Willy Østreng, editor

Synergies

Interdisciplinary Communications 2003/2004

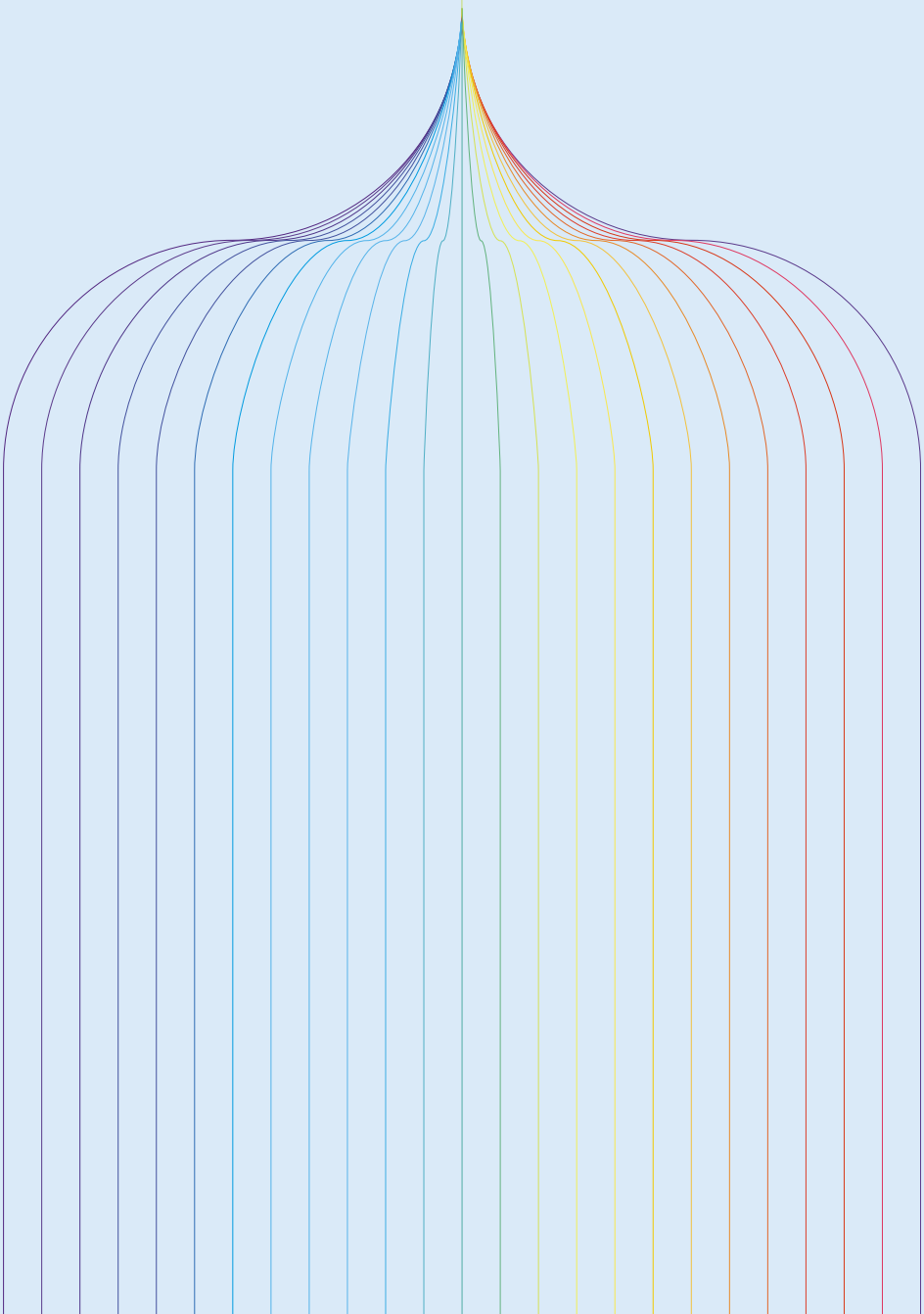
Centre for Advanced Study
at the Norwegian Academy of Science and Letters

CAS OSLO

Willy Østreng, editor

Synergies

Interdisciplinary Communications 2003/2004



Centre for Advanced Study in Oslo

The Centre for Advanced Study (CAS) is an independent private foundation. The Centre was established by the Norwegian Academy of Science and Letters in 1989, but its activities did not commence in full until the autumn of 1992. Its purpose is to promote basic research and interdisciplinary theoretical research on the highest international academic level within the humanities/theology, the social sciences/law and the natural sciences/medicine/mathematics. The Centre's academic activity is of a long-term nature and is to be permanent and academically independent vis-à-vis political and economic influences and the influence of research policy.

Outstanding researchers from Norway and abroad are nominated for one-year stays to engage in research in the Centre's premises in the Norwegian Academy of Science and Letters' mansion in Oslo. The activities are organized in three groups – one in the humanities, one in the social sciences and one in the natural sciences - each with from six to ten members whose affiliation is long-term. In addition come numerous researchers who spend shorter periods conducting research, altogether some 40–45 researchers of 10 to 15 nationalities a year. Each group is planned and organized around a unifying theme and headed by one or more outstanding researchers. The groups have no other obligations than their own research. They receive administrative and financial support from the Centre in formalized cooperation with five Norwegian universities and one high-level research college, i.e. the University of Oslo, the University of Bergen, the University of Tromsø, the Norwegian University of Science and Technology in Trondheim, the Norwegian University of Life Sciences in Ås and Norwegian School of Economics and Business Administration in Bergen. The Centre has a Board appointed by the Norwegian Academy of Science and Letters, the Universities and Colleges Council and the Research Council of Norway. The administration is taken care of by a staff of four full-time and two part-time employees and headed by a Scientific director.

Centre for Advanced Study
at the Norwegian Academy of Science and Letters
Drammensveien 78
0271 Oslo
Norway
Telephone: +47-22 12 25 00
Fax: +47-22 12 25 01
Email: cas@cas.uio.no
Internet: http://www.cas.uio.no

Content

Background and Foreword

The Centre for Advanced Study (CAS) in Oslo has two overriding long-term objectives. The first aims at enhancing the quality of Norwegian basic research to the highest international level and standard. Here the call is for specialization and penetration in depth – to benefit basic disciplinary science. The other aims at promoting the same quality and achieving the same level of excellence when it comes to interdisciplinary research. Here the call is for wholeness and integration in breadth as well as in depth – to benefit basic complex system science. The two objectives relate equally to the humanities, the social sciences and the natural sciences and both are supposed to find expression within and between the three fields of academe.

The CAS has no enduring thematic profile. The profile of the Centre is compositional, in that humanists, social scientists and natural scientists are present at all times. This opens up interesting opportunities for interdisciplinary activities, not least in combination with the fact that the premises of the CAS are physically restricted and the logistical structure is one of oneness. This means that the CAS is located in one building, has one seminar room, one luncheon room, one administrative staff and one Scientific director. Such a setting is new to most of the CAS' guest professors, who are used to working in different faculties located in different buildings and in different departments located on different storeys, using different seminar rooms, auditoria and canteen facilities. On campus, professors are separated by lawns (Keep off the grass?), asphalt (pretty hard on the feet?) and floors (connected by steep stairs and out-of-order elevators?) so it takes some effort to overcome these obstacles in order to meet with colleagues in other departments. At the CAS the fellows are merged into one multidisciplinary faculty and share all facilities so that no effort is required to overcome any obstacles. This mix of academic specialties, the physical closeness of the groups and the oneness of the infrastructure make the CAS an ideal arena for science dialogue across disciplinary boundaries and academic fields.

In 2003/04 the CAS took stock of this situation and decided to organise a series of luncheon seminars at which the fellows were invited to give presentations on their specialities to plenary sessions of all the groups with the aim of fostering a feeling of professional and social community through dialogue. Three groups were in action. The humanists were working to further *A New Understanding of the Mental*, the social scientists aimed at developing *A Comprehensive Model of Human Memory*, whereas the group of natural scientists were concerned with *Food-webs, Stoichiometry and Population Dynamics*. Each and every one of these groups could from its specialist stand shed some light on one or both sides of the relationship between body and mind, biology and soul, and matter and spirit. Thus, this relationship became the least common denominator for the groups and the seminar.

The format of the seminar was quite relaxed in that each presentation should not last for more than 30 minutes followed by an equally long discussion session. And, equally important: the seminar should only take place once a week. The presentations soon proved that the 'memory

group' and the 'mental group' had some interest in common when it came to concepts relating to knowledge and memory. The mental group, guided by a philosophical approach, introduced and used concepts such as *implicit* and *explicit knowledge*, *tacit knowledge*, *self-knowledge* and *implicit perception*, whereas the memory group, who applied a psychological approach, used concepts like *knowing* and the *feeling of knowing*. These conceptual variations relating to the concept of knowledge gave rise to most interesting discussions between the involved disciplines and groups. One of the psychologists concluded that the variety of concepts and approaches introduced in this setting had resulted in an extension of his international network and made useful contributions to his own research. Another example: In one of the presentations by the 'stoichiometry group' the decision was made to deviate from what had originally been planned and instead introduce the concept of *biological and genetic determinism* because the theme was assumed to be "of some relevance … (to) the philosophy group." One of the biologists stated in writing that he was grateful "… to the lunch group at the CAS for a variety of fascinating conversations during *middag*", indicating that the discussions initiated at lunch was continued during dinner. More examples could be cited. But to cut this already too long background note short, a summing up is required: the luncheon seminar turned out to be instrumental in creating a feeling of both social and professional community between the groups. At the same time it produced an interdisciplinary atmosphere for the clarification of concepts of significance in theory-building in more than one discipline and across group boundaries. Last but not least: The suggestion to produce this booklet of condensed versions of the 24 presentations was not put forward by the Scientific director, but by the fellows themselves. One stated reason for this was that such a publication would prove useful in the work of the groups after they had ended their stay with the CAS.

One last clarification: The first article in the booklet: "The Gribbin Syndrome and the Entities of Knowledge Integration" was never part of the seminar series. It was included to argue against the widespread assumption that specialization in basic research – which relates to the first objective of CAS – is the opposite of complex system science – which relates to the second objective of CAS. As argued in the article, specialization is the prerequisite of integrated research, implying that the two objectives are compatible, not contradictory.

Oslo, January 2005
Willy Østreng
Scientific director and Editor

Acknowledgement

This booklet involves the work of many individuals. Bjarne Røsjø, prosject leader at *Faktotum Informasjon A/S* has coordinated the work between the designer, CAS and the printer, whereas Ketill Berger at *dEDBsign* is responsible for the attractive design of the booklet. The language editing has been competently done by Dr. Patrick Nigel Chaffey, and Unn Haaheim Hagen, Marit Finnemyhr Strøm and Maria Sætre have assisted were and when need be. To these individuals, the editor would like to express his appreciation.

WILLY ØSTRENG:

The Gribbin Syndrome and the Entities of Knowledge Integration

The purpose of this introductory article is to discuss the diehard assumption that the increasing specialization of science acts counter to the need for understanding complex systems. Or, in the context of this booklet: Is specialized basic research the antithesis of integrated science?

In the twentieth century, disciplinary science has been so successful that outsiders sometimes picture it as a kind of monolithic corporate organization, like IBM or Microsoft. This imagery has nourished the popular belief that "scientists look alike, act alike, think alike and speak the same jargon (Weiner, p. 198)". No imagery could be more wrong. Science has become the very opposite – a 'Tower of Babel', where few, if any, speak the language of the others, and no one seems eager to learn more languages than their own disciplinary mother tongue. The accumulation of scientific knowledge has been so overwhelming that no one can hold it within "the horizon of a single mind (Weiner, p.198)". The truth of the matter is that even the brightest of scientists no longer manage to keep abreast of the total knowledge accumulation within their own discipline, not to mention the crossing of disciplinary boundaries. The burden of disciplinary knowledge has grown overwhelmingly and it has been doing so for a long time. More than a hundred years ago, Professor Karl Pearson at the University of London pinpointed the situation with this heartfelt sigh: "Scarcely any specialist of to-day is really master of all the work which has been done in his own comparatively small field. Facts and their classifications have been accumulating at such a rate, that nobody seems to have leisure to recognize the relations of sub-groups to the whole (Pearson, p. 17)."

Scientific director and Editor
Willy Østreng
Centre for Advanced Study
at the Norwegian Academy of Science
and Letters

Units of topical Specialization

Since this utterance, the process has picked up more than exponential pace. In order to cope, the scientific community has deliberately divided disciplines into smaller and more manageable entities, and the smaller disciplines into sub-disciplines, which have been further broken down into *units of topical specialization*, which are continuously getting lighter in weight, thinner in scope and deeper in penetration. Although, the disciplinary organization of science still persists, the professional frame of reference and identity is more with the smaller units of topical specialization than with the traditional disciplines. Today a discipline is composed of clusters of specialties – units of topical specializations – that form the micro-environments where research and communication take place. For the sake of illustration: By the year 1987, there were 8,530 definable fields

of knowledge in the sciences, and three years later, roughly 8000 research topics were sustained by specialized networks within the natural sciences alone (Thompson Klein (96), p. 42).

The Gribbin Syndrome

According to the noted science writer, John Gribbin, the educated elite has become *overspecialized*, and he anticipates great dangers: "The habit of specialists in any one area of science is to focus more and more narrowly on their special topic, learning more and more about less and less, until eventually they end up knowing everything about nothing. It was in order to avoid such a fate that, many years ago, I chose to become writer of science rather than scientific researcher (Gribbin, p.1)." Gribbin's concern is that modern science, if the present course is not corrected, will ultimately leave society with a breed of scientists who are so specialized that they have no one to speak to, no one to discuss with, no one to talk to, no one to learn from and no one to report to. They may end up in a state of secluded professionalism – as isolated islands of knowledge without bridges erected between them. This situation has been vilified in the public discourse as a negative force that only promotes fragmentation and specialization *in absurdum*. There are two basic reasons to doubt the validity of such an assumption.

Gribbin and his critics

The first has been aptly depicted by Bruno Latour who argues that an isolated specialist is a contradiction in terms because no one can specialize without the concurrent autonomization of a small group of peers. By the term autonomization he means the corrections/directions provided by the way in which a discipline, a profession, a clique or an 'invisible college' becomes independent and forms its own criteria of evaluation and relevance. To make individual progress, specialists depend on the critique provided by this process. In substantiating his point, Latour refers to the observation that scientists who are totally on their own doing field research in isolated parts of the world, never stop "speaking in a virtual arena of colleagues with whom they constantly argue in absentia as if the wooded landscape had been transformed into the wooden panelling of a conference room (Latour, p. 102)." Physicists, in particular, are singled out as having the habit of constantly talking to each other at the blackboards with no one else present (Weinberg, p.19). In other words: A specialist needs other specialists to talk to, to disagree with, to convince, to argue with, to be stimulated by, to quarrel with, to despise and look up to. This virtual arguing seems to be the dialectic of scientific progress. Thus, the likelihood is that specialists will never specialize in absurdum because that will deprive them of someone to relate to. The human psyche will see to it that specialization stops short of seclusion.

The second reason to doubt the Gribbin syndrome relates to an observation made by the Norwegian biologist Johan Hjort in the 1920s stating that the deeper we go into a problem, "the more do we feel that it is really itself part of a whole great structure which science and thought has erected (Hjort, p. 4)." The assumption is that units of topical specialization of one disciplinary origin may have significant features in common with component entities in other disciplines, and accordingly may have the potential to promote greater interdisciplinary understanding. Recently,

Barnes et al. argued that detailed work in science is never intelligible purely by reference to the esoteric conventions and concerns of the specialty in which it is performed. It always has significance for allied or opposed specialities, and is always liable to evaluation as an element of science generally and an instance of what is conventionally accepted as science (Barnes, Bloor and Henry, p.155)." What is said here is that increasingly, specializations overlap and transcend disciplinary boundaries. And the observation is that specializations have fostered a number of interactions as disciplinarians approach one another's borders, and that most border crossings occur at the level of specialties and not at the boundaries between *entire* disciplines (Thompson Klein-96, p.42). Here interdisciplinarity seems to depend on specialization, and the deeper and narrower it gets, the better the conditions for synthesis become. In this perspective, the units of topical specialization turn out to be the number one scientific integrator, challenging the artificial integrity of disciplinary borders. The corollary is that specialized basic research is not the antithesis of complex system science, but its foremost precondition.

Specialization and integration

The implication is that the Gribbin syndrome, *knowing everything about nothing*, may give rise to a new kind of scientific attitude in terms of *interdisciplinary curiosity* and *skill* in terms of *interdisciplinary practice*. The reasoning goes like this: No discipline is engaged with the entirety of another discipline. Disciplines actually interact at the trading zones to be found in many locations between units of topical specialization. At these zones disciplines either overlap, touch, mix or merge, easing interdisciplinary exchanges and interactions when it comes to concepts, methodological tools, insights and theories. As the units of topical specializations become deeper, smaller and narrower, the mental/intellectual distance between the units of topical specialization of other disciplines pertinent to the object of study becomes less and less. The smaller the unit, the greater the likelihood that disciplinary boundaries will be transcended to highlight the complexity of a particular unit of specialization. In this vein of thought, the emergence of complexity leads to the gradual erosion of boundaries of the special branch (Pilet, p. 634). The specialities have become vital sites for cross-fertilization between disciplines – for interdisciplinary exchange and integration.

By the time a scientist knows "everything about nothing" he or she has actually exhausted the possibilities within his/her own discipline to learn more. The only way to expand one's own understanding is to prey and harvest on the turfs of adjacent disciplines pertinent to one's own specialization. Specialization in basic research thus holds a definite potential for fostering interdisciplinary research on the tiniest topics. The cross-fertilization between disciplinary units of topical specialization may, in due time, spill over to broader areas of expertise and gradually provide more favourable prerequisites for interdisciplinary research on grander scales. As was the case at the time of Galileo, generalizations were possible because the amount of knowledge was comprehensible within 'the horizon of one mind.' To a certain extent, the same may apply to the units of topical specialization where the amount of multidisciplinary knowledge is comprehensible within the horizon of at least teams of

specialists. Here science may have come full circle. Thus, the Gribbin syndrome may prove to be an effective medication for modern science to foster wholeness from particulars.

References:

Barry Barnes, David Bloor and John Henry: *Scientific Knowledge. A Sociological Analysis*, Athlone, London, 1996

John Gribbin: *Almost Anyone's Guide to Science*, Yale University Press, New Haven, London, 1998

Johan Hjort: *The Unity of Science. A Sketch*, Gyldendal, London, Copenhagen, Christiania, 1920

Julie Thompson Klein: *Interdisciplinarity. History, Theory and Practice*, Wayne State University Press, Detroit, 1990

Julia Thompson Klein: *Crossing Boundaries. Knowledge, Disciplinarities and Interdisciplinarities*, University Press of Virginia, Charlotteville and London, 1996

Bruno Latour: *Pandora's Hope. Essays on the Reality of Science Studies*, Harvard University Press, Cambridge, Massachusetts, London, 1999

Karl Pearson: *The Grammar of Science*, Everyman's Library -939, J.M. Dent & Sons LTD, London, 1937

Paul-Emile Pilet: "The Multidisciplinary Aspects of Biology: Basic and Applied Research" in *Scientia*, vol. 74, no. 116, 1981

Steven Weinberg: *Dreams of a Final Theory. The Scientists Search for the Ultimate Laws of Nature*, Vintage Books, last edition, New York, 1993

Jonathan Weiner: *The Next Hundred Years*, Bantam Books, New York, 1990

JERKER RÖNNBERG

Cognition, Communication and Disability

The basic premise of this paper is that disability research – as a problem-based domain with high interdisciplinary demands – needs at least two types of knowledge integration for its future development of methods and concepts: vertical and horizontal.

A meta-theoretical choice which has been fruitfully applied within the Swedish Institute of Disability Research (SIDR) is what is called a critical realist perspective (cf. Bunge/Harré/Bhaskar). One important feature of critical realism is that it supposes that reality is stratified in different levels of description and explanation, none of which can easily be reduced to any other. The ontological premise thereby deviates from an empiricist view which advocates a more direct, one-to-one truth correspondence between reality and concept; it also deviates from a too one-sided social constructivist conception of disability-related phenomena, where the status of the relation between concept and reality is constructed and relativistic (Danermark, 2003).

Professor Jerker Rönnberg,
Department of behavioural sciences,
University of Linköping, Sweden.
CAS Fellow 2003/04.

Although this is true, the critical realist perspective adopted within SIDR includes some features of both extremes: the SIDR-perspective focuses on theory-driven explanations of what causal *mechanisms* may underlie phenomena and behaviour in different contexts and under different conditions. Theoretical integration among levels of explanation therefore becomes one important feature of the research endeavour within the SIDR. This especially applies to phenomena in everyday life and to research on participation in everyday activities. A further consequence of this general approach is that greater interdisciplinary cooperation across faculty borders will be required in order to generate satisfactory descriptions of disability and ability. New concepts and new methods, and re-combinations of existing concepts and methods, will have to be advanced to provide concepts that have the potential of connecting different levels of explanation. This kind of *vertical knowledge integration* is one means of proceeding in disability research. *Horizontal knowledge integration* is complementary to vertical knowledge integration and proceeds by attempting to test the generality or precision of concepts across disabilities and social contexts. The notion of working memory was proposed as one example of a concept that has proven useful in both the "vertical" and "horizontal" senses.

Complex working memory capacity, that is, the ability to maintain and process several pieces of information "on-line" for purposes of problem-solving or language understanding, represents a crucial cognitive resource in many cases of communicative disability. The example of GS was intro-

duced. GS is a deaf person who has developed a method for speech understanding that we call *tactiling* (Rönnberg, 1993). Tactiling denotes that GS uses his hand to pick up vibrations from the collarbone and neck of the speaker, which he combines with what he can visually extract from lip movements and facial expression. When GS became deaf (after having meningitis at the age of 8), he spontaneously learned that when he held his hand on the *throat* of his mother, he suddenly understood her lip movements when she read stories to him. Subsequently, he made the social adjustment of placing his hand on the *shoulder/neck* of the person he is communicating with.

What is remarkable about GS is that he is able to communicate at a nearly normal speed by means of tactiling. This kind of expertise is in part explained by his highly capacious working memory – a feature of his cognitive profile that he has in common with two other experts from our database on cognitive descriptions of individuals who vary in speech understanding skills (Lyxell, 1994; Rönnberg et al., 1999). Working memory resources are assumed to aid in retrospectively resolving ambiguities in a dialogue that are due to information being misperceived or simply missed due to the hearing loss, a process which is achieved by mentally filling in missing pieces of information on-line, and to serve as a basis for predictions of future exchanges in the dialogue. To be able to achieve that with efficiency, a large simultaneous processing and storage capacity is demanded. GS is extremely well equipped in this respect.

In vertical terms, GS happened to discover a method by which he could optimize the perceptual qualities of speech understanding. However, the *perceptual level of description and explanation*, i.e., the relative efficiency of the combination of tactilely and visually perceived phonemes, is not sufficient. When perceptual efficiency is combined with high working memory capacity, spoken elements can be processed even more effectively, given the functional role of working memory in dialogue comprehension suggested above. Still, this *cognitive explanation* is not sufficient to grasp the way GS communicates via tactiling. At a *social level of explanation*, it is easy to observe that GS has a well-developed strategy of approaching the talker, starting to communicate with him or her, while rather unobtrusively placing his hand on the speaker. Rarely do people understand that GS is deaf until after a several encounters. His way of behaving demands social competence, otherwise he risks being misunderstood. Again, it may be the case that he is supported by his capacious working memory, allowing him to be strategic, minimizing the number of repetitions and clarifications needed, and perhaps also minimizing what is needed in terms of "hands-on" tactiling. Thus, there are several levels of description and explanation of the tactiling method used by GS. A tremendous general working memory capacity has the potential of providing an explanatory mechanism that connects several levels of description and explanation.

Horizontally, it can be stated that this general mechanism serves to characterize and generalize to other expert speech-readers with other communicative habits and backgrounds: Case MM (Rönnberg et. al., 1999), who is a native bilingual, and Case SJ (Lyxell, 1994), who is a pure visual speech-reader. GS and SJ have post-lingual onsets of deafness (at 8 and 13 years, respectively), while MM has a congenital moderate hearing impairment.

We know that the concept of working memory capacity has been successfully used as a compensatory mechanism that can be trained in children with attention-deficit hyperactivity disorders (Klingberg et al., 2002). Training of working memory and its components has also been used for intervention in children with dyslexia and children with cerebral palsy. Working memory is a central concept when it comes to comparative studies of speech, sign and script understanding, both neuro-physiologically and cognitively (Rönnberg, 2003a,b), and it has been applied and generalized to other conditions such as Parkinson's disease and multiple sclerosis.

In a similar vein, other concepts, such as theory of mind (ToM) (Peterson & Siegel, 2001), have been used successfully in horizontal and vertical senses. The ability to grasp other peoples' intentions and reasons for communication and behaviour is fundamental to all role-taking in dialogues, and the lack of this ability has traditionally been one of the diagnostic features of children with autism. However, the study of ToM capabilities in deaf children and children with cerebral palsy has shown that a lack of ToM capabilities is due not to a kind of neurally defective module situated in the frontal lobes, but to a lack of opportunity to take part in meaningful social exchanges. Deaf children who use sign language as a native language do not have problems with ToM tasks, but children with cerebral palsy who do not have an intelligible articulation of speech, do show problems. Without true communicative opportunities, lack of feedback to one's own thoughts and emotions will in the long run also jeopardize one's ability to "read" and understand other persons' thoughts and emotions.

Many other examples exist in the current interdisciplinary development of disability-related concepts. It is ventured that a continuous interplay of vertical and horizontal explanatory mechanisms will help develop disability research towards a more mature knowledge domain. This is also the long-term goal of the SIDR.

References

Danermark, B. (2003). Different approaches in assessment of audiological rehabilitation: a metatheoretical perspective. *International Journal of Audiology*, *42*, S112–S117.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clicical and Experimental Neuropsychology*, 24, 781–791.

Lyxell, B. (1994). Skilled speechreading – a single case study. *Scandinavian Journal of Psychology, 35*, 212219.

Petersen, C.C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf, and normal hearing children. *Psychological Science*, *10*, 126–129

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clicical and Experimental Neuropsychology*, 24, 781–791.

Rönnberg, J. (1993). Cognitive characteristics of skilled Tactiling: The case of GS. *European Journal of Cognitive Psychology, 5*, 19–33.

Rönnberg, J., Andersson, J., Samuelsson, S., Söderfeldt, B., Lyxell, B., Risberg, J. (1999). A speechreading expert: The case of MM. *Journal of Speech, Language and Hearing Research, 42*, 5–20.

Rönnberg, J. (2003a). Working memory, neuroscience and language: Evidence from the deaf and hard of hearing. In M. Marschark & P. Spencer (Eds). *Oxford Handbook of Deaf Studies, Language, and Education*: Oxford University Press.

Rönnberg, J. (2003b). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology, 42*: S68–S76.

JOHN PERRY

Personal Identity, Memory and the Self

In this paper, I will contrast two concepts of personal identity, the philosophical and the psychological. Then I'll develop my account of self-knowledge. In the course of this I'll explain the concept of memory that I think is crucial to developing our sense of ourselves as persisting beings, and contrast it with other ways we can remember things about ourselves. I'm very interested in the light our colleagues here in the Center involved in the memory project can shed on this distinction.

The Philosophical problem of personal identity

Professor John Perry,
Department of Philosophy,
University of Stanford, USA.
CAS Fellow 2003/04.

The traditional problem of personal identity for philosophers is this: under what conditions are Person A and Person B one and the same person? This can be a practical problem because we have inadequate knowledge of events. The practical problem of personal identity often arises in the judicial system. The prosecutor claims that the defendant, the person sitting in the courtroom, is the very same person who committed the crime, at a different place and a different time. The problem confronting the jurors is one of knowledge, of knowing the facts; it is, as philosophers say, epistemological.

If the jury had a complete video of everything that happened in all the relevant parts of the world – maybe this would require more than a video, perhaps some assemblage of hyperlinked digitized videos produced by a system of video cams spread throughout America as a part of some future edition of the Patriot Act – they could probably be quite sure of the right answer. They would just rewind the video until they got to the crime, follow the movements of the criminal on the video or linked videos covering the different regions of the world into which he wandered, and see if the criminal ended up coming into the courtroom and sitting at the defense table.

The philosopher is more likely drawn to what might be called metaphysical issues, issues that may remain after all of the facts are, in some sense, known.

Suppose that as the jury follows the career of the criminal, call him Roscoe, he does the following. He goes to a completely up to date brain-science facility, where brain scientists have developed a technique for duplicating brains. The hope is that a person with some brain deterioration can have a new brain manufactured, made of sounder material, which will be psychologically indiscernible from the original. That is, when replaced, the new brain will give rise to the same beliefs and desires and memories and intentions as the old one; the headaches will disappear

and the once inevitable strokes won't occur, but the intentionality will be the same as before. Roscoe has his brain duplicated. He has his original brain and his body destroyed and the duplicate brain put into a different body, Jeff's body. Jeff has just been declared brain dead, although his other organs are in fine shape. The criminal actually did this just to confuse things and make it hard to trace his movements. He swears the neurosurgeons to secrecy, but they don't cooperate.

The survivor of this operation leaves the hospital and ends up in the courtroom. He admits having memories, or at least something very much like memories, of committing the crime. But his lawyer claims that the criminal actually slipped up, and committed suicide. A human being is an animal, and this is a different animal, a different human. The defendant is actually Jeff, with a brain transplant. He is no more the criminal than he would be had he gotten the criminal's liver or heart. What we have here, the lawyer argues, if Jeff, a man who had a terrible injury, and who, though saved by a miracle, has lost all of his memories, in their place having delusions of a criminal past. Jeff is to be pitied, not punished. He calls some philosophers as expert witnesses (paying them less, no doubt, than other expert witnesses charge)–Bernard Williams say.[1]

The prosecutor is undeterred. He also calls expert witnesses, perhaps John Locke or Sydney Shoemaker. They explain that our concept of a person is not really a concept of an animal, but of a certain sort of informational-action system, one that our person theory fits. These philosophers maintain that the person theory actually gives us a new concept of a continuing thing, one that conceivably could breach the bounds of bodily identity. Persons are systems that pick up information from experience, develop and sustain goals, and apply the information to achieve their goals. Such systems require a certain causal basis, some hardware on which the relevant data is stored and the relevant programs run. Usually this is provided by a single human body. But that is not a necessary requirement. Look, he may intone, we recognize the possibility of having the same person without having the same body when we talk of survival in heaven or hell, or reincarnation. These may be religious fantasies, but they show that it at least makes sense to have the same person when we don't have, in any ordinary sense, the same body or the same animal. Our criminal figured out a way of surviving the death of his body. The defendant is not Jeff, with a new brain and delusions, but Roscoe, with a new body and a duplicate brain.

Identity

Let's spend a little time on the concept of identity.

Identity versus Similarity

The concept of personal identity is a special case of what is sometimes called numerical identity. The relevant concept of identity is expressed in various ways, "are identical," "are one and the same" etc. If X and Y are identical, in this sense, there is just one thing that is both X and Y. So if the cows Bossie and Trixie are one and the same, if they are identical, then there is just one cow, called both "Bossie" and "Trixie." English is confusing in various ways. Almost all the words for numerical identity are also used to convey similarity. E.g., imagine now we have two cows, one named "Bossie" and the other named "Trixie". They are both guernseys,

both give the same amount of milk, both are somewhat ornery when milked. We might say, "Bossie and Trixie are the same," meaning that they are very similar or very much alike. Maybe the farmer liked Bossie so much he went looking for as similar a cow as he could get, he wanted one just like Bossie. We might say he wanted the "same cow" or even "the identical cow."

