motion, we also see the advent of nerve cells that are invoked in the organization of such motion. As the bodies became more sophisticated, the necessary “machinery” to control such movements had to implement complex interaction patterns that can be instantiated in very short time frames, down to fractions of a second. To do this with chemical processes was a big challenge that led evolution to the invention of neurons, axons and synapses, permitting a new degree of speed and flexibility.

If the implementation of sophisticated movement was the decisive driving force behind brain evolution, we should have closer look at the *evolution of movement sophistication*, hoping that it can give us clues about the major factors that decisively shaped what a brain does and how it achieves it.

Evolution of movement sophistication

A basic reference point for the evolution of movement sophistication is movement in the absence of any active control. This type of movement is observed, when an inanimate body such as a piece of rock is tumbling down a hill. Contrary to first expectation, already such kind of movement can become remarkably complex from the interplay of Newton’s laws of motion, laws of friction and the possibility of temporary or permanent deformations of the moving bodies. Even nowadays, such motions are highly non-trivial to replicate accurately in simulation; a major reason is that even tiny errors in modeling friction and elasticity can rapidly become amplified by the dynamics to give rise to macroscopic errors that can make the movement grossly unphysical. Even in the absence of such errors, Newtonian dynamics is sufficiently rich to lead to sometimes amazing “emergent” properties. Well-known manifestations are the formation of patterns on sand-dunes (many moving sand grains), or the de-mixing of

muesli into different grain sizes under vibration. And the example of the well-known “passive walkers” – entirely passive leg-pair-constructions that can convert potential energy on a ramp into downward-moving periodic walking motions – demonstrate that passive mechanisms can even sustain motion resembling strikingly some non-trivial forms of biological motion.

The evolution of nervous systems enabled organisms to modulate purely passive motion in increasingly sophisticated ways. The earliest nervous systems had to deal with swimming motion within water. Such motion strongly resembles “free movement” since the surrounding water offers only smooth viscous forces and a very homogeneous contact situation between body and environment. Neural controllers for this type of motion have been found to act as central pattern generators that can impose much of the required gross motion in a feedforward fashion, with parsimonious use of sensory feedback.

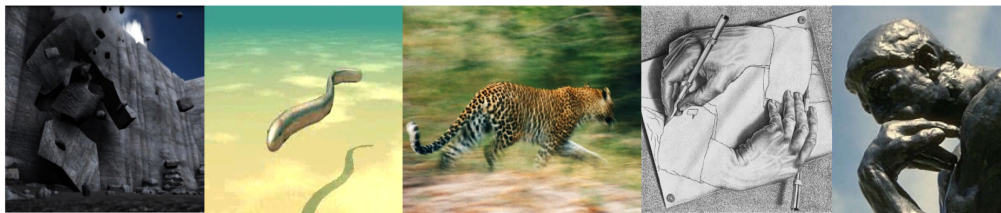


Fig.1 Five stages in the evolution of movement sophistication: inanimate motion (left), control of periodic motions in homogeneous environments (middle left), ego-motion in highly variable environments (middle), complex manipulator control (middle right), communication and thinking as control of mental objects (right).

When creatures left their aquaous environments, they faced a dramatic complexity increase in the contact situation between their body and the environment due to discontinuous transitions between free motion and various types of rigid body contact. In engineering, a major strategy to cope with such situations is to combine several controllers, each capable for handling one regime, plus suitable “machinery” for detecting or prediction transitions and switching control behaviors suitably. For very rigid bodies, this requires extremely fast responses to keep impact forces within manageable bounds. Since neural response times are much too slow for this, biological solutions are based on the use of elastic structures, which complicate the control problem from an engineer’s point of view, but allow to realize controllers with more slowly reacting components.

By and large, all walking motions are about coordinating a number of legs to advance the body across a more or less planar ground roughly perpendicular to the direction of gravity. When the complexity of the ground increases, the control problem becomes more difficult since a simple periodic leg coordination pattern soon becomes

inadequate and legs may slip, hit against obstacles or get trapped in holes. To cope with such challenges, the controller must somehow obtain information about non-uniformities and be able to select from a significant repertoire of suitable corrective motions.

Even richer variability in the mechanical interaction between a set of actuators and a mechanical counterpart occurs in grasping and manipulation with multifingered hands. One may speculate that the corresponding increased challenge for the control system that enables hands to grasp and manipulate many different kinds of objects paved the way for brains towards thinking and communication.

Both of these latter two abilities are characterized by *the manipulation of mental (instead of physical) objects*; such capability appears as a natural generalization of a manipulation of physical objects and for using them as tools to control other physical objects.

We may ask, what are the main innovations that are connected with the above evolutionary steps? From an implementation perspective, we may notice five major levels:

- a fixed, rigid body characterizes the motion of many lifeless objects
- introducing elasticity and joints leads to articulated bodies and dramatically increases the sophistication of possible movements
- addition of actuators enables active movements
- sensors for feedback control open richer ways to adjust movements to environmental conditions
- finally brains provide highly flexible mappings between sensors and actuators to enable biological motion

Why are hands interesting

From a control perspective, we may argue that biological control can be very grossly “clustered” into three main types:

egomotion: control of one’s own physical state w.r.t. environment

grasping: control of the physical state of some object(s) in the physical environment

communication and thinking control of the mental state of oneself and of others.

