

STUDIES IN
LANGUAGE
COMPANION
SERIES 92

Mental States

Volume 1: Evolution, function, nature

Edited by
Andrea C. Schalley
Drew Khlentzos

JOHN BENJAMINS PUBLISHING COMPANY

Mental States
Volume 1: Evolution, function, nature

Studies in Language Companion Series (SLCS)

This series has been established as a companion series to the periodical *Studies in Language*.

Editors

Werner Abraham
University of Vienna

Michael Noonan
University of Wisconsin-Milwaukee USA

Editorial Board

Joan Bybee
University of New Mexico

Ulrike Claudi
University of Cologne

Bernard Comrie
Max Planck Institute, Leipzig

William Croft
University of New Mexico

Östen Dahl
University of Stockholm

Gerrit J. Dimmendaal
University of Cologne

Ekkehard König
Free University of Berlin

Christian Lehmann
University of Erfurt

Robert E. Longacre
University of Texas, Arlington

Brian MacWhinney
Carnegie-Mellon University

Marianne Mithun
University of California, Santa Barbara

Edith Moravcsik
University of Wisconsin, Milwaukee

Masayoshi Shibatani
Rice University and Kobe University

Russell S. Tomlin
University of Oregon

John W.M. Verhaar
The Hague

Volume 92

Mental States. Volume 1: Evolution, function, nature
Edited by Andrea C. Schalley and Drew Khlentzos

Mental States

Volume 1: Evolution, function, nature

Edited by

Andrea C. Schalley and Drew Khlentzos

University of New England

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Mental states / edited by Andrea C. Schalley and Drew Khlentzos.

v. cm. -- (Studies in Language Companion Series, ISSN 0165-7763 ; v. 92-93)

Includes bibliographical references and index.

Contents: 1: Evolution, function, nature -- v. 2. Language and cognitive structure.

1. Psycholinguistics. I. Schalley, Andrea C., 1972- II. Khlentzos, Drew.

P37.M363 2007

401'.9--dc22

2007033516

ISBN 978 90 272 3105 5 (hb : set : alk. paper)

ISBN 978 90 272 3102 4 (hb : v. 1 : alk. paper)

ISBN 978 90 272 3103 1 (hb : v. 2 : alk. paper)

© 2007 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Preface	VII
List of contributors	IX
1. Mental states: Evolution, function, nature <i>Drew Khlentzos and Andrea C. Schalley</i>	1
2. Lithic design space modelling and cognition in <i>Homo floresiensis</i> <i>Mark W. Moore</i>	11
3. “As large as you need and as small as you can”: Implications of the brain size of <i>Homo floresiensis</i> <i>Iain Davidson</i>	35
4. <i>Homo</i> on Flores: Some early implications for the evolution of language and cognition <i>Michael J. Morwood and Dorothea Cogill-Koez</i>	43
5. Evolving artificial minds and brains <i>Pete Mandik, Mike Collins and Alex Vereschagin</i>	75
6. Multi-agent communication, planning, and collaboration based on perceptions, conceptions, and simulations <i>Peter Gärdenfors and Mary-Anne Williams</i>	95
7. The modal-logical interpretation of the causation of bodily actions <i>Hiroyuki Nishina</i>	123
8. Do we access object manipulability while we categorize? Evidence from reaction time studies <i>Anna M. Borghi, Claudia Bonfiglioli, Paola Ricciardelli, Sandro Rubichi and Roberto Nicoletti</i>	153
9. Speaking without the cerebellum: Language skills in a young adult with near total cerebellar agenesis <i>Alessandro Tavano, Franco Fabbro and Renato Borgatti</i>	171

10. Ontologies as a cue for the metaphorical meaning of technical concepts <i>Helmar Gust, Kai-Uwe Kühnberger and Ute Schmid</i>	191
11. Anti-realist assumptions and challenges in philosophy of mind <i>Drew Khlentzos</i>	213
12. Vagueness, supertranslatability, and conceptual schemes <i>Arcady Blinov</i>	233
13. Visual representation in a natural communication system: What can signed languages reveal about categorisation across different modes of representation? <i>Dorothea Cogill-Koez</i>	247
14. Hidden units in child language <i>Stephen Crain, Takuya Goro and Utako Minai</i>	275
Name index	295
Subject index	299
Table of contents of Volume 2	303

Preface

This volume is the first of a two-volume collection on mental states. The contributions to this volume focus on evolutionary and functional aspects of certain mental states in an effort to understand their nature, whereas the second volume is concerned with the question what language and language use reveals about cognitive structure and underlying cognitive categories.

Questions that are addressed in this volume include: (i) how early did cognitive states of a sort rich enough to support communication and planning appear in the evolutionary history of hominids?; (ii) is it possible to infer the existence of sophisticated cognitive states from evidence of tool use?; (iii) how do mental states represent situations or events or actions?; (iv) how can we theoretically model mental states?; (v) how can we simulate mental states and their functions?; and (vi) what insights can conceptual categorisation – both linguistic and non-linguistic – give us into the organisation and structure of the mind and hence of mental states?

The volume contains contributions from psychologists, linguists, artificial intelligence researchers, neuroscientists, archaeologists and philosophers, bringing together scholars from the diverse fields of cognitive science, or more specifically, the study of language and cognition. This reflects the provenance of the chapters, most of which were presented at the *International Language and Cognition Conference*, held in September 2004 at Pacific Bay Resort in Coffs Harbour, Australia.

We would like to express our gratitude to the reviewers from the diverse areas represented in these volumes. Without their expertise we could not have put together such a broad-ranging, interdisciplinary compilation. We would also particularly like to thank Vicki Knox and Michael Roberts for their tireless work in the preparation of the manuscripts, and the Language and Cognition Research Centre of the University of New England for its financial support.

Andrea C. Schalley and Drew Khlentzos

List of contributors

Arcady Blinov

Language and Cognition Research Centre
School of Humanities
University of New England
Armidale NSW 2351
Australia
ablinov@une.edu.au

Claudia Bonfiglioli

Dipartimento di Scienze della
Cognizione e della Formazione
Universita' di Trento, Polo di Rovereto
Via Matteo del Ben, 5 I-38068 Rovereto
(TN) Italy
bonfiglioli@form.unitn.it

Renato Borgatti

Department of Neurorehabilitation 1
Scientific Institute "E.Medea",
la Nostra Famiglia
Bosisio Parini (LC) 23842
Italy
renato.borgatti@bp.Lnf.it

Anna M. Borghi

Department of Psychology
Viale Berti Pichat 5
40127 Bologna
Italy
annamaria.borghi@unibo.it

Dorothea Cogill-Koez

Language and Cognition Research Centre
School of Behavioural, Cognitive and
Social Sciences

University of New England
Armidale NSW 2351
Australia
dcogill@une.edu.au

Mike Collins

Department of Philosophy
CUNY Graduate Center
New York NY 10016
United States of America
mikecollins165@comcast.net

Stephen Crain

Macquarie Centre for Cognitive Science
Macquarie University
North Ryde NSW 2113
Australia
scrain@maccs.mq.edu.au

Iain Davidson

Language and Cognition Research Centre
School of Humanities
University of New England
Armidale NSW 2351
Australia
Iain.Davidson@une.edu.au

Franco Fabbro

Neurolinguistics Unit
Scientific Institute "E.Medea",
la Nostra Famiglia
San Vito al Tagliamento (PN) 33078
Italy
franco.fabbro@pp.Lnf.it

Peter Gärdenfors
Cognitive Science
Lund University
Kungshuset, Lundagård
SE – 222 22 LUND
Sweden
peter.gardenfors@lucs.lu.se

Takuya Goro
Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park MD 20742
United States of America
takuyag@umd.edu

Helmar Gust
Institute of Cognitive Science
University of Osnabrück
Albrechtstr. 28
49076 Osnabrück
Germany
hgust@uos.de

Drew Khlentzos
Language and Cognition Research Centre
School of Behavioural, Cognitive and
Social Sciences
University of New England
Armidale NSW 2351
Australia
dkhlentz@une.edu.au

Kai-Uwe Kühnberger
Institute of Cognitive Science
University of Osnabrück
Albrechtstr. 28
49076 Osnabrück
Germany
kkuehnbe@uos.de

Pete Mandik
Cognitive Science Laboratory

Department of Philosophy
William Paterson University
263 Atrium Building
300 Pompton Road
Wayne NJ 07470
United States of America
mandikp@wpunj.edu

Utako Minai
Cognitive Neuroscience of Language
Laboratory
Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park MD 20742
United States of America
utako@wam.umd.edu

Mark W. Moore
Language and Cognition Research Centre
School of Humanities
University of New England
Armidale NSW 2351
Australia
mmoore2@une.edu.au

Michael J. Morwood
Language and Cognition Research Centre
School of Humanities
University of New England
Armidale NSW 2351
Australia
mmorwood@une.edu.au

Roberto Nicoletti
Dipartimento di Discipline della
Comunicazione
Università di Bologna
Via Azzo Gardino, 23
40122 Bologna
Italy
nicoletti@dsc.unibo.it

Hiroyuki Nishina

Faculty of Liberal Arts
Saitama University
255 Shimo-Okubo, Sakura-Ku
Saitama City 338-8570
Japan
hnishina@titan.ocn.ne.jp
hrnshn@mail.saitama-u.ac.jp

Paola Ricciardelli

Dipartimento di Psicologia,
Università di Milano-Bicocca
Edificio U9
Viale dell'Innovazione 10
20126 Milano
Italy
paola.ricciardelli@unimib.it

Sandro Rubichi

Dipartimento di Scienze Sociali,
Cognitive, Quantitative
Università di Modena e Reggio Emilia
Via Allegri 9
42100 Reggio Emilia
Italy
rubichi@unimore.it

Andrea C. Schalley

Language and Cognition Research Centre
School of Behavioural, Cognitive and
Social Sciences
University of New England
Armidale NSW 2351
Australia
andrea.schalley@une.edu.au

Ute Schmid

Faculty of Information Systems and
Applied Computer Sciences
University of Bamberg
Feldkirchenstr. 21
96045 Bamberg
Germany
ute.schmid@wiai.uni-bamberg.de

Alessandro Tavano

IRCCS "E.Medea"
Via Cialdini 29
33037 Pasian di Prato
Italy
alessandro.tavano@pp.Lnf.it
alessandro.tavano@libero.it

Alex Vereschagin

Cognitive Science Laboratory
William Paterson University
Wayne, NJ 07470
United States of America
philosopher684@yahoo.com

Mary-Anne Williams

Innovation and Technology Research
Laboratory
Faculty of Information Technology
University of Technology
Sydney NSW 2007
Australia
Mary-Anne@it.uts.edu.au

CHAPTER 1

Mental states

Evolution, function, nature

Drew Khlentzos and Andrea C. Schalley

This volume presents a rich diversity of views from researchers in cognitive science and associated disciplines – archaeology, artificial intelligence, linguistics, neuroscience, philosophy, and psychology – on the nature, function and evolution of “mental” or “cognitive” states. A quick glance at the titles of the contributions and/or the disciplinary backgrounds of the contributors might lead one to suspect little commonality in theoretical interests. However, this would be a mistake. Although the contributions differ markedly in approach and methodology, common questions about mind and cognition unite them.

One question, addressed in the first three chapters of this volume, concerns the *evolution* of cognitive states: How early did cognitive states of a sort rich enough to support communication and planning appear in the evolutionary history of hominids? What types of evidence could be marshalled to decide this question – is it possible to infer the existence of sophisticated cognitive states from the evidence of tool use, for example?

A different question concerning mental states is addressed in the four chapters that follow. This question concerns the *function* of cognitive states: How do mental states represent situations or events or actions? What types of representation must the brain process to achieve this feat? It is extraordinarily difficult to break free of the Cartesian intuition that the mind is wholly distinct from anything physical, the brain included. Descartes notoriously thought that animals were mere automata, not possessed of mentality at all. Even though contemporary work on animal intelligence makes this view seem implausible, Descartes’ challenges have become no easier to answer: What exactly is a mind? Which creatures have one?

The remaining six chapters deal, in one way or another, with the insights that conceptual categorisation (both linguistic and non-linguistic) can give into the organisation and structure of the mind. Although approaches range from the purely theoretical to the experimental, all of these chapters seek some unifying theoretical account of the *nature* of categorisation, and thus the nature of mind.

With these general observations behind us, let us turn to specific issues raised in the individual chapters. Since the dramatic discovery in September 2003 of *Homo floresiensis* on the island of Flores by Michael Morwood and his team, debate has raged over whether this tiny hominin, nick-named “the hobbit”, was a new species or not. How did a creature half the size of an adult human with a brain smaller than a chimpanzee’s manage to traverse the deep waters separating the island of Flores from the mainland? If this tiny hominin is a new species, from which hominin did it evolve? Could these diminutive hominins really have hunted massive prey such as *Stegodon* or were they merely opportunistic scavengers? How could they have planned such hunts without sophisticated forms of communication only a language makes possible?

From the start a vigorous sceptical view has maintained *Homo floresiensis* is something of a misnomer: The skeletal remains found at Liang Bua were those of a microcephalic form of modern human. However, this theory is hard-put to explain the latest findings of stone tools, remarkably similar to the ones found in 2003 together with the “hobbit” skeletal remains – the tools date to 880,000 years ago, more than 600,000 years before *Homo sapiens* appeared on the evolutionary scene.

The first three chapters – presenting conflicting assessments and therefore providing a stimulating basis for future discussions – focus on *Homo floresiensis*’ cognitive abilities as inferred from the design of their stone tools. Mark Moore in his chapter “Lithic design space modelling and cognition in *Homo floresiensis*” develops a sophisticated theoretical model that permits one to infer cognitive abilities from stone flaking. On the assumption that increases in the complexity of the flake unit are reliable indicators of advancing cognitive ability, Moore argues that *Homo floresiensis* most probably did not possess a high level of cognitive ability. The reason? *Homo floresiensis* simply chained together the simplest types of flakes without any attempt to stack them hierarchically as can be expected from a creature capable of advanced thinking and planning.

The question of *Homo floresiensis*’ small brain size in comparison to that of other hominins and the inferences it sustains to cognitive ability is directly addressed by Iain Davidson in his chapter “As large as you need and as small as you can’: Implications of the brain size of *Homo floresiensis*”. Davidson argues that the evolutionary pattern of changes in hominin brain size is not a simple “onwards and upwards” trend. Rather, it is a punctuated pattern explained by the interaction between body size and brain size. These two factors in turn need to be given a selectionist explanation, Davidson contends. As large brains are expensive to maintain there is a definite selectionist pressure against their emergence. The most plausible explanation for the disappearance of hominins with small brains is that they could not compete with the cognitive advantages conferred upon their rivals by their larger brains, Davidson argues. However, in the isolated environment of

Flores, no selectional pressures emerged to counteract the default choice of smaller body and brain size. What are the implications for the cognitive abilities of *Homo floresiensis*? Davidson thinks this can only be answered by looking to the behaviour of *Homo floresiensis*, in particular, their tool use, their use of fire and their hunting.