Note that in the numerical sense of identity, the sense in which there is just one thing, the idea of identical twins makes no sense. If they are identical, they are not twins; if they are twins, they are not identical. "Identical" in "identical twins" doesn't mean numerical identity, but similarity, or perhaps coming from a single egg.

Logical Properties of Identity

From now on I'll use "identity" in the sense of numerical identity unless I indicate otherwise. The logical properties of identity are simply consequences of the idea of just being one thing. For example, if you just have one thing, it has all the properties it has:

- If x is identical with y, and y has property P, then x has property P. [The indiscernibility of the identical]

  Further:
- If x is identical with y, y is identical with x (Symmetry)
- If x is identical with y, and y is identical with z, then x is identical with z (transivity)
- Everything is identical with itself, that is, for all x, x is identical with x (Reflexivity)

Identity and Time

The Greek philosopher Heraclitus got tenure for saying that you can't step in the same river twice, because new waters are always flowing in. This is deep and profound, but not quite right. Of course you can step in the same river twice, although as you do so, you won't be stepping in exactly the same water, at least if the river is flowing at any rate at all.

If we just say that when you step in the same river at two different times, it will not be exactly similar as it was before, it doesn't sound quite so profound.

Suppose that the Cayster is full of muddy water on Monday, but clear on Tuesday. Then don't we have the problem? How can one river have different properties at different times, given the principle we called the indiscernibility of the identical?

We just have to be careful. The same river has the property of containing muddy water Monday, and also the property of containing clear water Tuesday. If we include the time in the property, there is no problem.

Even if we speak in the normal tensed way, there is no problem if we are careful. The principle of the indiscernibility of identicals implies,

If x and y are identical, x has all the properties y has, and x had all the properties y had, and x will have all the properties y will have. But it doesn't imply, If x and y are identical, x *has* all the properties y *had*..."

Suppose Heraclitus stands in the clear Cayster on Tuesday, and says, "I stepped in this very river, the identical river, one and the same river, yesterday, and then it was muddy." From this he can infer that the river he is standing has clear water, and had muddy water, the day before, and that

the river he stood in yesterday had muddy water in it then, and has clear water in it now." But he shouldn't have concluded that it can't be the same river he is standing in today as he was standing in yesterday.

Continuity, Causation and Identity

The concept of identity is applied to everything, concrete objects, abstract objects (like numbers and properties), contrived objects (like the sequence consisting of the Eiffel Tower and Bob Dylan), clouds, wind currents, and so forth.

Persons belong to the very general category of concrete things, things which have a position in space and endure through time. It is often thought that the identity conditions of concrete things amount to spatial temporal continuity. Why is the coin in my pocket now the same one I put in there this morning? Because there is a spatio-temporal continuous path that stretches from spatiotemporal position of the coin this morning to the spatio-temporal position of the coin in my pocket now, and every point along this path is or was occupied by a coin. This is certainly something we at least expect of concrete objects, and it is the reason we usually think we can establish identity by establishing such a continuous history – as we imagined our jury doing in the case of Roscoe the criminal.

For most concrete things there is also an element of direct causality built into our concept. Technology provides a lot of ways of giving the illusion of a concrete thing although what we really have is a spatio-temporal connected succession of different things, made to provide the illusion of a single thing. For example, if I type an "s" in this file, and then go back and insert some spaces, I will think of the "s" I type as moving to the right along the line. This "s" isn't really a single concrete thing, but a succession of things made to give the appearance of a single thing. (Of course, it is a single *succession*, but a succession isn't a concrete thing, and a succession of "s"'s isn't an "s"). The similarity of the first s and the second s doesn't result from the usual sort of direct causality that makes a concrete thing look pretty much the same from instant to instant, even if it moves a little. Rather, one thing is annihilated and another put in its place by the editing program. I'll call this virtual identity.

In the case of the succession of letters, we don't really have continuity. That would require that between any pair of s's in the series there was another overlapping s. So maybe we can distinguish between virtual identity and real identity on that basis. On the other hand, are we sure that we really have continuity in the case of ordinary objects? It isn't really something we can observe. If the scientists at SLAC or CERN tell me that we don't really have temporal continuity, but that the careers of physical objects turn out to be full of little temporal gaps, I'd have to believe what they say. So I think we need to appeal to a concept of direct causality. The position, and the characteristics, of each successive stage of a physical object are explained by the position and characteristics of the earlier stage.

Ordinarily, we expect concrete things to change in gradual ways, unless there is a particular event that results in a lot of changes. I expect the coin in my pocket now to look pretty much the same as the one I put in my pocket this morning. Of course, if some time during the day I took it out and put it on a railway track and let a train flatten it, then it won't. That change will be explained, however, by the way the coin was, and the pres-

sures that the train exerted on it. The careers of concrete objects have a characteristic shape, each stage explained by how they were, and what happens to them.

This applies to humans in their physical aspects. You will expect me to look pretty much the same tomorrow as I do today, unless I get run over by a car or undergo cosmetic surgery or something like that. The similarity isn't due to some outside agency or program that is keeping track of how the successive John Perry's the worlds sees ought to look. It's just a consequence of the way people develop. Of course if people look too much the same as earlier stages of themselves, where the earlier stages are considerably earlier, that also requires explanation. If the person in question lives in Los Angeles, we assume cosmetic surgery.

Our concept of the identity of a person fits into this general scheme, even though the psychological characteristics of persons, their beliefs, desires, and traits, are much different sorts of properties than the shapes and sizes and appearances of (merely) physical things. Even if we adopt a Lockean theory of personal identity, and allow that we may have the same person even if we do not have the same animal, or as Locke puts it, allow that we can have the same person when we don't have the same man, we will have not abandoned entirely our ordinary conception of identity as grounded in the direct causation of basic similarities or explicable differences in the important properties of the object in question.

Psychological identity

Now I want to consider a different, and perhaps more common, sense of "personal identity."

When a psychologist or an ordinary man (i.e., not a philosopher) talks about the identity of a person they do not have in mind mainly something that could be decided by fingerprints or a driver's license picture, but an enduring structure within the person, his or her own individual combination of beliefs, goals, habits, and traits of character and personality, the pattern that as we might say, *makes* the person who he is.

Of particular importance is the sense the person has of himself. What properties does this person think are true of him? Which ones are most important to him? How does he see this as fitting into a narrative of his life? A psychologist might have a person rank the properties he or she takes himself or herself to have in importance. Which properties can they not imagine not having? Can this man imagine being a woman? Would it matter a lot? Can this philosopher imagine being an accountant? Can this neuroscientist imagine being a philosopher? Does this mother find it incomprehensible that she should not be a mother, or is it an accident in her life? Would being different in these ways destroy a persons's sense of who she or he is, and fracture the narrative of her or his life? Or could they be accommodated within the basic picture of himself that the person has? The most important, basic, inalienable facts about a person are more or less what the psychologist might think of as his or her identity.

Selves and the sense of identity

A word we often use in connection with a person's identity is "self". The concept of self involves both philosophical and psychological identity.

Some philosophers think of selves as rather mysterious immaterial entities. Sometimes selves are identified with the souls of Christian

theology, or the essential natures that are passed along in reincarnation, or some noumenal object that exists beyond normal space and time, outside of the causal realm, and joins, in some Kantian way, with the primordial structure of reality to create the world as we know it. I don't think such mysterious notions of the self are required to understand the person theory. I think that a self is just a person, thought of under the relation of identity. But that sounds mysterious enough, so let me explain.

Consider what it is to be a neighbor. A neighbor is just a person, thought of as having the relation of *living next to* to some person in question. A teacher is just a person, thought of as having the relation of "teaching" to some student. A father is just a person, thought of under the relation of *being the father of*. People play important roles in other people's lives, and we give these roles titles: neighbor, teacher, father, spouse, boss, and so forth.

But we play an important role in our own life. I have a relation to myself that I don't have to anyone else, identity. Self is to *identity*, as neighbor is to *living next door to*. It is a way we think of ourselves. The basic concept of self is not of a special kind of object, but as a special kind of concept, that we each have of ourselves.

We each have a very special way of thinking about our self, that is, thinking about the person who we are, via the relation of identity. We have a *self-notion*, a concept of ourself as ourself. I want to say a bit about this key concept, about a person's sense of who they are, of their own identity.

Perhaps its a little unclear what I'm looking for. Sometimes the best way to find something is to first consider a case where it is absent, and then see what is missing.

Castaneda's war hero

Now a sort of paradigm case of someone who doesn't know who they are, and in that sense lacks a sense of identity, and has a diminished self-concept, is someone who has amnesia. Here I am thinking of a certain kind of amnesia, which may only exist, in its most perfect and full-blown state, in fiction and in philosophical examples. This is a person who, as a result of a bump on the head, has no idea who they are. One assumes that the knowledge is somewhat still in the brain, waiting to be released by another fortuitous bump on the head, or maybe surgery, or maybe just time.

I'll use an example from the great late philosopher Hector-Neri Castaneda. He imagines a soldier – call him Bill – who having performed many brave deeds in a certain battle, is injured, loses his dog-tags, and awakens with amnesia. Not only does he not know who he is, no one else does either. He is clearly a soldier, however, and clearly due all the rights pertaining thereto, so he is hospitalized, cured of everything but his amnesia, and goes to Berkeley on the GI Bill. In the meantime, Bill's feats during the battle have become well-known. People don't know what became of him and assume he is dead and his body unrecovered somewhere. He is awarded many medals posthumously.

For the time being let's concentrate on Bill, lying in the hospital, not knowing who he is. Now of course there is a sense in which he *does* know who he is. He can say, "I am me." Suppose Bill feels a pang of hunger, and sees a piece of chocolate cake on the tray in front of him. Does he

wonder, into whose mouth this morsel should be put, in order to relieve *his* pang of hunger? No. He knows that he is the person who is feeling the pang of hunger, and the person whose arm he can control more or less at will, and the person who has a mouth which he can't see right below the nose the tip of which he can see, and he knows how to direct the fork and the cake into that mouth. He knows that he is sitting in a room on a bed, with a window out onto a lawn, maybe with a radio and some magazines on the stand beside him. So, he really knows a great deal about himself. Still, compared to the rest of us, he has a very diminished sense of self. He doesn't have memories from which he can construct a narrative about why he is where he is. He doesn't know what values, what commitments, what beliefs, what actions led him to this hospital room.

Also, since he doesn't know his own name, he can't exploit *other people's* knowledge of who he is. He can't exploit public sources of information about himself. This is something we all rely on. If I forget my phone number, I can look it up in the Directory. I find out something about myself in exactly the same way as you would find out the same fact about me. Indeed, there are lots of things that make it into the public conception of us, that we can't discover in any other way.

In contrast, all of the knowledge Bill has about himself, in the hospital (or almost all), he acquires by what I will call, somewhat ponderously, "normally self-informative ways of knowing about a person". That is, when you see an object by holding your head erect and opening your eyes, the object you see will be in front of someone. Who? You. Normally, at least, this is a way of finding out what is going on in front of the person who is doing the seeing. If you feel a pang of hunger, someone is hungry, and will have their hunger relieved if food enters their mouth and makes it to their stomach. Who? You.

Why do I say "normally"? Maybe some day brain scientists will invent a little device that will send message from one person's eyes to another person's optic nerves, so that the second person can directly see what is front of the first. This might have some military utility. Old, frail, jittery, demolition experts can guide the movements of young, healthy, steady, inexperienced ones, as they defuse bombs. These experts will then have a cognitive burden that is not placed on most of us. They will have to keep track of whom it is they are getting information about the immediate environment of visually. Most of us don't have to do that.

Now consider Bill's act of extending his arm, grabbing his fork, breaking off a piece of cake, and shoving it in his mouth. I'll call that a "normally self-effecting way of acting". Moving in that way is a way anyone can shove a piece of cake they see in front of them in their own mouths, a way of feeding themselves. Again, normally, because we can dream up cases where it wouldn't work.

I'll repeat my favorite example here. At the end of Alfred Hitchcock's movie "Spellbound" J. Carroll Nash holds a gun pointed at Ingrid Bergman, who is leaving his office, having just exposed his plot to frame his patient, Gregory Peck, for murder. We know who Nash will shoot if he pulls the trigger: the person in front of him. Shooting a gun pointed like that is a way of shooting the person in front of you. Then we see Nash's hand turn the gun around. The front of the gun barrel fills the whole screen. He fires. Whom does he shoot? Himself. Firing a gun held like that is a normally self-shooting way of acting. But suppose that Nash had a

donut-shaped head. Then it would be a way of shooting the person behind him. It's only a contingent fact that we don't have donut shaped heads. That's why we need to say "normally."

So Bill, even with his amnesia, has a good deal of self-knowledge, in a perfectly reasonable sense.

Bill proceeds to Berkeley, where he ends up getting a graduate degree in history, writing, for his dissertation, a biography of the war hero who gained his fame at the very same battle from which Bill woke up with amnesia. He doesn't figure out for quite a while that he is the war-hero, that his dissertation is actually autobiography.

Now the point of this is that Bill knows a great deal about a person, who happens to be him. In a sense, he knows a great deal about himself, for he knows a great deal about a certain person X, and he is X. But that's not what we would ordinarily say. We would say something like this: Bill knows a great deal about the person he happens to be, but he doesn't know much about himself.

Types of memory

In fact, even when Bill finally figures out that it is him he is writing about, we might be reluctant to call what he is writing an autobiography. One important thing Locke emphasized was that we have a special access to our own *past* thoughts and actions. We remember them – but we can remember the past thoughts and actions of others, too. I can remember that Elwood used to think that poison oak was edible; I can remember the time Elwood ate some poison oak.

But in the case of my own thought and action, I not only remember that someone did something, or that someone thought something. I remember thinking and doing things. Shoemaker calls this remembering from the inside. Our access to our own past thoughts and actions is phenomenologically and logically different than our memories about what others have thought and done. Remembering what one did and thought isn't *like* remembering what someone else thought and felt. And in the case of others, there is always the question of *who*? I remember someone eating poison oak, but was it Elwood? But if I remember eating poison oak, it was me that was doing the eating.

Once Bill figures out that he is the war hero, he can assimilate all the facts he has learned about his own to past into his own self-notion, his own conception of who he is. But he still won't be related to these things in the normal way, the way we expect of an autobiographer. He will know that he did these things, but he won't remembering doing them.

A similar distinction applies to our knowledge of what we will do in the future. I can know, or at least have a pretty well-grounded belief, what you intend to do, what you will do. But when I know what I am doing, what I am trying to do, what I intend to do, and in those ways, what I will do, it is based on a different way of knowing, a way each of us knows something of his own future; again, it is knowledge from the inside.

A case like Bill's is pretty fantastic, but the underlying moral is generally applicable. It is a fact about the complex informational world we live in, that we have lots of ways of getting information about ourselves that are not normally self-informative.

The notion that Bill was able to have of himself, even when he didn't know who he was, was his *self-notion*. Self-knowledge, in the ordinary sense,

is knowledge of ourselves attached to our self-notion. Knowing facts about the person you happen to be, as Bill did when he wrote his dissertation, isn't enough. If we know who we are, if we know our own names, we can incorporate what others notice and know about us into our own self-conception. We do this all the time. And in fact most of us are very concerned about what we might call our *public identities*. This is the shared conception of us, that others have. It is what our mothers and fathers and sons and daughters and colleagues and bosses and employees think of us. It is what is written next to our names in the newspaper or the college catalog, or on the vita on our web page. For many issues, it is a better source of information about ourselves than any normally self-informative method of knowing.

In fact, for many of us, perhaps for most of us, some very important building blocks of our own identity, our own self-conception, come from the outside, from assimilation into the "I" of the "me"; that is, by adopting as part of our self notion opinions about ourselves that originated with the insights, or mistakes, of others. My parents tell me that I am like my grandfather, that I am a thinker not a doer, and that becomes part of my self-conception.

As we construct our public identities, we rely on the help of others. Public identities are a bit like works of art, or publications; they are accomplishments, that take on a life of their own. And of course they need not be unique. I may be one person in the eyes of my surviving cousins, who meet every so often in Nebraska and reminisce about our grandmother and grandfather, and uncles and aunts and parents and each other. A somewhat different person in the eyes of my colleagues. And so forth. My self-conception, the picture of myself that animates me and explains how I act and react, may change subtly, or not so subtly, in different situations.

So I have a sense of my own identity. Here we see this other use of "identity". What is my identity? It is my own self-concept, the things I think hold true of me. A lot of this information I get from present perception: I think I am sitting in a chair, typing on a lap-top, listening to dixieland music, looking out the window at a rainy day. Some of it I have from memory. And some of it I have from what others have told me about myself, and from applying general information about people to myself.

Let me close by reiterating the basics of my account of self-knowledge:

- Each person has a special, dedicated, notion, his self-notion. This notion collects information acquired in normally self-informative ways, knowledge about his own mental and bodily states, and about what the world around him is like, and what he has thought and done in the past, and will do, or at least plans to do, in the future.
- Our self-notions also serve to collect information we get about ourselves in other ways, as long as we recognize that it is ourselves that the information is about. I read in the email notice of the conference what time I will be giving a paper, and where. I pick up information about myself under the name "John Perry" which is the same way that others get information about me.
- Normally we expect a person to have a very complex self-concept, full of things that he has learned about himself in the past, both in normally self-informative ways and as a result of what others tell him about himself. We expect his desires and goals to be based not simply

on urges and needs that he has now, that he can discover by present feeling and introspection, but also on memories of the past and goals adopted in the past.

- All of our actions are ultimately motivated by information that is stored in, or connected with, our self-notions. This information can motivate normally self-effecting actions. And all of our actions, however unselfish, and however remote we intend their consequences to be, come down to moving our limbs and other bodily parts in various ways, intended to bring about wider and wider changes, in virtue of the circumstances we are in.

References

Perry, John. 2000. *Knowledge, Possibility and Consciousness*. Cambridge: MIT Press.
Perry, John (ed). 1974. *Personal Identity*. Berkeley: University of California Press.
Perry, John. 2002. *Identity, Personal Identity, and the Self*. Cambridge: Hackett Publications.

Notes

1 Essays by Williams, Locke, and Shoemaker on personal identity can be found in Perry, 1974.

JONATHAN KNOWLES

Varieties of Naturalism[1]

Many philosophers call themselves naturalists and mean by this that they have the goal of accommodating traditional categories of philosophical inquiry – belief, consciousness, personhood, knowledge, free will etc. – within the naturalistic or physicalistic world-view of modern science. Within this camp, we find much variation in detail. One important source of variation consists in what is regarded as the legitimate concepts of natural science in terms of which one might effect a reduction of the (as one might generally refer to them) mental concepts, some philosophers allowing only extensional concepts (i.e. no modal notions like 'necessarily'), some also causal-nomological ones, some these plus natural teleological ones. Another important source of variation concerns the status of the proposals offered (or that these philosophers recommend should be offered – here we have a further source of variation that concerns whether one argues for the programme or actually tries to carry it out!). Thus, some see the reductive theories as upshots of conceptual analyses of mental concepts, whereas others (probably the majority these days) see them more like empirical theories (this does not have to mean 'bridge laws' in the manner of so-called 'classical' reductionism). What all these philosophers have in common, however, is their belief that there is an important reductive project to be carried out, and that philosophy's and/or cognitive science's chief role should be to carry it out.[2]

Professor Jonathan Knowles,
Department of Philosophy,
Norwegian University of Science
and Technology, NTNU.
CAS Fellow 2003/04.

Many other philosophers are not naturalists in this sense because they believe that there are certain features of our mental life that resist capture by scientific ideas. This is something we can know to be the case by reflection on our everyday modes of explanation and understanding, a reflection which enunciates a special place for the mental as autonomous from the understanding of the world that science gives us. This second kind of philosopher should be immediately contrasted with those who agree that science will never explain certain features of our mental life, but that this is due to inherent, though ultimately contingent limitations in the cognitive capacities of human beings: Just as rats will never understand calculus, so we will never understand (to take three central examples) how pain can be a brain state, how scientific belief-formation can be computationally tractable, or how language-use can be both creative and appropriate to circumstances. The thinkers who have argued for this kind of 'cognitive boundedness' have not seen it as implying the availability of an alternative mode of understanding of the things we will never understand scientifically. This is just what the second kind of philosopher thinks is available. However, many of these would also balk at being called non-naturalists, carrying with it as it does implications of belief in supernat-

ural entities or tracts of reality. As they see things, there is no reason not to see our common sense explanatory practices, centrally involving persons, mental states and their actions, as latching onto thoroughly natural features of the world – namely persons, their mental states and their actions – and no reason to think that the felicity of doing this is dependent on, or can be further illuminated by, some deep theory about what these things are – a theory that at best could show our common sense practices to be more or less in order. This way of thinking is also apt to stress the distinction between *Verstehen* and *Erklären* introduced by 19th century German anti-positivists as a way of backing up the distinctiveness of common sense psychological explanation without reneging on a materialistic ontology.[3]

My preferred variety of naturalism is a middle way between these two. Like the first kind of naturalists (the reductionists), I am impressed by the achievements of modern science, and see the understanding we gain from this enterprise as providing our best hope for knowledge and understanding *tout court*. However, like the second kind of naturalists (the anti-reductionists), I do not see this commitment to science as entailing commitment to the project of reducing mental concepts or phenomena to physical ones; thus, unlike the reductionists, I do not take it for granted that cognitive science's or philosophy's proper aim is giving significantly reductive theories. Such reductive projects, though certainly a part of science, are by no means exhaustive of or criterial for it. Wholesale reduction in modern science is a fairly remote ideal in view of the 'exploded' nature of modern physics and the rise of autonomous biological science. Further, there are (at least in my view) no good metaphysical arguments against the credentials of unreconstructed mental properties – mental properties *as such* – in the natural world. A good naturalist should accept the primacy of science, but also relinquish reduction as the criterion of the real. Doing otherwise will simply smack of *a priori* dogmatism inimical to the scientific spirit.[4]

Unlike the anti-reductionists on the other hand, I do not think there is any way of hermetically sealing off some realm of understanding to which the mental concepts are proprietary – of definitively or *a priori* ruling out the possibility of reduction to some more basic level, any more than there is with any other special science. There are two aspects to my scepticism on this score, one concerning the idea of a common sense understanding or world, the other the distinction between *Verstehen* and *Erklären*. To start with the latter, my scepticism can again be broken down into two sub-arguments: On the one hand, if one seeks to demarcate *Verstehen* (viewed as the understanding proprietary to intentional states and action) from *Erklären* by relating the latter to science and trying to point to various features of scientific theories that *Verstehen* allegedly lacks, one will fail. For example, if it is claimed that scientific theories explain by causal laws, then, if one means strict laws, this will be false, whilst if one means non-strict laws, it will be true, but then also a feature that applies to *Verstehen*. On the other hand, if one seeks to demarcate by pointing to something distinctive about *Verstehen*, such as, as is typical, the *normative* character of its explanations, one will have to contend with the fact that meaningful explanation of actions involving what *ought* to be the case must

always be relativised to the agent's *belief* about what ought to be the case, in a way that makes the normative element *as such* disappear in the explanation.[5]

The idea of common sense also fails to seal off psychology from science. If humans universally employ an explanatory scheme for understanding the mental states and actions of others, this is most naturally seen as a folk theory – alongside that we have of physical bodies – that we absolutely can compare to real science, and maybe seek thereby to correct and/or precisify. Though the latter will probably obtain to a lesser extent than in the case of folk physics (not surprisingly), the adequacy of folk psychology is nevertheless a contingent, empirical fact. Moreover, we can seek to understand the existence of such folk theories in relation to a more inclusive theory about their cultural or evolutionary origin (this does not presuppose reductive ambitions). Finally, the idea of a common sense ontology – that the world of ordinary middle-sized objects, persons and their actions is radically distinct from that physical science describes – is not without plausibility, but it does not insulate common sense from science, for this ontology, or at least a subset of it, is one that some sciences also presuppose and seek to understand – notably biological science. Given this overlap, it would seem odd if for some reason the categories relevant to understanding humans were radically incommensurable with those of the animal world. (Language of course complicates our lives enormously, but the question here is whether there is some wholly autonomous human realm that might demand a special kind of understanding, and to that the answer seems clearly negative.)[6]

In sum, I am neither a reductionist nor an anti-reductionist naturalist, but, I believe, a naturalist in the true sense of the word: We can only seek an understanding of the world within our best ongoing enterprise, which enterprise is roughly that of modern science, including cognitive science. Though successful reductions are part of that enterprise, and part of what makes it exciting, they are not criterial for it. My naturalism is thus compatible with those who hanker after substantive psychological theorising in a way that may also vindicate aspects of our common sense psychology. In the end, however, no idea can effect an absolute divide between science and the mental such that the latter will necessarily require a distinctive and non-scientific mode of understanding.

Notes

1 A more extended exposition of the ideas in this paper was presented at the XV Inter-Nordic Philosophical Symposium, Science: A Challenge to Philosophy? in Helsinki, May 2004, and I would like to thank the audience there for their comments. Thanks also to Jennifer Hornsby for feedback at a seminar at the CAS that has led, I hope, to a dialectically tighter presentation.

2 For different varieties of reductive projects, cf. F. Jackson From Metaphysics to Ethics (Oxford UP 1988), who espouses a conceptualist variety of reductionism); J. Fodor Psychosemantics (MIT Press 1987), who offers a more empirically-based theory of intentional states and content; A. Goldman 'What is justified belief?' (reprinted in H. Kornblith, ed., Naturalizing Epistemology, 2nd edition, MIT Press 1997), who offers a reductionist theory of justified belief; and D. Papineau Philosophical Naturalism (Blackwell 1993), who seeks to give a naturalistic account of just about everything from intentional content to mathematics that nevertheless allows itself the idea of natural teleology.

3 Two of the most influential naturalists of this second kind in the philosophy of mind are John McDowell (cf. his Mind and World, Harvard UP 1994) and Jennifer Hornsby (cf. her Simple

Mindedness, Harvard UP 1997). In epistemology, a somewhat similar view is espoused by Susan Haack in Evidence and Inquiry (Blackwell 1993). The idea that we are cognitively bounded with respect to certain 'mysteries' of the mind has been forcefully argued for by Noam Chomsky, cf. e.g. Reflections on Language (Fontana 1976).

4 My recommendation of a generally scientific outlook is reminiscent of that of W. V. O. Quine's (cf. e.g. his Theories and Things, Harvard UP 1981), but Quine is also an avid physicalist of a kind that destroys much of the attraction and, I think, consistency of his naturalism. The kind of pluralistic view of science alluded to in this paragraph is defended by John Dupré The Disorder of Things (Harvard UP 1993) and T. Crane & H. Mellor 'There is no question of physicalism' (Mind 1990). For arguments against the need to vindicate mental properties, cf. J. Knowles 'Does intentional psychology need vindicating by cognitive science?' (Minds and Machines 2001).

5 These arguments, and others, are presented more fully in J. Knowles 'Is folk psychology different?' (Erkenntnis 2002).

6 Some philosophers, influenced by behaviourist thinking, regard mental phenomena as somehow 'subordinated' to language and linguistic behaviour (e.g. Quine, Dummett, Davidson). This might furnish another way of carving out a special place for mental explanation (assuming the public nature of language) – but only insofar as the mental would thereby fail to be the robustly real phenomenon I am here taking it to be.

JAMES J. ELSER

Biological Stoichiometry: an Ecological Perspective on the Complexity of Cancer

Like many fields of the life sciences, cancer biology is an enormously complex and exponentially expanding field, involving work ranging from the molecular biology of oncogenes to environmental epidemiology. Most people are aware than cancer mortality rates are declining. This is primarily due to public health efforts (e.g. smoking reduction) that have reduced cancer incidence itself. What most people don't know is that survival rates for various types of cancers *once incurred* show only relatively modest improvements during recent decades (Clegg et al. 2002), despite many billions of dollars in expenditure worldwide during recent decades. For some cancers such as lung cancer (Patz et al. 2000, Henschke et al. 1999), no changes in survival rates have been noted despite recent advances in early detection. This situation is somewhat depressing and suggests that cancer biology is in need of new ideas and perspectives to achieve important breakthroughs in the understanding of cancer so that prevention and therapy might be significantly improved. It would be tempting to think that the enormous wave of information emerging in the post-genomics era will provide a solution. While this may indeed be the case eventually, consider the picture in Figure 1 (top), a *very* simplified map of a cellular metabolic system. This map is itself an intimidating picture and the prospect of understanding how the genome, in concert with its interactions with the extracellular environment, produces this map is a task of truly intimidating complexity. To a large extent, then, unlocking the secrets of cancer amounts to finding a needle in a haystack. Seen in this way, the promise of the post-genomics era is the promise that the haystack will become bigger! It might be argued, then, that the key to the way forward in this endeavor is not by generating more data but instead the key is in developing conceptual and theoretical tools that can help make order out of the bewildering chaos of data emerging from the "omics" factories.

Professor James J. Elser,
School of Life Sciences
Arizona State University, USA.
CAS Fellow 2003/04.

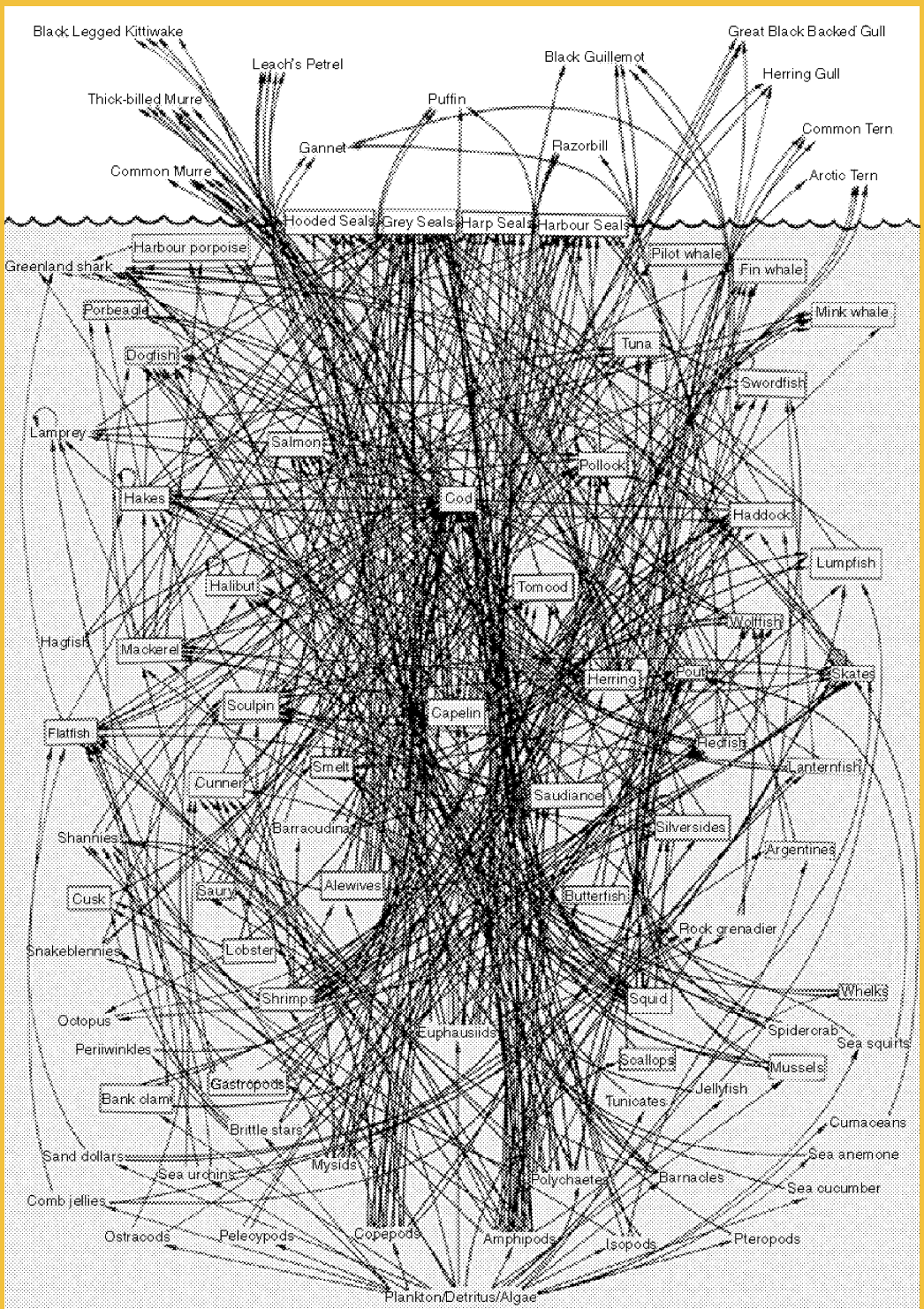
Now consider the picture in Figure 1 (bottom). This is also a greatly simplified map of a network of interactions, in this case the network of feeding interactions in an oceanic food web. Like the metabolic map, many of the details are missing but the picture remains highly intimidating. How can one make sense of such a system in order to better understand its dynamics? Recent work in the fields of ecosystem ecology and life history evolution has produced a set of ideas and analytical frameworks, in the form of "ecological / biological stoichiometry" that

has helped make sense of food web complexity and may help in making sense of cancer. In this paper I'll briefly attempt to show that this stoichiometric framework has relevance for understanding tumor biology. In this view, the goal in cancer therapy is to assure that the host (patient) wins in this ecological competition (that is, the tumor is eliminated) or, at the least, that there is a long-term stable coexistence in which the host maintains an acceptable level of health (that is, damage to normal tissue is minimized). By applying a stoichiometric perspective to better reflect the multivariate material demands and transactions of the players, health care professionals might be better able to turn the tables of competition in the patient's favor. We need to understand the functional ecology of the evolving tumor in its host habitat.