It appears that the need to control *hands* acted as a major driving force for the required “transition technology” between stage 1 and stage 3 above. If this assumption is correct, research into hands and their control should have a pivotal role for understanding many aspects that form the basis of cognition.

In any case, hands belong to our most important interfaces to the world. At a purely physical level, hands support us with an enormously rich repertoire of complex physical interactions with objects in our environment. Before the advent of machines, pretty much everything that was human-made was made by human hands. Although machinery and modern automation have significantly altered this picture, this modern “competition” to our hands has only been successful by building highly specialized devices for every required operation; the superb generality of our hands (and the control behind them) is still unparalleled even in our most sophisticated robots. It is precisely this generality and flexibility that we have to understand in order to elevate the capabilities of our machines (and in particular, of our robots) from the current level of mere automation to a level that might deserve the term “*cognitive interaction*”.



Fig.2 “Hand homunculus” depicting body parts in proportion to the required brain effort for their control.

The route to our highly developed manual capabilities was one of co-evolution between hands and brain and is likely to have posed a major challenge to brain evolution. We may witness a similar co-evolution in today’s robots: the advent of increasingly sophisticated hand designs pushes the development of algorithms to make adequate use of their capabilities; the experiences gained therefrom enable and reinforce the creation of yet better hand hardware.

This development puts research questions into our focus that had played only a relatively marginal role before humanoid robot hands were available: how to coordinate the constrained movements of a large number of “interaction surfaces” in order to control the motion of an object in various ways? How to gain and exploit rich contact information during this process? How to explore and learn new physical interaction patterns? How to create “action-based” object representations that are more strongly based on how we can interact with an object instead of how we can passively perceive it? How to recognize “interaction affordances” of other objects in ways that enable us to use them as tools enhancing the versatility of our hands still further? And, finally, how to make manual actions a natural part of cognitive interaction with the environment, including language and gestures.

This (highly incomplete) list of questions is strong evidence that research into manual action has the potential of a “Rosetta Stone” for understanding significant parts of cognition, and that hands can provide us with a “window into the brain” that is very different and complementary from the “windows” we are usually focusing on.

Grasping lab

In robotics, the availability of increasingly sophisticated robot hands [Bicchi00] acts as a strong driving force for studying ways to endow robots with more enhanced capabilities for manual action. While the Utah-MIT hand [Utah86] was a kind of yardstick design for a long time, the recent decade has seen a surge of developments towards lighter and more flexibly useable hands. The characteristics of some major contenders are summarized in Table 1. Systems like these begin to provide us with “output devices” to reach beyond simulation when trying to test ideas about the synthesis of manual actions or when aspiring to turn such ideas into practical utility.

Since most “natural” hand actions tend to involve bimanual interaction, an ideal setup should comprise a pair of interacting arms. The high effort to set up such systems makes such platforms even nowadays still a scarce resource. Among the few existing bimanual systems with advanced hands, the perhaps most widely known platforms are at DLR [DLR2Arm], NASA [Robonaut], and the Dexter system [Dexter] using two non-anthropomorphic Barrett hands.

Model	fin- gers	joints	active DOFs	act. type	Ref
Shadow	5	24	20	pn.	[Shadow]
Robonaut	5	22	14	el.	[Robonaut]
GIFU-III	5	20	16	el.	[Gifu]
DLR-II	4	18	13	el.	[DLRII]
Utah-MIT	4	16	16	pn.	[Utah86]
Barrett	3	8	4	el.	[Barret00]

Table 1: Specifications of some dextrous robot hands (el.=electrical, pn.=pneumatic actuator type).

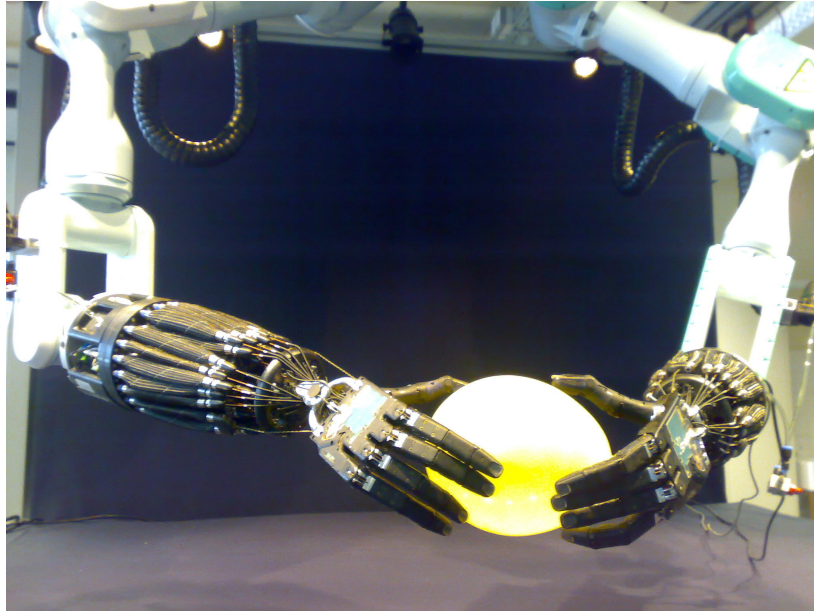


Fig.3 Bimanual system with two Shadow Hands mounted on 7-DOF PA-10 arms for positioning.