In their contribution “*Homo* on Flores: Some early implications for the evolution of language and cognition” Michael Morwood, the discoverer of *Homo floresiensis*, and Dorothea Cogill-Koez investigate the implications of the Flores find for the evolution of language and cognition more broadly, drawing on diverse evidence from different disciplines. They integrate a meta-theoretical discussion on models in evolutionary theory and stress the risky nature of any inferences to cognitive capacity from the fossil record of bones and stone implements. Nonetheless they are hopeful that the mass of diverse evidence from “animal communication to zoogeography” will impose enough specific constraints to filter out in the end all but the uniquely correct explanatory hypothesis.

Putting the special case of *Homo floresiensis* to one side we are left with some unresolved general questions, in particular about the correlation between brain size and cognitive ability: Are bigger brains invariably better cognitively? What minimum brain size is necessary to sustain higher cognitive processing? These are questions about the function and nature of mental states, and it is these issues that are explored in very different ways in the next four chapters.

In “Evolving artificial minds and brains” Pete Mandik, Mike Collins and Alex Vereschagin argue for the need to posit mental representations in order to explain intelligent behaviour in very simple creatures. The creature they choose is the nematode worm and the behaviour in question is chemotaxis. Many philosophers think that a creature’s brain state or neural state cannot count as genuinely mental if the creature lacks any awareness of it. Relatedly, they think that only behaviour the creature is conscious of can be genuinely intelligent behaviour. When the standards for mentality and intelligence are set so high, very few creatures turn out to be capable of enjoying mental states or exhibiting intelligent behaviour. Yet the more we learn about sophisticated cognitive behaviour in apparently simple organisms the more tenuous the connection between mentality and consciousness looks.

If there is a danger in setting the standards for mentality and intelligence too high, there is equally a danger in setting them too low, however. Many cognitive scientists would balk at the suggestion that an organism as simple as a nematode worm could harbour mental representations or behave intelligently. Yet Mandik, Collins and Vereschagin argue that the worm’s directed movement in response to chemical stimuli does demand explanation in terms of certain mental representations. By “mental representations” they mean reliable forms of information about

the creature's (chemical) environment that are encoded and used by the organism in systematic ways to direct its behaviour.

To test the need for mental representations they construct neural networks that simulate positive chemotaxis in the nematode worm, comparing a variety of networks. Thus networks that incorporate both sensory input and a rudimentary form of memory in the form of recurrent connections between nodes are tested against networks without such memory and networks with no sensory input. The results are then compared with the observed behaviour of the nematode. Their finding is that the networks with both sensory input and the rudimentary form of memory have a distinct selectional advantage over those without both attributes.

Even if it is too much to require mental states to be conscious, there is still the sense that there is more to mentality than tracking and responding to environmental states. One worry is that there is simply not enough plasticity in the nematode worm's behaviour to justify the attribution of a mind. A more important worry is that the nematode does not plan – it is purely at the mercy of external forces pushing and pulling it in the direction of nutrients. In this regard, it is instructive to compare the behaviour of the nematode worm with the foresighted behaviour of the jumping spider, *Portia Labiata*. *Portia* is able to perform some quite astonishing feats of tracking, deception and surprise attack in order to hunt and kill its (often larger) spider prey. Its ability to plot a path to its victim would tax the computational powers of a chimpanzee let alone a rat. It has the ability to plan a future attack even when the intended victim has long disappeared from its sight. *Portia* appears to experiment and recall information about the peculiar habits of different species of spiders, plucking their webs in ways designed to arouse their interest by simulating the movements of prey without provoking a full attack. Yet where the human brain has 100 billion brain cells and a honeybee's one million, *Portia* is estimated to have no more than 600,000 neurons!

How large must a brain then be to sustain mental states? Having looked at evolution theory and the construction of neural networks, approaches that promise to shed further light on this issue include those that simulate and model merely the functionality: Artificial intelligence researchers aver brains are not strictly necessary – all that is required is the relevant functional capacity of the brain. In their contribution “Multi-agent communication, planning and collaboration based on perceptions, conceptions and simulations”, Peter Gärdenfors and Mary-Anne Williams develop a conceptual framework for grounded communication between agents which they test in the domain of soccer-playing robots. These robots need to assess and then communicate to their fellows rapidly-changing information about the location and movement of the ball, opposition players and other team members.

The Gärdenfors-Williams framework makes use of the idea of *conceptual spaces* to model the changing conditions governing items in the robots' world. The

concepts in these conceptual spaces categorise information geometrically by means of similarity to given exemplars. Similarity is a matter of the distance between a given item and an exemplar in a multi-dimensional similarity space on this view. Concepts are represented by sets of convex regions in various domains. Gärdenfors and Williams distinguish concepts or mental representations that are *cued* to current environmental goings-on from those that are *detached* from such proceedings. Conceptual spaces then bridge sensorimotor information to detached representations of various sorts.

An interesting feature of their theoretical framework is the crucial role *simulations* play in it – and that they do so on different levels in their theoretical framework. Indeed, Gärdenfors and Williams hold that simulation underwrites all higher thought processes. In the case of their soccer-playing robots, each robot must form a world-model that has to be properly *grounded* to ensure success in collaboratively planned actions.

In the following chapter “The modal-logical interpretation of the causation of bodily actions” Hiroyuki Nishina provides a formal model of the various rotations and turns an agent’s body makes when performing various actions. He builds on Ray Jackendoff’s work and David Marr’s 3D model. Nishina uses modal-logical formulae that are evaluated in relation to rotation/turn functions to describe the characteristic movements associated with various actions. A causal relation ties one set of movements to another in the performance of a typical action. Nishina then attempts to show that the conceptual structure underlying action verbs such as *kick* really does correlate with the modal-logical representation of the turning/rotational microstructure of such actions.

It would be an interesting exercise to see whether Nishina’s modal-logical representations of the rotation of various joints could be incorporated into the simulations of Gärdenfors and Williams’ soccer-playing robots to refine prediction of other robots’ movements as well as control of their own.

Anna Borghi, Claudia Bonfiglioli, Paola Ricciardelli, Sandro Rubichi and Roberto Nicoletti in their contribution “Do we access object manipulability while we categorize?” set out to experimentally determine whether information on the affordances of objects (the types of interactions we can have with them) is indeed part of our conceptual representation of them. They designed two experiments to test subjects’ ability to categorise things. They construe their results as evidence that the brain responds preferentially to information about object manipulability rather than to information about an object’s function, and – given that their experiments are based on drawings or words – provide evidence that recognition of pictures or words denoting objects automatically activates motor representations associated with these objects.

In their chapter “Speaking without the cerebellum: Language skills in a young adult with near total cerebellar agenesis” Alessandro Tavano, Franco Fabbro and Renato Borgatti present the results of a case study of a person born with cerebellar agenesis who gradually acquired adequate communicative skills over a twenty year period. They point out that whilst the part played by the cerebellum and other subcortical structures such as the basal ganglia in the execution, coordination and timing of motor actions has long been appreciated, it is only recently that researchers have realised these structures also have an important role in language-learning, and, indeed, other forms of higher cognition. Precisely what that role is, however, remains unclear. Perhaps it is restricted to the coordination and sequencing of those processes that go to make up language comprehension and production? If so, the primary function of these subcortical structures would be one of sustaining and promoting the efficiency of the language-processing network through greater computational capacity.

The alternative hypothesis is that the cerebellum makes a specific contribution to language-processing. Tavano, Fabbro and Borgatti note there are two separate types of learning involved in language acquisition: procedural and declarative. Declarative knowledge is explicit and propositional. Our knowledge of the lexicon is declarative. Knowledge of phonology and morphosyntax on the other hand is largely implicit and resistant to verbal formulation – knowledge acquired in the learning process, without reflection on the principles governing it. The aim of the study of the patient G. R. was to determine whether language-learning was possible in the presence of cerebellar agenesis and to investigate the grammatical and lexical limitations on acquisition in the absence of the usual procedural-learning support provided by the cerebellum. They conclude that even though many of G. R.’s cerebellar functions seem to have been reassigned to cerebellar networks, this relocation is far from complete and that normal language acquisition is very far from being an exclusively neocortical concern.

A distinctive feature of human cognition is the effortless ability to construct analogies and interpret metaphors. How do we do this? As Helmar Gust, Kai-Uwe Kühnberger and Ute Schmid point out in their chapter “Ontologies as a cue for the metaphorical meaning of technical concepts”, metaphors have no general form, they can connect quite disparate domains and they are highly context-sensitive. Yet everyone immediately understands Romeo’s utterance “Juliet is the sun”.

Gust, Kühnberger and Schmid focus on metaphors in the technical domain drawn from Information Technology – such as *mouse*, *buffer*, *browser*, *virus* etc. Classical accounts of analogical reasoning posit mappings between the source and target domains grounded on extensions of properties and relations. Gust, Kühnberger and Schmid demur – they suggest an analogical relation associates source and target and that this relation explains the meaning of metaphors. They propose

a formal theory, Heuristic-Driven Theory Projection, HDTP, formulated in a many-sorted first-order language to model analogical reasoning. HDTP permits the computation of metaphorical meanings from the analogical relations it models using background knowledge of the ontologies of source and target domains to reduce the complexity of the computations.

The hope of cognitive science is to provide a wholly scientific or naturalistic account of the mind. If something as elusive as metaphorical meaning can prove amenable to a computational approach, that goal might seem a little closer. Yet there are certain sorts of mental states that seem highly resistant to any scientific reduction. These are states associated with immediate conscious experience. Thus our experience of pain and other bodily sensations together with our experience of external stimuli such as colours, sounds, tastes, touch are often taken to show that *pace* the aspirations of cognitive science, the mind is *not* part of the natural order but is distinct from it. This is because of the difficulty of finding a physical property with which to identify the colour red or the bitter taste of lemons, for example.

A famous argument designed to establish this conclusion goes by the name of the Knowledge Argument. This argument purports to prove that colour experience puts us directly in touch with non-physical sensory qualities or qualia. As philosophical arguments go, it is disarmingly simple and direct: Since a colour-blind superscientist could master all the physical facts about colour experience without knowing how red things characteristically look to normal colour perceivers, this phenomenal feature of colour experience could not be anything physical. Drew Khlentzos argues in his contribution “Antirealist assumptions and challenges in the philosophy of mind” that in spite of its apparent simplicity the argument rests on some dubious verificationist (or “antirealist”) assumptions. When these are exposed and questioned the argument is nowhere near as compelling as it appears, he contends.

An important question about the human mind is whether our categorisations or conceptual schemes are the only ones possible. Edward Sapir and Benjamin Lee Whorf famously denied this. Far from being inevitable, our categorisations are mere artefacts of our language and culture, they averred. Moreover, they claimed different cultures nurtured very different conceptions of such central notions as time, causation, agency, the self and morality, conceptions which were reflected in their languages. This is the Sapir-Whorf hypothesis of linguistic relativity, a hypothesis that has been enormously influential in linguistics, anthropology and sociology. In his famous paper “On the very idea of a conceptual scheme” the American philosopher Donald Davidson took issue with the idea that there could be radically different conceptual schemes.

Building on a thought-experiment of George Simmel’s, Arcady Blinov in his contribution “Vagueness, supertranslatability and conceptual schemes” argues

that Davidson is mistaken. Blinov imagines beings who do not share our colour experiences but instead have direct phenomenal experience of the same electromagnetic radiation that underwrites human colour experiences. He argues that their language, “Simmelian”, is not translatable into any human language. Nonetheless it is “supertranslatable”, he maintains – we can at least pair *sets* of Simmelian sentences, if not the individual Simmelian sentences required for *bona fide* translation, with English sentences such as *This surface is red*. Blinov goes on to recommend supertranslation as a method for modelling the pervasive phenomenon of vagueness in natural language.

Dorothea Cogill-Koez provides a different perspective on the universality of categorisation in her chapter “Visual representation in a natural communication system”. The question she investigates is whether human categorisation is constant across signed and spoken languages. She argues against the tendency to assume that, on the one hand, verbal representations of spoken languages reflect mental categories and, on the other hand, that the analogue structure of pictorial representation deployed by signed languages merely captures visual similarities between sign and referent. This is a false dichotomy, she contends, since the signs of signed languages are not mentally processed by means of matching the physical structure of their intended referents with any visual image of that referent and since the connection between verbal representations and mental categories is not immediate.

Focusing on signs used to represent visual spatial situations and events, “classifier predicates”, Cogill-Koez notes that whilst these signs do convey information about the visual-spatial properties of their referents without any conceptual preprocessing, nonetheless across the various signed languages classifier predicates assume standard discrete forms – standard handshapes and movements are deployed to depict idealised prototypes plausibly associated with distinctive mental categories.

The theme of mental categorisation is further pursued in the final chapter of the volume “Hidden units in child language” by Stephen Crain, Takuya Goro and Utako Minai who present some intriguing linguistic evidence for a nativist explanation of childrens’ understanding of a certain class of expressions common to all languages. This is the class of logical particles of the language: In English, these are words such as *not*, *and*, *or*, *every*, *only*. Crain, Goro and Minai focus on the interactions between negation *not* and disjunction *or* and between negation and conjunction *and*. In particular, they wish to know whether *or* and *and* in natural language retain the usual Boolean interpretations they receive in classical logic. If they do, then embedding them within the scope of an overt negation operator such as *not* or a covert one such as *only* ought to produce some systematic results. The simplest of these go by the name of De Morgan’s Laws of which there are two, one for negated disjunctions:

$$(I) \quad \neg(X \vee Y) :: \neg X \wedge \neg Y$$

and one for negated conjunctions:

$$(II) \quad \neg(X \wedge Y) :: \neg X \vee \neg Y$$

In these two laws, ‘ $\neg$ ’ is the sign for negation; ‘ $\vee$ ’ is the sign for disjunction, ‘ $\wedge$ ’ is the sign for conjunction and ‘ $::$ ’ is the sign for logical equivalence; ‘ X ’ and ‘ Y ’ are variables ranging over propositions. Crain, Goro and Minai call one direction of the first law the “Conjunctive Entailment” of disjunction in the scope of negation and one direction of the second law the “Disjunctive Entailment” of conjunction in the scope of negation.

Semantic Nativism predicts the existence of a default mental setting for the meanings of the logical particles prior to the acquisition of any specific natural language. To an empiricist Semantic Nativism might sound implausible but for logical expressions the view has considerable appeal – prelinguistic children appear to reason and if so they must be able to perform logical transitions of some sort. Moreover, there are some good evolutionary reasons for Nature to set the constraints governing the interpretation of logical particles *not*, *and* and *or* at their Boolean readings – for the Boolean versions of these operators correspond to simple set-theoretic operations of complementation, intersection and union, respectively. Hence, if there were languages with overtly non-Boolean interpretations of *or* and *and* for which young children assumed the default Boolean readings in the learning process, Semantic Nativism would have received some impressive empirical support.