Stoichiometry: into ecology and beyond.

Biological stoichiometry is the study of the balance of multiple chemical elements in biological systems (Elser et al. 2000b). It is an extension of the theory of "ecological stoichiometry", an approach developed in ecosystem ecology to better understand ecological dynamics in terms of the material balance of interacting organisms in the environment (Reiners 1986, Hessen 1997 Elser and Urabe 1999, Sterner and Elser 2002); this framework of ecological stoichiometry has formed the basis of our working group at the CAS. Motivating the development of ecological stoichiometry has been the realization that different organisms can contrast strongly in their elemental composition, with greatest attention given to the macroelements carbon (C), nitrogen (N), and phosphorus (P). It is now known that these differences have major implications for the ecology of these organisms (Sterner and Elser 2002). In particular, P-rich animals are unusually sensitive to the P-content of their food, suffering strong declines in growth and reproduction when consuming food with low P-content and vulnerable to erratic population dynamics and possible extinction in environments that do not supply sufficient P. Thus, the relative stoichiometric requirements of a species appear to be a key aspect of its ecological "niche" (Figure 2A, bottom).

These major ecological implications of body C:N:P stoichiometry have motivated ecologists to ask why, at both the proximate physiological and at the evolutionary levels, different taxa or growth stages are characterized by contrasting C:N:P ratios (Elser et al. 1996, 2000b). A primary hypothesis under consideration by evolutionary ecologists interested in stoichiometric patterns is called the "Growth Rate Hypothesis" (GRH; Elser et al. 2000b). In this hypothesis (Figure 2A), high P-content in biomass (low C:P and N:P ratios) is caused by increased allocation to P-rich ribosomal RNA necessary to achieve rapid rates of growth or development. This implies that species that have evolved high growth rate lifestyles with high P-demands are more likely to face ecological constraints due to insufficient supplies of P from the environment or diet and, thus, that there is an unavoidable trade-off in the evolution of a rapid growth rate strategy. Empirical evidence supporting the GRH is accumulating (Elser et al. 1996, Main et al. 1997, Vrede et al. 1998, Elser et al. 2000a, 2000b). Researchers are also seeking to understand the genetic underpinnings of the GRH. For example, Elser et al. (2000b) propose that particular differ-



A simplified food web for the Northwest Atlantic

Figure 1. The immense complexity confronted by two branches of biology. Right: a greatly simplified map of the network of cellular metabolism (www.genome.ad.jp/kegg/). Left: a greatly simplified map of the network of feeding relations in an ecosystem (Lavigne 1996). Ultimately, food webs (left) are the outcome of dynamic interactions among various organisms that acquire resources from the abiotic environment and each other in order to drive their metabolism (right) and leave offspring. What unites these two figures is that both systems must obey physical-chemical constraints that are captured in stoichiometric theory.

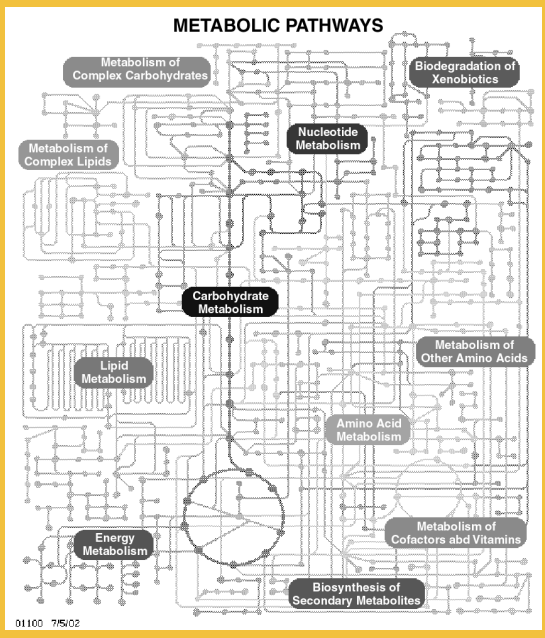
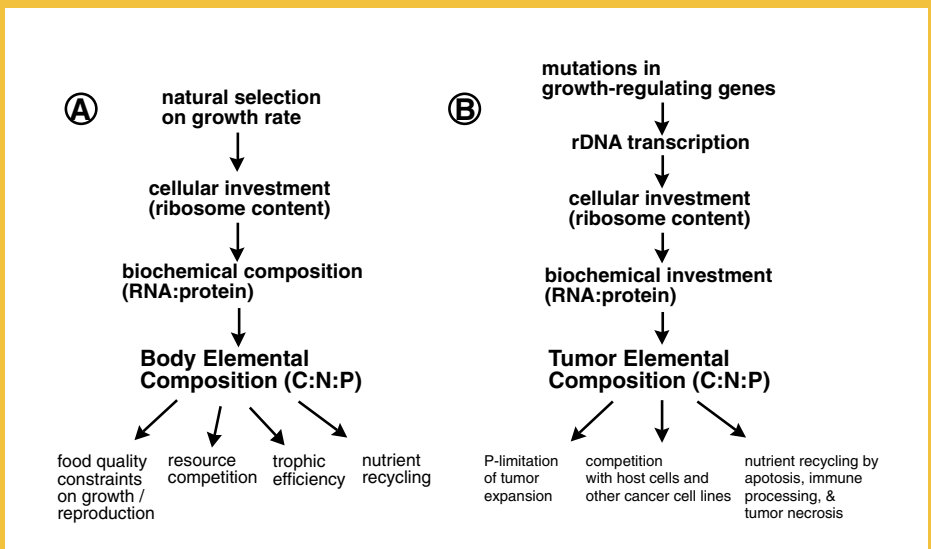


Figure 2. The stoichiometric growth rate hypothesis (GRH) as applied in evolutionary ecology (A) and cancer biology (B). In ecology, the importance of organismal carbon:nitrogen:phosphorus (C:N:P) stoichiometry for ecological interactions is indicated at the bottom of the figure while the biochemical and evolutionary determinants of organism C:N:P ratios are seen at the top. With respect to cancer biology, the ideas and prior findings synthesized in this paper suggest that the stoichiometric approach may also be plausibly applied to better understand tumor dynamics. From Elser et al. (2002).



ences among biota in the structure of ribosomal RNA genes (rDNA) are associated with the ability to produce the high rRNA phenotype necessary for rapid growth.

In our emerging work we are trying to apply the concept of biological stoichiometry, and more particularly, the growth rate hypothesis, to tumor biology (see Figure 2B). Since tumors generally are tissues with abnormally high growth rates and it is generally the case that malignancy is proportional to tumor proliferation rate, it seems that cancer biology provides an opportunity to test the GRH. It also seems possible, then, that stoichiometric theory as developed in ecology might offer important insights into factors regulating the outcome of the interaction between tumor and host. Space precludes any detailed discussion here; details are available via Elser et al. (2003). Here I recap the main points of that paper. First, a consideration of published literature does indicate that increased rRNA levels and rates of rRNA production and ribosome biogenesis are associated with the development of tumors ("tumorigenesis"). It also seems that genetic mechanisms connected to the rDNA and in the regulatory pathways leading to the rDNA are associated with tumor development and proliferation. Thus, important cancer-causing genes ("oncogenes") are connected to the machinery of cellular proliferation. Finally, while data are extremely limited, available evidence does indicate that tumors have unusually high P demands (as a result of high levels of P-rich RNA) and that there is an association between P metabolism and cancer development. A new grant from the USA's National Science Foundation and National Institute of Health is now getting started and will provide an opportunity for me to test these hypotheses explicitly.

Biomedical researchers have begun to issue pleas for conceptual and theoretical approaches to make functional sense of the intimidating mass of information surrounding cancer etiology and dynamics (Hanahan and Weinberg 2000, Gatenby and Maini 2003). The same can probably be said of many difficult problems in biological medicine. The situation is unlikely to improve much in the face of the blizzard of information emerging from high-throughput sequencing machines and microarray readers, as high-throughput of data doesn't necessarily produce high-throughput of understanding. This challenge is similar to the daunting prospect long faced by ecologists and evolutionary biologists confronting the vast biodiversity of living species present in nature and in connecting that diversity to the functioning of those species in energy flows and biogeochemical cycling. Just as ecologists cannot include all biological diversity in their theoretical models but must instead focus on key interactions that capture most of the major mechanisms, cancer theory cannot encompass the entirety of the genetic and protein diversity underlying tumor biology. Instead cancer biologists need conceptual clarity and theoretical tools of intermediate complexity to identify key mechanisms (Gatenby and Maini 2003). Perhaps the same conceptual framework, biological stoichiometry, now aiding ecologists in understanding food web dynamics will be of use to cancer biologists or to others working to understanding the enormous biological complexity emerging in the post-genomics era.

Acknowledgments

I'm grateful to my host Dr. Dag Hessen, to Dr. Willy Østreng for asking me to write the synopsis, and to the lunch group at the CAS for a variety of fascinating conversations during *middag*.

Literature Cited

Clegg, L. X., F. P. Li, B. F. Hankey, K. Chu, and B. K. Edwards. 2002. Cancer survival among U.S. whites and minorities: a SEER (Surveillance, Epidemiology, and End Results) population-based study. *Arch. Intern. Med.* 162:1985–1993.

Elser, J. J., D. Dobberfuhl, N. A. MacKay, and J. H. Schampel. 1996. Organism size, life history, and N:P stoichiometry: towards a unified view of cellular and ecosystem processes. *BioScience* 46:674–684.

Elser, J. J., T. Dowling, D. A. Dobberfuhl, and J. O'Brien. 2000a. The evolution of ecosystem processes: ecological stoichiometry of a key herbivore in temperate and arctic habitats. *J. Evol. Biol.* 13:845-853.

Elser, J. J., Y. Kuang, and J. Nagy. 2003. Biological stoichiometry of tumor dynamics: an ecological perspective. *BioScience* 53:1112–1120.

Elser, J. J., R. W. Sterner, E. Gorokhova, W. F. Fagan, T. A. Markow, J. B. Cotner, J. F. Harrison, S. E. Hobbie, G. M. Odell, and L. J. Weider. 2000b. Biological stoichiometry from genes to ecosystems. *Ecology Letters* 3:540–550.

Elser, J. J., and J. Urabe. 1999. The stoichiometry of consumer-driven nutrient cycling: theory, observations, and consequences. *Ecology* 80:735–751.

Gatenby, R. A., and P. K. Maini. 2003. Mathematical oncology: Cancer summed up. *Nature* 421:321–321.

Hanahan, D., and R. A. Weinberg. 2000. The hallmarks of cancer. *Cell* 100:57–70.

Henschke, C. I., D. I. McCauley, and D. F. Yankelevitz. 1999. Early lung cancer action project: overall design and findings from baseline screening. *Lancet* 354:99–105.

Hessen, D. O. 1997. Stoichiometry in food webs – Lotka revisited. *Oikos* 79:195–200.

Lavigne, D. M. 1996. Ecological interactions between marine mammals, commercial fisheries, and their prey: unravelling the tangled web. *in* W. A. Montevecchi, editor. Occasional Paper Number 91: Studies of high-latitude seabirds. 4. Trophic relationships and energetics of endotherms in cold ocean systems. Canadian Wildlife Service, St. John's Newfoundland.

Main, T., D. R. Dobberfuhl, and J. J. Elser. 1997. N:P stoichiometry and ontogeny in crustacean zooplankton: a test of the growth rate hypothesis. *Limnol. Oceanogr.* 42:1474–1478.

Patz, E. F., P. C. Goodman, and G. Bepler. 2000. Screening for lung cancer. *N. Engl. J. Med.* 343:1627–1633.

Reiners, W. A. 1986. Complementary models for ecosystems. *Am. Nat.* 127:59–73.

Sterner, R. W., and J. J. Elser. 2002. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. Princeton University Press, Princeton, N.J.

Vrede, T., T. Andersen, and D. O. Hessen. 1998. Phosphorus distribution in three crustacean zooplankton species. Limnol. *Oceanogr.* 44:225–229.

KAREN HELEN WILTSHIRE & MAARTEN BOERSMA

40 years of Pelagic Time Series at the Biologische Anstalt Helgoland: Evaluation and New Concepts

Since Victor Hensen in 1887 sought to answer the question of what the fish production level was in the ocean and how this was related to organic matter via primary production,relatively little has changed. We still pose similar questions, albeit maybe in a less linear fashion: related more to food web structures and organism interactions.

Hensen applied agricultural paradigms to the study of fisheries and presented us with his concepts in the field of quantitative marine ecology. He introduced the term plankton to describe all living or dead matter floating in the water column („was im Wasser treibt, ob hoch, ob tief, ob todt oder lebendig"). Other scientists at the time such as Haeckel, Schütt, or Dohrn were more interested in the functional and organism diversity. Ernst Haeckel coined the term "ecology" from the metaphorical Greek base Oikos and logos and implemented his holistic ideas on the island of Helgoland (Figure 1).

Dr. Karen Wiltshire,
Alfred Wegener Institute for Polar and Marine Research, Germany.
CAS Fellow 2003/04.

Dr. Maarten Boersma,
Alfred Wegener Institute for Polar and Marine Research, Germany.
CAS Fellow 2003/04.

The work of such scientists and the emergence of an increased interest in marine biology worldwide facilitated the foundation of the "Königliche Biologische Anstalt" on Helgoland in 1892. Thus, an era of intensive investigations began in the German Bight/ North Sea. Even before the formal foundation of the Institute, from 1873 onwards salinity and temperature measurements of the water column were carried out on a daily basis. This can be seen as one of the historically most important long-term marine data bases. However, unlike the present, where our interests lie in determining the effects of global warming, the original interests resulting in this time series were based more on curiosity regarding natural history.

Although the long-term investigations were interrupted briefly by war and the evacuation of the island, the Helgoland time series were restarted and are still continuing. Since 1962 long-term monitoring of biological, chemical and physical parameters has continuously been carried out at Helgoland Roads. Samples are taken each working day for water temperature, salinity, nutrient concentration and phytoplankton biomass at the cable buoy (Kabeltonne) station (Figure 2).

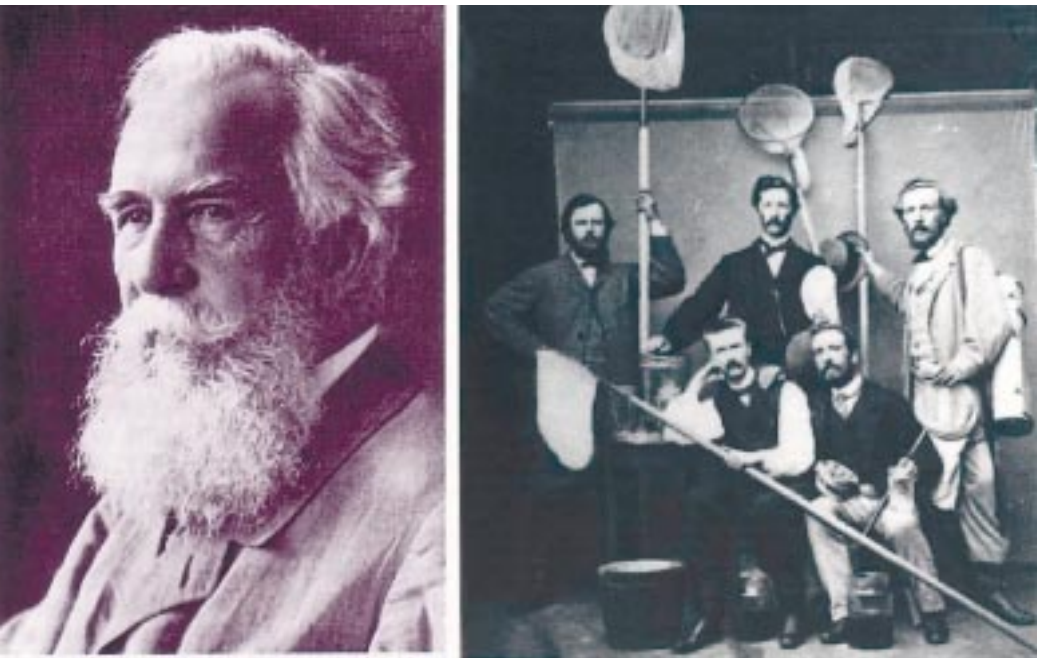


Figure 1: Ernst Haeckel left, and right with various colleagues and the then "newly invented" plankton nets in September 1885 on Helgoland (Archiv der Biologischen Anstalt Helgoland).

In the nineteen seventies the main worry in the German Bight was related to the problems associated with pollution and more specifically eutrophication. Indeed, the time series did show steady increases in nutrient loading at Helgoland which then levelled off in the eighties and nineties. However, a concurrent change in the phytoplankton biomass which was postulated has never been shown conclusively.

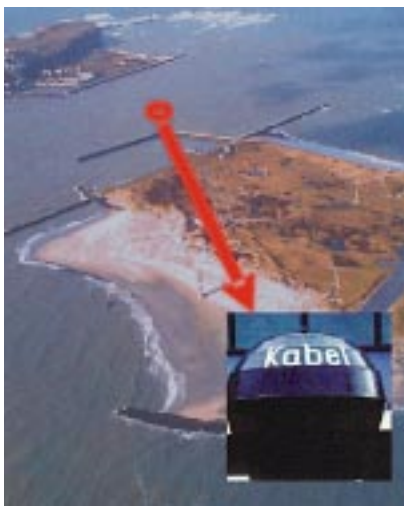


Figure 2: The cable buoy and the long-term sampling site at Helgoland.
Photo P. Mangelsdorf

In the meantime, the main interest in using the Helgoland data set has moved to global warming investigations.

The fingerprints of global warming on terrestrial animals and plants have been well documented. In contrast, for aquatic systems, we have little information on how the warming trends of the last 10–20 years have affected them. However, knowledge of how climate warming affects phytoplankton, in its important position at the base of the aquatic food chain, is vital. Hence, we analysed the temperature and phytoplankton data from Helgoland Roads time series, showing that a variable temperature trend over the past 125 years has culminated in a

warming trend of 1.1°C since 1962. Very cold winters (with a minimum of -1°C or less) occurred about every 10 years up to 1944, but notably only once since 1960 (Figure 3).

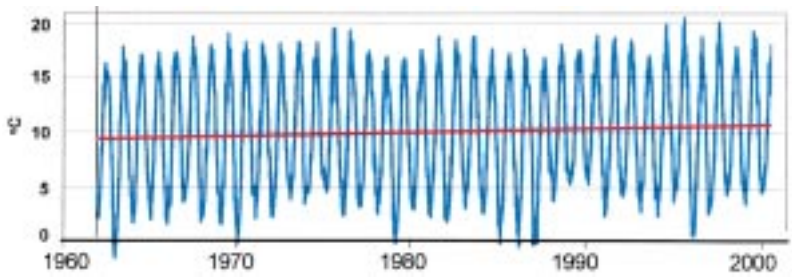


Figure 3: The water temperatures at Helgoland Roads 1962–2001

These data were combined with our phytoplankton counts to show a warming-related shift in phytoplankton succession thus making life cycle/food resource mismatches likely. As a result of this evidence we have at the BAH established a new project to investigate the interactions of pelagic organisms in more detail: the Helgoland Foodweb Project. With this we have established a new approach to dealing with marine food webs. A group of biologists including the disciplines of ecology, biochemistry, ichthyology, microbiology and physiology are working together to investigate the interactive role of microalgae, zooplankton, bacteria, larval fish in their chemical and, to a lesser degree, physical environment.

In this project we are following a three-pronged approach:

1. we monitor phytoplankton, zooplankton and physical parameters on a daily basis at Helgoland Roads and use this data for understanding trends and patterns.
2. we isolate key organisms from the water column during driving events such as algal blooms and in the laboratory try to understand interactions between organisms and to verify observations made in the field.
3. we try to re-apply the knowledge we have gained in the laboratory to further our understanding of our systems.

Ultimately we wish to understand our pelagic system better. We hope to be able to evaluate the patterns and occurrences of plankton. We need to know for example how our spring bloom of microalgae functions as this

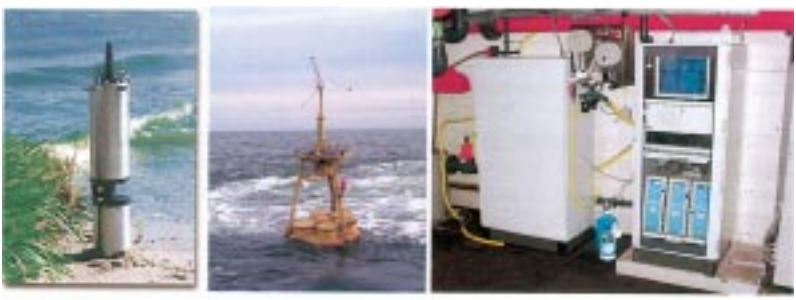


Figure 4: Examples of automated measurement technology augmenting the Helgoland Roads Long Term data set: automated sensors, buoys and Ferryboxes.

drives much of our early year ecology. This means that certain finely tuned timings of organism co-occurrence and sequential occurrence will be disrupted.

Long term data sets were subject to a lot of criticism about 10–15 years ago. As a result, many were discontinued. The BAH managed to weather this storm and now the Helgoland Roads Time Series is one of the most used series. It is being augmented with new and more efficient technologies such as automated sensors, buoys and Ferryboxes (Figure 4). This introduction of new technologies and rigorous quality control of Long Term Data series is vital to providing a continuum and a useful heritage to the next generation. Only in such a manner can our global fragility be assessed and monitored successfully.

CESARE CORNOLDI, ROSSANA DE BENI

Asymmetries between Perception and Mental Imagery (Imagery and Blindness)

When we speak of mental images we refer to representations of objects in our mind. Typically mental images refer to visual representations. However, according to a constructive view of mental imagery (e.g. Cornoldi, De Beni, Giusberti & Massironi, 1998), a mental image is less modality-specific than the corresponding perception. Furthermore, the content of the mental images evoked may consist of more layers than the corresponding sensory perception. Thus we could imagine a dear friend or the long hours before obtaining the result of an exam with deep emotional involvement, and hence produce a specific emotional mental image, but also reconstruct it as a mental image involving many modalities simultaneously.

Professor Cesare Cornoldi,
Department of Psychology,
University of Padova, Italy.
CAS Fellow 2003/04.

Professor Rossana De Beni,
Department of Psychology,
University of Padova, Italy.
CAS Fellow 2003/04.

The imagery process is strongly related to memory, thinking and perception. To find an exhaustive definition of mental images is not an easy task. In fact a number of different interpretations of the concept of mental images have been proposed. Holt (1964) observed that a mental image refers to all the subjective awareness experiences with an almost-sensitive modality, that is not only perceptual. Unlike perception, imagery is a mental process, difficult to ascribe to an exact stimulus-situation.

Based on sensory experience, we may represent the object in the knowledge system. Information about what we have experienced is registered in our memory system and then retrieved every time we need it. In many circumstances of our life the retrieval of mental images from memory represents a useful tool, as, for example, when we orient ourselves in a familiar environment using a visual mental map, or when -before leaving for a holiday- we imagine the optimal luggage arrangement in the boot of the car. The great use of imagery in everyday life and the close relationship between imagery and other cognitive processes justify the interest in this topic.

Mental images have been considered very similar to percepts in our mind. A debate is still running on the hypothesis that imagery and perception share common mechanisms. The analogical view of identity between perception and imagery was sustained by results deriving from the execu-

tion of certain mental operations using imaginal and perceptual representations. The mental rotation of 3D visuo-spatial configurations (Shepard and Metzler, 1971) or the dot-localization task using imagined stimuli showed that visual mental images are also subjected to the constraints of the physical world.

One interesting field of investigation as regards the relation between imagery and perception is the replication in the imaginal modality of certain phenomena effects or illusions obtained in the perceptual modality. This area is particularly open to criticism. Under imagery conditions different visual illusions may be obtained, such as the Mueller-Lyer illusion (Heller et al., 2002), Ponzo, Hering and the Wundt illusion in higher imagers (Wallace, 1984). However, Intons-Peterson and McDaniel (1991) reported a series of asymmetries between imagery and perception regarding the distance and magnitude estimations, relative contrast (brightness), structural factors, mental rotation and the role of knowledge. Giusberti et al. (1992) considered the different subjective experiences linked to imagery and perception and found that visual images and visual percepts differed in vividness ratings, and that visual perception involved more automatic and pre-attentive processes, while visual images generation implicated the involvement of controlled and non-automatic processes. Further studies by the same authors revealed that, when participants do not know the perceptual effect and/or the image is not based on a preceding perceptual exposition, the visual illusion is not present at the mental imagery level. For example the `pop-out` effect was examined by contrasting the representations of a reversed or inclined letter in a matrix made with the same letter under three different conditions, perceptual, memory perceptual and imaginal. Here are the instructions for the imagery situation described above (situation 1) and for another situation 2.

Table 1 – Instructions for the generation of mental images

Instructions
Situation 1
Imagine 5 rows each having 5 capital letters 'T'
Consider the second 'T' in the second row
Compare the vividness of a T within this context according to these two different conditions:
a) the 'T' is inclined 45 degrees
b)the 'T' is reversed 180 degrees
c) Which 'T' is more vivid and better stands out from the other 'T's?
Situation 2
Imagine a circle surrounded by other circles according to these two different conditions:
a) tthe circle is surrounded by 8 other circles which are a little bit smaller
a) tthe circle is surrounded by 8 other circles which are a little bit larger
In which case does the circle appear larger to your mind\s eye?

In the first two conditions of Situation 1 it was the inclined letter that 'popped out', and it did so very clearly, producing the most vivid representation. On the other hand, in the imagery representation the reversed letter appeared more vividly than the inclined one. Giusberti et al. (1998) confirmed the asymmetry between perception and imagery by using the Ebbinghaus (situation 2, the circle is larger in case a, i.e. when it is

surrounded by smaller circles, but the effect is more evident in perception that in mental imagery) and Ponzo illusions (the superior line is longer in perception than in mental imagery) (examples of the materials used for studying these asymmetries are given in Figure 1 and Figure 2).

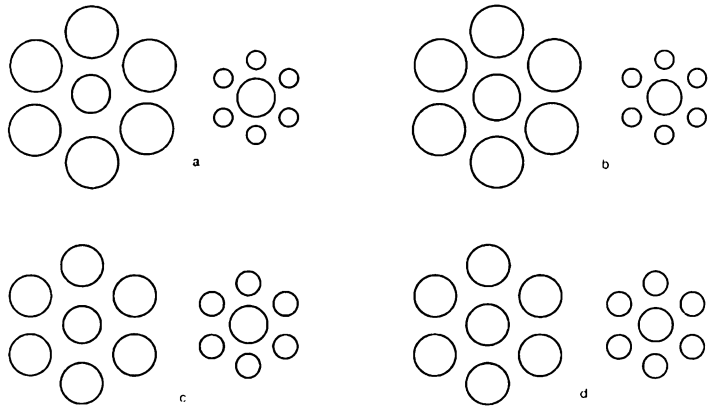


Figure 1. Examples of stimuli used for the Ebbinghaus illusion.

Giusberti: European Psycologist, Volume 3(4). December 1998. 281–288

Other results, problematic for an analogical theory of mental imagery, have been obtained with congenitally totally blind people who are able to do as well as the sighted on many imagery tasks (e.g. De Beni & Cornoldi, 1985). Thus given the contradictory results emerging from different studies the conclusions in this particular field are as yet difficult to draw. However, the distinction proposed by Cornoldi et al. (1998), between a visual trace, sharing characteristics with perception, and a generated image, with different properties, seems able to take into account the different results.

The possibility of "seeing" an object with our mind's eye seems to require that we have first experienced it in the perceptual world. This may be considered true if we refer to memory images, but we could also create original and totally new representations in our mind, not based on real perceptual representations. This assumption was sustained also by Hobbes, who affirmed that the generation of images is due to a combination, often new and original, of percepts stored in memory. The memory images can be generated on the basis of information retrieved from long-term memory. Cornoldi et al. (1998) called this type of mental image 'a generated image' and distinguished it from a representation directly derived from a recent experience or from a well-learned sensory pattern, called 'a visual trace'. According to their constructive view, generated images are the result of the combined synthesis of long term memory information coming from different sources and may be penetrated by beliefs, emotions, and conceptual knowledge.

The interest in studying mental imagery in congenitally totally blind people is motivated by different goals. Firstly, it allows us to establish the contribution of visual perception in the generation of mental images. Secondly, it shows whether and how it is possible to use mental imagery in order to improve memory performance in congenitally totally blind

people. Third, if, on the one hand, a series of tools is suitable for discriminating performances between sighted and blind individuals, on the other hand, the same tasks can be used in the assessment of individual differences in order to highlight the different nature of the WM processes, in particular the distinction between the passive maintenance and the active manipulation and transformation of information.

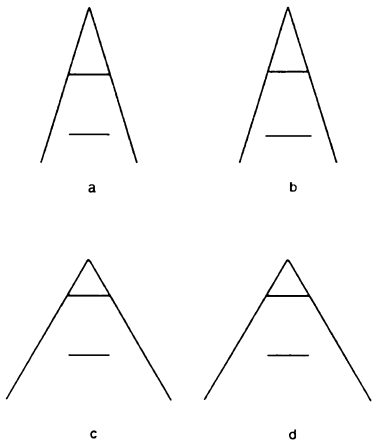


Figure 2. Examples of stimuli used for the Ponzo illusion.