The recently completed Bielefeld research platform is depicted in Fig. 3. Featuring two anthropomorphic Shadow Hands with 20 DOF each, mounted on Mitsubishi PA-10 arms, it comprises a total of 54 independent degrees of freedom. 24 Hall sensors per hand provide accurate joint angle feedback to control the 80 miniature solenoid on-off valves that adjust air in- and outflow into the pneumatically driven “muscle”-like actuators transmitting their forces via tendons to the fingers. The system is complemented with a 4-DOF mobile camera head for visual monitoring of the workspace. In the final setup each manipulator will be additionally equipped with 186 tactile sensors distributed over the finger pads.

Despite still far away from the capabilities of human hands, platforms like these begin to cross the critical threshold beyond which one can begin to study issues of advanced manual action in a robotics setting.

Model-based Approaches to Grasping

A primary task for hands is grasping of objects. Synthesizing and controlling grasps for articulated robot hands has a long history [BicKum00, OSC00] While a large part of classical manipulator control focuses on the treatment of contact-free motion, grasping is intrinsically concerned with the rich usage and management of complex contact situations. A suggestive idealization to deal with such situations is to model

the interaction between hand and object as a number of point contacts through which the hand imparts forces on the object. This idealization allows to treat a grasp as the determination of a set of suitably positioned contact points on the object plus the assignment of a force to each contact point, subject to the constraints that (i) the geometric positions and the assigned forces must balance the total external force and torque that acts on the object, (ii) the direction of each contact force lies within the “friction cone” that delimits the directions that avoid slipping, and (iii), the chosen locations and forces are realizable for the given hand design.

For a long time, this strongly geometry-based formulation of the grasping problem has dominated the research in the field. Its mathematical “cleanliness” has been a major attraction, inviting analysis about the existence of “force closure” grasps allowing the fingers to resist any directional forces or torques exerted on an object as long as their magnitude remains within a specified bound and each contact point is characterized by a known and non-zero friction constant, or the more restrictive “form closure” grasps, which achieve a similar effect even when the object is arbitrarily “slippery”. Other researchers used the paradigm for the formulation of measures of grasp quality that measure the stability of a grasp in terms of the six-dimensional “volume” spanned by all “wrenches” (i.e. force-torque pairs) that a grasp can resist without loosing the object. With such measures, finding a grasp becomes a constrained non-linear optimization problem, requiring a search in the space of contact positions and assigned forces subject to the constraint of kinematic feasibility and with the chosen grasp stability measure as the optimization criterion [BoFiHi02]. Initial attempts to solve this problem had to resort to more or less drastic approximations to make the task computational feasible until the discovery that the problem can be formulated as a semi-definite optimization problem, together with the development of good statistical sampling approaches, brought a big thrust in computational efficiency for its solution [BoFiHi05].

Despite these success stories, implementations of grasp abilities on real robot hands continues to be strongly lagging behind the development of the theory. One reason for this is that the above-sketches approaches are based – even within their idealizations – on a significant amount of detailed knowledge about the object geometry and the friction conditions at the contact point locations. However, in reality an accurate object model often is not available, or is only known approximately. This is even more true for the friction conditions at the contact points. A second reason is that the underlying idealization is not really realistic. Unfortunately, real finger contacts depart from an ideal point contact in several important ways: the finger tips are deformable, leading to an extended contact surface allowing to impart not only a force but also a torque at each contact. Moreover, the shape of this contact surface is a function of pressure, direction and the surface shape in the vicinity of the contact. Both aspects have strong effects on the way we grasp objects: while a point contact on an edge

or on an object corner would be ill-defined in its effect, the elastic properties of our fingers make us often prefer precisely such “singular” object locations, since they offer us a better control of the contact situation.

Avoiding the clockwork fallacy

Somewhat paradoxically, although the above-mentioned deviations from the point contact model facilitate the formation of good grasps, their inclusion in a proper algorithmic treatment causes a great deal of additional difficulties [BicKum00]. While it is still rather straightforward to extend the contact point model to more complex contact types (such as contact surfaces) as long as these remain rigid, a proper treatment of deformable fingertips still is largely beyond what is computationally feasible today. Furthermore, better contact models must also include dynamic friction, which is different from static friction and whose consistent treatment, e.g. for accurate physics-based simulations, is still only partially solved today.

When effects that facilitate the solution of a task appear as a major difficulty within the chosen representation framework, we should become wary that the chosen framework is perhaps leading into a wrong direction. The strength of the model-based approach results from a view that sees the world and the embedded processes as a “big clockwork”. As soon as all parts of this clockwork are identified and their properties are known, the operation of the clock can be figured out in great detail. But this may be a characteristic of clockworks in the first place, and at least the physics of the last century has provided us with striking examples that the assumption of “clockwork-ness” for the world can meet unexpected limitations – even within the confines of classical mechanics.