Crain’s team’s finding is that Japanese and Chinese children in the process of learning their native languages do indeed appear to adopt a Boolean interpretation of their disjunctive operators *ka* and *huozh*, respectively. This is in spite of the fact that both *ka* and *houzh* are “positive polarity items” that cannot be embedded within the scope of local negation. Thus, whilst the English sentence *the pig did not eat the pepper or the carrot* licenses the Conjunctive Entailments (i) the pig did not eat the pepper and (ii) the pig did not eat the carrot, the Japanese equivalent *Butasan-wa ninjin ka pi’ iman-wo tabe-nakat-ta* licenses neither of these entailments. For although this sentence literally says ‘the pig did not eat the pepper or the carrot’ what it actually means is ‘the pig did not eat the pepper *or* the pig did not eat the carrot’. That is, the Japanese disjunction *ka* takes wider scope than negation *nakat* irrespective of its surface position in the sentence. The unequivocal finding of Crain’s team is that both Japanese and Chinese children interpreted negated disjunctions as licensing Conjunctive Entailments even though their adult counterparts did not. Contextual and scope ambiguities conspire to unsettle our confidence in the interpretation of even the simplest sentences containing nested

logical operators. Crain, Goro and Minai's chapter contains a subtle and engrossing discussion of these issues and a persuasive case for Semantic Nativism. In so doing, it advances our understanding of the structure of the human mind.

Lithic design space modelling and cognition in *Homo floresiensis*

Mark W. Moore

Stone flaking is achieved through integral sets of geometrical identifications and motor actions collectively referred to as the “flake unit”. Early trends in technological evolution involved elaborating the internal complexity of the flake unit and later trends involved elaborating the way that flake units were combined. Studies by developmental psychologists suggest that internal and external increases in complexity reflect advances in cognitive ability. *Homo floresiensis* combined the simplest type of flake units by arranging them in chains rather than stacking them hierarchically. Thus *Homo floresiensis* lithic technology does not indicate high levels of cognitive ability.

1 Introduction

In this chapter I will infer the cognitive abilities of *Homo floresiensis* by examining the way the hominin flaked stone into tools. Two things are necessary to accomplish this: First, a model that relates stone flaking to cognitive ability, and second, an understanding of the ways *H. floresiensis* modified stones. I recently developed such a model – I call it the “lithic design space model” – and applied it to the results of an analysis of stone artifacts recovered from Liang Bua Cave in direct association with *H. floresiensis* (Moore 2005; Moore & Brumm 2007; Moore et al. Forthcoming).

It is no exaggeration to say that *H. floresiensis* is one of palaeoanthropology’s great challenges. *H. floresiensis* is a complex mixture of “primitive” and derived physical features that arose, according to one view, in a context of extreme isolation and dwarfing (Brown et al. 2004; Morwood et al. 2005). If so, generally agreed-upon evolutionary trends in the genus *Homo* are turned on their head with *H. floresiensis* (Lahr & Foley 2004). For instance, assuming that the ancestor of *H. floresiensis* was *Homo erectus*, we see on Flores the dramatic shrinking of absolute

brain size through time (Brown et al. 2004; Falk et al. 2005), defying trends in punctuated brain growth that occurred in Africa (e.g. Rightmire 2004).

Much has been made in the press of the “sophistication” of *H. floresiensis*’s stone tools (Jacob et al. 2006; Kohn 2005; Martin et al. 2006; Wong 2005). If, as is assumed, *H. floresiensis*’s initial technology resembled the stone tools associated with *H. erectus* on Java (cf. Brumm et al. 2006; Moore & Brumm 2007), not only did the hominin retain its technological abilities in the face of selective pressures on Flores, but it in fact increased those abilities. That *H. floresiensis* could manage to substantially shrink its brain yet retain the cognitive ability to make and use stone tools is telling us something important about evolution in our genus. However, to make sense of this issue we must first have a way of inferring cognition from stone tools.

2 Lithic design space

The very earliest stone tools in Africa were made by a process called “freehand percussion” (Schick & Toth 1993). I model the aspects of striking a flake by freehand percussion as the “basic flake unit”. Following from Greenfield’s (1991) study comparing linguistics to motor actions, I use a tree structure to portray the basic flake unit (Figure 1). The basic flake unit includes three elements: 1) an ideational element involving the identification of crucial geometric variables on the core; 2) three action elements done in response to the identification and resulting in the correct positioning of the core; and 3) a fourth action element involving the articulation of the two hands to remove the flake (cf. Pelegrin 1993: 311).

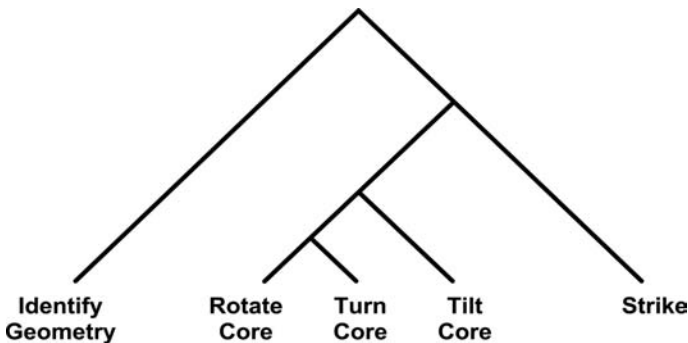


Figure 1. A tree structure model of the basic flake unit. To successfully remove a flake, the knapper must recognise the appropriate geometry on a flakable stone, must correctly adjust the core with the hand securing it, and must coordinate the actions of both hands to accurately and forcefully strike the core platform with a hammerstone

The ideational element refers to a hominin's ability to recognise an essential geometrical relationship consisting of three features (Figure 2). First, an area of high mass must occur on the face of the stone. Second, a suitable platform surface must occur on a different face of the stone but adjoining the high mass. And third, the first two features must be positioned at an acute angle (less than 90°) to one another. Brain-imaging shows that the recognition of this geometrical relationship recruits part of the brain's superior parietal lobes in adult *Homo sapiens*, the area correlated with cognitive representations of the external world. In contrast, the active aspects of striking flakes recruit areas of the brain that correlate with generalised motor tasks (Stout, Toth & Schick 2000).

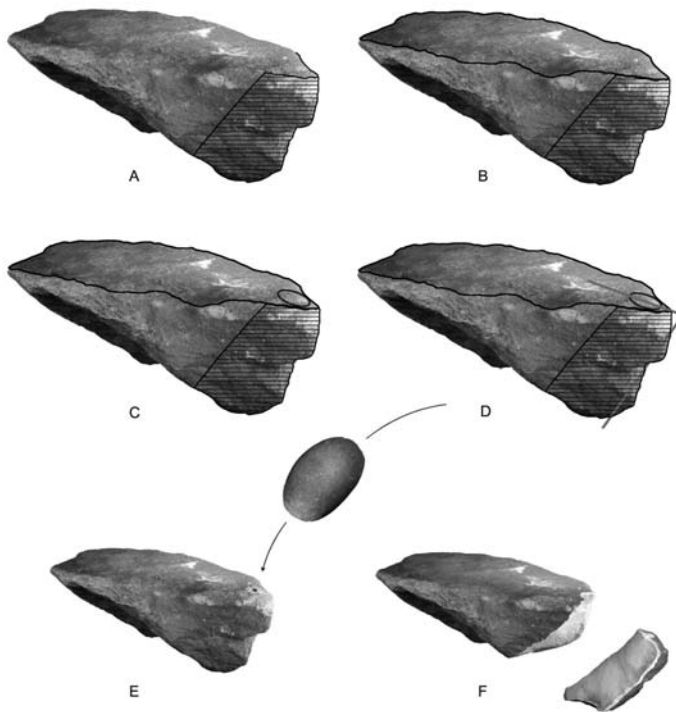


Figure 2. The crucial geometrical relationships for successful flake removal. a) High mass must occur on the face of the stone. b) A platform surface must be identifiable on a different part of the stone from the high mass. c) A portion of the platform surface must adjoin the high mass. d) Features matching a-c must be positioned at an acute angle to one another. e) The motor actions of the hand securing the core must be coordinated with the actions of the hand swinging the hammerstone to ensure the blow lands in the proper location. f) A flake is removed

Acting upon the geometrical relationship invokes three actions. First, the core must be rotated until the platform surface is positioned for striking. This will, in many cases, require rotation of the stone between faces to get the orientation correct. Second, the core must be turned from left to right or right to left (e.g. Toth 1985). This action positions the core relative to the arc of the hammer so that: 1) the intended point of force application is behind the high mass, and 2) the force delivered to this point will propagate through the mass. And third, the core must be tilted in relation to the hammer arc so that the downward and outward forces are delivered in the correct ratio to one another (see Crabtree 1972: 15–16). The actions of geometrical adjustment involve the non-dominant hand in modern humans. Striking the flake requires the articulation of both hands.

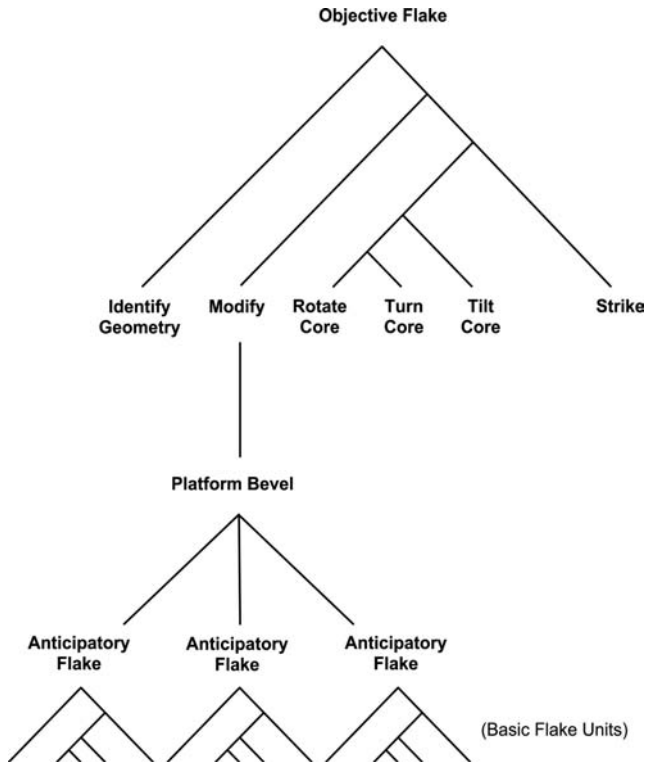


Figure 3. A tree structure model of the complex flake unit. The knapper must recognise that the platform/mass geometry on the core can be modified by preliminary flaking, triggering the construction of a lower tier. Geometrical adjustment occurs at the lower “anticipatory” tier by applying basic flake units. This prepares the core geometry for striking the “objective” flake in the upper tier

The basic flake unit is applied to zones of exploitable high mass as they are identified on the core. An elaboration in knapping complexity entails removing flakes to adjust the platform angle and enable the removal of otherwise unexploitable high mass. Removing flakes to adjust the platform angle – these are “anticipatory” flakes in the sense that they anticipate “objective” flake removals – involves repeatedly applying the basic flake unit within a structure subordinated to the process of removing the “objective” flake. I refer to the resulting two-tiered structure as the “complex flake unit” (Figure 3). The stimulus to construction of the lower tier is a realisation in the upper tier that geometrical relations can be created or improved by flaking. I have shown this as an arm of the tree structure labelled ‘modify’. As I have modelled it, identification in the upper tier applies across tiers because it is mandatory for the knapper to recognise that platform preparation is necessary before the lower tier actions are carried out. Once the upper tier identification is made, a series of lower tier identifications are triggered that drive the flake removals necessary to modify the platform angle.

The complex unit can be “elaborated” by grinding the platform edge and using a soft hammer, such as a bone, to strike off the flake. I refer to this as the “elaborated flake unit” (Figure 4). Platform abrasion strengthens the surface on which the blow will be landed. A different set of actions – distinct from the basic unit – are necessary to enact platform grinding. For example, Young and Bonnicksen (1984) have shown that platform grinding by the knappers they studied involved two “behaviour variables”: ‘rub’ and ‘shear’. Platform rubbing and shearing are controlled by core rotation, turning, and tilting actions similar to those used in removing flakes, but the guiding identification is very different.

An analogy is to think of flake units as “bricks”. The basic flake unit, then, might be a mud brick, the complex flake might be a low-temperature fired brick, and the elaborated flake unit is, say, a specialty product like an aerated concrete block. In this analogy, the different flake units are equivalent to differences in the *composition* of the bricks. I will return to this analogy below.

Up to this point I have described aspects of the internal organisation of flake units. The next step is to look at different strategies hominins used to combine units, what I call the “architecture” of a reduction sequence.

Application of a flake unit can be visualised as a simple two-step algorithm:

identify high mass → apply the flake unit.

The simplest way of combining flake units is to link them in series, like this:

(identify high mass → apply the flake unit) → (identify high mass → apply the flake unit) → (identify high mass → apply the flake unit)... etc.

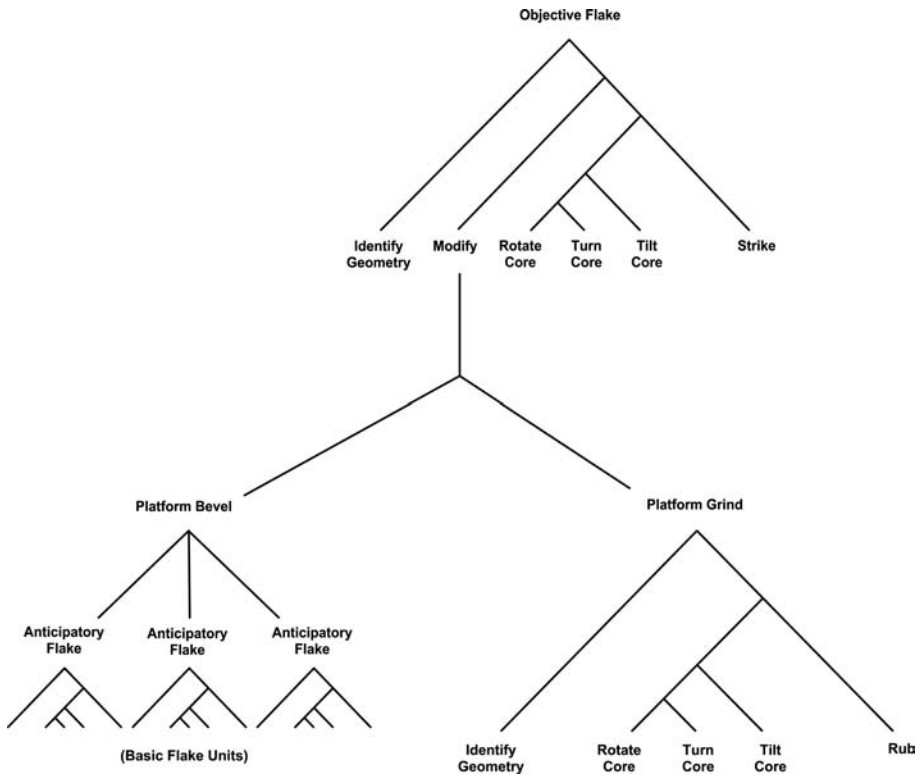


Figure 4. A tree structure model of the elaborated flake unit. Preliminary platform/mass geometrical relationships are modified by two processes in the lower tier. First, platform geometry is adjusted by applying basic flake units, and second, the platform edge is strengthened by grinding. The upper tier involves striking the objective flake using a soft hammer

A simple chain of flake units like this is non-hierarchical (although, as noted above, there may be an internal hierarchical relationship to the elements making up the unit). In each application of the algorithm, the knapping process begins anew, and the “identify high mass” part of the algorithm guides the progressive reduction of the stone. Since the removal of a flake reorganises the distribution of high mass on the core (Figure 5), the reapplication of the algorithm is always to a novel situation.