Giusberti: European Psycologist, Volume 3(4). December 1998. 281–288

One of the first objectives that cognitive psychologists had to achieve was to design memory tasks requiring mental imagery strategies in which congenitally blind people could be compared with sighted individuals. This issue has been tackled by several researchers. For instance, it has been found that blind people can use interactive mental images in order to retrieve paired words (e.g. Zimler and Keenan, 1983), although it seems that they need a longer time in order to create their mental images (Kerr, 1983).

Other researchers investigated the qualitative differences characterising mental images processed by blind and sighted people. Those studies highlighted a large number of similarities in the mental representations created with and without visual knowledge, pointing out that some properties are not necessarily due to the visual experience (Marmor, 1978; Zimler and Keenan, 1983). For instance, it was found that in carrying out mental rotation tasks, the performance of blind people is subject to the same characteristics as the performance of the sighted ones: the time taken in order to judge the identity of two spatial configurations differently oriented in the third dimension depends on the performance.

The development of new specific experimental tasks largely contributed to the investigation of mental imagery in congenital blindness. For instance, a meaningful contribution came from a methodology that firstly requires a tactile exploration of stimuli and then the creation of a mental representation that is stored in visuo-spatial working memory. By using this experimental procedure many authors have shown that tactile exploration of a pattern of stimuli is sufficient to generate mental images in congenitally blind people.

In our studies we have investigated the nature of mental images and of visuo-spatial processes in blind people by evaluating the effects of different stimulus types and experimental instructions. One of our goals was to investigate the role played by subjective experience in mental imagery and the effect of imagery value in recalling three different categories of nouns. We used 1) concrete words evoking some mental image of objects experienced by the blind (HI, High Imagery value); 2) abstract stimuli for which it was harder to create a link with personal knowledge of the world both for blind and sighted people (LI, Low Imagery value); 3) stimuli that could evoke a mental image but that could not be associated with a direct personal experience, such as for the item 'spaceship' (HINE, High Imagery Not Experienced) (Cornoldi, Calore and Pra Baldi, 1979). Blind people judged HINE stimuli as having a low imagery value. However, data on their memory performance were not so clear. Indeed non-sighted participants recalled a greater number of LI names, whereas in other conditions their performance was poorer than that of sighted people. Furthermore, the accuracy of performance in blind people depended on whether the recall was intentional or incidental.

The pattern of results found in HINE condition was later on investigated by De Beni and Cornoldi (1988) and Tinti and colleagues (1999). These researchers replicated the early findings confirming that blind people have difficulty in recalling HINE words, but, if we take into account their limited knowledge of the world, their performance was less impaired than could be hypothesised.

In another series of research tasks differences between sighted and blind participants in using mental imagery were investigated by comparing their performance in recalling sequences of verbal information. In a study by De Beni and Cornoldi (1985), we asked sighted and blind participants to use the loci mnemonics— which consists in imaging to locate a series of stimuli along a well known pathway — in order to retrieve a list of twenty single, pairs and triplets of words. We found that also non-sighted people took advantage in this task of imagery instructions, because they preferred to use mental representations instead of using the rehearsal strategy in order to recall the items list. Moreover, congenitally blind participants could generate complex interactive mental images, but they had difficulty in creating a single mental representation containing different stimuli. The difficulty met by the blind in memorising triplets was not due to a generic increase in the memory request. In fact, according to Tinti and colleagues (1999) when complex mental images are processed and used in an auditory format non-sighted people are not impaired.

References

Cornoldi, C., Calore, D. and Pra Baldi, A. (1979). Imagery ratings and recall in congenitally blind subjects. *Perceptual & Motor Skills*, 48, 627-629.

Cornoldi, C., De Beni, R., Giusberti, F. and Massironi, M. (1988). Memory and imagery: A visual trace is not a mental image, in M. Conway, S. Gathercole et C. Cornoldi (Eds), *Theories of Memory* (vol. 2, p. 87-110), Hove, UK; Psychology Press.

Cornoldi, C., and Vecchi T. (2003). *Visuo-spatial working memory and individual differences*. Hove: Psychology Press.

De Beni, R., & Cornoldi, C. (1985). The effects of imaginal mnemonics on congenitally totally blind and on normal subjects, in D. Marks and D. Russel (Eds), *Imagery* (vol.1, p. 54-59). Dunedin, New Zealand: Human Performance Associates.

De Beni, R., & Cornoldi, C. (1988). Imagery Limitation in Totally Congenitally Blind Subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 650-655.

Giusberti, F., Cornoldi, C., De Beni, R. & Massironi, M. (1992). Differences in vividness ratings of perceived and imagined patterns, *British Journal of Psychology*, 83, 533-547.

Giusberti, F., Cornoldi, C., De Beni, R., & Massironi, M. (1998). Perceptual illusions in imagery. European Psychologist, 3, 281-288.

Heller, M.A., Brackett, D.D., Wilson, K., Yoneama, K., Boyer, A., Steffen, H. (2002). The haptic Mueller-Lyer illusion in sighted and blind people, *Perception*, 31, 1263-1274.

Holt, R.R. (1964). Imagery: The return of the ostracized. *American Psychologist.* 19: 254-264.

Intons-Peterson, M. and McDaniel, M.(1991). Symmetries and Asymmetries between Imagery and Perception. In", C. Cornoldi and M.A. McDaniel (Editors), "*Imagery and Cognition. ,* (pp. 47-76). New York: Springer

Kerr, N.H. (1983). The role of vision in "visual imagery" experiments: Evidence from congenitally blind. *Journal of Experimental Psychology: General, 112*, 265-267.

Marmor, G.S. (1978). Age at onset of blindness and the development of the semantics of colour names. *Journal of Experimental Child Psychology*, 25, 267-278.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science*, 171, 701-703.

Tinti, C., Galati, D., Vecchio, M.G., De Beni, R. and Cornoldi, C. (1999). Interactive auditory and visual images in the blind. *Journal of Visual Impairment and Blindness*, 93, 579-583.

Wallace, B. (1984). Apparent equivalence between perception and imagery in the production of various visual illusions. *Memory and Cognition, 12*, 156-162.

Zimler, J., and Keenan, J.M. (1983). Imagery in the congenitally blind: How visual are visual images? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 269-282.

OLAV GJELSVIK

Agency: What is it Agents Exemplify?

Introduction

One might think that out deeds are our acts, but not everything we do amounts to actions. One issue is therefore this: How do we distinguish our actions from other things we do.

Donald Davidson gave the standard answer to this question: Our actions are those things we do intentionally, i.e. those events that can be described in such a way that they can be characterized as intentional under some description.

This Davidsonian reply is thin. It leaves out a further elucidation of what we mean by being "intentional", and it does not say much about the role of the agent in the production of the action. There is a further issue here. All physical actions are bodily movements, Davidson claims. What is the content of his claim?

Professor Olav Gjelsvik,
Department of Philosophy,
University of Oslo.
Group leader CAS 2003/04.

There is a difference between *what we do*, which might be melting some chocolate, and *what happens to the chocolate*, the melting of the chocolate. The first thing is something we do. What we do is captured by a transitive reading of the verb melting, what happens to the chocolate is captured by an intransitive reading.

Implication: There is a corresponding difference between a transitive and an intransitive reading of moving a hand, and the hand's movement. The second is something that happens to the hand, the first is something we do.

This is controversial. Those who like this distinction, normally belong to one camp among action-theorists. That is the camp of the conceptually rich. There are broadly speaking two conceptions of agency today: The conceptually rich and the conceptually poor. Let us now turn to them.

The conceptually poor: D. Dennett

What we mean by intentional: We are intentional systems.

Intentional systems are organisms that can be successfully described in the specific way that uses concepts like belief and desire. Successfully here means that we gain a lot of predictive power we could not otherwise have. In a thermostat we achieve nothing, in animals and people we do.

This marries into a causal-functional naturalism in the philosophy of mind: beliefs and desires are to be individuated by and identified with the causal-explanatory roles they play. Intentional behaviour/action is then explainable by reference to such beliefs and desires. This is the main picture we meet in decision-theory and social science.

Actions are then bodily movements (in the intransitive sense) caused by belief-desire pairs ("reasons") *in the right sort of way*. The concept of reason is thought of as something that can be reduced to a belief-desire pair – the concept is in this sense psychologised.

Advantage: It keeps things simple and straight.

Problems:

One cannot say anything about what the right sort of way is. This fact cannot really be explained, and this might be thought of as disturbing.

One seems possibly not to have available any contact with reasons as normative, and the contact with the normative is at best achieved through the contents of the beliefs and desires. But if one only has a causal-functional approach to content, the contact with the normative may be completely lost.

The role of the agent is limited to being the subject to whom these mental states are ascribed. The further role of the agent, if there is one, is then hidden in the phrase of causing the bodily movements in the right sort of way. But it can only amount to a role in the causal history of the action; there is no role in the action itself, the product of this history.

The conceptually rich: E. Anscombe:

Agents act and exemplify agency in their (transitive) bodily movements when two requirements are satisfied:

a) The agent has non-observational knowledge of what is being done (under a description). The phrase "what is being done" here appeals to the transitive sense of moving a part of one's body. What one knows can in central cases be thought of as the intention in moving that part of one's body.

b) This requirement accounts for the sense of "intention". It goes like this: If asked why one did it, one can give a reason. The reason is what motivates the intention by virtue of what supports a practical conclusion in favour of doing it. One knows what one is doing in this sense.

The agent is the agent of this (transitive) bodily movement, and is that by virtue of knowing non-observationally and practically what is being done, and knows that by virtue of being the subject of the piece of practical reasoning behind the action. The agent therefore exercises a sort of control, but it is rational control.

Note the following:

This conception of reason is not psychologised. It is a normative concept, and is seen as having a causal role insofar as it can be appealed to in an explanation of behaviour. The concept of reason is *sui generis*, even if it can be employed as a causal-explanatory concept. Its being *sui generis* means that it cannot be captured by causal-functional roles.

Nothing is said to the effect that what is known needs to be conscious, or directly present to consciousness. What you know is not necessarily transparent in that sense.

Problems: A very heavy burden is placed on the concept of knowledge, and the normative concept of reason. One might here be explaining the clear by the obscure.

Advantages:

There is a clearer role for the agent in agency – the agent is present in the action by controlling it rationally.

The notion of reason is irreducibly normative even if causal.

One has no need to speak of causing something the right sort of way. This is taken care of by the link between the reason and what one knows non-observationally.

Concluding perspective:

The conceptually rich view is much better. The need to introduce the irreducible "right sort of way" shows that one is trying a reduction of action to (intransitive) bodily movement with the right kind of cause, where no reduction can be had. Perhaps more importantly, there are many reasons with little or no causal-explanatory role. They are still reasons. We must give up a causal-functional approach to reasons; that approach can, at best, only capture reasons that do causal-explanatory work.

The resources of the conceptually rich approach are striking. The distinction between the transitive and the intransitive readings of bodily movement is available for all naturalists.

Non-observational knowing what you are doing can, I think, be ascribed to creatures with little sensitivity to reasons in the normative sense. Knowledge is naturally seen as prior to belief anyway. Animals have it abundantly.

Reason-sensitivity comes in degrees as humans mature, and can be thought of as a gradual concept.

Positive claims: Biology can interact profitably with the conceptually rich approach, not only with the conceptually poor.

Psychology might be interested in this for many reasons. We could try these two conceptions out on a number of pathological cases from psychology. My bet is that the conceptually rich approach will be helpful when conceptualizing both development stages and failures, and a number of puzzles and pathologies.

Social science in general will benefit from buying into a rich and substantial approach to reason and normativity. As it is practiced today it invites a deep scepticism about normative truths (it is all a matter of taste, they think), and this does great harm.

Philosophy will be challenged by this approach, because it goes against dominant ways of thinking about mind and body. But in my view it relates fruitfully to philosophical issues like freedom of the will.

DAG O. HESSEN

Genetic Determinism

While not being the topic of the Biology Group at the CAS this academic year, the concept of biological or genetic determinism nevertheless has some relevance to the theme of the philosophy group, and it is thus tempting to bring this up for discussion in the CAS seminar series. The topic itself raises a number of associated questions on the meaning of both "genetic" and "determinism". I have no intention of fully exploring these concepts here, but I want rather to make some remarks on different understandings of the concept.

Surely the understanding of *determinism* is by no means straightforward. By claiming, in a mechanistic tradition, that everything is based upon causal links, one could argue for universal determinism (in the tradition of the search for "the theory of everything"), and thus contend that genetic determinism itself is rooted in a chain of physically determined events. It may be fruitful however to distinguish between some conceptual levels of determinism or indeterminism to see where genes fit in. One may draw a distinction between *hard determinism* in the Spinozan sense, where fatalism could be seen as the religious analogue. The more widespread *soft determinism* would argue for a causal link from A to B, but not from A to Z due to a series of unpredictable interactions of chaotic and indeterministic happenings. Within natural sciences, most discussions on determinism will be in this realm. Irrespective of the understanding of determinism, the issue at stake is to what extent recent insights in biology and genes may shed new light (or provide challenges) on concepts such as free will, rationality, morality and responsibility. The fact that *physical traits* are to a major extent rooted in genes is not very controversial, so the question is rather the role of genes for mental capacities and personality. While (some) biologists have been accused of advocating a deterministic view of individual personality, it is worth recalling that there is a strong philosophical as well as religious tradition for the belief in innate characteristics (e.g. Augustine and Schopenhauer, "once a thief, always a thief").

Professor Dag Hessen,
Department of Biology,
University of Oslo.
Group leader CAS 2003/04.

Since the mental has a physical basis, it should not be very controversial either to state that a number of mental attributes have a fairly obvious genetic causality. But what about personality then? In many ways the discussion of genetic determinism builds on the old controversy of nature versus nurture in humans where socio-biology and later evolutionary psychology traditionally emphasize inherited and evolutionary traits as key tools for understanding human acts and motives, the point of departure for the idea that individual destiny is "in the genes", in James Watson's (one of the discoverers of the DNA double helix structure) famous "Central Dogma". This dogma claims a one-way instruction from

gene to protein. Watson also phrased the famous "Once we believed that our destiny was in the stars, now we know that it is, to a major extent, in the genes". Since these claims were made, the interplay between genes and their products, and not least the interplay between genes have been further disentangled (yet still far from fully understood), and have not supported strong genetic determinism. First of all, there is no 1:1 ratio between genes and "products", most characteristics are influenced by a large number of genes in a complex interplay. This was not least obvious after the discovery that the human genome was made up of no more than some 30,000 genes as opposed to the previously assumed ~100,000.

Note that there are different levels of genetic determinism, the common or shared and the individual. One could thus speak of human nature as being like this or that, also there could be a kind of determinism assigned to sex, "race" or group, and of course there is determinism at the individual level. The strong belief in racial difference and the very belief that there existed such a thing as well-defined races in the 1920s and 30s build on a (wrong) assumption of ethnic determinism. The same can be said about female versus male capacities. Typically when female admission to the Faculty of Medicine at the University of Oslo was discussed in the 1880s, it was claimed that " … the female nervous system and strength would in general not be able to handle these comprehensive studies and the demanding mental work …". Today there are more than 50 % female students … Another example of sexual determinism can be drawn from the writer Bret Easton Ellis who in an interview in the newspaper Aftenposten in 1999 stated that " … men are aggressive, wild, active; this is rooted in their biology, their physiology. They are especially obsessed by the primitive aspect of life, it is about dominance …". If we accept that there is such a thing as "human nature", that is important, is this a fixed nature? Is there a "good" or "bad" nature, are we altruistic or selfish, and are there general constraints on human preferences due to in-built, evolutionary, genetic preferences?

The brief answer to these questions is that there surely is such a thing as human nature, but not *a* human nature. First and foremost we are extremely flexible, and while this flexibility must also be rooted in nature, it makes the search for stereotypes into a sport with a major risk of failure. Certainly some cultural universals may be identified, but nurture seems to be superimposed on nature in this regard.

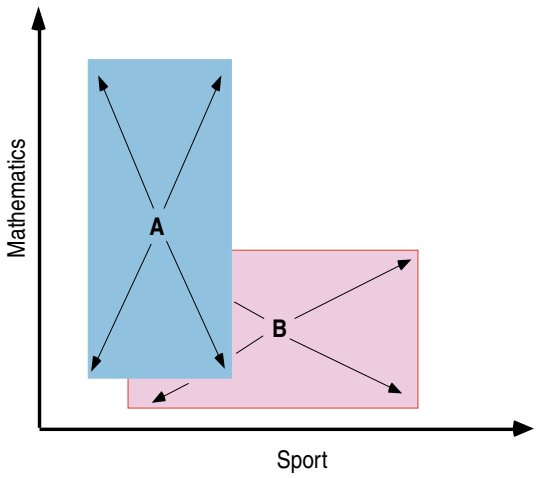
Now then, let us turn to the question of *individual* determinism. To what extent are physical and mental individual attributes fixed by genes. Are there genes "for" IQ as well as for athletic performance, are there genes "for" health and disease, genes for crime or care? I.e. can we really talk about genetic destiny? When it comes to physical attributes and health, the answer is – to a certain extent – yes. We have seen a boom in recent books with titles like "Your genetic destiny" and there are regularly articles in magazines and newspapers expressing the same message: there is indeed a genetic destiny. A number of chromosomal defects or gene variants (mutated alleles) have been identified that predispose us for certain types of illness, some fatal, others not, some with high likelihood, others with low. While some of these can be modified or even cured by life style or diet, there is nevertheless some kind of determinism (again depending on the definition) embedded in e.g. receiving three copies of chromosome 21, causing some 94% of all carriers of Down's syndrome. A large

number of other examples could be provided. The problem is, however, when this is transformed into some sort of general genetic fatalism: "There is nothing I can do, it's all in the genes". A strong belief in an individual genetic fate could, a priori, cause a corresponding lack of individual responsibility.

Perhaps the most dramatic manifestation of this kind of genetic destiny has popped up in the courtroom where we have seen (from the USA) examples of procedures where prosecuted murderers have been claimed innocent by the Defense since they came from families with "crime genes". While these arguments may seem far-fetched, there are examples of types of clearly inheritable traits that promote antisocial behaviour. Needless to say, however, this is a dangerous line of reasoning, and it contradicts the view of individual freedom and responsibility.

I have for amusement over time collected "Genes for …" headlines that regularly crop up, suggesting some sort of individual, genetic determinism for almost all kinds of disease or personality: "Genes for being a good mother", "Genes for faithfulness", "Genes for terror" and perhaps the best: "Genes for bad luck" where the absolutely indeterministic has been assigned genetic determinism. Well, genes *are* important, but not nearly as important as suggested by these headlines, whether or not the desk or the scientist is to blame.

Perhaps the most important aspect is that genes do not *prescribe* immoral or amoral acts, and are anyway no (strong) excuse for such acts (i.e. Moore's arguments on "the natural fallacy" stands firm). The message today is that genes and environment play in concert, it is not a matter of nature or nurture; both are needed to gain insight into human motives and acts. Genes "code" evolutionarily for mental capacity and flexibility that can override primary, biological goals. In a strict biological sense, humans are quite often highly irrational to the extent that we do not always try to optimize individual fitness.



For most of us, our genetic baggage certainly imposes some physical and mental constraints, but there is plenty of room for flexibility within the genetic frame. And the other way around, not even the very best "ski-genes" would be of much use in Africa - or the *potentially* new Einstein would probably not unfold his capacities growing up in the backyards of Rio. These arguments could perhaps be summarized in the figure above: along the axis of skills, there are some limitations that have a genetic causality. In the example above, A has more talent for numbers than sports and vice versa for B. Within these borders we can try to optimize our performance and/or explore our freedom and responsibility.

TORE HELSTRUP

Cognitive Control: By What or Whom?

"Systems" has been a central catchword in modern attempts to understand cognition. Cognition is a system, if not a system of subsystems (of sub-subsystems of …). In more than one respect the systems notion turns out not to be overly precise. At least all parties seem to agree that the systematism depends on some constraints applying differently to what is within compared to what is outside the system. To understand cognition is to understand the bottlenecks that constrain the flow of information within the system in question.

Several cognitive systems have been suggested, in particular memory systems. Information is assumed to be processed in a short-term working memory system, recruited either from the external world through a sensory register, or from long-term internal information storage. Information is then processed and stored in different subsystems, each with somewhat different structural and functional characteristics, e.g. representational formats. The explanatory challenge is to come to grips with the intriguing coordination that characterizes mental life. What binds all things together? How are information processing systems controlled? What prevents information processing activities from competing and contradicting one another?

Professor Tore Helstrup,
Department of Psychology,
University of Oslo.
Group leader CAS 2003/04.

Such questions are not new in psychology, although suggested answers have seldom been agreed upon. The suggestions have been sought from a variety of sources. For our purpose, let us focus on four lines of reasoning: about the question of agency, about control directions, about control levels, and about the nature of mental events. The answer we shall attempt suggests that a satisfactory solution to one of these questions must also answer the other three questions, and that all four questions must be treated together.

Traditionally the agency question has been mainly discussed within the psychology of personality. The research emphasis has not been on how a person controls mental events, so much as on attempts to understand what determines the formation of a person. The gambit has been an assumption that knowledge about this formation process would bring along knowledge about the person's agency functions. So far the knowledge about person formation is not impressive. With few exceptions theories of cognition leave out the person completely. Yet by more than common sense folk psychology one "knows" that persons are agents of cognition. What is more doubtful is how relevant studies of person formation will be

to answering the cognitive control question. A person may be responsible for his or her mental events without being responsible for his or her person formation.

The situation is different when it comes to the question about control direction. Although not much is to be found in current cognitive theorizing about the source of control, it is common parlance to specify whether the control is directed bottom-up or top-down. Alternative terms for the same distinction are "data driven" versus "conceptually driven", or "sensory" versus "cognitive" processing. These distinctions make no commitments as to what or who is exercising the control, but implicitly suggest that there is not an abundance of possible control directions.

According to the theorists, apparently cognitive control might be found in stimulus conditions. Behaviourists have no trouble with this assumption. If not found there, it must be sought in the cognitive system. The Gestalt psychologists would agree on that. One cannot but wonder whether control after all must lie in coordination of the two control directions, which seems to be another way of saying that one cannot have directions without a source. The direction assumptions remain empty so long as the agency question is circumvented.

Control levels can be conceived of as levels of processing (Craik & Lockhart, 1972). Superficial registration of sensory task attributes demands less directed efforts than a deeper semantic processing. One sees the close relation to the directions interpretation of cognitive control. The level interpretation hooks up more directly with the energizing aspect of control, e.g. by effort assumptions.

The control we have referred to so far is hardly controversial, but not of much interest before it is made clear what is assumed to be controlled - presumably the units of cognition. However, what are the proper units of mental analysis is far from settled. Before we proceed further with the control question, consider three unit metaphors: points, connections and organisations.

One popular type of mental unit can be conceived of as constituting points in a mental space, e.g. ideas or concepts. Such units can vary in clarity or vividness, but scarcely along a dimension like truth. The strength of a unit of this kind may be measured by differentiation from other point units. Clear units stand out. Vague units dissolve into other units.

Strength interpretations favour connections as mental units. Most famous is the S-R unit. Associative strength, or in behaviourist terminology "habit strength" between stimulus and response, is supposed to be built up by gradual reinforcements, and presumably to be weakened over time. The strengthening process was assumed to be empirically demonstrated. The nature of the weakening process turned out to be far more difficult to determine. Associations connect, but can they (truthfully) represent?

Organisations have structure. The strength of an organisation reflects its structural features. A stable structure is more resistant to change than unstable ones. Organisations vary in structural goodness and complexity, not directly along strength dimensions. Complex organisations like scripts or texts can presumably represent, and thus be assigned, truth values.

Units like the three types mentioned have never been observed, and probably never will be. They function more like "mental models" for the theorizer, than as observational guidelines. Concepts like 'percept',

'memory trace' or 'thought' can be modelled as any of the three mentioned unit types. However, how one comes to theorize about perception, memory or thinking is not unaffected by one's implicit or explicit unit modelling.

Those who subscribe to a mental philosophy that the complex must be explained by its constituting elements are inclined to base their research on what we have termed mental points or connections. Others believe that elements must be organised, e.g. to carry meaning. Their research will naturally be anchored in assumptions about structured units (cf. Thagard, 2000).

When one conceives of mental events in terms of points or connections (or similar entities), one is prone to seeing external forces working on mind. The processing is seen as bottom-up directed. The control is "of mind". One can conceive of mind in terms of organisational units, and still believe that the control is external. But with this conceptual frame the chances are greater that mental life may be seen as controlled "by mind". The organisational idea invites self-organizing interpretations.

Our three mental units urge a static conception of mind. Agents are analyzed into states described in terms of points, connections and organisations, where state changes are understood by reference to information processing by levels and directions. What if this somewhat static frame of reference is replaced with a more dynamic one? (cf. Juarrero, 1999).

Probably it is more correct to consider mental events as including the agent as a proper part, than to think of them as occurring "in" the subject (cf. Gallese & Metzinger, 2003). A mental event might for instance be described by reference to a series of attributes (e.g. agent, place, time, intentions, actions, main objects, …) that vary along representational dimensions (e.g. having verbal, visual, motor or emotional values) and develop (change) over time (cf. Zacks & Tversky, 2001). At any point in time a mental event will have a certain organisational structure which produces informational feedbacks and feedforwards influencing its course (including "memories" and "plans"). Events of such complexity are scarcely "represented" in mind. They are probably simply unfolding. Mind is the going on of such events.

William James (1890) is well known for his "stream of consciousness" model of mental life. According to this line of thinking, mental life is never at rest, always in development. In the references made to the "stream of consciousness (SOC) model" the stream is tacitly treated as taking place within the person's mind. Further, the SOC model is commonly interpreted as representing ongoing thinking, or ongoing perception, or ongoing memorizing, etc. In other words thinking, perception and memory are treated as elements of mind, elements that can occupy the mental slot called stream of consciousness. From a vantage point like this the stream of consciousness can rightly be questioned in terms of its control. The stream aspect is conceived of as a phenomenological experience, not reflecting genuine mental units.

We have suggested that mental life is a pulsating stream analogous to James' SOC model (Helstrup, in press). But we believe that this stream is far from unitary. A number of processes is always going on. Some are very short lived. Others last for longer periods, and some for life-long stretches. At any time there are cognitive sub-processes starting up, terminating or in the process of development. As pointed out by Blumenthal (1977) there

will always be rapid integrations of such sub-processes. Blumenthal emphasized that integrations take place here and now across all forms of ongoing cognition. We assume that integrations will also take place over time, binding together intentions for the future with memories of the past, and binding these together with what is going on at present. The integrations are the organisational forces responsible for the structure of mind, perhaps in line with the suggested field forces of the Gestalt psychologists. Binding is the modern conception of integrative organisation (cf. Cleeremans, 2003).

Only a part of the processes that are going on will at any time be integrated, vertically and horizontally, and become manifested as conscious integrations that are experienced as the stream of thought. Most processes will be unconscious, but still be highly influential on cognition.

From this point of view the person is the mental integration that takes place. The person is not a static entity, but a dynamic process. The person seen in this way is partly self-organizing, and by feedback and feedforward able to influence the integrations that constitute the person.

Hence mental life is partly controlled by the unconscious mechanisms that are not manifested in conscious integrations. Partly mental life is controlled by the dynamic person. Personal control, in this sense, is dependent on well functioning mechanisms. There is no opposition between control by mechanisms and control by the person, since the person is part of the developing cognition. The control is not "in the person". The person is (among many cognitive aspects) control. From this perspective the person is a controlling mind process, not a static substance controlling mind as a separate entity (cf. Prinz, 2003).

Cognitive control should thus not be seen as executed by separate entities. We are in control of mental life in terms of being persons, and we are persons in terms of the same control.

References

Blumenthal, A.L. (1977). *The process of cognition*. Englewood Cliffs, NJ: Prentice-Hall.

Cleeremans, A. (Ed.). (2003). *The unity of consciousness*. Oxford: Oxford University Press.

Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Gallese, V. & Metzinger, T. (2003). Motor ontology: the representational reality of goals, actions and selves. *Philosophical Psychology*, *16*, 365–388.

Helstrup, T. (in press). *Personlig kognisjon* (Personal cognition). Bergen: Fagbokforlaget.

James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.

Juarrero, A. (1999). *Dynamics in action*. Cambridge, Mass: MIT Press.

O'Connor, T. (Ed.). (1995). *Agents, causes, events*. Oxford: Oxford University Press.

Prinz, W. (2003). Emerging selves: Representational foundations of subjectivity. *Consciousness and Cognition*, *12*, 515–528.

Roessler, J. & Eilan, N. (Eds.). (2003). *Agency and self-awareness*. Oxford: Oxford University Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, Mass: MIT Press.

Tulving, E. & Craik, F.I.M. (Eds.). (2000). *The Oxford handbook of memory*. Oxford: Oxford University Press.

Zacks, J.M. & Tversky, B. (2001). Event structure in perception and conception. Psychological Bulletin, *127*, 3–21.

ROBERT PTACNIK

Omnivory – the Strategy of Eating your Competitor

Definition of omnivory

In a wider sense, omnivory means the feeding of an organism on several trophic levels. E.g., many birds feed on seeds and fruits (lowest trophic levels) as well as on small animals.

A special case of omnivory is intra-guild predation (IGP, Holt and Polis, 1997). Here the terminal consumer feeds on another heterotrophic consumer, and on the prey of this consumer (see Fig. 1). As a consequence the so-called intermediate consumer suffers from competition for food with the terminal consumer, as well as being prey to the terminal consumer.

Research Fellow Robert Ptacnik,
IFM-GEOMAR,
University of Kiel, Germany.
CAS Fellow 2003/04.

Given the obvious disadvantage for the intermediate consumer, IGP was assumed to be an exception in nature, leading to the extinction of the intermediate consumer sooner or later. This assumption was confirmed by mathematical models showing destabilizing effects of omnivory in simple food web models. This 'finding' supported the view of natural food webs as linear chains, which prevailed in ecology until the early 1980s (e.g. Pimm, 1982). As methods of food-web analysis became more efficient, it became obvious that natural food webs are non-linear, interwoven networks. At the same time, omnivory and IGP were found to be widespread in terrestrial as well as in aquatic food webs.

Modelling helped to find conditions for coexistence of intermediate competitors and their prey. A major prerequisite is that the terminal consumer benefits substantially from the intermediate competitor (Diehl, 2003). In this situation, a strong decline of the intermediate consumer is to the disadvantage of the omnivore, promoting the coexistence of both consumers.

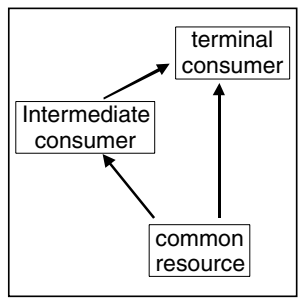


Fig.1. Scheme of intra-guild predation. The intermediate consumer suffers from both competition and predation by the terminal consumer.

Food quality of different trophic levels

The elemental composition of plant and animal biomass varies considerably, with the major differences in the carbon (C) : nitrogen (N) and carbon : phosphorus (P) ratios. Plants are characterized by high C:N and C:P, whereas animals are characterized by low C:N and C:P ratios. This is because carbon is the major component of structural material (cellulose,

lignin) in plants. As a result, the food of herbivores (animals feeding on plants) has a much higher C:(N, P) ratio than their own biomass, making especially P and N the growth-limiting nutrients in the food of herbivores. On the other hand, carnivores feed on prey with an elemental composition similar to their own biomass. Because the organic carbon in the prey is not only used for assimilation of biomass, but also as a source of energy, a considerable part of the carbon will be respired, making carbon the limiting nutrient in purely carnivorous organisms.