This does not mean that we wish to argue for new physical phenomena in robot grasping; however, there may be a fundamental *information problem*: it may forever remain unfeasible to know the mechanical interaction parameters of most common objects to a sufficient precision to make “clockwork approaches” work (a somewhat related experience has been made in the computer vision community more than a decade ago when it turned out the visual input usually does not provide sufficient information for an accurate 3D-reconstruction of the seen object).

Evidently, there are also good reasons for adopting “clockwork approaches”. One reason is that there are perfect domains where a “clockwork approach” works best since all the necessary information can be provided. However, an entirely different driving force may result from the nature of our own cognition. Whenever we attempt to solve a problem by conscious analysis, we have to cast it into a format that is *communicable*. This is an elementary prerequisite for a division of labor within a team and, therefore, deeply engrained in every activity that we do in a social context.

However, it also seems very important when we just wish to unfold a problem “for ourself”: we have to find some “explicit format” which we can “put before our mental eyes” in order to perform the analysis. We thus may suspect that our strong bias to “clockwork-approaches” does not only stem from their inherent strength in suitable domains; a very deep cause may lie in the nature of our (conscious) thinking and our (conscious) way of communication, which both heavily rely on “mechanistic” representations involving explicit, symbol-like entities and relations among them.



Fig.4 : Common visual concepts, such as the visual concept of a rose, arise only through the participation of a very substantial number of image elements (left: 200 pixels; middle: 1000 pixels; right: 10k pixels).

Remarkably (and perhaps fortunately!) this “explicit mode” of problem solving is not our only mode of operation: many feats of our brain (including vision, language processing, motor control and – alas – grasping) occur in their major part at a “prerational level” [Prerat00] without giving us any awareness of their underlying “inner mechanisms”. When we look at the picture of a rose, we can immediately assemble thousands of pixel elements into the percept of a rose – and we also fail to instantiate the percept of a rose when we try to reduce the pixel number to a range of item numbers that is typical for our mode of explicit communication (e.g. a few ten to hundreds).

Therefore it is likely that we face the challenge of developing methodologies of problem solving (a term heavily connected with classical AI!) that can avoid the “clockwork fallacy” and reach beyond the constraints of carefully handcrafted representations that optimize human readability in the first place.

To address this challenge, we have at our disposal at least three mutually reinforcing strands of developments: the “connectionist” type of approaches (and their modern generalizations that emphasize the concept of creating and shaping dynamical systems); evolutionary algorithms, and finally the increasing awareness of the role of embodiment. All three strands of developments are characterized by methodologies that create implicit “representations” that can be highly functional, but not necessarily “well to read” for our human mind.

A “Topological Approach” to grasping

Looking for “non-clockwork” approaches to grasping, insights about the characteristics of human grasping behavior can provide useful clues for solutions. There have been a number of taxonomies of human grasps [Cutko89, SaFISo02] the gross topology of the hand shape and the contact situation with the grasped object. They converge in a gross subdivision of grasps into a small number of basic grasp types, such as “power grasp”, “precision grasp” or “pinch grasp”.

The importance of the gross hand shape for grasping has also become apparent in behavioral studies of the process of grasping. A major characteristic is an early preshaping of the hand to prepare the final part of the approach phase and the closing of the fingers around the object [Cast05, SaFISo02]

Imitating this strategy for a robot hand offers an attractive alternative approach to grasping that can avoid the “clockwork fallacy” [HPSSF00, KraChr03, RHSR07] the biology-inspired approach does not rely on any prior selection and optimization of contact points; it neither requires detailed friction information. Instead, it emphasizes a process view [Jean84] in which the essential phases of the grasp are primarily shaped according to situational features that are largely topological in their nature and, therefore, very robust.

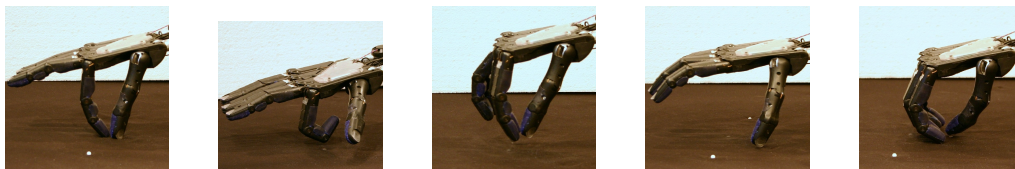


Fig.5 : Hand preshapes used as initial conditions to generate object grasps depicted in Fig. 6

The basic underlying idea is to view grasping as the *creation of a “cage”* around the object. To this end, a suitable hand preshape must guarantee that the fingers attain a good initial position for the closure phase during which the fingers “wrap around” the object. Closure motion of a finger segment stops when its further movement meets any significant mechanical resistance. Ideally, such resistance should be sensed with tactile sensors attached to the digits and to the palm; in our current set up, we were working without such sensors and instead monitored the progress of the movement through positional feedback from finger joint angle sensors.