Note that chaining together flake units does not *require* any planning ahead. The mass on the core face progressively changes in predictable ways with repeated flake removals, but, to chain flake units together, the knapper does not have to make this prediction or even understand that such prediction is possible. A flake unit can be chained to a prior unit even if the knapper believes each flake removal begins with a “new” stone. Thus chaining flake units together can be “mindless”. If units are applied repeatedly to the *same* stone, sets of overlapping scars are created, and the

resulting products can look deceptively like highly-designed tools. This is a particular danger when complex or elaborated flake units were chained together because the process creates so many overlapping flake scars on different core surfaces. The situation is similar to natural selection in biological evolution, where a mindless algorithm results in complex and highly patterned structures (Dennett 1995).

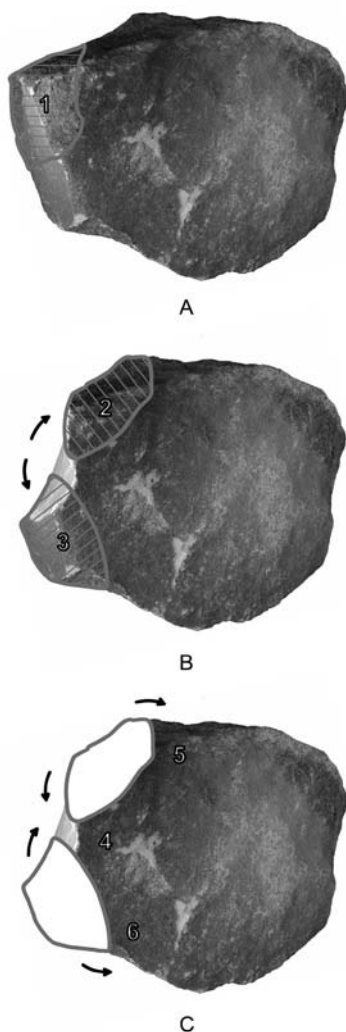


Figure 5. Redistribution of mass on a core as flakes are removed. a) Platform view of high mass in zone 1. b) Removal of zone 1 has isolated high mass in zones 2 and 3. c) Removal of zones 2 and 3 would isolate high mass in zones 4, 5, and 6. The removal of flakes laterally offsets and isolates zones of high mass (arrows)

Returning to the brick analogy, chaining flake units together is equivalent to placing bricks edge-to-edge (Figure 6). Patterns might be achieved by this, but the strategy used to combine the bricks is still based on placing them adjacent to one another. We can refer to this as “footpath” architecture. “Building” architecture involves stacking bricks vertically. An upper row cannot be placed until a lower row is constructed. In knapping, this means applying flake units early in the tool-making process so that a specific arrangement of core mass occurs later in reduction. Stone removal is done in sequential stages to achieve a particular result.

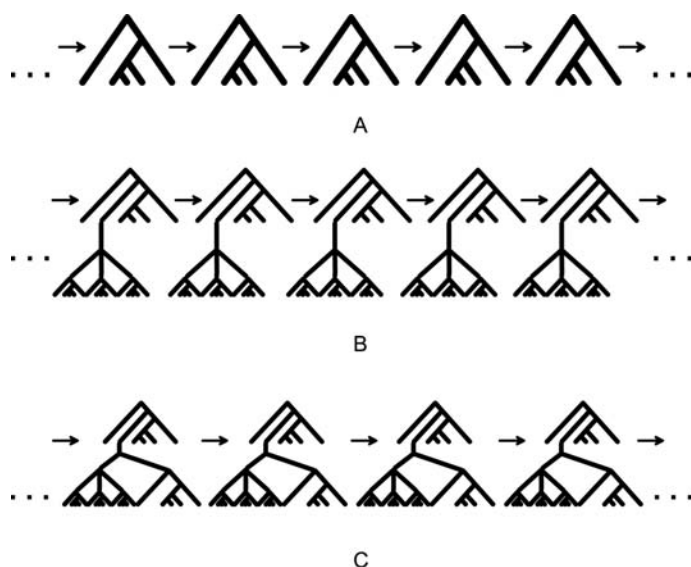


Figure 6. Simple chains of a) basic flake units, b) complex flake units, and c) elaborated flake units

For instance, Aborigines at Camooweal, Queensland, used a stage-based method to produce triangular pointed flakes for use as spearpoints and fighting knives (Figure 7). This involved preparing the core face and the platform surface right at the outset of core reduction and knapping in relation to these prepared surfaces (Moore 2003). In this case, basic flake units were combined hierarchically to create the mass relationship necessary to add on more basic flake units to detach the desired macroblades. Much as buildings are constructed by different sorts of brick bonds, there are myriads of ways in which flake units can be hierarchically arranged.

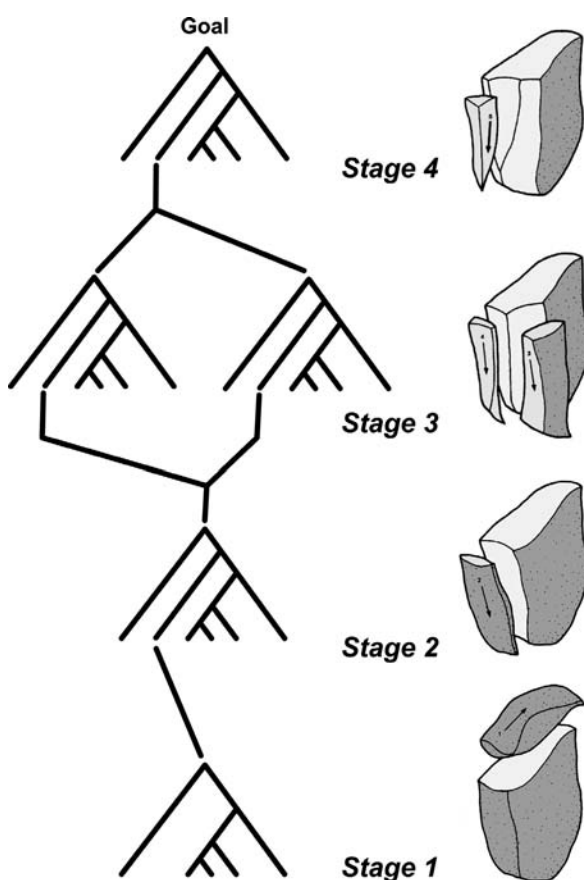


Figure 7. Blade manufacture by Aborigines in northwest Queensland, Australia (after Moore 2003). Basic flake units were arranged hierarchically to achieve the goal of producing pointed blades. The core platform was prepared in Stage 1, the core face was prepared in Stages 2 and 3, and the pointed blade was struck in Stage 4. Subsequent reduction involved repeating Stages 3 and 4

3 Lithic design space and cognition

Patricia Greenfield and others have correlated the development of motor skills and language in human infants and compared this to motor skills demonstrated by other primates (Conway & Christiansen 2001; Frigaszy et al. 2002; Greenfield 1991, 1998; Greenfield & Schneider 1977; Greenfield, Nelson & Saltzman et al. 1972; Johnson-Pynn et al. 1999; Piñon & Greenfield 1994). Specifically, Greenfield (1991) documented changes in motor strategies used by human infants to

combine nesting cups and compared this to the way infants combine sounds.¹ Greenfield concluded that infants mostly used a “pairing” strategy of combining both cups and sounds, with a hierarchical strategy emerging at about 20 months. Changes in combination strategy are proposed by Greenfield to reflect a gradual differentiation in Broca’s area and the creation of networks to the anterior parts of the prefrontal cortex (Greenfield 1998: 60). Adult chimpanzees and other primates do not use a habitual hierarchical strategy in combining motor skills (Greenfield 1991: 545; see also Conway & Christiansen 2001; Gibson 1990: 98; Johnson-Pynn et al. 1999). Greenfield believes that this means the brains of these primates are undifferentiated, more like the brains of modern human infants. The implication is that brain differentiation arose after our evolutionary divergence from chimpanzees (Greenfield 1991).

Greenfield’s correlation of brain evolution with linguistic and motor action strategies offers one way that the design space model can inform research into cognitive evolution. Of course, Greenfield and other researchers gave primates cups to combine, whereas knappers created flake units before they combined them. The lithic design space model tracks combination strategies at both of the levels discussed previously – the internal flake unit level and the flake unit architectural level – and hierarchical thinking occurs at both levels. At the internal level, knappers recognised a complex geometrical relationship of hierarchically-arranged surfaces. For the single-tier basic flake unit, acting on this relationship did not require a hierarchical arrangement of motor actions. However, to build complex and elaborated flake units with two tiers, the knapper visualised a complex geometrical relationship that only partially existed, and then acted on this visualisation by applying hierarchically-arranged motor actions. Importantly, the complexity of the strategy used to build a flake unit was not necessarily reflected in the complexity of the architectural strategy used to combine flake units with each other. As we have seen in the Camooweal example, simple flake units could be combined hierarchically. Conversely, elaborated flake units could be combined serially.

Brain differentiation leads to the phenomenon of “automaticity”. Automaticity refers to behaviour that is unconscious and routinised (Givón 1998: 154–155). Greenfield notes that “as skill increases, the lower levels of [hierarchical] organisation became automatic and conscious attention came to be addressed to the higher levels” (1998: 160). This is the phenomenon seen where students learning touch-typing first focus their attention on individual letters, then, as letters become

1. MacNeilage (1991), Swann (1998: 156), and Tomasello (1991) argued that a better homologue to motor development is found in word combinations, rather than phoneme combinations, because early words are not produced by combining phonemes. Greenfield later accepted this argument (1998: 161).

automatic, the focus of conscious attention shifts to words, and as words become automatic, the focus shifts to sentences or thoughts. This is reflected in brain development as Broca's area comes under the control of the left prefrontal region, resulting in the automaticity of grammar (Greenfield 1998: 159–160). My observations indicate that student knappers achieve automaticity in enacting flake units relatively early in the learning process. Knapping “expertise” comes from mastering the hierarchical combination of flake units to achieve certain effects (cf. Callahan 1979: 37–38; Whittaker 2004).² Applying such expertise requires active, conscious attention and makes greater demands on cognitive powers than automatic actions (Savage-Rumbaugh & Lewin 1994: 240).

4 Lithic design space and the standard story of stone technology

Stone technology emerged with the basic flake unit fully mastered (Delagnes & Roche 2005: 466–467). This means that early hominins could accurately recognise the geometrical relationship component of the unit and could articulate their motor actions to act upon this relationship. As we have seen, hierarchical thinking is inherent in applying the basic flake unit. Architecturally, however, the stone tools made from 2.6 million years ago to about 500,000 years ago could have been produced by chaining together basic flake units. That is, footpath architecture dominated the way that our earliest ancestors combined flake units, and these units were always basic ones. However, much as a footpath might be a simple line of bricks or might consist of geometrical patterns, it is possible that early knappers may have achieved different tool outlines by simple chaining (cf. Wynn 1989). This is a thorny issue because, as Davidson (2002) contends, the outlines archaeologists infer from tools might in fact be a product of the way archaeologists sort artifacts and not evidence of hominin “intent”. The important point here is that if “form” was in fact imposed on stones at this early date (cf. Mellars 1996: 25), it was done by chaining together basic flake units.

Something changed by about 500,000 years ago. We see for the first time evidence for innovation in flake units. This takes two forms. First, the complex flake unit appeared; hominins acted on the realisation that they could modify the geometrical relationship on a core by anticipatory flaking. And second, an elaborated flake unit emerged; hominins started grinding platforms and using soft percussors

2. This highlights the key difficulty of using experiences of modern knappers to infer aspects of early hominin cognition. Since it is not possible for modern humans to divorce themselves from hierarchical thinking and doing, modern knappers cannot simulate the non-hierarchical flake unit combinations practiced by early hominins.

such as bone (Bordaz 1971: 24–27; Bordes 1968: 51–52; e.g. Wenban-Smith 1989).³ In our analogy, hominin knappers invented more sophisticated bricks. But despite these elaborations, knappers still placed the bricks edge-to-edge according to simple footpath architecture.

Hierarchical combination of units in stone toolmaking – the staging of flake removals – does not occur until perhaps 70,000 BP (e.g. Wurz 2002), although there may be a few somewhat earlier contenders (Bar-Yosef & Kuhn 1999). We do not commonly see hierarchical architecture in the archaeological record until after 40,000 BP.⁴

5 The stone toolkit of *Homo floresiensis*

A necessary starting point for discussing the lithic technology of *H. floresiensis* is to address the question of whether this hominin in fact made the stone artifacts at Liang Bua. The evidence correlating stone artifacts with their hominin maker is always circumstantial, and evaluating this evidence is complicated when two or more species lived in a region. Nevertheless, on present evidence, the case for a *H. floresiensis* toolmaker at Liang Bua is about as good as it gets. Liang Bua contains up to 4 metres of cave deposits containing abundant flaking debris associated exclusively with the bones of *H. floresiensis*. There are no modern human remains in these lower deposits. *H. floresiensis* disappears from the deposits about 11,000 years ago, correlating with a thick ashfall in some parts of the cave and a discontinuity in the stratigraphic sequence in other parts of the cave. Modern human skeletons appear in the deposits for the first time above the discontinuity (Morwood et al. 2004, 2005). The most parsimonious interpretation of the physical evidence – a similar argument underpins assessments of hominin toolmakers in Africa and elsewhere – is that the stones were flaked by the hominin found in association with them, in this case *H. floresiensis* (Reynolds 2004).

The methods and results of the Liang Bua lithic analysis are reported in detail elsewhere (Moore 2005; Moore & Brumm 2007; Moore et al. Forthcoming). What follows is a brief description of the technology.

3. Flake struck with a soft hammer are shaped differently than flakes struck with a hard hammer (Hayden & Hutchings 1989; Pelcin 1997).

4. The Levallois method is traditionally interpreted as a hierarchical method of combining flake units to make stone flakes of predetermined size and shape (e.g. Dibble & Bar-Yosef 1995). However, I believe the Levallois method is a chain of complex flake units that mimics highly designed tools (Moore 2005: 226–230; see also Sandgathe 2004).

H. floresiensis sourced its stone from water-rolled cobble deposits of the nearby Rae Racang River or its terraces. Stones selected for knapping consisted of volcanic, tuff, or a marine chert of moderate to high flaking quality. *H. floresiensis* reduced large cobbles while at the stone source into flakes measuring on average about 40–50 mm in maximum dimension or slightly larger. The hominins selected some of these flakes to carry to the cave, along with small hammerstones of coarse, tough volcanic material. We call the transported flakes “flake blanks”.

Flake blanks were reduced into smaller flakes within the cave. Most reduction was done by a combination of three knapping techniques within a “small flake production loop”. The most common technique was freehand percussion, with flakes driven across the flattish surfaces of the flake blanks (Figure 8).