This difference in the elemental composition of animals and plants is clear-cut in terrestrial ecosystems. In the open-water areas of aquatic ecosystems, 'plants' are represented by unicellular algae (phytoplankton). Being very small and unicellular, these organisms have only minor requirements for structural material, and thus have lower C:(N, P) ratios than plants. As a result, their consumers (zooplankton) can be limited either by mineral nutrients (P, N) or by organic carbon of their prey, depending on the prey species and its physiological state.

In addition to limitation by 'macronutrients' (C, N, P), more complex biochemical compounds can represent a limiting factor in the prey of consumers. Especially some polyunsaturated fatty acids (PUFAs) and amino acids cannot be synthesized by many heterotrophic organisms, but must be consumed with their food. Such biochemical compounds are found in low concentrations in many plant species (incl. some phytoplankton), but not in animals.

Given the nutritional differences between vegetable and animal prey, it is not surprising that some primarily herbivorous organisms enrich their diets by ingesting other heterotrophs.

The example of calanoid copepods

Calanoid copepods are tiny crustaceans (0.5-2 mm in size; related to crabs and crayfish), that live in the open-water zone of lakes and seas. They represent the major trophic link between unicellular organisms (incl. phytoplankton) and planktivorous fish in most marine and in some freshwater ecosystems. While they were originally believed to be herbivorous, research in the last decade has revealed that heterotrophic protists comprise a considerable part of their diet. These heterotrophic organisms constitute high-quality food for the copepods with a high content of nitrogen, phosphorus (low C:(N,P) ratio) and biochemical compounds.

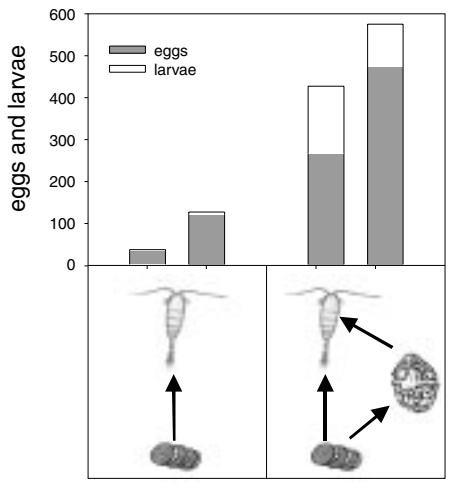


Fig. 2. Cumulative sum of eggs and larvae produced by each of two groups of copepods that were fed either on a pure phytoplankton diet (left side), or on a mixed diet with a protist as intermediate consumers (right side). Phytoplankton was represented by a diatom, *Skeletonema costatum*, the protist by a heterotrophic dinoflagellate, *Gyrodinium sp.* (Ptacnik 2003).

In a simple experiment, copepods were either fed on a pure phytoplankton diet, or on a mixed diet of phytoplankton and a protist, that was itself feeding on the phytoplankton (intermediate consumer, compare with Fig. 1). The reproductive success (eggs and larvae after 6 days of incubation) is displayed in Fig. 2. The obvious difference in reproductive success of the copepods in this experiment was most likely a result of a higher concentration of PUFAs in the heterotrophic protists compared to the phytoplankton (Ptacnik, 2003).

Omnivory in humans and our influence on natural food webs

The growing human population obtains its food from almost every ecosystem on earth. Humans are omnivores that need an animal fraction in their diet to obtain well-balanced nutrition. However, by virtue of our having become such a dominant species, our food pattern affects the fate of various ecosystems.

The strong decline in piscivorous (fish-eating) fish in the world's oceans is maybe the most dramatic example of our influence on natural food webs. Industrial fisheries first focused on large, piscivorous fish. As they were reduced to low levels, mesh-sizes of trawlers were reduced and smaller fish are now harvested ('Fishing down the food web', Pauly et al. 1998). As a consequence, industrial fisheries are threatening large predatory fish nowadays in two ways, by direct reduction, and by reduction of their prey (Fig. 3). Today, the stocks of large piscivorous fish in the world's oceans are approximately 10 % compared to pre-industrial levels (Myers and Worm, 2003).

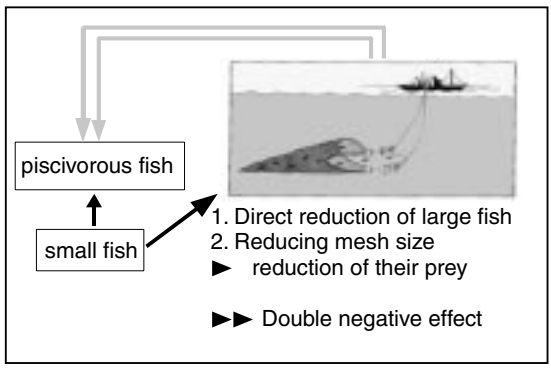


Fig. 3. Direct and indirect negative effects of industrial fisheries on large piscivorous fish.

References

Diehl S. 2003. The evolution and maintenance of omnivory: Dynamic constraints and the role of food quality. *Ecology 84* (10): 2557–2567.

Holt RD, Polis GA. 1997. A theoretical framework for intraguild predation. *Am. Nat.* 149 (4): 745–764.

Pauly D, Christensen V, Dalsgaard J, Froese R, Torres F jr. 1998. Fishing Down Marine Food Webs. *Science* 279 (5352): 860–863.

Pimm SL. 1982. *Food Webs*. Chapman & Hall, New York.

Ptacnik, R. 2003. *Omnivory in planktonic food webs: a study on the impact of mixotrophic flagellates and microzooplankton on food web dynamics and productivity*. Dissertation, Institute for Marine Research, Kiel.

Myers RA, Worm B. 2003. Rapid worldwide depletion of predatory fish communities. *Nature* 423: 280-283.

Thingstad TF, Havskum H, Garde K, et al. 1996. On the strategy of ''eating your competitor'': A mathematical analysis of algal mixotrophy. *Ecology* 77 (7): 2108–2118.

MAGNE DYBVIG

Modern Theories of Consciousness: Some Alternatives

The problem

Many philosophers have believed that it will forever prove impossible to explain mind in terms of matter. The famous philosopher, mathematician and physicist Leibniz (1646 – 1716) expressed this fundamental intuition thus: "Supposing that there were a machine whose structure produced thought, sensation and perception, we could conceive of it as increased in size with the same proportions until one was able to enter into its interior, as one would into a mill. Now, on going into it one would find only pieces working upon one another, but never would one find anything to explain perception" (Monadology, §17)

Professor Magne Dybvig,
Department of Philosophy,
Norwegian University of Science
and Technology, NTNU.
CAS Fellow 2003/04.

Nowadays the problem is often formulated in terms of brain and consciousness: Can consciousness be explained in terms of brain processes? We know a lot about how mental processes depend on neural processes. But what we may call "Leibnitz' problem" – that no matter how exhaustively you study the brain as a material mechanism, it seems that you will only find further material phenomena, not consciousness – is still with us.

The relationship between mind and brain: The main positions

In modern times, before the 20th century, the most popular interpretation of the mind-brain relationship was some version of *dualism*. It claims that mind is essentially non-physical. The brain is the place where this non-physical reality interacts with physical reality. The reason why you cannot "see" the mind when you inspect the brain is that the methods of inspection are adapted to the observation of material phenomena, and not to the observation of immaterial phenomena like e.g. thoughts. So what you can inspect using the methods of the natural sciences, is at most the *correlates* of consciousness, not the conscious itself.

In the 20th century, a series of *materialist*, or *physicalist*, alternatives to dualism have been developed. The main positions are (philosophical) *behaviourism*, the *identity theory*, *functionalism* and *eliminativism*.

Behaviourism: According to (philosophical) behaviourism the mind is simply the behaviour, or dispositions for behaviour, that an organism exhibits. The brain is not the mind, but the mechanism that enables mind – i.e. the underlying mechanism that enables the complex behaviour which *is* the mind. And the reason why you cannot observe mind by

simply observing the brain is not that mind is something immaterial. The reason is that you are so to speak looking in the wrong place – at the mechanism that makes mind possible, not at mind (the behaviour) itself.

Identity theory: A frequent objection to behaviourism is that we think of mind not as the behaviour itself but as what *causes* and *regulates* behaviour. And what causes and regulates behaviour are brain states; so mental states *are* brain states according to the (neural) identity theory. This mind-brain identity must be accepted as a kind of scientific truth, comparable to e.g. the identity of light and electromagnetic waves. So the states that you inspect when you inspect the brain *are* (some of them) mental states – it is only that you will not *recognise* them *as* mental states until you have developed the right 'theoretical spectacles'.

Functionalism: An objection to the identity theory is that mental phenomena, e.g. pain, can be realised in the brain in many different ways, depending on what kind of organism we are talking bout. According to functionalism, mind is not brain states, but something more abstract – namely the functional states the brain can be in. Anything (e.g. a complex robot, or an extraterrestrial being) with inner states that performed the right functions would have a mind, even if it did not have a biological brain. In functionalism the relationship between brain and mind is often compared to the relationship between hardware and software. And the reasons why you cannot observe mind by just observing brain processes, is that you are not focusing on a sufficiently abstract level – you are like an engineer who does not understand a computer because he only sees the electronic *hardware* and not the *software* (i.e. the set of programmed functions) that runs on this hardware.

Eliminativism/instrumentalism: What is common to behaviourism, identity-theory and functionalism is a belief that mental phenomena are real phenomena that can, in the end, be described in terms taken from the natural sciences (including biology) – either as behaviour, or neural states, or functional states. Eliminativism maintains that this is not the case – our common sense conception of mind is a *theory* of mind ("folk psychology") that is basically wrong, so that nothing corresponds to mental phenomena "in the real world". A correct theory will only refer to brain states and behaviour, not mind. Mind is at most *a useful fiction* (instrumentalism); and the reasons why you cannot observe the mind by observing the brain, are simply that the mind does not exist – there is no mind to observe.

Different phenomena may require different theories

None of the theories mentioned above have been generally accepted among philosophers working on the mind-brain relationship. Many look on themselves as some kind of materialists (or "physicalists"). Few are fully-fledged dualists, but elements of such a position can also be found in contemporary philosophy – notably the following two points: Consciousness cannot be completely reduced to brain processes, and the study of it requires (in addition to methods found in the natural sciences) some special methods – a special kind of *self-observation* (introspection, or "phenomenological descriptions") and perhaps some kind of *interpretation* of behaviour (similar to the interpretation of texts).

It may be that different theories fit different types of mental phenomena. In philosophy it is for instance usual to distinguish between states that have a kind of semantic content, similar to the content of

words (e.g., both a thought and an assertion can have the content 'it is raining'), and states that lack such a content (sensations). The first are sometimes called propositional attitudes (or intentional states), while the latter are called qualia. It may be that the understanding of *propositional* attitudes requires some kind of interpretation, and that a full understanding of qualia is impossible without some kind of introspection. It has also been claimed that while propositional attitudes can be understood in functional terms, qualia require a dualist theory – they are irreducibly mental. And some materialists (who reject dualism) have claimed that a functional theory may be true for propositional attitudes, while some kind of identity theory is true for qualia.

The answer may also depend on how we conceive the relationship between mind and brain

Traditionally philosophers have thought of the relationship between mind and matter either in terms of identity ('the mind is nothing but brain states and/or behaviour') or in terms of causality ('mind is different from brain states, but somehow caused by brain states'). Lately it has been proposed that it would be better to think of the relationship as a kind of *supervenience*-relationship. Mental states supervene on brain states if it is impossible to have a change of mental states without *some* change in brain states. Or conversely: Complete similarity in brain states entails complete similarity in mental states. Such a relationship implies that the mental is a kind of function of the brain even if it should prove impossible to formulate exact causal laws for how mind depends on the brain.

It has also been pointed out that *individual* mental events (e.g. the pain that I feel just now) can be identical with individual brain events (e.g. the firing in C-fibres going on just now) without *the properties* of mental events necessarily being identical with neurological properties. The first type of identity is called "token identity" while the latter is called "type identity". If this view is accepted one can for example say that the pain I feel is in fact token-identical with some brain event, while it has properties (e.g. 'being a throbbing pain') which *cannot* be identified with neurological properties (though they probably *supervene on* such properties). Such a view is often called *non-reductive* physicalism, and may be considered a kind of compromise between a physicalist and a dualist position.

Understanding mental phenomena may require a 3-stage functionalist strategy

Among the physicalist theories mentioned above, it is probably functionalism that has had most adherents in recent years. If you accept such a theory, it is natural to claim that the study of mind has three stages or dimensions:

Stage 1: Formulate a functional analysis of the mental phenomenon in question

Stage 2: Describe a psychological mechanism that implies ("implements") the function

Stage 3: Describe a physiological mechanism that implies ("implements") the psychological mechanism

Let us take 'consciousness' as an example. What does it mean that a state is conscious rather than unconscious? If you are a functionalist, you will try to describe the characteristic way in which a conscious (in contrast

to an unconscious) state functions. One proposal is that I can immediately make use of what I am conscious of, and that I can use it for many different purposes. So a functional analysis of consciousness could be: Consciousness = immediate & global availability.

Next you might ask what kind of cognitive mechanism implements (makes possible) this immediate and global availability. A possible proposal is e.g. that mind is organised around a 'global workspace' (a kind of 'working memory'; Baars 1988) and that a piece of information is conscious at a certain moment in so far it forms part of the global workspace at that moment.

The third question will then be what kind of *neurological* mechanisms implement the *psychological* mechanism (e.g. 'global workspace' or 'working memory') which implements *consciousness*, functionally defined. A proposal is that some kind of synchronized firings of neurons engaged in related tasks may play a key role in integrating ("binding") and storing information in working memory – or Baars' 'global workspace' – so that they become immediately and globally available in action and speech (for a brief survey of some cognitive and physiological mechanisms that might implement consciousness, see Chalmers 1996).

Is there an ambiguity in the concept of consciousness?

Though the above proposal for a 3-stage functionalist analysis of consciousness is highly speculative, one can at least see how consciousness 'in principle' could be a necessary consequence of the way in which the brain is organised – *if* a functionalist analysis of consciousness is possible. But is such a functionalist analysis really possible? It has been claimed (e.g. Block 2002) that we must distinguish between two concepts of consciousness. The first is consciousness as what one has *access to* ("access-consciousness"); the other is consciousness as *experience* ("phenomenal consciousness"). The first may be analysed in functional terms (more or less as suggested above); the other cannot be so analysed.

If one accepts this distinction, one may also claim that "consciousness" covers different phenomena that require different theories, as suggested above. One possibility is to combine a functionalist theory of access-consciousness with some kind of dualist theory (or non-reductionist materialist theory) of phenomenal consciousness. Another possibility is to combine a functionalist theory of access-consciousness with an identity-theory of phenomenal consciousness, where the latter is taken to consist in the occurrence of sensations ("qualia").

References

Baars, B. J. 1997. *In the Theatre of Consciousness*. Oxford: Oxford University Press.
Block, N. 2002. Concepts of Consciousness. In Chalmers 2002.
Chalmers, D.J. (ed.) 2002. *Philosophy of Mind*. Oxford: Oxford University Press
Chalmers, D. J. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
Leibniz, G.W. 1961. *The Monadology*. In Leibniz: Philosophical Writings. London: J.M. Dent & Sons.

LARS-GÖRAN NILSSON

Memory and Genes – is there an Association?

In recent years, there has been an increasing interest in examining the role of genetic markers when it comes to cognitive functions. Researchers in behavioural genetics have made great efforts to make this area known to researchers in cognitive psychology. Plomin (1999; Plomin & Crabbe, 2000; Cardon, Faulkner, DeFries & Plomin, 1992) have been some of the most influential contributors in this respect. This new development has aroused great interest among cognitive psychologists, who now see genetics as an important tool to extend current knowledge about cognitive functions. The basic question in this kind of research is to determine whether, and, if so, to what extent genes can be associated with various cognitive functions. If such associations can be established, the question is to find out what the underlying mechanism is.

Professor Lars-Göran Nilsson,
Department of Psychology, Stockholm University, Sweden.
CAS Fellow 2003/04.

In this paper, I will describe some studies that have focused on potential associations between some genes and human memory function. Nilsson et al. (1996) studied associations between six serum protein polymorphisms and two forms of memory, episodic and semantic memory. These polymorphisms were complement C3, haptoglobin, properdin factor B, orosomucoid, group-specific components, and transferring C. We predicted that complement C3 and the acute-phase reactant haptoglobin should be of special interest as immune response factors. As expected we found strong associations between these two markers and episodic memory suggesting that immune response factors may be of importance in preserving episodic memory. In the haptoglobin system, there was evidence of a primary phenotypic association involving heterozygotes. An association involving heterozygotes indicates that linkage disequilibrium with alleles at other loci influencing memory function is unlikely. The association with complement C3 alleles may be due to either linkage disequilibrium or functional involvement at the protein level. It is noteworthy that the genetic associations demonstrated in this study hold for episodic memory but not for semantic memory. These two memory systems differ in several ways. Episodic memory is responsible for remembering personal events that are defined in time and space. For example, in order to remember what was served for dinner last Saturday, one has to travel backwards in time to recall where one had dinner. The temporal and spatial cues emerging in doing this may then help in retrieving what was served for dinner. This travel backwards in time requires a conscious recollection of the dinner episode on Saturday. Such a conscious recollection of a certain study episode is not required for semantic memory, which is a memory system for general knowledge. For

example, in order to recall that the chemical formula for regular table salt is NaCl, it is not necessary to travel backwards in time to a certain episode in school when this piece of information was, most likely, first encountered. It is more likely that the response is generated by knowing implicitly that table salt is a chloride of sodium. At any rate, the finding of this dissociation between episodic and semantic memory is an important result for memory theory because it adds to other dissociations between episodic and semantic memory, thereby providing converging evidence for the differentiation between these two memory systems.

Another gene that we have studied in our own laboratory is ApolipoproteinE (APOE). This gene is located on chromosome 19. Its primary role is to influence the metabolism of lipids, primarily cholesterol. There are three alleles of this gene: ε2, ε3 and ε4. The most common form is ε3 occurring in about three-fourths of the population; ε2 and ε4 occur in about 10% and 15% of the population, respectively. The three alleles of APOE form six possible genotypes, 22, 23, 24, 33, 34, and 44. Allele ε4 is a risk factor for cardiovascular disease in middle age and for Alzheimer's disease in old age. The pathophysiological mechanism for the "bad" ε4 allele is not yet fully understood. One claim is that ε4 does not protect key neuronal structures from excessive phosphorylation, which leads to neuronal degeneration. Another claim is that APOE is involved in the continuous synthesis and repair of cellular membranes. Stone et al. (1998) demonstrated that the ε2 and ε3 isoforms of APOE serve an important role in this repair work, whereas the ε4 allele is less successful in this work. Persons with the ε2 and ε3 alleles receive necessary neuronal protection and are much less likely to develop cardiovascular disease and Alzheimer's disease.

In the Betula Study (Nilsson et al., 1997) we excluded persons with dementia and cardiovascular disease in order to examine whether the three APOE alleles have any direct influence on memory functions in healthy individuals in adulthood and old age. The results in several studies (Nilsson et al., 2001, 2002, in press) demonstrated that indeed the carriers of the ε4 allele show a lower episodic memory performance than carriers of the ε2 and ε3 alleles in old age (65-80 years). In Nilsson et al. (in press), we were able to demonstrate two additional, important findings. One finding was a dose effect showing that carriers of two ε4 alleles performed at the lowest level. Specifically, carriers of two ε4 alleles fail more profoundly in acquiring and recollecting episodic information than carriers of one ε4 allele, who in turn fail more than carriers of non-ε4 alleles. Although such a dose effect may support the notion that APOE has a direct effect on cognitive function, the dose effect per se does not necessarily differentiate this hypothesis from the alternative hypothesis of more preclinical dementia cases among the carriers of the ε4 allele despite the fact that care was taken to minimize the risk of including preclinical cases. However, the hypothesis of a differential potency of the APOE alleles in triggering (directly or indirectly) self-initiated cognitive processing needed for demanding memory tasks is certainly compatible with the dose effect demonstrated here. Another finding was that middle-age carriers of the ε4 allele showed a better performance than carriers of the ε2 and ε3 alleles in episodic memory tasks, and especially so in tasks requiring recall rather than recognition of information. The explanation of this unexpected result is not yet clear. The findings would seem to

suggest that a gene may have different functions at different stages in life. From an evolutionary perspective, it is difficult to imagine that one allele of a gene would have as its only function to lower cognitive function in old age. The present data may indicate that the ε4 allele indeed has some other, yet unknown, function, which is prominent earlier in life. A speculative hypothesis is that the ε4 allele may have a basic positive effect on the organism in early years and that this positive effect has a cost to the organism by means of exhaustiveness. When life expectancy was lower, this effect of the ε4 allele was not observed or noticed, but as life expectancy increases, it is increasingly devastating to the human mind and body. Obviously, more research is needed to explore this issue further.

A third approach in the Betula Study to examine this issue is about transmitter related genes. In the prefrontal cortex, the catechol O-methyltransferase (COMT) gene is essential in the metabolic degradation of dopamine, a neurotransmitter implicated in cognitive functions. In a 5-year longitudinal analysis, de Frias et al. (in press) examined the effect of a polymorphism in the COMT gene on individual differences and changes in memory in adults aged 35-85 years. De Frias et al. reported that carriers of the Met/Met genotype (with low enzyme activity) performed better on episodic memory, as compared to carriers of the Val allele (with higher enzyme activity). The COMT gene was not significantly related to semantic memory. Division of episodic memory into its recall and recognition components located the difference with respect to episodic recall; no gene-related differences were observed in recognition. The memory dissociation is parallel to that observed with carriers of the ε4 allele of APOE. The effect of COMT on memory was similar for middle-aged, young-old, and old-old adults and held across a 5-year period. Thus, the COMT gene is another candidate gene for memory functioning in adulthood and old age.

References

Cardon, L.R., Faulkner, D.W., DeFries, J.C., & Plomin, R. (1992). Multivariate genetic analyses of specific cognitive abilities in the Colorado Adoption Project at age 7. *Intelligence,16,* 383–400.

de Frias, C. M., Annerbrink, K., Westberg. L., Eriksson, E., Adolfsson, R., & Nilsson, L.-G. (in press). COMT gene polymorphism is associated with declarative memory in adulthood and old age. *Behavior Genetics.*

Nilsson, L.-G., Adolfsson, R., Bäckman, L., Cruts, M., Edvardsson, H., Nyberg, L., & Van Broeckhoven, C. (2002). Memory development in adulthood and old age: The Betula prospective-cohort study. In P. Graf & N. Ohta (Eds.), *Lifespan development of human memory.* (pp. 185–204). Cambridge: The MIT Press.

Nilsson, L.-G., Adolfsson, R., Bäckman, L., Cruts, M., Small, B.J., & Van Broeckhoven, C. (in press). The influence of APOE status on episodic and semantic memory: Data from a population-based study. *Neuropsychology.*

Nilsson, L.-G., Bäckman, L., Nyberg, L., Erngrund, K., Adolfsson, R., Bucht, G., Karlsson, S., Widing, G., & Winblad, B. (1997). The Betula prospective cohort study: Memory, health, and aging. *Aging , Neuropsychology and Cognition, 4,* 1–32.

Nilsson, L.-G., Sikström, C., Adolfsson, R., Erngrund, K., Nylander, P.-O., Beckman, L. (1996). Genetic markers with high versus low scoring on episodic memory tasks. *Behavior Genetics, 26,* 555–562.

Nilsson, L.-G., Van Broeckhoven. C., & Adolfsson, R. (2001). Genetic Contributions to Individual Differences in Memory Performance. *European Psychologist, 6,* 264–271.

Plomin, R. Genetics and general cognitive ability. *Nature, 402 (Suppl),* 25–29.

Plomin, R., & Crabbe, J. (2000). DNA. *Psychological Bulletin, 126,* 806–828.

SVEIN MAGNUSSEN

What do People believe about Memory?

Beliefs about the limits and the reliability of human memory are important because they govern our judgements and evaluations of the stories other people tell in everyday and forensic contexts. Psychologists and neuroscientists have been studying memory for more than a century. To what degree have the results of scientific research been incorporated in the psychological folklore? Do people typically nurture ideas about memory that conflict with the current knowledge, or do they have scientifically realistic ideas? To have some tentative answers to these questions, we carried out a nationwide telephone survey in two steps, asking representative samples of 1000 adult Norwegians a set of general questions about memory (Magnussen et al., 2004), selecting questions that are frequently asked by the media.[1]

Professor Svein Magnussen,
Department of Psychology,
University of Oslo.
Group leader CAS 2003/04.

We started with two questions that memory experts are frequently asked. First, do you think it is possible to train memory? Weekly magazines publish articles on memory enhancing techniques – how to improve your memory – most of which are rephrasing of the various mnemonic techniques, but sometimes the mnemonic techniques are presented as suitable for overall memory improvements. The scientific literature on memory expertise indicates that the superior memory of experts in the various fields, such as chess and sports, is limited to domain-relevant information and does not carry over to other fields (Tulving & Craik, 2000). Thus scientists would tend to answer that memory cannot be exercised in this way. However, when we probed this "muscle" concept of memory, the results showed that an overwhelming majority of the participants, 94%, believe that the memory capacity can be trained, and only 2% were sceptical. Closely linked with the idea of memory exercising, is the question of whether long-term memory has a limited storage capacity or is limitless. This question is illustrated by the textbook anecdote of the professor of ornithology who stopped learning the names of his students because each time he learned the name of a student he forgot the name of a bird. However, memory science is not aware of any limitation on the amount of information the brain is able to store and retrieve. Most classical papers on the memory for large amounts of information suggest that human long-term memory is virtually limitless. Recent evidence that the brain is continuously forming new synapses and even growing new neurons (Gould et al., 1999) suggests a system that might be expanding according

to the needs. Whatever memory researchers might believe, the results of the survey show that a majority of the participants (69 %) believed there was a limit to memory.

How does memory change across the life span? We asked four questions. First, what do people believe about the memory of small children as compared with the memory of adults? The scientific evidence is quite clear, the memory reports of children aged 3-6 years are basically correct if they are questioned properly, but contain fewer details than do the stories of older children and adults (Peterson, 2002). On this question, the public does not agree with science. A large majority (75 %) believed that small children's memory was at least as good as the memory of adults, and 38 % of the participants even thought it was better. This is interesting, given the daily experience parents have that children do not tell very much of what happened in the kindergarten or in school, and when asked what they did, a frequent answer is "we played". Second, how well do adults remember their early childhood? The concept of childhood amnesia refers to the inability of adults to remember anything from the early years of life Rubin (2000), usually before three years of age, although each of us may possess a grey zone with memory glimpses and vague images before genuine episodic memories emerge (Peterson, 2002). It is, however, unlikely that public belief when it comes to early memories would be shaped by the results of memory research; rather it is influenced by the informant's own childhood memories and would therefore conform to science, which it did. Very few informants (1 %) believed it was possible to have memories from birth onwards, and a few more believed that it was possible to have memories from the first year; in fact the public is more conservative than science, as more than 50 % of the participants believed that no memories were available before four years of age. This might be a little surprising, given the frequent articles on age regression in the popular media and the current popularity of various regression exercises offered at courses, seminars and non-professional therapies. Obviously, the overwhelming majority of the readers and listeners remain soundly unconvinced by such claims.

Episodic memory is the last form of memory to develop and the first to decline in old age, the latter fact being the target of many jokes about "Alzheimer light" among adults when something has slipped from the mind. We asked 1000 participants to judge their own memory performance over the last five years – had it become better or worse – and another 1000 participants to tell us at what age they believed memory started to decline. The results revealed interesting discrepancies between ratings of one's own memory, the general belief in time of onset of memory decline, and the objective finding from large-scale studies of memory changes in the adult life span. People have an unsupported pessimistic view of their own memory. Forty-three percent of the participants between18-29 years of age reported that their memory had declined, a similar proportion of participants between 30-44 years reported a decline, and this figure rose to 50% for participants aged 45-59 years and to 62% for participants over the age of 59. However, when an equivalent sample of participants were asked when they thought age decline stared, only 6% believed it started before 30 years of age, and more than 50% of the participants, irrespective of their own age, believed it started after the age of 50. The results of empirical research suggest, however, that the general change in perform-

ance on episodic and semantic memory would not be noticeable until well after 60 years of age (Nilsson, 2003). When healthy young to middle-aged people claim memory problems, it is probably mostly due to misattribution of the normal memory problems all people have rather than genuine age changes.

On September 10, 2003, the Swedish foreign minister Anna Lindh was stabbed to death in a shopping mall in Stockholm, in front of many people. How well will these witnesses later remember this tragic and dramatic event? Are traumatic events remembered better or more poorly than ordinary events? The answer is not obvious. On the one hand, it might be argued that such events are frequently fast moving and that observations are unreliable, or that the drama of the event would lead to emotional activation that might interfere with, or block observational capacities and memory encoding. Therefore, the memories of emotional, dramatic events might be dim. Or it might be argued that such memories are so frightening that they are not allowed into consciousness, they are "repressed". On the other hand, it might be argued, as current memory researchers do argue, that emotional activation might act by focusing attention and facilitating encoding of attended details, which would lead to enhanced memory for some aspects of the event at the expense of other aspects, but leading to vivid subjective memories of the traumatic event. The results of empirical studies appear to be quite straightforward. Traumatic events are better remembered than ordinary events both by children and adults, even if the memories are subject to similar distortions as memories of ordinary events (McNally, 2003). The results of the survey show that the majority of informants agreed with science, with 70% responding "better" and only 11 % responding "worse" on the simple question whether dramatic events were better or more poorly remembered that non-dramatic events.

We also asked two questions that were directly aimed at probing the idea that frightening events might be repressed. One of these specifically mentioned the self-reported amnesic murderer – between 25 and 70 % of suspects of violent killings claim no memory of the event (Parkin, 1997) – and asked whether the participants believed that such claims were real or faked. The idea that traumatic memories are blocked from consciousness can be traced to the psychoanalytic concept of repression, originally formulated to explain the blocking of painful childhood memories from conscious recollection. However, the concept of repression does not belong to the arsenal of mechanisms of forgetting in current memory research (Tulving & Craik, 2000), as it does not stand the test of relevant real-life studies of traumatized individuals (Goodman et al., 2003); indeed trauma-induced psychogenic amnesia is extremely rare, if it exists at all. Rather, studies of war veterans, some of whom may themselves have committed gruesome acts, and of victims of such acts, suggest that these memories persist all too well (McNally, 2003). The participants were split in half, with a small majority voting for faking, and interestingly the number of participants who believed the amnesia was faked was higher among participants who had only completed elementary school (20%) than among participants in possession of a university degree (46%). Sometimes folk psychology beats intellectual speculation.

References

Goodman, G.S., Ghetti, S., Quas, J.A., Edelstein, R.S., Alexander, K.W., Redlich, A.D., Cordon, I.M. & Jones, D.P.H. (2003). A prospective study of memory for child abuse. New findings relevant to the repressed-memory controversy. *Psychological Science*, 14, 113–118.

Gould, E., Reeves, A.J., Graziano, M.S.A. & Gross, C.G. (1999). Neurogenesis in the neocortex of adult primates. *Science*, 286, 548–552.

Magnussen, S., Andersson, J., Cornoldi, C., De Beni, R., Endestad, T., Goodman, G., Helstrup, T., Koriat, A., Larsson, M., Melinder, A., Nilsson, L.G., Rönnberg, S. & Zimmer, H. (2004). What people believe about memory. Submitted for publication

McNally, R.J. (2003). *Remembering trauma*. Cambridge, MA, Harvard University Press.

Nilsson, L.-G. (2003). Memory function in normal aging. Acta Neurologica Scandinavia, 107 (Supplement 179), 7–13.