Conceptually, this approach can be viewed as the preparation of a suitable initial condition (the hand preshape, together with its positioning relative to the object) for a subsequent “attractor dynamics” (the finger closure phase) with termination events (object contact) for the simultaneously occurring constituent movements of the individual fingers.

Main determinant of the resulting grasping strategy is the *hand preshape* (together with its positioning) that defines the initial condition for the closing phase. We have found that already a very restricted set of five hand preshapes (all finger precision grasp, two finger precision grasp, power grasp, two-finger pinch grasp and three finger “special” grasp) is sufficient to achieve successful grasping for a wide range of common objects [RHSR07, Roet07]. Grasping requires that each object is assigned to one hand preshape (this is currently not automated; the “correct” preshape for an object is commanded by a human operator making a corresponding hand gesture in front of a camera) and that the robot hand is positioned in a known distance and orientation to the target object. Therefore, object position and orientation has to be determined, e.g. by a vision front-end.



Fig.6 Example grasps (left) of the Shadow Hand with the algorithm from [Roet07] for a benchmark collection of 21 common household objects (shown on the right).

To evaluate the performance of the resulting grasps, we carried out a benchmark using a representative set of 21 benchmark objects depicted in Fig. 6. While research into grasping dates more than four decades ago, so far there exists not yet any established benchmark procedure for evaluating grasps and grasp success. The benchmark suite and the procedure (outlined below) to measure grasp success involves only objects that are readily available everywhere and may be a first step towards a more systematic and generally accepted benchmarking procedure for grasp evaluation.

For each object, the robot had to make ten grasp attempts, each of which consisted in grasping and lifting the object. Each attempt that resulted in a successful lifting of the object was counted as a success, otherwise as a failure. Table 2 offers a summary

of the measured performance and shows that the majority of objects can be grasped with a high success rate; but some objects (such as the bunch of keys or the pencil) are still too demanding for the described strategy: in the case of the keys the many moveable parts preclude reliably grasping, while the pencil is too slim to allow the fingers to attain a stable “force opposition” configuration when picking it from the table. Note that in both cases also the “clockwork-type” approach would be highly unlikely to be successful: neither could it remedy the size problem in the case of the pencil, nor would a grasp point determination on the basis of a detailed modeling of the movements of the keys appear feasible.

Obj	Succ	Grasp	Obj	Succ	Grasp
1	10	Force (+III)	11	7	Pre
2	10	(special)	12	6	III
3	10	(+III,-Force)	13	7	Force
4	10	Force	14	7	III (+Force,-Pre,-II)
5	10	Force (+III)	15	6	Force (-III)
6	10	Force (+III)	16	5	III
7	9	Force (-III)	17	4	Pre
8	8	III (+Force,-Pre,-II)	18	3	Pre
9	8	Force	19	4	Pre
10	9	Force	20	0	III
			21	0	Pre

Table 2: Performance of “topology-based” grasping strategy for 21 benchmark objects depicted in Fig.6. Column “Succ” gives number of successful grasping trials out of 10 attempts; Column “Grasp” indicates best grasp preshape (with less optimal alternatives in brackets).

There are also important details to take into account for the described “topology-based approach”. A major factor affecting grasp success turned out to be the proper timing of the finger contacts on the object. Experiments with human subjects have revealed that humans tend to maximize the simultaneity of their finger contacts on the object. This strategy helps to minimize “unwanted” object movements during the finger closure phase.

A similar timing could be obtained for the robot grasps by a proper adjusting of the initial finger positions in the pregrasp. The work in [Roet07] shows how a physics-based simulation of the finger closure phase can be used to achieve an appropriate pregrasp optimization for each object, however, at the price of then requiring at least approximate geometry data for each object. A second significant factor is the optimization of the thumb opposition. For the sake of brevity, we refer the reader to [Roet07] for the technical details.

The robustness of the resulting grasping strategy is not only apparent in its good generalization properties to novel objects but also in its portability to a different hand design. While usually most grasping algorithms are only evaluated on a single hand system, we took the availability of an older, three-fingered hand system as an opportunity to test how well the algorithm works with such a more restricted manipulator. While the overall grasp success rate decreases as a result of the reduced “caging” that can be provided with three fingers (instead of five on the Shadow Hand), even the three finger hand was able to successfully grasp the majority of the benchmark objects depicted above.

From Grasping towards Manual Intelligence

Grasping an object is only one of a large array of sophisticated capabilities connected with our hands. Once a child has mastered grasping, many other sensori-motor competences come into view: alternating between different grasps according to what an object is used for; coordinating both hands for performing complex operations that are too difficult for a single hand alone; learning the numerous interaction patterns such as guiding, supporting or adjusting when we arrange rigid or non-rigid objects in purposeful fashion; acquiring important action-based concepts such as that of a container that unite interaction patterns with mechanically very different types of objects; and mastering tools such as spoons, knives or keys to enable operations that would be difficult or impossible with our hands alone.

This list could be continued much further, and each of its entries denotes a highly non-trivial type of interaction pattern which is in most cases rather routine for human hands, but with very few exceptions well beyond the state of the art in current robotics [[Robonaut05](#)].