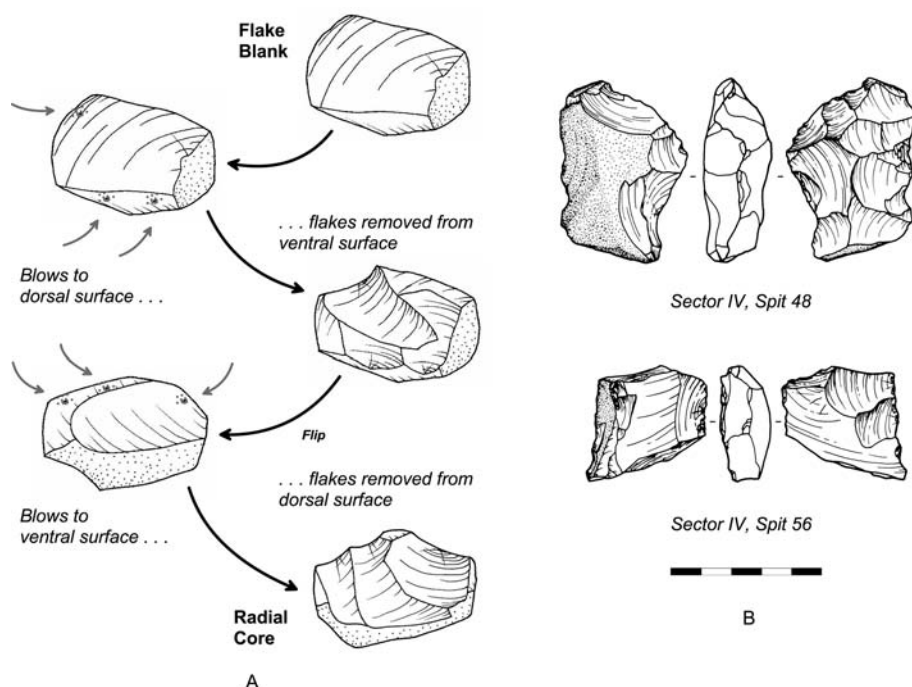


Figure 8. Freehand planar percussion by *Homo floresiensis*. a) Schematic model showing how flakes were struck from flake blanks by delivering blows roughly parallel to the plane defined by flake blank length and width. b) Flake blanks from Liang Bua Cave reduced by freehand percussion. Scale 50 mm

Flake removals were usually to both the dorsal and ventral surfaces of a flake blank, resulting in a “bifacial radial core”. These cores are bifacial in the sense that flakes were removed from both faces, and radial (or, more accurately “centripetal”) in the

sense that flakes were driven off various parts of the blank's perimeter and tend to be oriented towards the centre. Flakes were struck invasively, extending past the centre line of the blank, or non-invasively, with scars limited to the blank's margins. Sometimes blanks were rotated in such a way that flakes could be struck from the edges, known as the "burin" technique (Figure 9).

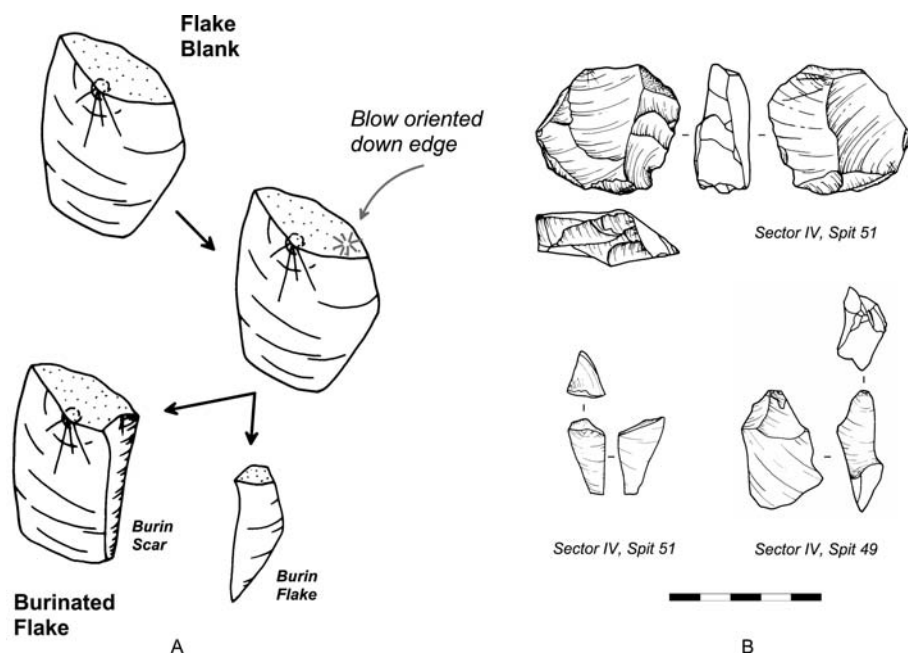


Figure 9. Burin percussion by *Homo floresiensis*. a) Schematic model showing how flakes were struck down the edges of flake blanks. b) Products of the burin technique from Liang Bua Cave. Top: core with burin scars on the distal edge. Bottom: burin flakes struck from the edges of flake blanks. Scale 50 mm

The "cross-axis truncation" technique involved placing the flake blank on an anvil and striking near the middle of the face, truncating the blank (Figure 10). The freehand technique dominated the reduction pattern, but switches to other techniques occurred frequently. Switches between techniques and/or before flipping the stone tended to be preceded by 1 to 3 blows. Knapping by *H. floresiensis* was usually a rather staccato event with short series of blows followed by major shifts in knapping gesture.

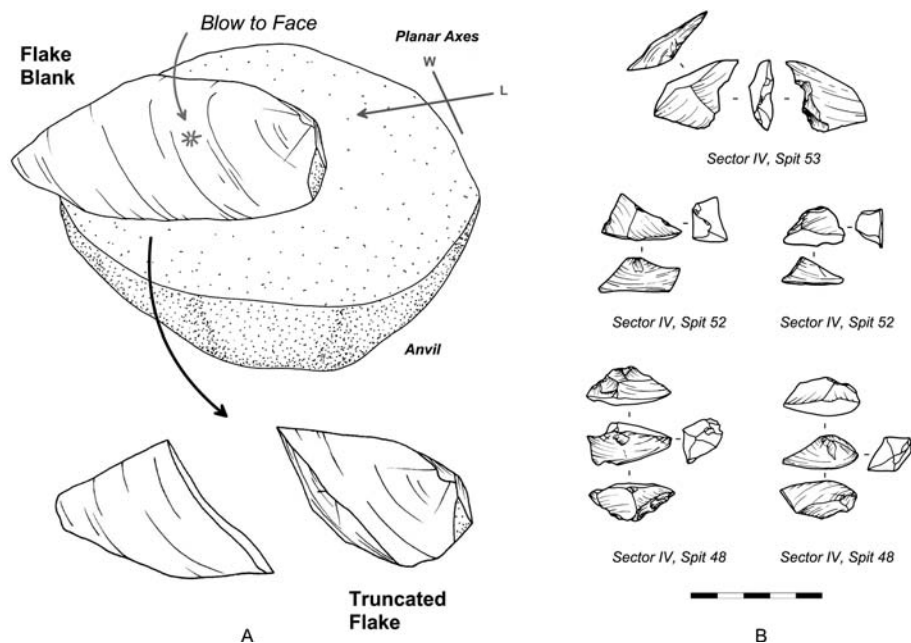


Figure 10. Cross-axis truncation by *Homo floresiensis*. a) Schematic model showing how truncation was achieved by delivering a blow to the face and breaking the core perpendicular to the plane defined by length and width. b) Products of the cross-axis truncation technique from Liang Bua Cave. Scale 50 mm

In rare cases *H. floresiensis* delivered up to 39 blows in a series without changing technique or flipping the stone. This is how the hominin made the tools we call “perforators” (Figure 11). Perforators are flake blanks that were retouched on opposite margins, creating a parallel-sided projection. Despite their apparent distinctiveness, we must be careful not to fall into the trap of imposing our own ideas of form onto these objects. For instance, other parallel-sided retouched tools at Liang Bua fail to resemble the “classic” perforators (see Figure 11). Classic perforators may be at the extreme end of continuous variation. Nevertheless, the way that these tools were produced is different from most reduction that occurred at Liang Bua.

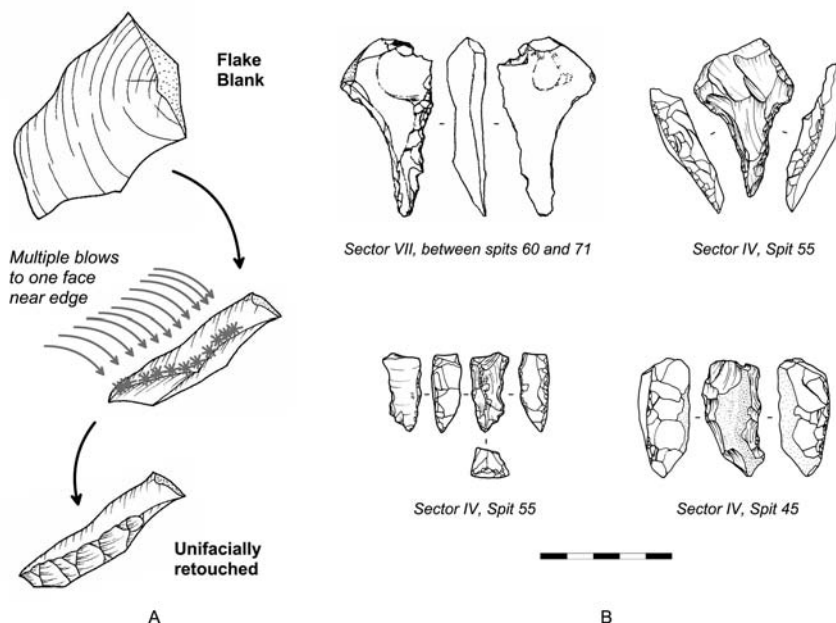


Figure 11. Unifacial freehand percussion by *Homo floresiensis*. a) Schematic model showing how one surface of a flake blank was retouched by multiple blows. b) Unifacially retouched flakes from Liang Bua Cave. Top and bottom left: “perforators”. Bottom right: flake retouched on parallel margins. Scale 50 mm

H. floresiensis also reduced flake blanks by the bipolar technique (Figure 12). In this technique, the flake is held edge-on onto an anvil and a blow is delivered straight down onto the uppermost edge. This causes the faces of the flake to shear away, producing many small flakes and shatter fragments. Bipolar reduction was applied separately from the small flake production loop, and, in some cases, small flakes from the loop served as blanks for applying the bipolar technique.

It appears that the purpose of the small flake production loop was two-fold: first, to produce sharp-edged flakes, and second, to produce strong, retouched edges. Radial cores probably served as both strong-edged tools and as a source for sharp-edged flakes. The perforators might imply that *H. floresiensis* sometimes made strong edges in a specific configuration. It seems likely that the bipolar technique was another approach to making sharp-edged tools.

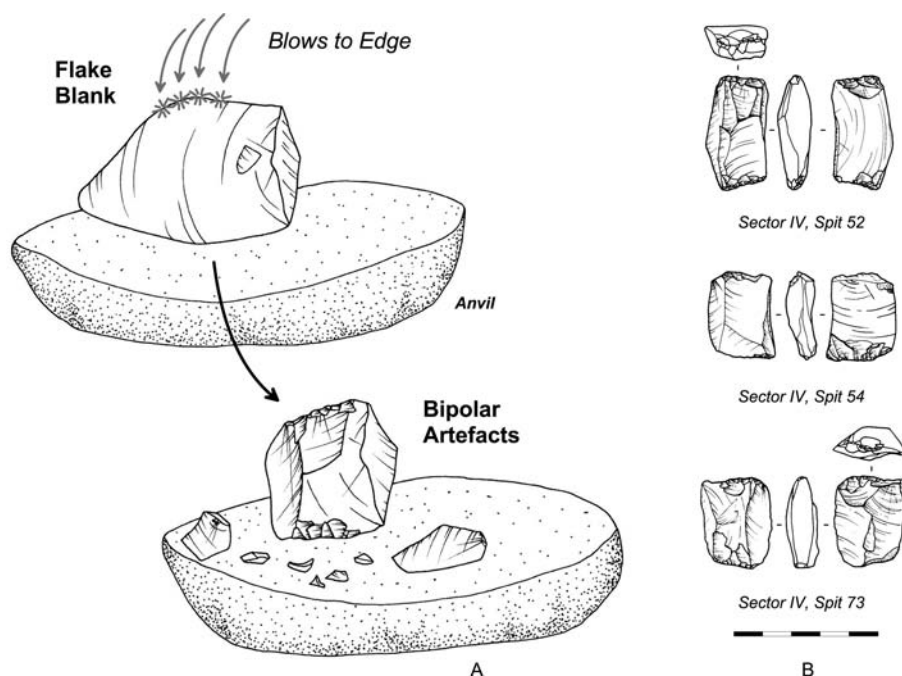


Figure 12. Planar bipolar percussion by *Homo floresiensis*. a) Schematic model showing how flakes were placed edge-on onto an anvil and the uppermost edge was struck, creating small flakes and shatter. b) Flakes from Liang Bua Cave reduced by bipolar percussion. Scale 50 mm

Blades – flakes that are twice as long as they are wide – occur in direct association with *H. floresiensis*. This has been viewed as evidence for modern human-like behaviour among *H. floresiensis* (Lahr & Foley 2004; Martin et al. 2006). Modern humans from the Late Pleistocene onward developed sophisticated methods of mass-producing blades based on the hierarchical stacking of flake units. Extrapolating backwards from these late assemblages, archaeologists have asserted that blades themselves are a good proxy measure of advanced, modern human-like cognition (Bar-Yosef & Kuhn 1999). However, blades can also be produced incidentally during virtually any knapping technique if the mass on the core was elongated. I use an analogy with cheese cutting to demonstrate this (Figure 13).

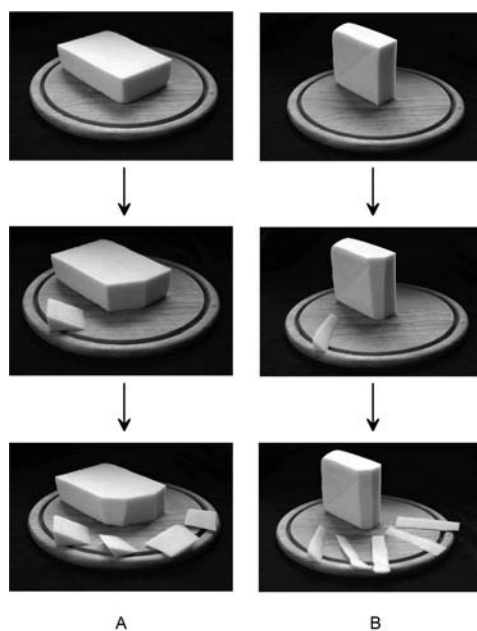


Figure 13. “Cheeseboarding” as a heuristic tool. a) Short, stumpy cheese slices created by cutting down the short axis of the block. b) Cheese “blades” created by cutting down the long axis of the block

If pieces are cut from a block of cheese through the short axis of the block, stubby, non-blade shapes are produced. However, if cheese strips are cut in the same way but from the long axis of the block, cheese “blades” are produced. The presence of blades in an assemblage *can* indicate higher-level cognition, but the way you determine this is by looking at the way knappers arranged flake units to produce them. If flake units were arranged hierarchically, a strong case can be made for advanced cognition. If not, the most parsimonious explanation is that blades were created fortuitously due to the morphology of the high mass targeted by the flake unit. Blades compose a small proportion of the assemblage at Liang Bua – about 7% – and there is no evidence that they were intentionally produced by stacking flake units hierarchically (Moore et al. Forthcoming).