Parkin, A.J. (1997). *Memory and amnesia*. Oxford, Blackwell Publishers

Peterson, C. (2002). Children's long-term memory for autobiographical events. *Developmental Review*, 22, 370–402).

Rubin, D.C. (2000). The distribution of early childhood memories. *Memory*, 8, 265–269.

Tulving, E & Craik, F.I.M. (2000). *The Oxford handbook of memory*. Oxford, Oxford University Press.

Notes

1 In the survey we avoided complicating the issue by calling attention to the various forms of memory that scientific taxonomies define, but accepted that memory is simply what people believe it is.

ANNIKA MELINDER

Who would believe in a Liar?

Although quite anonymous in Norway, witness psychology is both a practical and theoretical segment that has grown out of cognitive psychology. A significant branch within witness psychology is related to judgements of a person's credibility, such as whether emotional expressions are reliable signs of truth, or whether there is any typical behaviour associated with truth so that one might make predictions from a person's behaviour about his or her statements and testimony. Knowledge as such is important for the professional lie detector as well as in the private sphere when one wants to distinguish deception among otherwise credible individuals. Because we do lie. Research using students' diaries and people's notes has shown that students lie at least twice a day and that the average man adds a creative touch to the truth at least once a day (DePaulo et al., 1996).

Dr. Annika Melinder,
Department of Psychology,
University of Oslo.
CAS Fellow 2003/04.

Even if there are not any gender differences in the number of lies we tell, the content may differ so that women tend to lie more about their physical appearance, whereas men tend to exaggerate their earning potential (Memon, Vrij & Bull, 2003). To mention another disappointment: when people have described the most serious lie they have ever told to somebody else, overwhelmingly many report that the targets of these lies were romantic partners (Anderson, Ansfield, & DePaulo, 1999). Such lies were told to cover serious issues, such as infidelities. Would we really like to know about other's betrayals? If so, is there any method that can reliably separate the truth from a lie? Or that can point to who is credible and who is not?

Unfortunately, researchers have found no reliable behavioural sign of truth, nor have they found any significant verbal cues to deception. However, there are still influential experts who claim to be able to help categorize liars from truth-tellers by using simplistic systems. In the following we survey briefly what empirical evidence there is in order to state our abilities to detect deception.

The naïf view

Psychologists have discussed the question of emotional reactions and their relation to cognitive processes from the very beginning of this field, which this statement captures: " … the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion" (William James, 1884). Such ideas, which have indeed been empirically tested and gained some support, are the ground for the naïf view of how and why states of the mind, including deception, can be transformed into observable behaviour.

The underlying assumption is that lies evoke emotions (anxiety and fear), which are out of the range of behavioural control even if the person

can control the verbal communication. The second assumption is that deceptive behaviour demands more cognitive resources than truthful behaviour, which will consequently lead to less monitoring of one's own behaviour (Vrij, 2000). Following both assumptions, one will expect more nervous and anxious behaviour when a person is telling lies, than when she or he is telling the truth. Primarily one will predict more bodily reactions such as sweat and higher blood pressure, but also more changes in the pitch of voice and speech flow when people are telling lies. Note, however, that such reactions may be caused by other sources. For example, when people are accused of something of which they are not guilty, the accusation *per se* can produce an increased level of arousal, which will evoke the same bodily responses as described.

Deception theories that are based on the notion that "emotion (evoked by the state of stress that a liar is supposed to experience) leads to specific deceptive responses," depend on certain facts from empirically supported theories. However, these facts tend to be misused in certain deceptive frameworks (e.g., Inbau, Reid, & Buckley, 1986; Inbau, Reid, Buckley, & Jayne, 2001) so that specific behaviour is interpreted as an indication of deception (gaze aversion, fidgeting, placing hand over mouth, and postural shifts), whereas other behaviour is interpreted as a sign of truth (looking into the other's eyes). Of course, trusting wrong cues could lead lie detectors into false beliefs. When this happens in a forensic setting, an individual's rights may be endangered, as researchers have pointed out (Vrij, Semin, & Bull, 1996; Vrij, Edward, & Bull, 2001).

Professionals' expertise when detecting lies

Ekman and O'Sulllivan (1991) explored the assumption that untruthful intentions would be reflected in behaviour and detected by trained professionals. Thus, they let experienced professionals, such as secret service officers, judges, psychiatrists and polygraph operators see a video tape of different women who talked on-line about the feelings they felt when viewing either positive or aversive scenes. The women had been instructed to state how they felt (truth) in half of the cases, whereas they reported the opposite feeling (lie) with respect to what they felt in the other half of the cases. The professionals' task was to tell who was communicating their honest feelings, and who was not. Table 1 shows the disappointing results,

Table 1. Professionals' Ability to Detect Deception

	Truth	Lie	Total
DePaulo & Pfeifer (1986) (law enfor)	64	62	53
Ekman & O'Sullivan (1991) (secret service)			64
Ekman & O'Sullivan (1991) (polygraphers)			56
Ekman & O'Sullivan (1991) (police officers)			56
Ekman, O'Sullivan, & Frank (1999) (CIA)66		80	73
Ekman et al. (1999) (sheriffs)	56	78	67
Ekman et al. (1999) (law enforcement)	54	48	51
Koehnken (1987) (police officers)	58	31	45
Meissner & Kassin (2002) (law enfor)			50
Porter et al. (2000) (parole officers)	20	60	40
Vrij (1993) (police detectives)	51	46	49
Vrij & Graham (1997) (police officers)			54
Vrij & Mann (2001a) (police officers)	70	57	64
Vrij & Mann (2001b) (police officers)	51		
Total	**55**	**55**	**55**

Note. Percentage of Professionals' Ratings on Behaviour regarding Truth and Lies. The Table has been adapted from Vrij, 2002.

namely that even highly trained professionals perform badly. All groups – except the secret service officers – scored at chance level, a result that has been replicated in several studies and reviewed by Vrij (2002).

It is reasonable to argue that high stake situations evoke other aspects in the suspect's behaviour than laboratory research can model. Practitioners have therefore claimed that researchers should use realistic settings when studying professionals' capabilities. In line with this, Vrij and Mann (2001a) exposed experienced police officers to video films from press conferences of a person who was appealing to the public for help in finding a missing relative, or the murderer of a dead relative. In some of these cases, the person who had appealed to the public was subsequently found guilty of the crime committed. Thus some of the persons appealing were lying during the press conference and the task of the police officers was to judge whether or not the person appealing for help was guilty or innocent. And, voilà, the result does slightly improve although it is the detection of truth that contributes to the enhanced result, and not the detection of lies (Table 1). A point of interest is that criminals have shown themselves able to outperform students in terms of detecting lies, but not in detecting truth (Hartwig, Granhag, Strømvall & Andersson, 2004).

Reasons for incorrect beliefs

So far, we can note that 1) there is no such thing as directly observable behaviour that corresponds to deception, and 2) professional experience does not help in making correct evaluations. It would be interesting in this context to know what kind of guidance people, including professional lie detectors, use when cuing for truthful versus deceptive behaviour. It could be that people are taught the wrong cues. When people, again including professional lie detectors, are asked to describe what verbal and non-verbal behaviour they regard as reliable signs of credibility, the reports show an incredible similarity across respondents. Hence, if the person avoids eye contact, smiles (more insecurely), closes his/her eyes more frequently, and shows more bodily anxiousness such as nods and shaking the head, well then people tend to believe that the person is lying. When specifically asked about it, even experienced police officers reported avoidance of eye contact as the most frequent sign they used in the evaluation of a suspect's performance (Vrij & Mann, 2001). Compared to controlled studies that have been conducted of actual behaviour when lying and a meta-analysis of 116 empirical studies of 158 signs of lying (DePaulo et al., 2003), few of these subjective signs correspond to the objective indicators, as shown in Table 2.

Table 2. Cues Present during Deception

pitch of voice	>
illustrators	<
hand/finger movements	<
logical structure	<
unstructured production	<
quantity of details	<
contextual embedding	<
description of interaction	<
reproduction of conversation	<
unusual details	<
visual details	<
sound details	<
space details	<
time details	<
cognitive operations	>

Note. More (>) or less (<) while lying

In sum, professional lie detectors' and ordinary peoples' evaluations of credible behaviour seems to be based on stereotyped views which have no correspondence with objective facts. These stereo-

typed views largely influence peoples' own behaviour when they try to produce both truth and lies, and they guide us when we try to detect deception. If anything, a skilled liar knows what behaviour people in general regard as credible. Bill Clinton, when denying any sexual relationship with Miss Monica Lewinsky, looked firmly into the camera and eloquently assured viewers of his innocence.

Final remarks

Even if we know that people do not judge truthful versus lying behaviour correctly, research is inconclusive with regard to whether people actually perform in the same way when they are telling the truth as when they are telling a lie. Analyses of people's facial micro-expressions have shown that well trained experts can detect even minimal signs of emotional involvement. When trained to understand and observe a range of emotional expressions that are associated with correspondent muscle activation, Ekman, O'Sullivan and Frank (1999) showed that CIA agents and sheriffs did indeed score significantly above chance level (Table 1). Thus, evolutionary psychology postulates that strong emotions activate muscle actions in the face. When we become frightened for example, we tend to raise and pull together the eyebrows and to raise the upper eyelids and tense the lower eyelids. Narrowing of the lips and lowering of the eyebrows are equally descriptions of anger, and when the corners of our lips are pulled up, with creases in the skin below the eyes and crows-feet wrinkles beyond the corners of the eyes, then we are expressing (Ekman, 1992). In practice it is virtually impossible to observe micro-expressions, and even if we could, such expressions only mirror the emotional arousal, not the content of the person's thoughts.

If non-behaviour is unreliable as a detection tool, what about verbal behaviour then? Over the years researchers have developed check lists for analyzing statements and judging their reliability. An example of this type is the Content Based Criteria Analysis (CBCA), which includes 19 different scoring categories. The presence of the categories in a statement increases the likelihood that the statement is true. However, experimental studies are inconclusive regarding to what degree the instrument is a valid detection method. As can be seen from Table 2, cognitive operations (references to thoughts instead of perceptual features when recounting an event) and pitch of voice are the only two aspects that really increase when people are lying. People, including children, can produce detailed and elaborate stories about fictitious events that never took place, and, seen in this light, fabricate lies.

Even if we do not like to believe in a liar, there is not much in the way of scientific support for our selection of reliable signs of credible behaviour. Thus, at least once a day you will be betrayed.

References

Anderson, D. E., Ansfield, M. E., & DePaulo, B. M. (1999). Love's best habit: Deception in the context of relationship. In P. Philippot, R. S. Feldman & E. J. Coasts (Eds.). The social context of nonverbal behavior (pp. 372–409).

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129 (1)*, 74–118.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics and marriage*. New York: W.W Norton.

Ekman, P. & O'Sulllivan, M. (1991). How can we catch a liar? *American psychologist*, 46, 913–920.

Memon, A., Vrij, A., & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility*. Chiester: Wiley & Sons Ltd.

Hartwig, M., Granhag, P-A., Strømvall, L. A., & Andersson, L. O. (2004). Suspicious minds: Criminal's ability to detect deception. *Psychology, crime, and law*, 10, 83–95.

Inbau, F. E., Reid, J. E., & Buckley, J. P. (1986).). *Criminal interrogation and confessions*, 3rd edn. Baltimore, MD: Williams & Wilkins.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions*, 4th edn. Gaithersburg, MD: Aspen Publishers.

James, W. (1884). What is an emotion? *Mind*, 9, 188–205.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice*. Chichester: Wiley.

Vrij, A. & Mann, S. (2001). Lying when the stakes are high: Deceptive behavior of a murderer during his police interview. *Applied cognitive psychology*, 15, 187–203.

Vrij, A. & Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high stakes lies. *Psychology, crime and law*, 7, 119–132.

Vrij, A. Semin, G. R. & Bull, R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of nonverbal behaviour*, 20, 65–80.

Vrij, A., Edward, K., & Bull, R. (2001). People's insight into their own behaviour and speech content while lying. *British Journal of Psychology*, 92, 373–389.

BJØRN RAMBERG

Subjectivity, Emotion and Science

Introduction

Subjectivity seems to be quite different from the rest of the stuff in the world. Things that are endowed with subjectivity are special sorts of things; they feel, they think, and they act. They have, as we say, minds. Now I do not think that *mind* and *subjectivity* are synonymous terms. To my ear, ascribing *subjectivity* to some creature is to indicate an interested involvement in the world, an experiential point of view, and a stake in how the world unfolds. Perhaps these are properties also of minded beings, but if that is so, it is not, I think, simply a truth of semantics. Be that as it may; though I shall also touch on the topic of concepts and meaning, mere definitions of terms do not matter much for my main purpose in this context. I want to consider the relation between our practical experience and actual knowledge of subjectivity, on the one hand, and the rapidly developing scientific illumination of the capacities that subjectivity involves, on the other. I will suggest that if we are concerned with subjectivity, and interested in the relation between science, folk knowledge (our commonly shared core of assumptions and competencies) and philosophical inquiry, we ought to pay particular attention to the *emotions*.

Professor Bjørn Ramberg,
Department of Philosophy,
University of Oslo.
Group leader CAS 2003/04.

Empirical knowledge and conceptual change

Many of us take it that we have a *scientific* view of the world. This means at least that, though we may in fact be ignorant about much of what science has actually uncovered, we take it that the practice of science is the best route to knowledge of how things are. We take it that ordinary folk knowledge on various topics stands to be corrected by scientific discovery, as indeed much of it has been over the last centuries. However, as science advances and penetrates some domain of folk knowledge, this is not simply a matter of exchanging true (or at least systematically justified) belief for false opinion. Also our classifications, our concepts, are changed. Concepts as such are not true or false, though they may turn out to be empty (*phlogiston*). More often, however, conceptual revision rids us of concepts that are superficial, pointless, or based on misconceptions (*sublunar*). The great revolutions in natural sciences are all conceptual breakthroughs as well as empirical discoveries—indeed, the two are usually intrinsically related.

To say that our concepts change is to say that we begin to classify objects in some domain according to different properties. The history of biology, for instance, can be construed as a story about such change, as can some of the disputes among current theorists. The disagreement

between, for instance, Gould and Dawkins over evolutionary explanation, is less about which properties there are, and more about which properties biologically *matter* (Sterelny). To say that some property scientifically matters, is to claim that it has explanatory value. In the natural sciences, a key explanatory factor is *projectability* (Griffiths). To say that a property is projectable, is at least to say that it reliably co-varies with other properties of objects in the domain. What we want, scientifically, is to identify things by concepts that lock on to their systematically projectable properties. A scientific attitude, then, is to accept that while we still classify things in many ways for diverse purposes (*food*, *red*, *evil*, *square*), scientific classification has primacy when we want to say how things really are. With regard to physical elements, for instance, we have come to treat microstructure as having the final classificatory word. With respect to living matter, ancestry plays the essential role. Science, in such cases, progresses by homing in on *natural kinds* (Griffiths, Murphy).

Science and folk psychology

However, while we have little difficulty accepting such dynamics with regard to black holes and quarks, gold and H_2O, anemones and nucleic acids, things are not so simple when we turn to the matter of mind, to the properties of subjectivity. Past decades have seen tremendous progress in our scientific understanding of the biological mechanisms underlying many aspects of our mental lives. Psychology, biology, anthropology, and ethology have provided what some view as a revolution in our scientific understanding of ourselves and related beings. Yet there are philosophers who claim that the very nature of subjectivity is forever beyond the ken of any science with which we are now familiar (Nagel). The mind, they say, is in principle *irreducible* to the concepts of natural science. Others take the view that our folk-knowledge of mind is on a par with astrology and magic; as science progresses, our current terms of psychological interpretation and understanding will be replaced by quite different, genuinely explanatory concepts, typically taken to be fixed by the neural sciences (Churchland). When we want to talk about what the mind *really* is, say such thinkers, we have to talk about the brain, in terms that make perspicuous the connections between neural development, traits and events and what we call behaviour.

I think both these views are misguided. Subjective states may well be amenable to natural-scientific investigation, in pretty much the sense with which we are familiar. However, the fact that they are does not mean that folk-knowledge of mind must be rejected, or substantially revised in the name of scientific progress. The staunchly anti-reductivist attitude underestimates, it seems to me, the dynamic interaction between scientific concepts and ordinary categories. The scientistic view of mental categories as eliminable, by contrast, assumes that more precise neurobiological knowledge together with scientifically informed classification of behaviour will undermine our ordinary folk-psychological categories. The actual effect of such scientific success, however, may in fact be quite the opposite. An interesting example, worth looking briefly at before we turn to our knowledge of emotion, is the case of cognitive ethology.

Animal minds and human minds

The study of animal behaviour was dominated by strict behaviourist methodology long after the cognitive revolution took hold in human

psychology. There are many reasons for this, but one worth pointing out here is the lingering association of mental faculties with linguistic abilities. Thinking of *thought* as something intrinsically tied to *language* may have seemed to demystify the life of the rational mind, but it hampered our ability to conceive and explain the behaviour of "dumb brutes." Freed of this prejudice, cognitive ethology has over the last few decades evolved into a conglomerate of research programs. Partly as a result of dramatic improvements in neurobiological knowledge coupled with increased sophistication in evolutionary theory, concepts referring to cognitive, affective and communicative abilities are embedded in naturalistic accounts of animal behaviour (Allen and Bekoff; Bekoff, Allen, and Burghardt; Leary and Tangny). Animal minds are not black boxes but complex systems of representational abilities and affective responses.

So are human minds. They are, however, in some respects dramatically different from those of animals. The remarkably fine-grained and along some lines infinitely productive system of representation and communication that is human language, is indeed one extremely salient respect in which the human mind is unique. Moreover, while language is undoubtedly a biologically conditioned ability bestowed on us by evolution, there is good reason to think that the concepts we need to describe knowledge of meaning involve an understanding of norms of rationality (Davidson). If this is right, it has consequences for ordinary psychological states of the sort that we typically capture when we attribute particular beliefs and desires to each other, by means of linguistic content. When we language-users communicate and interpret each other by attributing what philosophers call propositional attitudes, we understand each other in terms of mental states that are tied to linguistic meaning. In so far as these states are captured through concepts sensitive to norms of rationality, there is no way to directly integrate them in a purely biological science of behaviour. Hence, conceiving of the mind in terms of propositional attitudes, states theoretically describable in decision-theoretic terms, appears to insulate genuine psychology from serious penetration by biological categories.

Emotion

The conception of mind as a set of action-explaining propositional attitudes is an idealization that captures important features of our mental lives and explanatory practices. It affords us, however, an impoverished approach to the nature of subjectivity and its relation to the terms of natural-scientific investigation of human experience and behaviour. Turning to emotions, our prospects are richer.

As we would expect, the great advances in empirical research into emotions have brought with them a number of classification proposals (Damasio; Ekman and Davidson; Griffiths). This diversity, however, may indicate more than the usual creativity displayed in the course of the rapid advancement of sciences into new domains. Careful examination of the nature of emotion as an explanatory, experiential and interpretative category shows that we identify emotions along a remarkably diverse range of axes (Elster). Particular sciences will take hold of at most some of these.

It has been reasonably well established that folk classifications of some states of emotion turn on properties that are quite robust, about which we can make discoveries, and with regard to which we may be corrected. The pan-cultural (Ekman; Griffiths) distribution of core features of what are

often called basic emotion attests to this. It suggests that some affective states are psycho-biological natural kinds, linking us to creatures without language, identifiable independently of fine-grained propositional attitudes. How such emotions arise in us and what they lead to, how we know them in others and ourselves, and what we are able to do by virtue of having and recognizing them—these are questions that various sciences are currently exploring. Yet, unlike other aspects of our nature that science can treat in this way, emotions are visible, indeed essential resources, within the context of ordinary interpretation itself. While we routinely explain behaviour rationalistically, in terms of the beliefs agents have and the goals that motivate them, we spontaneously understand behaviour also as expression of emotion, in a way that cannot be reduced to decision-theoretic terms. Furthermore, the interface between biological categories and affective states represents one end of an impressively wide spectrum; basic emotions may directly connect us with our biological natures and evolutionary roots, but the life of human emotion encompasses states that cannot be understood or perceived except in the context of propositional attitudes. Human emotions have, at least in significant measure, a *narrative* identity and significance, and depend on an ability to explicitly perceive actions and events as meaningful (Goldie).

I cannot show it here, but I believe that our knowledge of emotion in fact represents a competence with finely graded states across a spectrum, a spectrum tied at one end to biological response-mechanisms and at the other to uniquely human narrative abilities. This suggests that the diversity and heterogeneity of emotion-states is not arbitrary. Rather, the multifaceted folk-psychological category of emotion is a conduit between the dimensions of meaning, narrative and value, on the one hand, and the dimension of biological constraints and possibility, on the other.

References

Allen, Colin and Bekoff, Marc.1997. *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. Cambridge, MA: MIT Press..

Bekoff, Mark, Allen, Colin, and Burghardt, Gordon M. 2002. *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. Cambridge, MA: MIT Press.

Churchland, Patricia. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press

Davidson, Donald, 1984. *Inquiries Into Truth and Interpretation*. New York and Oxford: Oxford University Press.

Damasio, Antonio, 2003. *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Harcourt, Inc.

Ekman, Paul and Davidson, Richard J. (eds.), 1994. *The Nature of Emotion: Fundamental Questions*. New York and Oxford: Oxford University Press.

Elster, Jon, 1997. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.

Goldie, Peter, 2000. *The Emotions: A Philosophical Inquiry*. Oxford and New York: Oxford University Press.

Griffiths, Paul E., 1997.*What Emotions Really Are*. Chicago and London: The University of Chicago Press.

Leary, Mark R. and Tangney, June Price (eds.). 2003. *Handbook of Self and Identity*. New York: The Guilford Press.

Ledoux, Joseph. 2002. *Synaptic Self: How Our Brains Become Who We Are*. Harmondsworth: Penguin Books.

Nagel, Thomas. *What Is it Like to Be a Bat?* In *The Philosophical Review*, LXXXIII, 4 (October 1974): 435-50.].

Sterelny, Kim. 2001. *Dawkins vs. Gould : Survival of the Fittest*. Icon Books, UK.

JAN HARALD ALNES

Philosophy and Science

I shall sketch the views of two of the most influential philosophers of the last century, Willard van Orman Quine and Ludwig Wittgenstein, on the relationship between science and philosophy. In their talks here at the CAS, the philosophers have emphasized the fact that science *sets limits* to philosophy in the sense that one should be suspicious of philosophical positions which contradict, or are incoherent in the light of, established scientific knowledge.[1] The issue has been whether and to what extent these limits admit philosophical issues and philosophical reflection.

To Quine, the size of this "free space" is limited indeed. We begin with an extract from "Has Philosophy Lost Contact with People".[2] Quine has underscored that the great philosophers from the past, including Plato, Aristotle, Descartes, Hume and Kant "were scientists in search of an organized conception of reality"; and he continues:

Associate Professor
Jan Harald Alnes,
Institute of Philosophy,
University of Tromsø.
CAS Fellow 2003/04.

"Their search did indeed go beyond the special sciences as we now define them; there were also broader and more basic concepts to untangle and clarify. But the struggle with these concepts and the quest for a system on a grand scale were still integral to the overall scientific enterprise. The more general and speculative reaches of theory are what we look back on nowadays as distinctively philosophical (*Theories and Things*, pp 190 –191)."

Thus, philosophy used to be an integral part of science, and should continue to be so. As Quine views matters, philosophy presupposes science in the following two senses: It adopts the ontology of the sciences, and the evidence it invokes is of the same nature as that of the sciences themselves. Now, the ontological commitments of a theory can be determined as follows. One *paraphrases* the theory into first-order quantification theory and thereby determines the domain of the universal quantifier; this domain circumscribes the ontological commitment of the actual theory. The details of the second point are rather tricky, and Quine's view evolved over the years. Let me just observe that Quine characterizes scientific, or to be more precise, "sensory" evidence, by the use of such phrases as "the stimulation of sensory receptors", "surface irritations" and "the triggering of nerve endings". Note that this notion of evidence is *itself* a result of scientific development.[3] Quine's philosophical program is nicely captured by the title of his last book: *From Stimulus to Science*. This *very* long road, from stimuli, i.e. the triggering of nerve-endings, to science, i.e. a rather comprehensive theory about the world, with respect to the individual as well as to the culture to which she belongs, is to be accounted for by way of relying on the sciences themselves in the two ways mentioned. Quine mistrusts everything that smacks of a priori or "purely philosophical" reasoning – all reasoning takes place within the one big ongoing

scientific enterprise. Thus he rejects a number of the traditional philosophical tools, such as the classifications of truths into necessary or contingent and/or analytic or synthetic. Truth is truth, period.[4] The concluding passage of "Five Milestones of Empiricism" is telling:

> The naturalistic philosopher begins his reasoning within the inherited world theory as a going concern. He tentatively believes all of it, but believes also that some unidentified portions are wrong. He tries to improve, clarify, and understand the system from within. He is the busy sailor adrift on Neurath's boat (*Theories and Things*, p. 72).

We conclude so far by saying that although Quine's two-fold commitment to science is a metaphysical commitment, he objects to all other kinds of metaphysics.

Quine and Wittgenstein have many philosophical views in common. Both focus on the learning of language when they reflect on the philosophy of language, and both deny that meaning is a kind of abstract entity; in particular, they reject the idea that there are propositions in G. E. Moore and Bertrand Russell's sense of the term – a leading assumption of analytic philosophy since the turn of the 20th century. And both subscribe to holism in one version or another. In other respects, they are miles apart, and Wittgenstein, at least according to my reading, is even less of a metaphysician than Quine. This is particularly clear with respect to our present topic. For while the naturalistic philosopher is a busy sailor adrift on Neurath's boat, Wittgenstein's philosopher is not even on board. Wittgenstein's view is that philosophy is more or less completely independent of science, and he thinks that the philosopher from time to time is incoherent, confused and talks nonsense. Now, why is philosophy independent of science, and what kind of confusion are we talking about?

Wittgenstein returned to philosophy in 1929, after a 10-year break. A major aim of his *Tractatus*, the main work of his youth, had been to give the "*general form of propositions and language*" – one could certainly call this the *most* general aim that one can have in philosophy. But upon his returning to philosophy, this very aim becomes a main target of criticism. By now, the author of *Tractatus* accuses the philosophers, and foremost among them is no doubt his formerly close associate Bertrand Russell, of a "craving for generality"– they strive at formulating theories, and that makes them insensitive or even blind to the distinctiveness of the particular; they see similarities where one ought to look for differences. This shows up in a variety of ways, and the one to be focussed on here is the different kinds of, or uses of, concepts in science and in philosophy, respectively. In *Philosophical Investigations*, Wittgenstein introduces his notion of family resemblance. This notion plays a central role in Wittgenstein's response to his former self. He presents a number of different kinds of activities that we call "games", and he argues that the notion cannot be defined or explained by way of necessary and sufficient criteria; rather, he says "the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail" (*PI*, §66). And then in the next paragraph:

> I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. – And I shall say: 'games' form a family … (*PI* §67).

One has an option here: try to define "game", or accept Wittgenstein's argument that all such definitions would involve a more or less sharp break with our ordinary notion of a game. Here we shall pursue Wittgenstein's line and ask: What is the significance of this phenomenon, that of family resemblance? How can this distinction between kinds of concepts be used to draw a wedge between science and philosophy? As I understand him, Wittgenstein thinks that in the sciences the occurrence of such concepts play a relatively minor role, while in philosophy they play a decisive role.[5]

We need another notion introduced into the philosophical vocabulary by Wittgenstein, namely that of a language-game. *Very* roughly, we might say that a language-game is the product of a circumstance and a use of language; a language game consists of language-users and some material circumstance. Wittgenstein's point is that while scientific concepts are relatively stable from language-games to language-games, such is not the case with respect to concepts characterized by way of family resemblance. The most non-uniform of all such concepts are those of ethics and aesthetics, and he maintains that if you look for definitions corresponding to our concepts here, you will find none that satisfies you. Wittgenstein offers the following advice:

In such a difficulty always ask yourself: how did we *learn* the meaning of this word ("good" for instance)? From what sort of examples? In what language-games? Then it will be easier for you to see that the word must have a family of meanings (*PI*, §77).[6]

Thus, when the philosopher treats a family notion by way of theory, this theory simplifies or even goes against the very use of the notion. One thinks one is studying a phenomenon, but what one is studying is really the use of a word. In Wittgenstein's own phrasing: "Philosophical investigations: conceptual investigations. The essential thing about metaphysics: it obliterates the distinction between factual and conceptual investigations" (Zettel, § 458).

It is not clear to me whether Quine or Wittgenstein, or neither, is right. I believe that most philosophers, and certainly those at the CAS, take for granted that more meaningful work can be done in philosophy than is allowed for by Quine and Wittgenstein. And most, I think, are closer to Quine than to Wittgenstein. In closing, let me note that we have looked at two out of several ways of understanding the relationship between science and philosophy, and my tentative conclusion is simply that this issue is a philosophical one, and thus in no way neutral. But that might have been obvious from the outset.

References:

W. V. Quine: *Theories and Things*, Cambridge, Mass: Belknap Press, 1981.
W. V. Quine: *From Stimulus to Science*, Cambridge, Mass.: Harvard University Press, 1995.
L. Wittgenstein: *Tractatus Logico-Philosophicus*, London: Routledge, 1974.
L. Wittgenstein: *Philosophical Investigations*, Oxford: Blackwell, 1953.
L. Wittgenstein, *Zettel*, Oxford: Blackwell, 1967.

Notes

1 Let me note in passing that I shall not even make an attempt at spelling out the highly problematic "established scientific knowledge." (I understand it to contain both descriptive and normative aspects.)

2 The article is from 1978. It is Quine's response to Mortimer Adler's accusation that during the past half century, philosophy has been transformed into a new subject that no longer confronts questions of broad human interest, and no longer speaks to the ordinary man.

3 Quine needs some such notion of evidence in order to be an empiricist and to avoid "making scientific method …solely a quest for internal coherence" (Theories and Things, p. 39).

4 It should be evident that by way of the two restrictions we have looked at, Quine goes beyond Plato, Aristotle and the other philosophers that he mentions in "Has Philosophy lost Contact with People." But, as is made entirely evident in the essay with the informative title "Five Milestones of Empiricism," this is itself a result of the development of philosophy; cf. Theories and Things, pp. 67–73.

5 Family resemblance should be distinguished from related phenomena. It is not the same as vagueness. To be bald is a vague concept in the sense that there are persons about whom it is indeterminate whether it is correct to ascribe to them this property or concept. But still, when we reflect on the use of this concept we do not encounter a "complicated network of similarities overlapping and criss-crossing". An ambiguous term, on the other hand, is one with two or more distinct meanings, as for instance "bank", but neither of its distinct uses need have a "complicated network of similarities overlapping and criss-crossing".

6 I assume the parenthetical remark is directed at G. E. Moore's idea in Principia Ethica, that the main task of ethics or meta-ethics, is to give a correct general explanation of the concept 'good', applicable in every possible instance. Moore's view was that good is a simple, non-natural quality that certain things in the world exhibit.

TOR ENDESTAD:

Metaphors of Memory: to Reconstruct a Dinosaur

In our attempt to understand memory both scientists and laymen are forced to try to describe something unobservable. We can experience the effect of not remembering but have no introspective access to the mental processes that constitute a memory. In every day use of language to describe memory, a number of metaphors are in frequent use. We talk about memories that can be "lost", be "found", be searched for etc.