This is not only due to a lack of sensing or control at the level of existing hand hardware. In fact, we all know that a human can perform an amazing range of “manual” actions even when equipped with nothing more than a rigid mechanical hook. Examples like this make clear that a very essential ingredient for the scope of manual actions is the cognitive machinery that shapes and binds the low-level interaction patterns together. If our robots had only a tiny percentage of this cognition, they very likely could perform marvellous actions with the robot hands that exist – despite all of their shortcomings [[AmaWoo05](#)].

However, the immense scope of the required “manual intelligence” becomes clearer when we envisage the significant time used by children to learn the use of their hands. This time extends over many years and the involved effort appears comparable to that involved in the acquisition of language.

It is an intriguing question to what extent the acquisition of language and the manual

capabilities may be interrelated [Neil92] (and we will speculate on that issue below), but the case of sign language demonstrates impressively that the control of our hands can be at least as sophisticated as language itself. However, most uses of our hands are not connected with the task of explicit communication, but instead are involved in interaction patterns that we usually never need to speak of (and if we have to do so, we find it very hard to verbalize what our hands are doing).

Therefore, we may suspect that the nature of our “manual intelligence” is very different from the rational intelligence that was in the focus of classical AI for a long time. Given the insights that the attempts of classical AI to understand intelligence have provided, this difference may be very encouraging: by its strong grounding in physical interaction patterns while at the same time spanning the enormous “semantic spectrum” that reaches from low-level control to tool use and even to emotional, artistic and linguistic expression, “manual intelligence” may be “just the right kind” of intelligence to embark on for a research program devoted to a better understanding of cognition. Such a research perspective is also well in line with the increasing – although not undisputed – appreciation of the important role of embodiment for cognition.

Measuring Manual Intelligence

Envisaging manual intelligence as a research topic as well as an important capability for robots, we should at least find some ways to make it measurable. Currently, there exist practically no generally established benchmarking procedures, even when we restrict ourselves to rather well-defined capabilities such as object grasping. A tentative proposal within the EURON initiative is based on a bimanual Barrett hand system and proposes to evaluate grasp success for a number of (artificial) benchmark objects [Morales06]. A different benchmark, employing the set of 21 widely available household objects (shown in Fig.2), has been suggested in [Roet07] and has been used to compare grasp optimization schemes on two different robot hands [RHSR07].

Useful guidance for measuring manual intelligence might be provided from surgery, where the comparison of different training strategies with respect to their impact on the acquisition of manual skills in surgeons is an important issue [HamHal00]. For instance, manual skills in using a laparoscope have been successfully modelled as temporal force and torque profiles imparted on the instrument [RHRS01]. In the study of child development, a widely accepted procedure for measuring the development stage of motor skills is the Peabody Motor Development Scale [PDMS]. It has a part specifically focusing on grasping skills, featuring 26 different test tasks each of which is ranked on a nominal 3-scale. Another 72 tasks measure visuo-motor coordination. While the majority of these tests are probably still too hard for the level of manual intelligence of today’s robots, they might become useable in the near future when

robot hands can do more than now. Until then, these test designs might provide useful inspiration how to design manual skill benchmarks for robots, for instance, embracing instruction by demonstration as a natural part of any performance measurement.

Manual Intelligence and Language: Some Speculations

If we look at the multitude of research issues that have been in the focus of robot manipulation during its inception about forty years ago, we recognize that an overwhelming proportion of this research has been addressing relatively low-level issues of hand control. While this work has significantly contributed to the design of better robot hands, any research dealing with the realization of manual abilities beyond the basic acts of reaching and grasping has been very scarce at best.

Reminding us of the enormous abilities in a simple hook when “driven” with the right amount of cognition, we may have been too obsessed with developing the “right clockworks” at the wrong level to make our hardware “tick”.

Our hands are very general manipulators. We can arrange our fingers into many specialized “virtual tools” with very different and highly specialized patterns of operation. This makes it very hard to believe that manual skills can be derived from a small set of general principles: it seems that our repository of manual skills contains a large number of “inventions” made through our long experience of hand use. Interestingly, most of these “inventions” seem to be made by all people in a very similar way; however, at least some of the more sophisticated inventions would probably take too long to invent ourselves and are instead taught to us (or we pick them up through observation and imitation of others): tying knots, for instance, or many skills that make up the different craftsmanships.

This resembles language: a lot of structure seems not derivable from any deeper logics. It “just” has evolved to serve the purpose of communication, and the multitude of languages demonstrates impressively that there are many different solutions possible.

If we pursue the analogy with language further, we may find that much of our current research is still focused on making our main “articulatory instrument” useable at all. This is very reminiscent of the “babbling phase” during which a child learns to use its vocal tract to articulate a basic repertoire of syllables. It might be interesting to analyse early human manual actions from such a perspective: it would appear as a very reasonable economy principle if the brain used similar strategies (and/or circuits) to acquire phonetic and “manual” babbling abilities.

If we are willing to take the analogy with language development serious, the next challenge would be an identification and mastery of larger basic chunks of manual action with a similar building block role as words in language. There has already been

some intriguing work aimed at the identification of “basic actions units” of human movement [Schack04, GuFeAl05]

This work has shed some light on our understanding of how skillful movements might become generated from a more limited repertoire of basic primitives; at the same time it has enabled to optimize the performance level in sports by identifying which features in the mental organization of action units are indicative of expert performance and reinforcing those features through correspondingly focused training procedures [Schack06].