6 Cognition in *Homo floresiensis*

Where does the stone technology of *H. floresiensis* fit within the lithic design space model? *H. floresiensis* was an expert at striking flakes from stone. The hom-

inin was able to recognise the hierarchical geometrical relationship necessary for freehand flaking and was capable of the gestures necessary to act on this geometrical relationship.

There is no sign of an elaboration of the basic flake unit in *H. floresiensis*'s knapping techniques. Specifically, there is no evidence that *H. floresiensis* manipulated platform angles on cores by anticipatory flaking, and there is no evidence for platform grinding prior to striking off flakes (Moore et al. Forthcoming). *H. floresiensis* achieved all of its effects in stone by applying the basic flake unit.

H. floresiensis combined flake units in chains: the hominin practiced “foot-path” architecture in its stone flaking technology. The “perforators” may be important in this context because if they are evidence for a deliberately imposed form, this indicates an ability to produce relatively complex patterns in the way units were arranged. There is no evidence in the Liang Bua assemblage for the hierarchical combination of flake units. Evidence for hierarchical organisation does not occur in the vicinity of Liang Bua until the Neolithic in the context of adze manufacture (Figure 14). These tools were made by modern humans and were sometimes interred with burials on Flores (Verhoeven 1952).

This assessment places the Liang Bua assemblage squarely at the Oldowan grade of technological complexity (Moore & Brumm In press). The stone artifact evidence suggests a minimum level of cognitive ability similar to the earliest stone knapping hominins in Africa. This observation is consistent with an emerging view that *H. floresiensis* descended from an early hominin migration out of Africa, one involving a different species than *Homo erectus* (Argue et al. 2006; Falk et al. 2005: 244; Groves 2004; Morwood et al. 2005). If so, *H. floresiensis* carried its technology with it out of Africa and changed it little during the subsequent millennia of its evolutionary history.

In conclusion, *H. floresiensis*'s stone technology does not reflect an increase in cognitive ability as a response to selective pressure. Internal elaboration of the flake unit did not occur at Liang Bua like it did at early African *Homo* sites. Conversely, the stone flaking technology of *H. floresiensis* does not show a *decrease* in technological ability either (Moore et al. Forthcoming), and this is significant. It suggests two things. First, stone technology was a key element of the adaptation; selective changes to brain size, body size, body morphology, and behaviour worked around the persistence of stone technology. And second – perhaps paradoxically in light of the first point – a technological adaptation to the environment was not an effective shield against the selective pressures of island endemism.

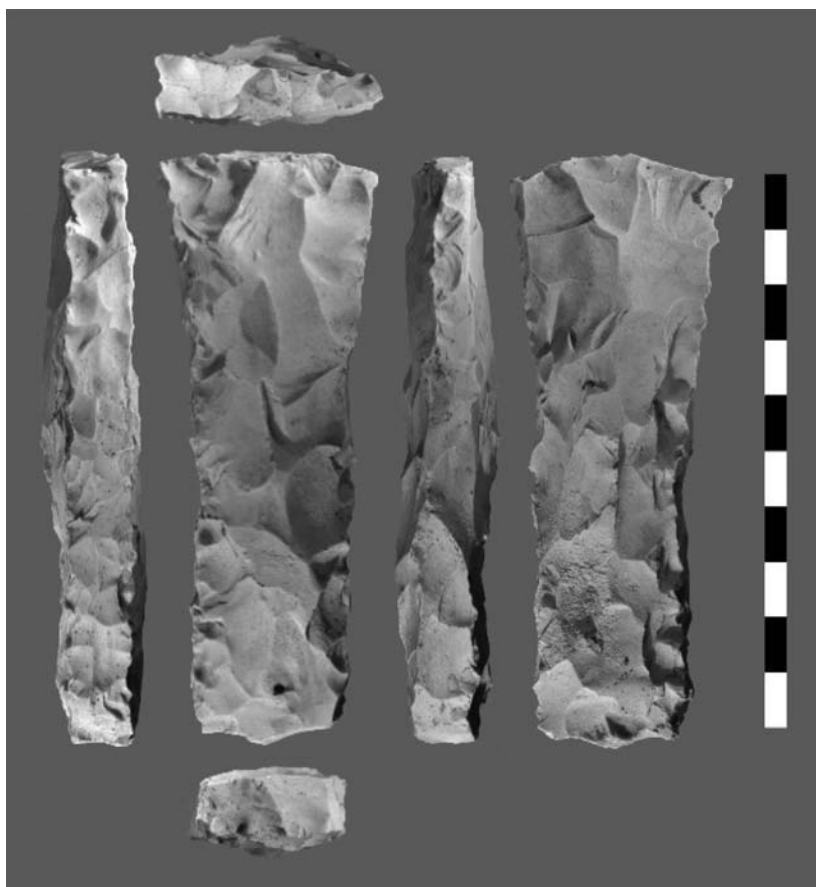


Figure 14. Square-sectioned chert adze blank from Golo Roang, a Neolithic site situated across the river from Liang Bua. Scale 100 mm

Acknowledgements

I thank Dorothea Cogill-Koez and members of the Language and Cognition Research Centre at the University of New England for the invitation to present this work. The ideas expressed in this chapter benefited enormously from conversations with Iain Davidson. Funding for the Liang Bua excavations was provided by an Australian Research Council (ARC) grant to Michael J. Morwood. The 2003–2004 excavations at Liang Bua Cave were conducted by the Indonesian National Centre for Archaeology (ARKENAS), with Radien P. Soejono (counterpart), Michael J. Morwood (counterpart), Thomas Sutikna, E. Wahyu Saptomo, Jatmiko, Sri Wasisto, Rokus Awe Due and Douglas R. Hobbs. The Liang Bua work was authorised by Dr Tony Djubiantono (ARKENAS Director). The research presented here was funded by an ARC Australian Postgraduate Award.

References

- Argue, Debbie, Denise Donlon, Colin Groves & Richard Wright. 2006. *Homo floresiensis*: Microcephalic, pygmoid, *Australopithecus*, or *Homo*? *Journal of Human Evolution* 51(4). 360–374.
- Bar-Yosef, Ofer & Steven L. Kuhn. 1999. The big deal about blades: Laminar technologies and human evolution. *American Anthropologist* 101(2). 322–338.
- Bordaz, Jacques. 1971. *Tools of the old and new Stone Age*. Newton Abbot: David and Charles.
- Bordes, François. 1968. *The old Stone Age*. New York, NY: McGraw-Hill.
- Brown, Peter, Thomas Sutikna, Michael J. Morwood, Radien P. Soejono, Jatmiko, E. Wayhu Saptomo & Rokus Awe Due. 2004. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431. 1055–1061.
- Brumm, Adam, Fachroel Aziz, Gert D. van den Bergh, Michael J. Morwood, Mark W. Moore, Iwan Kurniawan, Douglas R. Hobbs & Richard Fullagar. 2006. Early stone technology on Flores and its implications for *Homo floresiensis*. *Nature* 441. 624–628.
- Callahan, Errett. 1979. The basics of biface knapping in the eastern fluted point tradition: A manual for flintknappers and lithic analysts. *Archaeology of Eastern North America* 7. 1–180.
- Conway, Christopher M. & Morten H. Christiansen. 2001. Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5(12). 539–546.
- Crabtree, Don E. 1972. *An introduction to flintworking* [Occasional Papers of the Idaho State University Museum 28]. Pocatello, ID: Idaho State University.
- Davidson, Iain. 2002. The finished artefact fallacy: Acheulean hand-axes and language origins. In Alison Wray (ed.), *The transition to language*, 180–203. Oxford: OUP.
- Delagnes, Anne & Hélène Roche. 2005. Late Pliocene hominid knapping skills: The case of Lokalalei 2C, West Turkana, Kenya. *Journal of Human Evolution* 48. 435–472.
- Dennett, Daniel C. 1995. *Darwin's dangerous idea: Evolution and the meaning of life*. New York, NY: Simon and Schuster.
- Dibble, Harold L. & Ofer Bar-Yosef (eds.). 1995. *The definition and interpretation of Levallois technology* [Monographs in World Archaeology 23]. Madison, MI: Prehistory Press.
- Falk, Dean, Charles Hildebolt, Kirk Smith, Michael J. Morwood, Thomas Sutikna, Peter Brown, Jatmiko, E. Wahyu Saptomo, Barry Brunnsden & Fred Prior. 2005. The brain of LB1, *Homo floresiensis*. *Science* 308. 242–245.
- Fragaszy, Dorothy M., Amy T. Galloway, Julie Johnson-Pynn & Karen Brakke. 2002. The sources of skill in seriating cups in children, monkeys and apes. *Developmental Science* 5(1). 118–131.
- Gibson, Kathleen R. 1990. New perspectives on instincts and intelligence: Brain size and the emergence of hierarchical mental constructional skills. In Sue Taylor Parker & Kathleen Rita Gibson (eds.), *'Language' and intelligence in monkeys and apes: Comparative developmental perspectives*, 97–128. Cambridge: CUP.
- Givón, Talmy. 1998. Toward a neurology of grammar. *Behavioral and Brain Sciences* 21(1). 154–155.
- Greenfield, Patricia M. 1991. Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences* 14(4). 531–596.
- Greenfield, Patricia M. 1998. Language, tools, and brain revisited. *Behavioral and Brain Sciences* 21(1). 159–163.
- Greenfield, Patricia Marks & Leslie Schneider. 1977. Building a tree structure: The development of hierarchical complexity and interrupted strategies in children's construction activity. *Developmental Psychology* 13(4). 299–313.

- Greenfield, Patricia Marks, Karen Nelson & Elliot Saltzman. 1972. The development of rule-bound strategies for manipulating seriated cups: A parallel between action and grammar. *Cognitive Psychology* 3. 291–310.
- Groves, Colin. 2004. Some initial reactions to the publication of the discovery of *Homo floresiensis*. *Before Farming* 4(4). 2.
- Hayden, Brian & W. Karl Hutchings. 1989. Whither the billet flake. In Daniel S. Amick & Raymond P. Mauldin (eds.), *Experiments in lithic technology* [BAR International Series 528], 235–257. Oxford: British Archaeology Reports.
- Jacob, Teuku, Etty Indriati, Radien P. Soejono, Kenneth Hsü, David W. Frayer, Robert B. Eckhardt, Adam J. Kuperavage, Alan Thorne & Maciej Henneberg. 2006. Pygmoid Australo-melanesian *Homo sapiens* skeletal remains from Liang Bua, Flores: Population affinities and pathological abnormalities. *Proceedings of the National Academy of Sciences* (PNAS) 103(36). 13421–13426.
- Johnson-Pynn, Julie, Dorothy M. Frigaszy, Elizabeth Hirst, Karen E. Brakke & Patricia M. Greenfield. 1999. Strategies used to combine seriated cups by chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and capuchins (*Cebus apella*). *Journal of Comparative Psychology* 113(2). 137–148.
- Kohn, Marek. 2005. The little troublemaker. *New Scientist* 186. 41–45.
- Lahr, Marta Mirazón & Robert Foley. 2004. Human evolution writ small. *Nature* 431. 1043–1044.
- MacNeilage, Peter F. 1991. Linguistic and manual evolution. *Behavioral and Brain Sciences* 14(4). 568–569.
- Martin, Robert D., Ann M. MacLarnon, James L. Phillips & William B. Dobyns. 2006. Flores hominid: New species or microcephalic dwarf? *The Anatomical Record Part A* 288A(11). 1123–1145.
- Mellars, Paul. 1996. Symbolism, language, and the Neanderthal mind. In Paul Mellars & Kathleen Gibson (eds.), *Modelling the early human mind*, 15–32. Cambridge: McDonald Institute for Archaeological Research Monographs.
- Moore, Mark W. 2003. Australian Aboriginal blade production methods on the Georgina River, Camooweal, Queensland. *Lithic Technology* 28. 35–63.
- Moore, Mark W. 2005. The design space of lithic technology. Armidale, NSW: University of New England PhD thesis, Department of Archaeology and Palaeoanthropology.
- Moore, Mark W. & Adam Brumm. 2007. Stone artifacts and hominins in island Southeast Asia: New insights from Flores, Eastern Indonesia. *Journal of Human Evolution* 52. 85–102.
- Moore, Mark W. & Adam Brumm. In press. *Homo floresiensis* and the African Oldowan. In Erella Hovers (ed.), *Interdisciplinary approaches to the Oldowan*. New York, NY: Springer Science and Business Media.
- Moore, Mark W., Thomas Sutikna, Jatmiko, Michael J. Morwood & Adam Brumm. Forthcoming. Continuities in stone flaking technology at Liang Bua, Flores, Indonesia. *Journal of Human Evolution*.
- Morwood, Michael J., Peter Brown, Jatmiko, Thomas Sutikna, E. Wahyu Saptomo, Kira E. Westaway, Rokus Awe Due, Richard G. Roberts, T. Maeda, Sri Wasisto, and Tony Djubiantono. 2005. Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* 437. 1012–1017.
- Morwood, Michael J., Radien P. Soejono, Richard G. Roberts, Thomas Sutikna, Chris S. M. Turney, Kira E. Westaway, William J. Rink, Jian-Xin Zhao, Gert D. van den Bergh, Rokus Awe

- Due, Douglas R. Hobbs, Mark W. Moore, Michael I. Bird & L. Keith Fifield. 2004. Archaeology and age of a new hominin from Flores in Eastern Indonesia. *Nature* 431. 1087–1091.
- Pelcin, Andrew W. 1997. The effect of indenter type on flake attributes: Evidence from a controlled experiment. *Journal of Archaeological Science* 24. 613–621.
- Pelegrin, Jacques. 1993. A framework for analysing prehistoric stone tool manufacture and a tentative application to some early stone industries. In Arlette Berthelet & Jean Chavaillon (eds.), *The use of tools by human and non-human primates*, 302–314. Oxford: Clarendon Press.
- Piñon, Denise & Patricia M. Greenfield. 1994. Does everybody do it? Hierarchically organized sequential activity in robots, birds, and monkeys. *Behavioral and Brain Sciences* 17(2). 361–365.
- Reynolds, Tim. 2004. Some initial reactions to the publication of the discovery of *Homo floresiensis*. *Before Farming* 4(4). 4–5.
- Rightmire, Philip. 2004. Brain size and encephalization in early to mid-Pleistocene *Homo*. *American Journal of Physical Anthropology* 124. 109–123.
- Sandgathe, Dennis M. 2004. Alternative interpretation of the Levallois reduction technique. *Lithic Technology* 29(2). 147–159.
- Savage-Rumbaugh, Sue & Roger Lewin. 1994. *Kanzi: The ape at the brink of the human mind*. New York, NY: John Wiley and Sons.
- Schick, Kathy D. & Nicholas Toth. 1993. *Making silent stone speak: Human evolution and the dawn of technology*. London: Phoenix.
- Stout, Dietrich, Nicholas Toth & Kathy Schick. 2000. Stone tool-making and brain activation: Position Emission Tomography (PET) studies. *Journal of Archaeological Research* 27. 1215–1223.
- Swann, Philip. 1998. Greenfield on language, tools, and brain. *Behavioral and Brain Sciences* 21(1). 155–159.
- Tomasello, Michael. 1991. Objects are analogous to words, not phonemes or grammatical categories. *Behavioral and Brain Sciences* 14(4). 575–575.
- Toth, Nicholas. 1985. Archaeological evidence for preferential right-handedness in the lower and middle Pleistocene, and its possible implications. *Journal of Human Evolution* 14(6). 607–614.
- Verhoeven, Theodore. 1952. Stenen Werktuigen uit Flores (Indonesië). *Anthropos* 47. 95–98.
- Wenban-Smith, Francis F. 1989. The use of canonical variates for determination of biface manufacturing technology at Boxgrove Lower Palaeolithic site and the behavioural implications of this technology. *Journal of Archaeological Science* 16. 17–26.
- Whittaker, John C. 2004. *American flintknappers: Stone age art in the age of computers*. Austin, TX: University of Texas Press.
- Wong, Kate. 2005. The littlest human. *Scientific American* 292(2). 56–65.
- Wurz, Sarah. 2002. Variability in the middle Stone Age Lithic sequence, 115,000–60,000 years ago at Klasies River, South Africa. *Journal of Archaeological Science* 29. 1001–1015.
- Wynn, Thomas. 1989. *The evolution of spatial competence*. Urbana, IL: University of Illinois Press.
- Young, David E. & Robson Bonnichsen. 1984. *Understanding stone tools: A cognitive approach* [Peopling of the Americas Process Series 1]. Orono, ME: Center for the Study of Early Man, University of Maine.