Numerous metaphors have been suggested to guide our scientific understanding of memory. Memory has been said to be like a wax tablet, an encyclopaedia, a muscle, a telephone switchboard, a computer and a hologram. Theorists have proposed core-context units, cognitive maps, memory tags, kernels, loops etc. (Underwood 1972). However one feature is common to most metaphors about memory. They seem to be based on the idea of an organized space; a storage of some kind. The space might have a structure of networks with nodes or paths or hierarchies with localizations and classifications. The nodes or localizations represent verbal, perceptual, propositional or other entities of memories. This of course has a tremendous impact both on how we scientifically understand memory and how we talk about and understand memory in every day life.

Associate Professor
Tor Endestad,
Department of Psychology,
University of Oslo.
CAS Fellow 2003/04.

In these terms we talk about "storing" memories, of "searching" for and "locating" them. We organize our thoughts; we "look for" memories that have been "lost". If we are lucky, we "find" them. In memory research two contrasting metaphors have inspired the enquiry: the multiple store metaphor and the archaeology metaphor.

The multiple store metaphor

The idea that it is useful to describe memory as composed of separate "stores" has guided theoretical research on memory over the last 40 years. Based on an analogy with the stores in computers, most of the research has been based on the distinction that Atkinson and Shiffrin (1968) made between a short term and a long term store in memory. The short term store is believed to hold information over a short period of time. During this time it has a certain probability of being transferred into the long term store. In the original model this was linked to the kind of processing that was performed with the information. Repetition would be one way of increasing the probability that information were transferred. The long term store was thought of as an infinite space where information was kept until it was found and brought forward.

The idea of separate stores in memory has been derived partly from the various memory stores found in computers. This has also had implications for how we have conceptualized the kind of units that can be stored in memory. A great deal of emphasis has been on verbal word-like materials, where units are not unlike those found in high level computer programs. Implications of this model are that memory traces are static structures with active processes between stores. In this context, to retrieve is to locate a memory and select appropriate information. Memories are viewed as complete; all information in a trace is available at the same time.

The store metaphor has inspired laboratory research on memory to the extent that the Atkinson and Shiffrin model often is called "the standard model" of human memory. The model has been developed to include a multitude of stores and processes to account for the empirical data (figure 1).

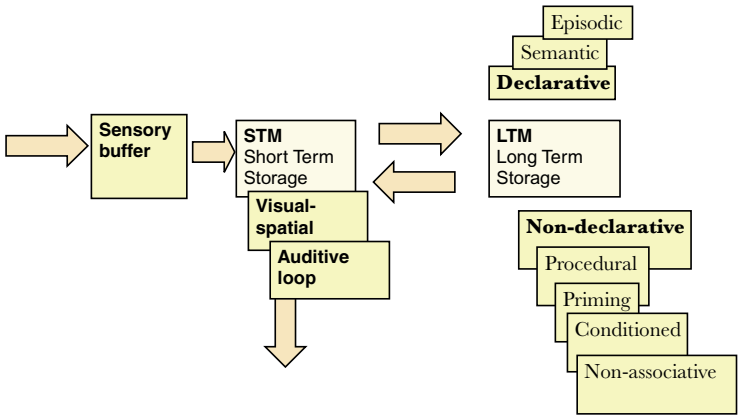


Figure 1. Extension of the Atkinson & Shiffrin model.

Several researchers have criticized the multiple store models for not capturing important memory phenomenon like false memories or incomplete memories. At the same time the multiple store metaphor has inspired researchers to perform well controlled experiments with lists of simple materials. Researchers observing memory phenomena in "real life" have questioned the validity of store inspired research. As an alternative Neisser (1967) proposed a different metaphor to account for everyday memory.

Memory as archaeology

In parallel with the store metaphor another metaphor encapsulating different aspects of memory has been developed. Bartlett (1932) observed that episodes are remembered in terms of common knowledge. This common knowledge Bartlett believed was a structure similar to schemata. The basic idea behind this metaphor is that remembering is some kind of reconstruction of memories from available information rather than a verbatim reproduction of the contents of memory. From remembering a general theme and some details we reconstruct our memory of an event, such as a story that has been told to us. Schema refers to an active organization of past reactions or of past experiences, which must be supposed to be operating in any well-adapted organic response." (Bartlett, 1932, p. 213).

These ideas were conceptualized by Neisser (1967) who stated the importance of the dynamic, reconstructive nature of human memory. He suggested that remembering "... likens the constructive work of a palaeontologist who uses a small set of bone fragments as well as general knowledge about dinosaurs and other similar animals in order to reconstruct and piece together the skeleton: "out of a few bone chips, we remember the dinosaur".

The implications of the archaeology metaphor are that memory traces are incomplete as opposed to the "all in one" structure in the store model. To remember is a process where memories are constructed, not found or selected. In this case to remember is a question of probability of match between an actual event and what is remembered (Koriat & Goldsmith, 1996).

Some remarks on metaphor and memory

There is no way to prove a metaphor wrong or right. Metaphors are conceptual tools that help us to understand phenomena in a more or less appropriate way. They provide a framework within which memory phenomena are analyzed and explained.

Metaphors highlight some aspects of a phenomenon and hide others. This means that it is important to recognize that research can be guided by metaphor to such a degree that we miss important attributes of a phenomenon. Both the multiple store metaphor and archaeology metaphor capture important aspects of memory. The store metaphor has guided laboratory research while the archaeology metaphor has guided studies of memory in everyday settings. Researchers guided by the store metaphor have studied memory driven by theory while researchers guided by the archaeology metaphor have studied memory as it occurs in a more phenomenological way. The two models seem to reflect fundamentally different ways of thinking about memory. As Koriat & Goldsmith (1996, pp. 186) argue, "… even if agreement could be reached about the memory phenomena that ought to be studied, the experimental procedures, and the appropriate context of enquiry, the two metaphors would still imply different perspectives for looking at and interpreting the data".

References

Underwood, B.J. (1972). Are we overloading memory? In A.W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C: Winston.

James, W. (1890/1983). *The principles of psychology*. Cambridge, MA: Harvard University Press.

Atkinson, R.C., & Shiffrin, R.M. (1968) Human memory, A proposed system and its control processes. *The psychology of learning and motivation: advances in research and theory*, vol. 2, ed. K. W. Spence. Academic Press.

Neisser, U. (1967) *Cognitive Psychology*. Appleton-Century-Crofts.

Bartlett, F.C. (1932) *Remembering*. Cambridge University Press.

Koriat, A., & Goldsmith, M. (1996). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. Behavioural and Brain Sciences, vol. 19(2), pp. 167–228.

MARIA LARSSON

Smell your Way Back to Childhood: Autobiographical Odor Memory

Introduction

The olfactory sensory system is our oldest sense and also unique among the senses in synapsing directly with the amygdala-hippocampal complex, the neural structures supporting basic survival functions such as memory and emotion. A growing body of evidence suggests that memories evoked by odors differ from other memory experiences. The scope of this brief overview is to provide some illustrations of these differences.

Associate Professor Maria Larsson,
Department of Psychology, Stockholm University, Sweden.
CAS Fellow 2003/04.

Autobiographical memory

Autobiographical memory is memory for the events of one's life. Research indicates that the autobiographical memory knowledge base may be described as three layers of knowledge: lifetime periods that span periods of years or decades (e.g., my life as a scientist); general events representing time periods of weeks and months (e.g., my time as a research fellow at the CAS); and finally event-specific knowledge comprising sensory-perceptual knowledge spanning periods of seconds, minutes, or hours (e.g., the spectacular CAS farewell dinner held at the Norwegian Academy of Science and Letters).

One approach to the study of autobiographical memory is the exploration of the age distribution of event-specific memories recalled across the whole life span. The typical procedure in this research is to expose subjects to different sensory cues (e.g., tar may be presented visually, verbally or as an odor). In instances when a memory is evoked, that must be well defined, the person is asked to date when the specific event took place. Extensive research on memory distributions across the life span using verbal cues has evidenced a remarkable stability across individuals (e.g., Rubin & Schulkind, 1997). The distribution comprises three unique components: childhood amnesia, the bump, and the recency effect. Childhood amnesia reflects the dramatic reduction in the number of memories reported from early childhood. In contrast, a substantially larger number of memories are recalled between the ages of 10 and 30, which has been termed the reminiscence bump. The third component, denoted recency, reflects better retention of events occurring from the last 10 years. The well-documented distribution of verbally cued autobiographical memories over the lifespan is displayed in Figure 1.

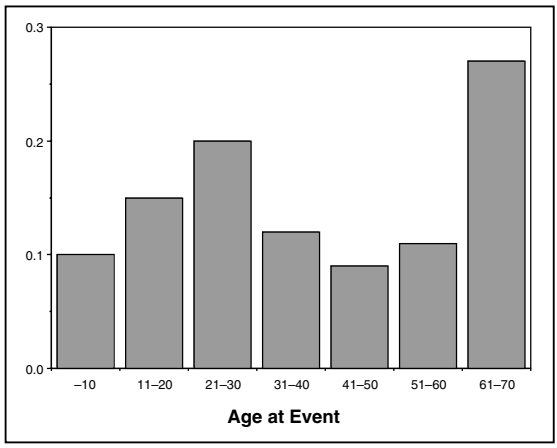


Figure 1. Distribution of verbally cued autobiographical memories over the lifespan.

A number of theories have been advanced to explain the "bump" phenomenon such as that the life period involves a high degree of dynamics illustrating that late youth and early adulthood is a time during which identity formation is a key process (e.g., education, marriage, children), resulting more memories being encoded. Another potential explanation is that the cognitive capacities are at an optimum level between the ages of 10 and 30. The relatively few memories reported from the childhood period are most likely due to the fact that the brain structures subserving personal memories are not fully developed.

As noted above, most of the available evidence on autobiographical memory is based on verbal cuing and knowledge is sparse regarding how other sensory cues influence recollection of personal events. In a recent study (Willander & Larsson, 2004), we found that memories evoked by odors differ from memories associated with verbal and visual information. Specifically, in contrast to the well-established memory peak in young adulthood, olfactorily evoked memories were associated with an earlier period in life – childhood. That is, when a memory was evoked by an odor (e.g., cinnamon, tar), the event was typically reported as having occurred when the person was between the ages of 5 and10 years. The memory distribution obtained for odors is shown in Figure 2.

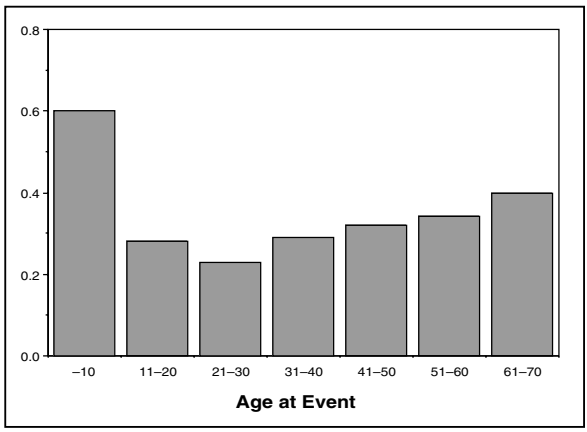


Figure 2. The memory distribution obtained from odors.

Furthermore, our participants also reported that the content qualities of autobiographical memories evoked by odors and words were different.

Odor-evoked memories were associated with stronger feelings of being brought back in time, thought of, and spoken of less often, as compared with memories evoked by the verbal cues. In addition, olfactorily evoked memories were more often experienced with an emotional connotation as compared with their verbal analogue.

Why, then, are odor-associated memories older? We know that the olfactory sense is our oldest sensory system, both from a phylogenetical and ontogenetical perspective. It is a system that is highly active early in our lives and our chemosensory knowledge is often based on events that we experienced early in life. For example, chemosensory learning is already present in the womb and infants interact primarily with the environment through the chemical senses: smell and taste. Even though we do not think about it in our daily lives, we know exactly how most objects and materials would "taste" and "smell" if we were to place them in our mouths (e.g., the sensation of sand, a plastic bag, or a cloth).

It is possible that each sense has a "critical" period during which information in that system is processed more effectively. These periods may in turn reflect the evolution of the sensory systems. We know that the proportion of the brain that processes olfactory information decreases with increasing complexity of a species' neural complexity. This may also be reflected in human development – from fetus, infant, child, to adult.

References

Rubin, D. C., & Schulkind, M. D. (1997). The distribution of memories across the life span. *Memory and Cognition, 25*, 859-866.

Willander, J., & Larsson, M. (2004). *Smell your way back to childhood: Autobiographical odor memory*. Manuscript submitted for publication.

ASHER KORIAT

Knowing and the Feeling of Knowing

An important property of memory that is evident in everyday experience is that the information that we can retrieve at any point in time represents only a fraction of what we actually know. However, even when we fail to recall a name or a word, we can still judge whether it is stored in our memory and is worth searching for.

Everyone is familiar with the "tip-of-the-tongue" (TOT) state that we sometimes experience, when we block on a certain word or name. William James (1893), one of the founders of modern psychology, described the TOT state as follows:

"Suppose we try to recall a forgotten name. The state of our consciousness is peculiar. There is a gap therein; but no mere gap. It is a gap that is intensely active. A sort of wraith of the name is in it, beckoning us in a given direction, making us at moments tingle with the sense of our closeness and then letting it sink back without the longed-for term. If wrong names are proposed to us, this singularly definite gap acts immediately so as to negate them. They do not fit into its mould. And the gap of one word does not feel like the gap of another, all empty of content as both might seem necessarily to be when described as gaps" (p. 251)

Professor Asher Koriat,
Department of Psychology,
University of Haifa, Israel.
CAS Fellow 2003/04.

What is fascinating about the TOT state is that it represents knowing about the unknown. It illustrates a discrepancy between the subjective conviction that we "know" the name, and our actual inability to produce it. Naturally, the question arises: How do we know that we know? This question, in fact, applies to memory in general, because by and large, we are quite accurate in our meta-cognitive judgements, that is in what we know about our own knowledge.

The feeling of knowing (FOK) has been traditionally regarded as a mystery, and has been typically discussed in connection with the concept of "intuition". This concept emerges, for example, in the study of creativity: The reports of highly creative people about their thought processes suggest that they often have the intuitive feeling that they are on the right track to a solution, and can even sense that they are about to reach that solution before they actually do so. In fact, descriptions of the creative process suggest that a great deal of the cognitive work goes on unconsciously (Ghiselin, 1952), and yet people can monitor these underground processes without actually being aware of them. This is very similar to the FOK that we have about some piece of information before we can retrieve it.

My work as well as that of others (Koriat, 1993; Metcalfe, 2000) has attempted to demystify the FOK. The classical account of the FOK (as well as other types of meta-cognitive judgements) is that it is based on accessing an internal monitor that "knows" whether the solicited information is present in memory. Thus, Hart (1965), for example, proposed the existence of a Memory Monitor that can inspect the contents of memory and detect whether the trace of a particular item is stored there or not. Therefore, when I am presented with a question (e.g., "what is the word for 'thanks' in Norwegian?") I first consult the Monitor to ensure that the information is stored in memory. The result is a subjective feeling of knowing like that associated with the TOT state. Of course, the existence of such a Monitor is adaptive, because it can save me the time and effort looking for something that is not there.

The advantage of this "trace access" account is that it can also explain why our feelings are accurate. The problem however, is that this account only pushes the question one step back because we have now to ask How does the Monitor know that I know?

An alternative account that has been gaining popularity is that meta-cognitive feelings in general are based on an inference, but this inference operates unconsciously: We are not aware of making any inference; we are only aware of its end product – a sheer subjective feeling. When we are asked a memory question, the FOK that we experience is based on a variety of subtle cues. One such cue is the overall accessibility of partial information that the question brings to mind. Even when we cannot remember a name or a word (e.g., *takk*), we might recall that it is a short word, that it contains k, that it has a certain feel, etc. In one study (Koriat, 1993) I have shown that such partial pieces of information contribute to the FOK whether they are correct or not. Thus, if I recall that the elusive word is long, and contains the letter f, I should still have a strong FOK about *takk*. This is because according to the accessibility account that I proposed, people have no privileged access to the contents of their memory over and above what they can retrieve from it, and furthermore, they have no way of knowing whether the partial information that comes to mind is right or wrong. All they can base their FOK on is the overall amount of partial information and the ease with which it comes to mind (for example, if the letter f jumps more easily to mind, one may have a stronger FOK about the Norwegian word for 'thanks' than if the same letter comes more slowly to mind).

According to this account, then, the accuracy of FOK judgements is not at all guaranteed, but depends on the validity of the cues on which they are based. Why is the FOK nevertheless generally accurate? This is simply due to a basic quality of memory: The information that comes to mind is much more likely to be correct than wrong. Therefore, the overall accessibility of information is generally predictive of correct memory.

However, in the exceptional conditions in which such is not the case, FOK judgements can be completely inaccurate. For example, when a question brings to mind a great deal of information (sometimes information that is not helpful for getting the answer) it might precipitate a strong unwarranted feeling of knowing. It has even been shown that some such questions can result in a TOT experience. For example, when presented with the question "What is the last name of the Canadian author who wrote the novel *The Last Bucket*?" a considerable proportion of people

report that they feel the answer is on the tip-of-the-tongue even though the question has actually no real answer (Schwartz, 1998). This feeling presumably stems from many activations that are produced by the question that have nothing to do with the answer. Possibly, when we search our memory, we cannot zoom in precisely on the correct answer in memory (even if it is there), and activations deriving from many related elements may contaminate our subjective feelings. Indeed, activations stemming from "neighbours" of the answer have been shown to create an illusion of knowing about the answer.

Thus, there is nothing magic about "intuitive feelings". In fact, I have shown that although by and large FOK ratings are valid predictors of one's actual memory performance, under some special conditions the correlation is negative: The more one feels that one knows the less likely is one actually to know (Koriat, 1995).

Research on the FOK has applied as well as theoretical implications. On the applied side, it is important to stress that people generally follow their intuitions blindly. Therefore the accuracy of these intuitions is critical. When we visit a doctor we would like to be sure that he can accurately monitor his own knowledge to the extent of requiring a second opinion when he (accurately) feels uncertain. In recent years we have learned a great deal about what factors contaminate one's intuitive feelings, leading them astray. Some people are bound to have illusions of knowing, failing to recognize their own incompetence. As Dunning et al (2003, p. 83) argued: "People tend to be blissfully unaware of their incompetence. This lack of awareness arises because poor performers are doubly cursed: Their lack of skill deprives them not only of the ability to produce correct responses, but also of the expertise necessary to surmise that they are not producing them".

The following proverb has something to say about individual differences in meta-cognitive accuracy:

He who knows not and knows not that he knows not is a fool; shun him.
He who knows not and knows that he knows not is ignorant; teach him.
He who knows and knows not that he knows is asleep; awaken him.
He who knows and knows he knows is a wise man; follow him.

On the theoretical side, the importance of meta-cognitive research is that it illuminates a particular mode of cognitive operation. In general, we distinguish between two levels of experience, each with its own mode of operation. The higher level involves an explicit mode of operation, characterized by relatively high degrees of consciousness and control: When we have to act, we consider various options in a conscious and rational manner, and control our behaviour accordingly. The lower level, in contrast, is relatively unconscious and involves automatic effects on behaviour: A variety of factors may affect the person's behaviour outside his/her consciousness and outside his/her control.

Within this scheme, meta-cognitive feelings are assumed to mediate between the two modes of operation. On the one hand, they are shaped by an inferential process that operates automatically and unconsciously to produce a sheer subjective feeling. On the other hand, once they are formed, they can serve as the basis of conscious, controlled action. Thus, they play the role of a go-between, allowing a transition between an unconscious-uncontrolled mode of operation and a conscious and controlled mode. Indeed, I have argued that the function of subjective

experience, in general, is that of augmenting self control, that is, of allowing some degree of control over processes that would otherwise influence behaviour directly and automatically, outside our awareness.

References

Dunning, D., & Johnson, K., Ehringer, J., & Kruger, J. (2003). why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*, 83–87.

Ghiselin, B. (1952). *The creative process*. Berkeley and Los Angeles: University of California press.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208–216.

James, W. (1893). *The principles of psychology*, Vol I. New York: Holt.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling-of-knowing. *Psychological Review*, *100*, 609–639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311–333.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149–171.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131*, 349–363.

Schwartz, B. L. (1998). Illusory tip-of-the-tongue states. *Memory, 6*, 623–642.

CARSTEN HANSEN

Truth: A Contemporary Philosophical Debate and its Bearing on Cognitive Science

Traditional conceptions of the significance of truth and its intractability

In contemporary philosophy, one finds widespread agreement about two claims concerning the concept of truth. First, that (A) truth is an absolutely central notion, indispensable in any attempt to give a philosophical characterization of ourselves and our place in the world. And this is so because it appears that we need the concept of truth, if we are to give an adequate account of the aims of science, the relations of language to the world, as well as logic – the character of sound reasoning – to mention some central cases. Now, the (apparent) centrality of the notion of truth lends urgency to the task of giving an account of the nature of truth. For, as the examples just noted seem to indicate, insight into the underlying essence of truth promises to shed light on just about every other aspect of our conceptual scheme (the fundamental concepts in terms of which we think of ourselves and our place in the world).

Associate Professor
Carsten Hansen,
Department of Philosophy,
University of Oslo.
CAS Fellow 2003/04.

However, a large number of contemporary philosophers would also agree that (B) truth has proved extremely resistant to elucidation. Not to put too fine a point on it, we hardly seem to have made any progress in giving an account of the nature of truth. The main alternatives, among traditional theories of truth, are correspondence and coherence theories of truth. They illustrate the difficulties that we face. The central idea of a correspondence theory can be expressed as the claim that

X is true iff X corresponds to the facts

The key idea of a coherence theory, on the other hand, is that

X is true iff X belongs to a sufficiently coherent and comprehensive system of beliefs or propositions

The problem with correspondence theories is that they make use of notions – like that of a fact – which presuppose the notion of truth, thus rendering them viciously circular. The central objection to coherence theories, on the other hand, is that there can be rival and divergent

systems of belief that are equally coherent – one containing p and the other containing its negation (not-p), for example. And since p and not-p cannot both be true, the coherence theory appears to be vulnerable to straightforward counter-examples. Basically, it seems that no one has yet been able to come up with a theory of truth which is neither false nor viciously circular.

Contemporary 'deflationist' or 'disquotationalist' views of truth

In a way, if you believe (A), (B) may not seem all that surprising: the reason why the notion of truth is so difficult to elucidate is simply that it is so fundamental. (There are no more basic notions in terms of which it could be analysed.) However, there is a distinctly modern view of truth – variously known as 'deflationism', 'disquotationalism' or 'minimalism' – which holds that the notion of truth is neither deep nor difficult. Rather, it is both easy to explain as well as 'metaphysically trivial' – which is to say that it has virtually no bearing on any important philosophical issue.

Deflationists take as their starting point something that is common ground between all parties to the dispute. But how we are to formulate this important area of agreement depends on a prior issue: whether we think of truth as a property of sentences (utterances) or of what those sentences express - propositions (as philosophers like to say). I shall circumvent a number of important technical difficulties, and crudely illustrate the matter in terms of propositions as well as sentences (utterances). For (almost) any proposition – that snow is white, for example – we would all agree that

the proposition that snow is white is true if and only if snow is white

Similarly, and suitably qualified, it seems hard to gainsay that

'Snow is white' is true if and only if snow is white

Sentences of this form are often called T-sentences, they are instances of the schema: (T) X is true if and only if p. There are some philosophically important exceptions where the schema does not hold – paradoxes and statements that are neither true nor false. For these cases, we would not – or should not – accept the corresponding instance of the (T)-schema. The deflationist knows that there are such cases, and suggests that we leave them aside (for the time being).

He or she then goes on to suggest that the expression 'is true' is exhaustively characterized (that is to say, implicitly or explicitly defined) by the totality of (acceptable) T-sentences. Furthermore, the deflationist holds that this body of sentences tells us all there is to know about truth. This is the characteristic, defining thesis of deflationism.

The deflationist's view of the point of 'true'

The deflationist's view of the nature of truth gives rise to an immediate objection. For if truth really is as trivial as they would have it, it seems a mystery why there should even be a word for it in the language. The deflationist responds by insisting that the truth-predicate subserves a real though strictly limited purpose. (It is important to see what this purpose is, for it gives rise to the dominant argument for deflationism.) Deflationists

often point to Quine as the philosopher who first clearly identified the utility of the truth-predicate, though this is in all likelihood incorrect. At any rate, the explanation that he did more than anyone else to promote this is that the point of the truth predicate is to enable us to say things about the world (non-linguistic reality) by talking about sentences.

Thus, instead of expressing myself directly, by saying that the sun is shining, for example, I could – if I wished – 'ascend to the level of sentences' and say that the sentence 'the sun is shining' is true. In this case, however, there would be no real point in doing so. The utility of the truth-predicate, we are told, stems from the fact that there are situations in which we are prevented – because of ignorance or memory failure, or whatever – from expressing directly the things we want to say about the world. Here, the truth-predicate earns its keep.

Suppose I have had a discussion with someone who said something with which I agreed. The next day, I may have forgotten exactly what it was that he or she said for which reason I cannot express it directly. Nevertheless, I can refer to that assertion using a form of words like 'what X said yesterday (about such and such), and then go on to re-affirm that assertion by saying 'What X said yesterday (about such and such) is true'. This last sentence stands in, or acts as a surrogate, for the proposition that I cannot now express.

The following is, perhaps, a more interesting case of this use of 'true': consider a theory about the physical world – such as a typical first order version of the Euclidean theory of space – which is not finitely axiomatizable (because it contains a number of axiom schemas each having infinitely many axioms as instances). Suppose one has grounds for rejecting the theory without knowing exactly which part of it to reject, or suppose that one accepts it, but regards it as contingently true. In the first case, one wants to deny the infinite conjunction of the axioms, and in the second case, to assert the possibility of the negation of this infinite conjunction. However, we are not in a position to complete the infinite conjunction, for which reason we cannot express ourselves directly. Instead, we make use of the truth predicate: In the first case, one will put the rejection by saying 'Not every axiom of this theory is true'. In the second case, one will express one's acceptance by saying 'It might have been the case that not every axiom of this theory is true'. (I have borrowed this example from Hartry Field, one of the leading proponents of deflationism.)

The case for deflationism

In fact, the deflationist claims that the aforementioned use of 'true' is the only thing we really need it for. And this gives rise to the following central argument for deflationism:

The point, and only serious purpose, of the truth predicate (our notion of truth) is the 'logical' one of providing 'alternative objects of our attitudes', in particular surrogates for infinite conjunctions or disjunctions. To fulfill this role, all that is required is that it be governed by the equivalence schema. Thus, there is simply no need to go beyond the deflationist account of truth: no reason to suppose that there is more to truth than what can be read off of the acceptable instances of the equivalence schema.

Assessing the plausibility of deflationism and seeing what is really at stake

In a full-dress assessment of deflationism, we would have to consider what is to be said for and against the operative premise (the first one), and whether the conclusion follows from the premises (once the content of the relevant propositions has been made clear). Here, I shall merely have space to bring to your attention the consequences of the deflationary conception of truth.

Considered in isolation, truth, as defined by deflationists, appears to be entirely trivial. What needs emphasis, however, is that adopting that notion as one's notion of truth has extremely wide-ranging consequences. The point is that someone who endorses the deflationist conception of truth is thereby committed to a highly controversial view of the nature of meaning (and mental content).

There is a dominant tradition in the philosophy of language and mind which takes the notion of a truth-condition to be a key notion in the philosophy of language (and mind). In the case of language, the underlying idea – which was perhaps first made explicit by Frege – is that to understand a sentence is to have grasped its truth-condition. That is to say, to know the meaning of a declarative sentence is to know what has to be the case for it to be true.

A deflationist speaking of truth cannot hold that the meaning of a sentence is to be identified with its truth-condition – for then there would be more to truth than is expressed by the T-sentences. What that means is that if you think that there are compelling reasons for holding on to a truth-conditional conception of meaning, those reasons would give you grounds for rejecting the deflationist view of truth. The deflationist, however, is not going to give up so easily. For there is, in fact, another general approach to the nature of meaning, which holds that meaning is to be explained in terms of something like use rather than truth-conditions. The most famous proponent, and maybe originator, of this idea is Wittgenstein - with whom we associate the slogan that 'meaning is use'.

Here, there is only time to note the following main points. First of all, we are in a position to see that the fundamental question facing theorists of truth is how to conceive of the relation between truth and meaning. Secondly, though 'use theories of meaning' have been much discussed over the past forty years or so, no one has to date presented a theory of that kind with anything like the level of detail and plausibility possessed by extant, truth-conditional theories. For this reason, as well as others, even a moderate claim on behalf of deflationism – that the balance of evidence is currently in their favour – is highly questionable.

Truth at the interface between philosophy and cognitive science

Hopefully, I have managed to show that the issue over deflationism is one that is of central concern to philosophers. However one might still ask whether the issue has any bearing on matters outside philosophy. To see that it does, one only needs a brief reminder of the kinds of things that psychologists and cognitive scientists of various stripes seek to understand, and of the terms in which their explanations are couched. For among the things that they – including researchers at the CAS – are actively seeking to explain are such phenomena as vision, memory and action. There is not a shred of doubt, then, that intentional notions, and the notion of

representational or truth-conditional content, crop up throughout these disciplines. The notions in question are clearly employed in the initial characterization of the phenomena to be explained, but they also figure crucially in the theories which seek to account for them. Modern cognitive science, then, is up to its neck in truth-conditional content.

Now, it may occur to you that this shows more than that there is a connection between ongoing scientific concerns and the abstruse philosophical debate about deflationism. For it may seem that an appeal to the cognitive sciences serves to give a decisive answer to the philosophical debate about truth. The very fact that there is a successful ongoing scientific practice which is committed to the use of the notion of representational content gives us sufficient grounds – so the thinking goes – for supposing that there is more to truth than deflationists maintain.

However things are a good deal more complicated than that. A successful ongoing scientific practice may give us strong *prima facie* reasons for thinking that deflationism's view of truth is wrong-headed. Nevertheless, a philosophically illuminating resolution of the issue over the nature of truth, and its relation to meaning, cannot simply appeal to ongoing scientific practice. For one thing, a philosopher will need to know more about the precise nature of the concepts that psychologists and scientists make use of, as well as what scientists are really committed to in their talk of meaning and content. Related to this point is the fact that one of the things that many philosophers find deeply puzzling – and rightly so, in my view – is an assumption shared by science and common sense. Namely, that mental states can have causal effects by virtue of their possession of representational content.

In other words, it is unclear exactly what science tells us about the nature of meaning – whether meaning needs to be explained in terms of truth-conditions or not. And for that reason, current scientific practice, on its own, cannot yield a decisive answer to the philosophical question about the nature of truth. On the other hand, it would, of course, be completely absurd to suppose that science should not make use of the notions that it does, until they have been clarified to the extent which would suit the interests of philosophers.

The point is rather that each camp stands to gain – though in different ways - from keeping abreast of the activities of the other. Philosophy is in many respects a second-order discipline. In the present case, this means that philosophers arguably have to study basic features of the cognitive sciences carefully, if they are to come up with well-grounded answers to their question about the nature of truth. The sciences, on the other hand, do not need to turn to philosophy in order to solve the tasks that they set themselves. The contribution of philosophy to the sciences lies rather in the possibility of a deeper, or reflective, understanding of the content of the claims advanced in scientific theories.