A database of human manual actions can become an important cornerstone for setting up an equivalent of a “vocabulary” in language. It would have to represent its entries not only at the trajectory level, but also at more abstract levels, such as in form of idealized physics-based simulation chunks, together with annotations at the task level. It might play a similar role for manual action research as the famous WordNet corpus [WordNet98] in linguistic research.

In the same manner as “knowing” a vocabulary of a language involves the (very non-trivial!) ability of associating objects, activities and relations by their “names” in the vocabulary, “knowing” a vocabulary of manual actions necessitates the ability of associating actions that “fit” into a given constellation of objects. Recognizing such “affordances” [Gibson79] requires to recognize a *possibility*, calling for a non-trivial extension to the familiar classification paradigm of pattern recognition.

Another key aspect is the *usage* of dictionary entries. Here once more language can offer guidance: using a word requires expectations of its effect on the mind of the receiver. Likewise, manual action units will have to become tied to representations about their effects on the involved objects. This level is very different from the (also very non-trivial) control issues to be solved for physically executing the manual action unit, and we already have pointed out that that level might more aptly be seen in analogy to the articulation of the sound patterns of a word.

We believe that only such deeply grounded representations of manual action primitives can form the necessary basis for considering longer action sequences (“sentences”). It is well known in linguistics that the words within a sentences are not primarily data, but instead very parsimonious “interaction instruments” to instantiate a desired mental picture in the mind of the receiver. This point has been brought to fore perhaps most succinctly in recent work that views language from the viewpoint of an “interaction game”, using game theory to analyze the utility of different choices for the next interaction step [Jaeger08].

Since language has evolved to enable talking about actions, it should be of little surprise that there is a close correspondence between sentences about actions, and the structure of the actions themselves. If we consider, e.g., the use of a tool such as a knife, we clearly can distinguish a “subject” (the knife), an “object” (the bread)

and a “predicate”, namely the activity of cutting. From an evolutionary perspective, the ability of hands to assemble structures in the physical domain appears as a very natural basis for its extension into the mental domain, enabling an assembly of mental structures when the interaction of hand action sequences with physical objects is replaced by the interaction of sentences with mental objects in the mind of the listener.

Using hands in a purposeful fashion to interact with and shape a wide range of differently structured physical contexts would then resemble the linguistic capability of conducting conversations on a wide range of subjects. At this level, manual intelligence would need to invoke many of the cognitive abilities that we so far primarily connect with language: recognizing agentship and roles, planning ahead, associating meaning to object constellations and many more. And although we must always commit ourselves to one single thread of real action, we also can imagine action alternatives before our mental eye, leading into analogies with the linguistic formulation of conditionals.

Conclusion

Taking serious the view that brain-like behavior is primarily about the active shaping of interaction, the study of manual actions, or, more ambitiously, of Manual Intelligence, should bear a great potential to reveal significant insights about brains and the replication of some of their functions in machines. While traditional robotics viewed most parts of the environment as obstacles and has been biased towards generating movement strategies that kept the robot away from any objects, research into manual action has to embrace contact and interaction as main modes of operation. This will shift our approaches for representing objects, situations and actions in a healthy way and spur novel approaches which have interaction in their center, calling for new ways to detect, recognize, decompose and control patterns that are no longer of a purely input nature. This appears to be a good move, since it is likely to refreshen also the interaction of our ideas and between us researchers ourselves.

References

- [AmaWoo05] St. Amant, R., and Wood, A. B. (2005). Tool use for autonomous agents. Proc. National Conf. on Artificial Intelligence (AAAI), pp. 184-189.
- [Utah86] C. Jacobsen, E. K. Iversen, D. F. Knutti, R. T. Johnson, and K. B. Biggers (1986) Design of the Utah/MIT dexterous hand. ICRA Conf. Proceedings, pp. 1520-1532