CHAPTER 3

“As large as you need and as small as you can”

Implications of the brain size of *Homo floresiensis*

Iain Davidson

In this chapter I discuss the brain and body size of *Homo floresiensis* in light of the interpretation of changes in both quantities for other hominins from three million years ago to the present. Considering selection against both large brains and against small brains emphasises the likely role of cognition in the evolution of brain size.

1 The pattern of brain size and body size changes among hominins

The general pattern of brain size changes during the course of hominin and human evolution is well known and variously interpreted (Aiello & Dunbar 1993; Dunbar 1992, 1993; Holloway 1983; Ruff, Trinkaus & Holliday 1997; Trinkaus & Ruff 1999).¹ Some scholars calculate statistics on the distribution so that an onwards and upwards trend can be shown (de Miguel & Henneberg 2001; Lee & Wolpoff 2003). But as Aiello and Dunbar (1993) show, the trend is not simply onwards and upwards. The continued appeal to an onward and upward model (e.g. Lewin & Foley 2004: 450–451) relies on the simplification of calculating averages for particular species and assigning average dates for these species. This, in turn, depends on the assignment of specimens to species.

De Miguel and Henneberg (2001) recently assessed the variation in estimates of cranial capacity (but not of dating). We can use these data to plot cranial capacity size for individual specimens against time and show that the pattern is not simply onwards and upwards (Figure 1) but punctuated, and that the assignments of specimens to species may not be the most appropriate way to look at the data.

1. Aiello and Dunbar (1993), for example, would use the estimated brain size to imply that *Homo floresiensis* lived in groups of about 60 individuals and spent about 15% of their time grooming!

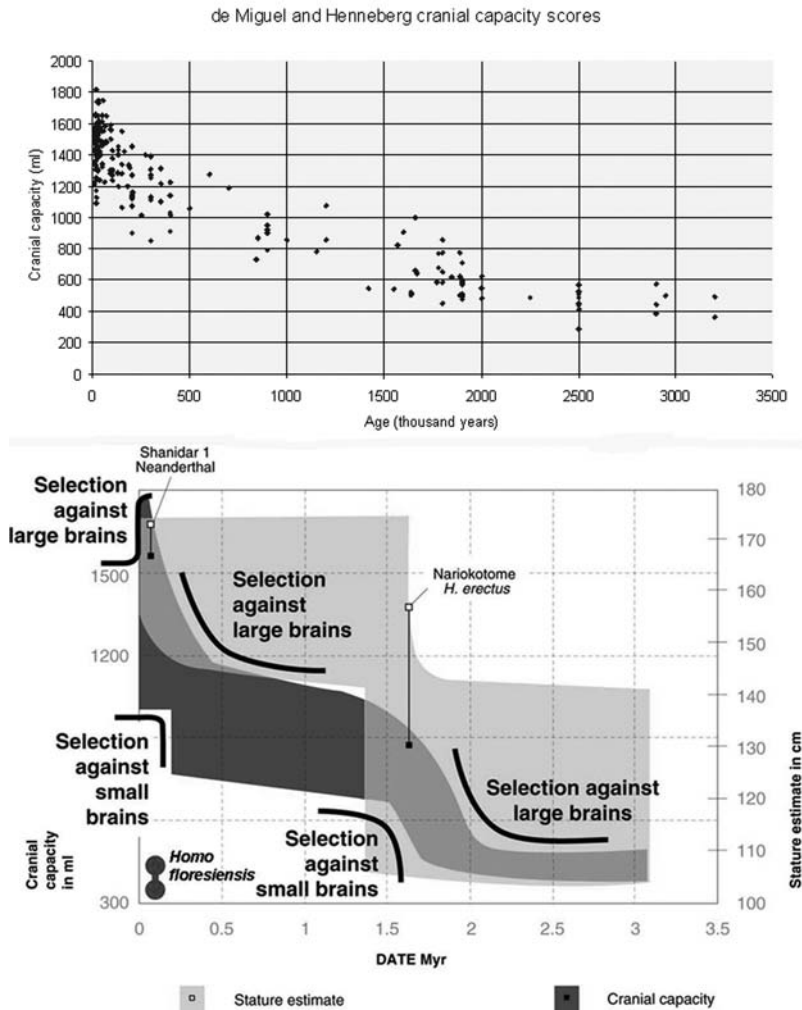


Figure 1. Top, pattern of changing cranial capacity (ml on y-axis) with time (thousand years on x-axis) using data from de Miguel and Henneberg (2001). Bottom, representation of data on cranial capacity against time (million years on x-axis) together with representation of data on stature (cm on right hand y-axis) (references in text), showing the key stages of selection against large or small brain size. Also shown are the estimates of cranial capacity and stature for *Homo floresiensis* (reference in text)

In a previous publication (Davidson 1999b) I summarised the situation in this way. Between 3 and 2 million years ago the range of cranial capacities of hominins was small (about 400–450 ml) and showed little or no change over the period. Between 2 million years and 1.5 million years ago cranial capacity increased both its range of

variation and its size (about 450–950 ml) when all hominins are considered together. At about 1.5 million years the smaller end of the range of variation disappeared,² and if modern scientists name the small-brained hominins as a separate taxon (probably the genus *Australopithecus*) they say that taxon became extinct. The new lowest values were about 650 ml. Between 1.5 million years and about 400 thousand years there was very little change in cranial capacity, though both lower and upper limits increased slightly. At 400 thousand years, the upper limits of hominin cranial capacity increased, reaching a maximum about 50 thousand years ago. About 200 thousand years ago, the smaller end of the range of variation again disappeared (or, if named as a separate species, that species became extinct).

One factor that needs to be taken into account is the interaction between body size and brain size, made more difficult by the fragmentary nature of the fossil record. Because the fossils are rarely anything like whole skeletons, skulls from which cranial capacities (hence brain sizes) may be estimated (and on which species identifications are based) are rarely found with postcranial specimens from which statures or body sizes are generally estimated. Data on stature are available from work by McHenry (1991) and Feldesman and colleagues (Feldesman, Kleckner & Lundy 1990; Feldesman & Lundy 1988). Without assigning specimens to species, and averaging both cranial capacity (as an estimator of brain size) and body size estimates, it is difficult to calculate Encephalisation Quotients as was undertaken in the original publication of *Homo floresiensis* (Brown et al. 2004).

The pattern of changes of stature is slightly different from that of changes in cranial capacity. On the evidence, the range of variation in stature remained fairly constant (at 1.05–1.40m) from before 3 million years ago until *after* the initial increase in range of cranial capacity had begun. About 1.6 million years ago, the range of statures increased quite dramatically (to 1.75m). Since 1.4 million years ago individuals of small stature (below 1.40m) have been rarely found. From 1.4 million to the present, stature has remained relatively constant in range and size (between 1.40–1.75m). Thus, the way in which change happened is similar to that for cranial capacity change: an initial expansion of range followed by the extinction of the lower end of the expanded range of variation. The significant difference, however, is that, before the discovery of *Homo floresiensis*, there was only one major change in stature during the course of hominin evolution.

2. Spoor et al. (2007) recently described a fossil with a small cranial capacity slightly later than 1.5 million years.

2 The selectionist explanation of brain size and body size evolution

In seeking to explain this pattern of changes, I have previously argued that the most important point is to consider selection against aberrant forms or values. Thus, we know that brains are expensive organs to maintain, impose risks to themselves and the mother during childbirth (Foley & Lee 1991) and generate heat dangerously (Falk 1986). There is, therefore, strong selection against large brain size. The brain will tend to be as small as selection can make it, but selection cannot reduce brain size below that needed to allow the animal to function. Hence the title of this chapter “as large as you need and as small as you can”.

For the first stage of increase in range of brain sizes among hominins, there are straightforward contributions towards an explanation for the relaxation of selection by appeal to the increasing evidence of the use of food obtained from animals and of the manufacture of stone tools (probably used in meat acquisition) (Noble & Davidson 1996). The dietary changes, in particular, may also have had an impact on body size. Some support could be found for this pattern of selection in the argument that the second stage of brain size increase corresponded to the improved nutrition that resulted from cooking food (Brace 1995), but Brace’s argument may be weakened by the apparent lack of influence on body size. It remains a possibility that this combination of increased head size with constant body size may be related to broader changes in the pattern of foetal and infant development (Davidson 1999b).

This narrative about increases in brain and body size during the course of hominin and human evolution is more or less traditional. Much less familiar is that it is also necessary to argue why hominins with small head and brain size became extinct. I previously argued that it would be consistent with other arguments Noble and I have developed (for example in Noble & Davidson 1996), if hominins of smaller brain size “succumbed to selective pressure in competition” with hominins “that had expanded brains, enhanced learning and complex sequenced actions” (Davidson 1999b: 259). In this context, the needs of hominins meant that small brains were not good enough for them. Given the lower nutritional requirements for the small-bodied, small-brained hominins, it seems likely that cognitive differences were crucial in such competition.

3 The brain and body of *Homo floresiensis*

Understanding the body size and brain size of *Homo floresiensis* should, I suggest, appeal to the same sorts of mechanisms as have been invoked to explain other patterns of change. Why were small body and brain size good enough for them

and not for late australopithecines? If they cannot be explained by the same mechanisms or processes, then these are probably not an explanation for the other changes either.

I suggest that in the relatively impoverished faunal environment of Flores, selection operated against large body and brain size. This is consistent with the selectionist interpretation. However, the most interesting feature is that the selectionist interpretation here suggests that small body and brain size will only survive if there is no competition from a similar organism. This seems to suggest that we might expect *not* to find evidence for a large bodied hominin contemporary with *Homo floresiensis*, and none has, so far, been found.

This conclusion makes it more unlikely that, after the initial colonisation of Flores by an ancestral hominin, there were many more arrivals. Morwood's work elsewhere in Flores has shown that hominins were making stone tools on the island by 800,000 years ago, but not much before that (Morwood et al. 1999; O'Sullivan et al. 2001). Figure 1 shows that by these dates, there were no hominins (so far found) with either small brains or small body size. This makes it more likely that the first crossings from Bali to Lombok-Sumbawa and from Lombok-Sumbawa to Flores were accidental and not part of a pattern of use of watercraft.³ I have argued previously that deliberate use of watercraft might be argued as an indication of fully modern behaviour, including language use (Davidson & Noble 1992), and why it appears unlikely that the first colonisation of Flores is an indicator of such behaviour (Davidson 2001). Morwood and I (Morwood & Davidson 2005) have suggested that the survival of individuals on rafts of natural vegetation following the tsunami of December 2004 is an indication of the sort of natural accident which would bring hominins across such water barriers.

Homo floresiensis survived on Flores at dates significantly after the first colonisation of Australia by fully modern people using watercraft (Davidson 1999a). There are two implications here: 1) this gives us the opportunity to calibrate the rate at which interaction between hominins of different cognitive abilities leads to the extinction of one of them; 2) the balance of evidence suggests that the first colonists of Australia may not have passed through Flores on their dispersion route that led them to Australia, suggesting that Birdsell's (1977) northern route through Sulawesi and Papua is a more likely path for the first Australians and New Guineans. It is conceivable, on this model, that the first modern humans in Timor (Veth, O'Connor, & Spriggs 2004) went there from Australia.

3. Morwood and van Oosterzee (2007) suggest that the initial accidental colonisation of Flores was more likely from the north.

4 Cognitive implications of the selectionist model for *Homo floresiensis*

Skeletal morphology has always been a poor guide to cognition, not least because of the uncertain relationship between cognitive abilities and brain size or morphology (Skoyles 1999). The cognitive abilities of *Homo floresiensis* can only be judged by the interpretation of its behaviour. This will come from the analysis of the stone tools (see chapter by Moore in this volume) and by evaluation of the suggestions by Morwood and his co-workers (this volume) of the use of fire and of hunting. Such evaluation has proved difficult in the past. The most famous example of early fire use, at Zhoukoudian in China, required detailed analysis before archaeologists could distinguish the role of hominins in the activity (Weiner et al. 1998). Mitchell (1997) has shown that even the recognition of hearths can be a very complex process. The word “hunting” can mean many things, but the mere killing of animals does not make it a fully modern behaviour. Rather what makes hunting a relatively recent activity in hominin and human evolution is the organisation of the activity and the targeting of prey categories (Stiner 1994). It will be very difficult to evaluate the archaeological evidence given that there is no modern evidence of the organisation or prey selection required in hunting pygmy elephants and giant reptiles.

All of this means that the discovery of *Homo floresiensis* raises more questions than it answers both for the interpretation of the find within Flores, but also beyond. In Flores, there are questions, first, of the cognitive abilities of the new hominin, but then of the evolutionary emergence of those abilities. In addition to the questions already raised, there are important issues about the process of evolution as the ancestor of this creature moved from the mainland of South-East Asia to Flores. In particular, the archaeology of Lombok and Sumbawa now beckons as one of the most interesting issues in world archaeology – did the differentiation of the new species take place entirely on Flores, or did it begin elsewhere? And when we know something about those anatomical associations, how did the cognitive abilities of the Lombok/Sumbawa ancestor compare with those of *Homo floresiensis*? In this chapter, I have argued that the interpretation of the evolution represented by the anatomy should be related closely to questions of selection, and questions of cognitive ability can only be addressed through parsimonious analysis of the archaeological record.

References

- Aiello, Leslie C. & Robin I. M. Dunbar. 1993. Neocortex size, group size, and the evolution of language. *Current Anthropology* 34. 184–193.