PASCAL ENGEL

Tacit Belief

Although – and probably because - the notion of belief is so widespread in the social sciences, in psychology and in philosophy, it is very often vague. There are at least two ways of ascribing beliefs to people: through their explicit avowals and through what they do. The latter criterion seems, however, to be the more reliable, for it often happens that we do not believe what we say we believe; in such cases, better look at what the person does. Minimally, a belief is a disposition to act in certain ways. This seems to be a necessary condition: of course many beliefs are so concealed that they never get out of our mouths, but it is extremely difficult to conceive of a belief which could never, at least potentially, influence one's mental life. The minimal condition for predicating a belief to someone is that it should at least be *able* to play some role within our psychology. If this is so, it seems correct to say that many, possibly most, of our beliefs are implicit or tacit, in the sense that, even if they do not manifest themselves in our actions and in our psychology, they *could* do so. Some, and probably most, of our beliefs have to be tacit, for cognitive life would be impossible if we had in front of our minds all the things that we believe. Our predicament would be that of Borges' character, Funes the Memorious, who kept *all* his memories and perceptions explicit, and whose mind was consequently like "a garbage can".

Professor Pascal Engel,
Department of Philosophy and Sociology,
University of Paris-Sorbonne Paris IV,
France.
CAS Fellow 2003/04.

The idea that some representations and mental states are tacit or implicit is quite common in psychology. Since Chomsky at least it is customary to talk of tacit knowledge of grammar; psychologists talk of implicit memory, and distinguish between procedural memory, which is implicit, and declarative memory (Schacter 1989); perception through some senses such as touch is largely implicit, whereas pointing to a target is not; in blindsight subjects report not being conscious of some information which they nevertheless use in some tasks; some capacities in development, such as the capacity to attribute mental states to others ("theory of mind") are largely implicit. "Tacit" is often another name for "unconscious". But the former notion is distinct from the latter if we suppose that a tacit state is one which can in principle be accessed. And there lies our problem: how is such information attributed and accessed?

In the case of belief, the problem is that there seems to be no limit to the number of beliefs one can ascribe to us when we perform one single action. If Sam runs to catch his train, he must certainly believe that his train has not yet left – otherwise he would not be running. But he must also believe that the station is not too far, that trains leave on time, that the platform is accessible, that trains are machines, that they do not go faster than the speed of light, etc. Where should we stop? The problem is not

simply, as it is often said, that in order to have a belief, one must have quite a lot of other ones. It is that if most beliefs are tacit, which beliefs are we allowed to ascribe to an agent in a given circumstance? Some are immediately present to consciousness. Others are tacit in the sense that they could come to consciousness if we attended to them. So one could try the following definition of tacit belief: *X tacitly believes that P if and only if X is disposed to believe explicitly that P.* This seems to account for the dispositional character of our beliefs. But it is open to obvious counterexamples. Thus, on this definition, when I ask you: "Did you believe that elephants did not wear pyjamas ?" or "Did you believe that Kant's left earlobe was smaller than the Sea of Tranquillity?", you will probably say yes, and believe explicitly the content of these sentences, but are these things that you *believed*, even tacitly? These beliefs probably never occurred to your mind before. Such "beliefs" rely on general knowledge, but others may rely on simple consequences of what one believes. For instance if you believe explicitly that Rome is north of Naples, it is likely that you believe implicitly that Naples is south of Rome. But tacit beliefs seem to be just pseudo-beliefs, and not beliefs at all: they play no role within our psychology. What is special about them, when compared to other dispositional beliefs, is that the very fact that you raise them creates a disposition to believe them. Normally, a disposition, such as fragility, manifests itself in the presence of a stimulus. We can imagine dispositions which go out just when they are about to manifest themselves: for instance a poison which every time it is about to act is prevented from doing so by a sudden antidote. David Lewis (1996) calls such dispositions "finkish". Here we have dispositions which are triggered by a single stimulus. Tacit beliefs, unlike ordinary dispositions, do not have a causal basis (which is usually memory: I do not *remember* that elephants do not wear pyjamas). And they are very different from the kind of implicit representations that psychologists talk about, for instance when they analyse the feeling of knowing or the "tip of the tongue" phenomenon (Koriat 1992). In the latter case, you feel that you know, but you do not have access to what you know. In the present case of tacit beliefs, you do not feel that you believe, but you have access to it! Hence we can be sceptical about tacit belief as defined above.

We should not conclude, however, that the notion of tacit belief is incoherent. It is too useful in psychology, for instance to explain common dissociations such as amnesia, to be dropped out . But we must find a way

of restricting its scope. One proposal (Maloney 1989) distinguishes two sorts of belief: those which are responsible for our actions, and those which are sensitive to evidence (this is more or less a version of the procedural/ declarative distinction). But this is not very helpful, for it begs the question: which is which? A better proposal is this (Crimmins 1992): *X tacitly believes that P if and only if it is as if X believes that P*, or as if one's cognitive dispositions were relevantly similar to one's believing that P. On this view a tacit belief can be ascribed on condition that it could be relevantly similar to a belief that one could have in the course of one's psychology. But counterfactual sensitivity (as if) is notoriously vague. Just the same, postulating the existence of an "extrapolator deducer" which would extract the appropriate beliefs is unsatisfactory, just like the postulation of a "supervisory attentional system" (SAS).

If we want to make sense of the notion of tacit belief, we certainly have to incorporate a proposal like Crimmins'. But we also have to take a different tack, and to notice three things.

1) Belief is a mental attitude which falls short of knowledge. This is a banality: beliefs, by definition can be true or false, whereas knowings are necessarily true. But it does not follow that our tacit beliefs, are, as it were, tacit *opinions*. They are tacit (dispositional) beliefs because belief in general aims at knowledge, and not simply at truth; to believe that P is to have an attitude towards P which is *as if* one knew that P, it is "botched knowledge"(Williamson 2000). If we take this perspective, tacit belief is necessarily connected with tacit knowledge. Indeed, it is because it belongs to our semantic memory and our general knowledge that elephants do no wear pyjamas, and that earlobes of people are smaller than seas, that one can ascribe to us such tacit beliefs. But of course we do not *believe* such things, we do not have any *attitude* towards them. It is only within a given cognitive task, in a certain contextual setting, and with respect to giving a certain kind of explanation, that we can ascribe tacit beliefs. 2) Knowing is not necessarily knowing that one knows. Tacit knowledge is a case at hand: if we have a tacit knowledge of the grammar of our own tongue, or of certain clues in our navigation in space, we do not know that we know. The same is true of belief. Some philosophers claim that to have beliefs one must have the concept of belief. If this means that our beliefs must be beliefs about our beliefs, this is false, not only for a number of creatures (animals, infants), but also for us in general. Beliefs are not necessarily reflexive, and neither are tacit beliefs. 3) Most of the time tacit knowledge is understood in terms of capacities and abilities. In Ryle's (1949) terminology, it is said to be "knowledge how", or practical knowledge, and not "knowledge that", or propositional knowledge. But this distinction is moot (Stanley and Williamson 2001). But belief is propositional, it is by definition an attitude towards a proposition, and there is no "belief how" in the

sense in which there is a knowing how (you can know how to make a chocolate mousse, but it seems odd to say that you *believe how* to make a chocolate mousse). So, if there is tacit belief, it is unlikely that it is procedural, if "procedural" means a form of knowledge how. Hence we should also revise the distinction between procedural and declarative knowledge as well.

References

Crimmins, M 1992 "Tacitness and virtual Beliefs", *Mind and Language*

Dienes, D. and Perner, J "A Theory of Implicit Knowledge", *Behavioral and Brain Sciences*

Koriat "The feeling of knowing, some metatheoretical implications", *Consciousness and cognition*

Lewis, D. 1996 "Finkish dispositions", *The Philosophical Quartely*

Maloney, C 1989 "It's hard to believe", *Mind and Language*

Ryle, G. 1949, *The concept of Mind*, London: Hutchinson

Shacter, D. 1989 "Implicit Memory: History and Current State", *Journal of Experimental psychology*

Stanley, J. and Williamson, T. 2001 "Knowing how", *Journal of Philosophy*

Williamson, T 2000, *Knowledge and its limits*, Oxford University Press

DAGFINN FØLLESDAL

Indexicals and the Mind

Indexicals are some small words in our language, such as "I", "here" and "now," which we use all the time, but which are still not properly understood. These words are currently engaging some of the best philosophical minds. At least one highly respected philosopher, David Kaplan, has devoted most of his life to understanding them.

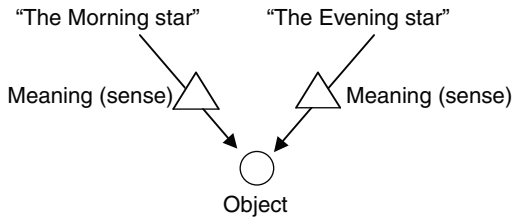
Why should one devote so much thought and energy to these words? We use them daily and do not experience them as problematic. However, questions about indexicals are like Augustine's questions about time: "What then is time? If no one asks me, I know; if I wish to explain it to one who asks, I know not." It took a while before philosophers started to notice that the indexicals are problematic. Only when they had developed fairly systematic approaches to how language functions in communication and how language relates to the world did they discover that the indexicals were recalcitrant. Peirce, Frege and Husserl were early explorers of the indexicals. Russell, Reichenbach and many others followed. Decisive leaps forward were made by David Kaplan and John Perry who is a member of our group, *Towards a New Understanding of the Mental*.

Professor emeritus
Dagfinn Føllesdal,
Department of Philosophy,
University of Oslo.
CAS Fellow 2003/04.

What then are these problems? I have already indicated that one does not notice the problems until one starts thinking systematically about language. This is one of the difficulties with philosophical problems in general, the same difficulty that Augustine experienced: To explain a philosophical problem is often just as hard as to propose a solution. In order to get at the problems raised by indexicals we first have to take a look at the traditional views on language and see how they apparently give us a good way of understanding how language works – until we get to indexicals.

Gottlob Frege (1848–1925), who created modern logic, saw that in order to get a satisfactory understanding of logic he had to know how language functions in communication and how it relates to the world. In particular, he needed to figure out how we can understand the relation of identity. He did so in a famous article, "On sense and reference" in 1892. Here he worked out with great precision a view that comes fairly naturally and had cropped up several times earlier in the history of philosophy: in the stoics, in many medieval philosophers and especially clearly in Bolzano 60 years before Frege. The idea is that each sentence and word in our language has a meaning, or sense, which determines what the expression refers to. An example Frege uses is "The Morning star" and "The Evening star". Each of these expressions refers to the same object, which is also referred to as "The planet Venus". However, it took a while before the old Babylonians discovered that these expressions refer to the same

object. The expressions express different meanings, or senses, the first of them that of a star prominent in the morning, the second that of a star conspicuous in the evening. When astronomers discovered that they were the same star, they discovered that the two expressions, although they have different meanings, still refer to the same object:



Frege held that the meaning of an expression determines its reference. A basic principle in his view on meaning is this:

If two expressions have the same meaning, then they have the same reference.

But not conversely, we just saw how two expressions with different meanings, "The Morning star" and "The Evening star", can relate to the same object. The meaning could be compared to a set of features that an object has to have in order to be the reference of this expression. Frege also compared the meaning of an expression to an aspect of the object. "The Morning star" and "The Evening star" express two aspects of one and the same star: the first that it is dominant in the morning, the second that it dominates in the evening.

Frege also used this trichotomy of expression, meaning and reference to explain how we communicate with one another. Briefly, a speaker or writer communicates by uttering or writing an expression whose meaning is grasped by the listener or reader.

This all works well until we get to the indexicals. Let us now see what happens. Let us start with the word "I". When I say "I", I refer to myself. But when you say "I", you refer to yourself. How can that be? According to Frege's basic principle, which was stated above, this implies that the word "I" has different meanings when it is used by different persons. This seems somewhat disturbing, but Frege was willing to bite the bullet. For reasons that we shall not discuss here, he even went so far as to contend that the word "I" expresses a special sense for each person that cannot be grasped by anybody else. This creates quite a problem for Frege's theory of communication, since a main purpose of the notion of sense for Frege is that we communicate by using words whose sense is the same for the speaker and the listener. Yet our awareness of ourselves is such an elusive notion that we may easily nod approvingly when Frege writes: "Now everyone is presented to himself in a particular and primitive way, in which he is presented to no-one else."

Let us now, however, consider the word "now". On different occasions when I say "now" I am referring to different times. However, does "now" change its meaning from one occasion of use to the next? The same with "here", "today", "yesterday" and so on. We cannot allege that what these words refer to is elusive, as Frege did in the case of the self. The view that indexicals change their meaning all the time seems utterly implausible.

So what shall we do? Three groups of approaches to the problem have been proposed, and we shall now briefly consider them one by one:

1. Classical theory of meaning (sense)	2. Hybrid expressions	3. Hybrid sense (Object is part of proposition) or: causal impingements
Frege (*sui generis* sense sense of "I")	Frege	a) Singular propositions b) Husserl c) Perry
Problem: Communication = grasping same sense	Problem: Reference to the object is still via sense	

The first of these approaches is the one we have discussed. Nobody accepts this as satisfactory. Also Frege saw that it would not do. However, there are some hints in his later writings, for example the following:

… the mere wording, as it can be preserved in writing, is not the complete expression of the thought; the knowledge of certain conditions accompanying the utterance, which are used as means of expressing the thought, is needed for us to grasp the thought correctly. Pointing the finger, hand gestures, glances may belong here, too.

The basic idea, which is explicit in the passages I have underlined, is that the expression, which expresses the meaning, consists not just of words, but also includes other items, such as pointing the finger, etc. In another passage Frege states that also "the time of utterance is part of the expression of the thought." This interpretation of Frege was proposed in 1977 and it was worked out in detail in 1982 and later years by the German philosopher Wolfgang Künne, who aptly called this the *hybrid* view: what expresses meaning is not just ordinary linguistic expressions, but in many cases a hybrid complex consisting of a mixture of words, physical objects and movements and even times and places.

This is an interesting view, but again it leads to difficulties. We shall not go into these difficulties here, but only note that they indicate that the basic idea of the classical theory of meaning does not work: reference is not determined by meaning, not even by the meaning of hybrid expressions. The difficulties seem to indicate that the notion of meaning that is so central in the classical theory cannot do the job alone, not even the rich meaning expressed by hybrid expressions. The objects themselves that we are referring to must be brought in, and in a different way from the way they were brought in as parts of hybrid expressions. One proposal has been that the meanings themselves, and not just the expressions, are hybrid, they contain objects of various kinds as their parts. Two philosophers who have contributed greatly to clarifying the situation are David Kaplan, whom I mentioned earlier, and John Perry.

Lately it has been discovered that the Czech-German philosopher Edmund Husserl, the founder of phenomenology, has a highly interesting discussion of indexicals in a manuscript from 1911. Long before Frege developed his hybrid expression view on indexicals Husserl was aware of problems that show that a more radical approach is needed. He gives the following example which is a devastating stumbling block for any attempt to save the classical theory of meaning:

But how is it, if on two heavenly bodies two people in completely similar appearances of the surroundings are considering "the same" objects and make "the same" judgments about them? Does not the "this" then in these two cases have a different meaning?

Sixty-four years later, in an article "The Meaning of 'Meaning' " (1975) Hilary Putnam proposed a very similar example, the Twin-Earth thought experiment, which like Husserl's was intended to show that there is something fundamentally wrong with the classical theory of meaning and reference. However, Husserl had very advanced and interesting ideas of how the problems could be solved within his phenomenological framework. At the core of phenomenology lies a broad notion of meaning that applies both to linguistic expressions and to our actions and activities of various kinds, including perception. According to this broad theory what we refer to by our words depends not only on their meaning, but also on things in our surroundings and how our body is affected by and located relative to them. Thus, for example, in the twin worlds, words with the same linguistic meaning refer to different objects by virtue of the fact that our body is related to different objects in the two worlds. So the reference of our words is determined not just by their meanings but also by our body and its relations to the world and the objects in it.

There are many striking similarities, but also important differences between Husserl's view on indexicals and that of Perry. There will not be time here to go into these here. However, let me conclude with a few words on how the study of indexicals is related to the theme of our group "Towards a New Understanding of the Mental": A proper treatment of the indexicals requires us to revise our conception of the mental and its relation to the world. The mental does not relate to the world through a purely mental connection, as envisaged by the classical conception of meaning, but through a complicated interplay where our body and the impingements of physical objects and events on our body play an important part.

TIMOTHY WILLIAMSON

Armchair Philosophy and Counterfactual Thinking

Philosophy's traditional method is thinking, without observation or experiment. Crude rationalists regard philosophy's a priori method as a virtue. According to them, it makes philosophical results especially reliable, because immune from perceptual error. Crude empiricists regard philosophy's a priori method as a vice. According to them, it makes philosophical results especially unreliable, because immune from perceptual correction. The two groups share the assumption that the a priori method of philosophy is profoundly unlike the a posteriori methods of the natural sciences. They focus on philosophers' appeals to intuition, particularly in the use of imaginary counter-examples to refute theories. How do such refutations work?

Professor Timothy Williamson,
Faculty of Philosophy,
University of Oxford, UK.
CAS Fellow 2003/04.

A paradigm is Edmund Gettier's 1963 refutation of the traditional analysis of knowledge as justified true belief. He sketches imaginary situations in which someone has what most people classify as a justified true belief that does not constitute knowledge. Such a situation does not refute the claim that knowledge actually coincides with justified true belief. What it refutes is the stronger claim, to which the traditional analysis is committed, that knowledge *necessarily* coincides with justified true belief. It does so through two claims. First, the situation is possible; it could have occurred. Second, if it had occurred, the subject would have had a justified true belief without knowledge. From those two claims, it follows that justified true belief is not sufficient for knowledge.

The second claim is best formulated as a counterfactual conditional about what *would* have happened *if* the situation had occurred, rather than as a claim about what *necessarily* follows from its occurrence, because examples cannot be described in complete detail; much background must be taken for granted. We envisage the descriptions as realized in ways that minimize departures from actuality in respects about which nothing is explicitly stipulated.

Of the two claims about the Gettier cases, only the second has been disputed. It corresponds to the 'intuition' that the subject in the Gettier lacks knowledge (but has justification). The second claim is uncontroversial, because the situations Gettier describes are mundane practical and physical possibilities, not far-out science fictions. In fact, I once brought about such a situation during a lecture. The real-life occurrence of Gettier situations makes almost no difference to the epistemology of Gettier's argument.

Asserting counterfactuals is not distinctive of a priori methodology. They are often highly contingent and asserted on a posteriori grounds. We

use them in our practical dealings with the world. Many counterfactuals are closely linked to causal statements. Counterfactuals follow from a posteriori natural laws: if it is a natural law that salt dissolves in water, then if this quantity of salt had been put in water, it would have dissolved. They also figure in the explanation of our evidence for many accepted empirical claims. We might support the claim 'There are no kangaroos on this island' with the counterfactual 'If there were any, we'd have seen some of them by now'.

We have a general cognitive ability to handle counterfactual conditionals. Sometimes we can reason from the antecedent to the consequent or to its negation, using background beliefs compatible with the antecedent. Often we seem to use a less formal or conceptually articulated process, perhaps involving some sort of mental simulation. We have no good reason to expect that the evaluation of 'philosophical' counterfactuals uses radically different cognitive capacities from the evaluation of ordinary 'unphilosophical' counterfactuals. Very often, the background knowledge needed to evaluate a counterfactual consists not of specific items of information acquired on specific occasions but of a more general sense of how things go, honed over long experience. Such a sense is typically not presented to the subject in usably verbal form.

The point is not that no distinction at all can be drawn between the a priori and the a posteriori, or that nothing falls on the a priori side, but that such a distinction lacks the significance with which it is often credited. In particular, we should not suppose it to imply major differences in reliability. If a priori intuition is understood as a distinctive cognitive capacity or pathology, it is not required for using imaginary counterexamples against philosophical theories or analyses. We have our ordinary capacities for making judgements about what we encounter, and a further capacity to evaluate counterfactuals by running those capacities 'offline'; that is already enough for philosophy to get going, without any need of a kickstart from a special faculty of intuition.

Of the two premises in Gettier's underlying argument, only the counterfactual conditional has been disputed. The other premise, that his imaginary situations are possible, is uncontroversial. In other philosophical examples, however, the possibility claim is also controversial, because the imaginary situations are more bizarre and impossible to bring about. Does the claim of their possibility rely on a distinctively philosophical faculty of intuition?

Even our capacity to assess claims of possibility and necessity can be derived from our capacity to handle counterfactuals. For possibility and necessity can be defined in terms of counterfactuals. Something is necessary if and only if whatever were the case, it would still be the case. Something is possible if and only if it is not such that whatever were the case, it would not be the case. Those definitions allow one to derive standard logical principles about possibility and necessity from standard logical principles about counterfactuals. Starting with the counterfactual conditional, we can build a promising theory of necessity and possibility. The capacity for thought about them cannot be isolated from the capacity for ordinary thought about the natural world, or excised without loss to the latter, for it is implicit in the latter.

Discussions of the epistemology of possibility often focus on imaginability or conceivability as a test of possibility: a notoriously unreliable or

circular test. Such discussions typically ignore the role of the imagination in evaluating counterfactual conditionals. In doing so, they omit the appropriate context for understanding the relation between possibility. The imagination is a standard means for running our cognitive capacities 'offline' in evaluating counterfactuals. The process is manifestly fallible and practically indispensable. We may have a special cognitive faculty or module dedicated to evaluating counterfactuals.

Thus investigation of the use of imaginary counterexamples in philosophy shows that they do not involve a special faculty of rational intuition or the illusion of such. They simply involve particular applications of general cognitive capacities – notably, the capacity to process counterfactuals – widely used throughout our cognitive engagement with the spatio-temporal world.

TOM ANDERSEN

Biodiversity and Ecosystem Services

Edvard O. Wilson once commented that although science has brought us precise answers to a range of exceedingly difficult questions, such as the number of stars in the universe, the masses of the earth and its neighboring planets, the number of genes in a virus, the mass of the electron, etc., we still do not know how many species of living organisms there are on earth – not even to the closest order of magnitude! This fact, which may appear surprising to many, has been a major motivation behind the rising interest in biodiversity the last 20 years. The term biodiversity itself is just as old, starting with a conference called the National Forum on Biodiversity in 1986 (Wilson 1988). Biodiversity is often represented as three hierarchically nested organization levels: those of genes, species, and habitats.

Associate Professor
Tom Andersen,
Department of Biology,
University of Oslo.
CAS Fellow 2003/04.

Genetic diversity refers to the level of variability among genes with the same or similar function in a population of organisms. By a population we here understand a group of organisms capable of exchanging genetic material. Genetic diversity is the basis for the ability of organisms to evolve and adapt to changing environmental conditions. Thus, inbreeding and loss of genetic diversity often reduce the likelihood of population persistence. Relevant examples are found both in the management of large carnivorous mammals and in stocking programs for economically important species like the Atlantic salmon.

Species diversity is what most people probably conceive as the amazing diversity of life: the morphological and functional variability among living organisms, although the species concept itself can become technically complicated. The most common species definition is attributed to Mayr (1970) "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups," but this definition becomes less useful when considering non-sexually reproducing (clonal) organisms, such as bacteria and protists.

Habitat diversity refers to the variability of physical environments to which different living organisms are adapted. The preservation of intact living habitats is the major premise for species persistence. Or, put in other words, it is impossible to conserve a given species without also taking the preservation of its natural habitat into consideration. Habitat destruction and fragmentation is by far the greatest threat to all three hierarchical levels of biodiversity today. Particular concern is given to the ongoing destruction of biodiversity "hot spot" habitats like rain forests and coral reefs.

The distribution of known species among phylogenetic groups is probably surprising to many: the biochemically and morphologically diverse groups of bacteria and protists (unicellular eukaryotes) are also the most species-poor ones with fewer than 100,000 described species altogether. Fungi and higher plants constitute about 350,000 species, while metazoas (multi-cellular animals) are by far the largest group with about 1,320,000 described species. Part of this discrepancy is probably due to the amount of effort devoted to investigating different groups of organisms, although techniques of molecular biology are today making it easier to discriminate between species and reconstruct phylogenies (evolutionary trees) for unicellular organisms. But the most obvious biological explanation is probably the evolution of sexual reproduction in multi-cellular plants and animals and how this paves the way for rapid adaptation and inter-sterility barriers between morphologically similar populations.

These are the species that have been collected, scientifically examined, and classified. What we do not know is how many species there are that have not yet been described. Extrapolations to the total number of species are hampered by the fact that the earth has not been explored uniformly – northern temperate regions, especially in the vicinities of universities have been investigated far better than others. Potentially species-rich habitats like the high canopies of tropical forests or the deep oceans have been appreciably under-explored. Consider that 2/3 of the earth is covered by oceans, that the mean depth of the oceans is 3700 meters, and that most of marine biological research has been focused on the upper few 100 meters of this. So it is perhaps not surprising that current estimates of the number of species of earth range from 2 to 100 million.

There are several reasons to believe that species diversity is currently being lost at an increasing rate and that human activity is the main cause of this. Habitat loss due to clear-cutting, agriculture, and urbanization is generally recognized as the major determinant of biodiversity loss. As populations need a minimum habitat area to maintain themselves, habitat fragmentation due to construction of roads, power-lines, etc also contribute strongly to this trend. Many populations are also negatively affected by local pollution and over-exploitation. The latter is most evident for economically important fish stocks. Invasions by non-native (exotic) species are often also cited as a major threat to biodiversity, especially in connection with ballast water released by marine transportation. But there is an ongoing and unresolved scientific discussion on this: some scientists argue that species invasions have throughout geological time always led to increasing species diversity (e.g., Rosenzweig 2001). Rapid climate change due to emissions of greenhouse gases has also been identified as a potential threat to biodiversity; where the critical issue is to what extent species are able to migrate fast enough to escape local changes in climatic conditions (e.g., Thomas et al. 2004).

A lot of research effort is being invested into quantifying how fast biodiversity is being lost, what the consequences are of this loss, and if there are services that healthy ecosystems provide us with, which are diminished by loss of biodiversity. Ecosystem services (Daily et al. 1997) are processes by which the environment produces resources that we often take for granted. Such as: moderating weather extremes, mitigating drought and floods, protection from erosion, regeneration and preservation of soils, cycling and transport of nutrients, protection from the sun's ultraviolet

rays, detoxification and decomposition of wastes, purification of air and water, dispersion of seeds, control of agricultural pests, regulation of disease-carrying organisms, pollination of crops and natural vegetation. For example, 80% of Mississippi River Valley wetlands have been lost by draining and channeling since 1940, leading to less capacity in the river system to absorb and buffer floods. The 1993 Mississippi River flood caused by lack of wetland buffering resulted in property damage estimated at 12 billion dollars. For example, 9 of the top 10 drugs used in the U.S. originate from natural plant products, and among the top 150 prescription drugs 87 originate from plants, 21 from fungi, 6 from bacteria, and 3 from vertebrates (snakes). For example, over 100,000 different animal species are involved in pollination of economically important plants (bees, butterflies, moths, beetles, flies, birds, and bats). One third of human food comes from plants pollinated by wild pollinators, making a value of 4-6 billion dollars per year by wild pollination services in the U.S. alone. These few examples illustrate that ecosystem services we take for granted can be of immense value to human existence and welfare.

We know way too little about how biodiversity loss can affect the integrity of these services. Stanford ecologist Paul Ehrlich has forwarded the Rivet hypothesis: compare an ecosystem to an airplane fuselage – it can stand a few rivets popping, but at the loss of a critical one the ability to remain flying is lost. Ehrlich's hypothesis is but one out of many competing ones concerning biodiversity loss and ecosystem integrity. The problem is that we do not know which one best describes the situation, and upon which of them we should base our decisions.

To take a step back; Darwin (1859) already had the intuition that "…if a plot of ground be sown with one species of grass, and a similar plot be sown with several distinct genera of grasses, a greater number of plants and a greater weight of dry herbage can be raised in the latter than in the former case …the truth of the principle that the greatest amount of life can be supported by the great diversification of life, is seen under many natural circumstances". In the other words, we should expect a diverse community to be more productive and stable than a monoculture.

This intuition remained unchallenged until the 1970s, when Robert M. May showed that mathematical models of ecological communities with many interacting species should actually be less stable and more prone to extinction than simple ones (May 1973). The shock of May's derivations led ecologists into a long discussion on how to actually describe stability in ecosystems (elegantly summarized by Grimm and Wissel 1997). In the last decade much effort has been devoted to empirical studies of relationships between species richness and ecosystem stability, productivity, and reliability.

The BIODEPTH project (Biodiversity and Ecosystem Processes in Terrestrial Herbaceous Ecosystems - http://www.cpb.bio.ic.ac.uk/biodepth/contents.html) involving large-scale field experiments in 8 European countries seems to have firmly established that there actually is a positive relation between species richness and several indicators of ecosystem function in grassland ecosystems (Tilman 1999). We still know little about how well these results generalize to ecosystem services in other habitats and on higher trophic levels. Given

the accelerating loss of biodiversity on our planet and the indisputable values of ecosystem services to mankind, these should be key research issues for the future.

References

Daily, G. C, S. Alexander, P. R. Ehrlich, L. Goulder, J. Lubchenco, P. A. Matson, H. A. Mooney, S. Postel, S. H. Schneider, D. Tilman, G. M. Woodwell. 1997. Ecosystem Services: Benefits Supplied to Human Societies by Natural Ecosystems. *Issues in Ecology*, no. 2

Darwin, C. 1859. *On the Origin of Species*. John Murray, London

Grimm, V. and C. Wissel. 1997. Babel, or the ecological stability discussions: an inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia 109*: 323–334

May, R. M. 1973. *Stability and Complexity in Model Ecosystems*, Princeton Univ. Press

Mayr, E. 1970. *Populations, Species and Evolution*. Harvard University Press

Rosenzweig, M. L. 2001. The four questions: What does the introduction of exotic species do to diversity? *Evolutionary Ecology Research* 3:361–367.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. Ferreira de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G.F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips S. E. Williams. 2004. Extinction risk from climate change. *Nature* 427: 145–148

Tilman, D. 1999. Diversity and production in European grasslands. *Science* 286:1099–1100

Wilson, E. O. 1988. *Biodiversity*. National Academies Press.

Synergies
Interdisciplinary Communications 2003/2004

Editor: Willy Østreng, Scientific director
Production: Faktotum Informasjon AS
Design: Ketill Berger, dEDBsign
Printing: Grafisk Senter Grøset AS
Copies: 800
ISBN: 82-996367-2-8

■ This booklet contains condensed versions of 24 presentations held at luncheon seminars at the Centre for Advanced Study (CAS) in Oslo in 2003/2004 by fellows from the research groups A New Understanding of the Mental (humanists), A Comprehensive Model of Human Memory (social scientists) and Food-webs, Stoichiometry and Population Dynamics (natural scientists).

The CAS is an ideal arena for science dialogue across disciplinary boundaries and academic fields, and the seminars have been instrumental in creating a feeling of both social and professional community between the research groups. They have also produced an interdisciplinary atmosphere for the clarification of concepts of significance in theory-building across group boundaries.

> "… specialized basic research is not the antithesis of complex system science, but its foremost precondition."
>
> *Willy Østreng*

> "People have an unsupported pessimistic view of their own memory."
>
> *Svein Magnussen*

> "… no one has yet been able to come up with a theory of truth which is neither false nor viciously circular."
>
> *Carsten Hansen*

> "Some people are bound to have illusions of knowing, failing to recognize their own incompetence."
>
> *Asher Koriat*

> "… cancer biology is in need of new ideas and perspectives."
>
> *James J. Elser*

Synergies
Interdisciplinary Communications 2003/2004

Centre for Advanced Study
at the Norwegian Academy of Science and Letters

Drammensveien 78
NO-0271 Oslo
Norway
Telephone: +47 22 12 25 00
Fax: +47 22 12 25 01
Email: cas@cas.uio.no
Internet: http://www.cas.uio.no
ISBN: 82-996367-2-8