- [Barret00] Townsend W (2000) The BarrettHand grasper – programmably flexible part handling and assembly *Industrial Robot* 27(3):181-188.
- [Bicchi00] Hands for dexterous manipulation and robust grasping: a difficult road toward simplicity *IEEE Trans. Robotics Autom.* 16(6):652-662
- [BicKum00] A Bicchi, V Kumar (2000) Robotic grasping and contact: a review *Proceedings ICRA'00* pp. 348-353
- [BoFiHi02] Borst C., Fischer M., Hirzinger G. (2002) Calculating hand configurations for precision and pinch grasps. *Proc. IEEE IROS 2002*, pp. 1553-1559.
- [BoFiHi05] C. Borst, M. Fischer, and G. Hirzinger (2005). Efficient and precise grasp planning for real world objects. In: *Multi-point Interaction with Real and Virtual Objects* (F. Barbagli, D. Prattichizzo, and K. Salisbury, eds.), *Tracts in Advanced Robotics* 18, pp. 91-111
- [Cast05] U Castiello (2005) The Neuroscience of Grasping *Nat Rev Neurosci* 6:726-736
- [Cutko89] Cutkosky M.R. (1989) On Grasp choice, grasp models and the design of hands for manufacturing tasks. *IEEE Trans. Robotics and Automation* 5(3), 269-279
- [DLRII] J. Butterfass, M. Fischer, M. Grebenstein, S. Haidacher, and G. Hirzinger (2004) Design and experiences with DLR Hand II, in *Proc. World Automation Congress, Sevilla*.
- [DLR2Arm] Ott C, Eiberger O, Friedl W, Bauml B, Hillenbrand U, Borst C, Albuschaffer A, Brunner B, Hirschmuller H, Kielhofer S, Konietzschke R, Suppa M, Wimbock T, Zacharias F, Hirzinger G (2006) A Humanoid Two-Arm System for Dexterous Manipulation. In *6th Humanoid Robots Conf.*, pp. 276-283
- [GuFeAl05] G Guerra-Filho, C Fermüller, Y Aloimonos (2005) Discovering a Language for Human Activity. In: *Proc. AAIL-2005 Fall Symposium*.
- [Gibson79] Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin, Boston.
- [Gifu] T Mouri, H Kawasaki, K Yoshikawa, J Takai, S Ito (2002) Anthropomorphic Robot Hand: Gifu Hand III *Proc. of Int. Conf. ICCAS2002*
- [HamHal00] Hamdorf J.M. Hall J.C. (2000) Acquiring surgical skills. *British Journal of Surgery* (87):28-37

- [HPSSF00] Hauck, A. Passig, G. Schenk, T. Sorg, M. Färber, G. (2000) On the performance of a biologically motivated visual control strategy for robotic hand-eye coordination Proc. IROS 2000, 3:1626-1632
- [Jaeger08] Jäger G (2008) Applications of Game Theory in Linguistics. Language and Linguistics Compass 2:1749-1767
- [Jean84] Jeannerod (1984) The timing of natural prehension movements. J. Motor Behavior 16(3), 235-254.
- [KraChr03] Kragic D, Christensen H.I. (2003) Biologically motivated visual servoing and grasping for real world tasks IROS 2003 Proceedings Volume 4:3417-3422
- [Morales06] A. Morales (2006) Experimental benchmarking of grasp reliability. <http://www.robot.uji.es/people/morales/experiments/benchmark.html>
- [Neill92] D MacNeill (1992) Hand and Mind: what gestures reveal about thought University of Chicago Press 1992
- [PDMS] M.R. Folio, R. R. Fewell (2000) Peabody Developmental Motor Scales PDMS-2 Therapy Skill Builders Publishing
- [Robonaut] F Rehnmark, W Bluethmann, J Mehling, RO Ambrose, Myron Diftler, Mars Chu, and Ryan Necessary (2005) Robonaut: The Short List of Technology Hurdles Computer 38:28-37
- [SaFISo02] M Santello, M Flanders, JF Soechting (2002) Patterns of Hand Motion during Grasping and the Influence of Sensory Guidance Journal of Neuroscience 22(4):1426-1435
- [Schack04] T Schack (2004) The cognitive architecture of complex movement. Int. J. of Sport and Exercise Psychology 2(4):403-438.
- [Schack06] T Schack and F Mechsner (2006) Representation of motor skills in human long-term memory. Neurosci. Letters 391:77-81
- [Dexter] Dexter - Mechanism, Control and Developmental Programming http://www-robotics.cs.umass.edu/Research/Humanoid/humanoid_index.html
- [Prerat00] Cruse H., Dean J. and Ritter H. (eds) (2000) Prerational Intelligence – Adaptive Behavior and Intelligent Systems Without Symbols and Logic. Vol 1-3, Studies in Cognitive Systems, i Kluwer Academic Publishers
- [RHRS07] Röthling F., Haschke R., Steil J.J. & Ritter H. (2007) Platform Portable Anthropomorphic Grasping with the Bielefeld 20 DOF Shadow and 9 DOF TUM Hand. IEEE IROS Conference Proceedings

- [Roet07] Röthling F. (2007) Real Robot Hand Grasping using Simulation-Based Optimisation of Portable Strategies Dissertation, Faculty of Technology, Bielefeld University
- [Robonaut05] F Rehnmark, W Bluethmann, J Mehling, RO Ambrose, Myron Diftler, Mars Chu, and Ryan Necessary (2005) Robonaut: The Short List of Technology Hurdles Computer 38:28-37
- [RHRS01] Rosen, J. Hannaford, B. Richards, C.G. Sinanan, M.N. (2001) Markov modeling of minimally invasive surgery based on tool/tissueinteraction and force/torque signatures for evaluating surgical skills IEEE Trans. Biomed. Engineering, 48(5):579-591
- [OSC00] AM Okamura, N Smaby, MR Cutkosky (2000) An overview of dexterous manipulation Proceedings ICRA'00 pp. 255-262
- [Shadow] Shadow Robot Company, The Shadow Dexterous Hand. [Online]. Available: <http://www.shadow.org.uk/products/newhand.shtml>
- [WordNet98] Ch. Fellbaum (Ed.) WordNet – An Electronic Lexical Database. MIT Press 1998.