- Birdsell, Joseph. 1977. The recalibration of a paradigm for the first peopling of Greater Australia. In Jim Allen, Jack Golson & Rhys Jones (eds.), *Sunda and Sahul*, 113–167. London: Academic Press.
- Brace, C. Loring. 1995. Biocultural interaction and the mechanism of mosaic evolution in the emergence of "modern" morphology. *American Anthropologist* 97. 711–721.
- Brown, Peter, Thomas Sutikna, Michael J. Morwood, Radien P. Soejono, Jatmiko, E. Wayhu Saptomo & Rokus Awe Due. 2004. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431. 1055–1061.
- Davidson, Iain. 1999a. First people becoming Australian. *Anthropologie (Brno)* 37(1). 125–141.
- Davidson, Iain. 1999b. The game of the name: Continuity and discontinuity in language origins. In Barbara J. King (ed.), *The origins of language: What nonhuman primates can tell us*, 229–268. Santa Fe, NM: School of American Research.
- Davidson, Iain. 2001. The requirements for human colonisation of Australia. In Ian Metcalfe, Jeremy M. B. Smith, Michael J. Morwood & Iain Davidson (eds.), *Faunal and floral migration and evolution in SE Asia-Australia*, 399–408. Lisse: Swets & Zeitlinger.
- Davidson, Iain & William Noble. 1992. Why the first colonisation of the Australian region is the earliest evidence of modern human behaviour. *Archaeology in Oceania* 27. 135–142.
- Dunbar, Robin I. M. 1992. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 20. 469–493.
- Dunbar, Robin I. M. 1993. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences* 16. 681–735.
- Falk, Dean. 1986. Evolution of cranial blood drainage in hominids: Enlarged occipital/marginal sinuses and emissary foramina. *American Journal of Physical Anthropology* 70. 311–324.
- Feldesman, Marc R., Geoff Kleckner & John R. Lundy. 1990. Femur/stature ratio and estimates of stature in Mid- and Late-Pleistocene fossil hominids. *American Journal of Physical Anthropology* 83. 359–372.
- Feldesman, Marc R. & John K. Lundy. 1988. Stature estimates for some African Plio-Pleistocene fossil hominids. *Journal of Human Evolution* 17. 583–596.
- Foley, Robert A. & Phyllis C. Lee. 1991. Ecology and energetics of encephalization in hominid evolution. *Philosophical Transactions of the Royal Society of London, Series B* 334. 223–232.
- Holloway, Ralph L. 1983. Human brain evolution: A search for units, models and synthesis. *Canadian Journal of Anthropology* 3. 215–230.
- Lee, Sang-Hee & Milford H. Wolpoff. 2003. The pattern of evolution of Pleistocene human brain size. *Paleobiology* 29. 186–196.
- Lewin, Roger & Robert A. Foley. 2004. *Principles of human evolution*, 2nd edn. London: Blackwell.
- McHenry, Henry M. 1991. Femoral lengths and stature in Plio-Pleistocene hominids. *American Journal of Physical Anthropology* 85. 149–158.
- Miguel, Carmen de & Maciej Henneberg. 2001. Variation in hominid brain size: How much is due to method? *Homo* 52. 3–58.
- Mitchell, Catherine. 1997. Interpreting charcoal patches in open-site archaeology: Need we be stumped for an answer? Armidale, NSW: University of New England BA (Hons) thesis.
- Moore, Mark W. This volume. Lithic design space modelling and cognition in *Homo floresiensis*.
- Morwood, Michael J. & Dorothea Cogill-Koez. This volume. *Homo* on Flores: Some early implications for the evolution of language and cognition.
- Morwood, Michael J. & Iain Davidson. 2005. Out of ancient disasters, forebears may have colonised new lands. *Sydney Morning Herald* (14th January). 17.

- Morwood, Michael J., Fachroel Aziz, Paul O'Sullivan, Nasruddin, Doug R. Hobbs & Asaf Raza. 1999. Archaeological and palaeontological research in central Flores, east Indonesia: Results of fieldwork 1997–98. *Antiquity* 73. 273–286.
- Morwood, Michael J., & Penny van Oosterzee. 2007. *A new human. The startling discovery and strange story of the "Hobbits" of Flores, Indonesia*. New York, NY: Smithsonian Books.
- Noble, William & Iain Davidson. 1996. *Human evolution, language and mind*. Cambridge: CUP.
- O'Sullivan, Paul B., Michael J. Morwood, Douglas Hobbs, Fachroel Aziz, Suminto, Mangatas Situmorang, Asaf Raza & Roland Maas. 2001. Archaeological implications of the geology and chronology of the Soa basin, Flores, Indonesia. *Geology* 29(7). 607–610.
- Ruff, Christopher B., Erik Trinkaus & Trenton W. Holliday. 1997. Body mass and encephalization in Pleistocene *Homo*. *Nature* 387. 173–176.
- Skoyles, John R. 1999. Human evolution expanded brains to increase expertise capacity. *Psychology* 10(2). <http://cogprints.org/2263/00/skoytarget.htm>.
- Spoor, Fred, Meave G. Leakey, Patrick N. Gathogo, Francis H. Brown, Susan C. Antón, Ian McDougall, Christopher Kiarie, Frederick Kyaho Manthi & Louise N. Leakey. 2007. Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. *Nature* 448. 688–691.
- Stiner, Mary C. 1994. *Honor among thieves: A zooarchaeological study of Neandertal ecology*. Princeton, NJ: Princeton University Press.
- Trinkaus, Erik & Christopher B. Ruff. 1999. Diaphyseal cross-sectional geometry of Near Eastern Middle Palaeolithic humans: The femur. *Journal of Archaeological Science* 26. 409–424.
- Veth, Peter, Sue O'Connor & Matthew Spriggs. 2004. Changing research perspectives from Australia's doorstep: The joint Australian-Indonesian Aru Islands initiative and the archaeology of East Timor project. In Tim Murray (ed.), *Archaeology from Australia*, 209–230. Melbourne: Australian Scholarly Publishing.
- Weiner, Steve, Qinqi Xu, Paul Goldberg, Jinyi Liu & Ofer Bar-Yosef. 1998. Evidence for the use of fire at Zhoukoudian, China. *Science* 281. 251–253.

CHAPTER 4

Homo on Flores

Some early implications for the evolution of language and cognition

Michael J. Morwood and Dorothea Cogill-Koez

The Indonesian island of Flores has yielded evidence of a new hominin species, *Homo floresiensis*, and an earlier species currently known from tools only. This chapter reflects on some of the implications the new finds may have for the evolution of language and cognition. In the light of relevant existing patterns of argument about early hominin morphology and behaviour, the Flores finds appear to weaken the argument for “big brains” as a sufficient cause of cognitive and linguistic abilities, to highlight, conversely, the significance of brain reorganisation events in human evolution, and to refocus attention back onto the very dawn of our genus as the point at which interestingly human patterns of communication and cognition are likely to have appeared.

1 Introduction

This chapter focuses on the significance of new finds on Flores for modelling the evolution of language and cognition. We briefly review a limited, relevant selection of existing lines of argumentation from the archaeological record as regards cognitive and linguistic abilities in ancestral hominins.¹ We then review new evidence

1. Immediately, some clarification on taxonomy must be given. Little is needed for this chapter. Following recent genetic models, “hominids” will include that branch which now terminates in ourselves, chimpanzee species, and gorillas. Within this broader group, “hominins” will refer to species on “our” sub-branch, which diverged from the rest some 5–7 million years ago. Within the hominins, fine discrimination of lineages is not needed for the purposes of this chapter, which therefore adopts the by now old-fashioned and more inclusive classification of *H. habilis*, *H. erectus*, “archaic” *H. sapiens* and fully modern *H. sapiens*. These older labels identify the species whose morphology or behaviour is of interest, without engaging in otherwise inescapable and soon-to-be-outdated discussion of the merits of the range of competing new classifications.

from Flores, Indonesia, on these same points. Some of the evidence is thought to relate to *Homo erectus*, one of the earliest species of the genus *Homo*; most of it pertains to a new species of hominin, *Homo floresiensis*.

The combination of morphology, behavioural evidence and dating as regards these two species on Flores is unprecedented. The finds appear to overturn some previously undoubted positions on the relationship between morphology and cognition in the hominin lineage. They may also help to pinpoint some key stages and events in the evolution of higher cognitive and linguistic abilities. Finally, in combination with standard models in evolutionary theory, current and future finds on Flores and in the region promise to provide new and welcome constraints on modelling in the evolution of language and cognition.

A second theme of the chapter relates to the issue of cross-disciplinary research. In the biological sciences generally, multifactoriality is a fact of life; any problem one wishes to address is subject to a huge number of influencing factors, many of them unknown. The spare elegance of irrefutable logic, based on limited known premises, can seem to the biologist a sweet but impossibly distant dream, while to the linguist or philosopher the life-sciences can seem content to construct straw models from frail likelihoods, piling uncertainty on uncertainty with gay abandon. The acknowledged problem posed by multifactoriality in biological sciences is exacerbated in disciplines like archaeology or evolution, which aim to reconstruct the past. Here the method of repeated, controlled experiments available to other biologists is impossible, and certainty is a luxury that one must simply learn to do without.

To the stern logician, if certainty is impossible, all is lost; what grounds are there to exclude any scenario at all? Perceived lack of constraints has repeatedly undermined the credibility of evolutionary scenarios; in practice, above all, in modelling the evolution of language and cognition. We put it that the only criterion is a strongly constraining one; relative degree of reasonableness. We argue that a reasonable evolutionary reconstruction is based on the convergence of all available lines of case-specific evidence, supplemented where necessary with the application of broad general patterns. This latter set of influences is given by observed regularities over the known history of life, from micro-organisms to whales. It is a not inconsiderable data set, and forms the basis of standard models in evolutionary theory such as will be applied in this chapter.

Repeatedly, then, the point will be made that such-and-such a conclusion is not inevitable – but that in light of all case-specific evidence to date, set in turn against standard patterns of evolutionary modelling, it is a rational, broadly-based position to take at this point. We make no apologies for the inevitable uncertainty of the result, and for the fact that, as regards hominin evolution, the very database

continues to change. What we outline is some of that change, and a few reasonable conclusions that can be drawn from it.

2 Some existing patterns of argument regarding hominin cranial morphology, behavioural complexity, and the dating of “human” cognitive and linguistic skills

2.1 Arguments from morphology

Prominent in all discussion of the evolution of “human” cognitive abilities has always been the issue of brain size as measured by fossil cranial capacity. The conviction that brain size relates to intelligence is as persistent as are the difficulties in determining exactly how it does so, and the debate has intensified rather than diminished recently.² The commonest measure used, however, is encephalisation quotient (EQ); the ratio of brain size relative to body size when allometric effects have been accounted for.³ Brain tissue is, of all body tissues, one of the most energetically expensive to maintain. Furthermore, with ever greater brain size relative to body size, reproductive costs also increase, as pregnancy becomes more energetically expensive, birth more risky and, after a certain point, the period of post-natal dependence also extends. Given these heavy costs there is little doubt that increased encephalisation must also be associated with substantial selective

2. Among the numerous alternative models, *H. floresiensis* may be less able to contribute to debates that focus on the relative proportions of internal structures of the brain. However, calculations of proportional structural changes that result from greater convolution of the brain surface may well be accessible and could prove extremely interesting, and the new species can already been seen to have great relevance to models that emphasise absolute brain size (Gibson, Rumbaugh & Beran 2001).

3. Allometric effects in general are the effects of “mere” change in size within a lineage (i.e. a group of related species, or branch of the evolutionary tree). For example, there is an extremely widespread pattern across the animal kingdom whereby when bodies are bigger, the head is also bigger – but not by the same percentage as the body. Hence, ever larger species tend to have *relatively* ever smaller heads, without affecting their intelligence in any obvious way.

The normal rate of size change of any body part X to any body part Y can be calculated for a lineage. Those species in the lineage which then stand out as departing from the resulting regular allometric rule of proportional change become especially interesting, because one can assume that evolutionary events have here altered the regular developmental patterns of growth of size of X or Y in this species. EQ itself, as originally conceived by Jerison (1973), is a measure of how much a species’ brain size departs from what would normally be predicted for a mammal of its body size, with an EQ of 1 representing a brain/body ratio that is exactly as expected for a mammal of that size.

benefits for its possessors, and the most obvious selective benefit delivered by increased brain size has always seemed likely to be increased intelligence.

Encephalisation and intelligence as conceptualised by our species do indeed appear to increase roughly in tandem across living mammalian species (Jerison 1973). Within major lineages of primates it appears to do so as well, with (for example) gorillas, with an EQ roughly 25% of ours, generally conceded to be outperformed by chimpanzees with an EQ about one third larger than their cousins, and in turn exceeded in intelligence by us, with our EQ about three times that of a chimpanzee (Holloway, Broadfield & Yuan 2004). Thus it has been reasonable to infer that the same pattern held in extinct members of this lineage as well; that increasing cognitive capacities accompanied the observed increases in EQ of hominin species across time.

The EQ attributed to a species varies quite widely with the mathematical formula adopted, the measurement method for cranial capacity, the body weight estimate, the specimens used, and of course the species to which the specimen is attributed (cf. Foley 1996; Holloway, Broadfield & Yuan 2004; McHenry 1982; Tobias 1987) – but the overall pattern is not in significant dispute. Lewin and Foley (2004) give a typical sequence in which *Australopithecus* (EQ = 2.5), at somewhat more than a chimpanzee's EQ of 2, yields to early *Homo* species of first *Homo habilis* (EQ = 3.1) and then *Homo erectus* (EQ = 3.5), then on through to *Homo sapiens* (EQ = 5.8).⁴ A further implication of this sequence in living and extinct hominins is that it is not unreasonable to estimate the relative intelligence of any newly-discovered extinct hominin by its EQ, yielding an unsteady but unidirectional increase in intelligence as hominins evolved. Finally, based on the picture to date, the “best bet” would be that any newly-discovered hominin species is far more likely than not to slot into the existing trend from earlier and smaller-EQ hominin species to ever more recent species with ever larger EQ and ever greater cognitive capacities.

A second physical correlate of our increased cognitive abilities is relative expansion and hence increased folding of the surface of the neocortex, giving our species its well-known relatively convoluted brain surface. In both ourselves and other species, known functions can often be mapped onto resulting physical landmarks on the surface of the brain. The convolutions associated with Broca's area, the Sylvian fissure that marks the upper boundary of the temporal lobe, or the lunate sulcus that in primates demarcates the primary visual cortex of the occipital lobe are relevant examples.

4. The recent slight drop in EQ within our species muddies this picture, but the broad pattern and common positions based on it are outlined here.

Such convolutions can in turn leave their impression on the inner surface of the skull. Hence, in a kind of inverted phrenology, endocasts of fossil crania can yield information about brain morphology, and thus also by inference about cognitive abilities in extinct species. The technique and individual applications of it are not without controversy, but it has been used, for example, in attempts to pinpoint the relative expansion of the temporal and parietal areas that marks modern *H. sapiens* (where it now supports, among other things, auditory perception of language, the ability to produce meaningful speech, and spatial abilities), and to track the emergence of enlargement in the region of Broca's area, which in ourselves supports aspects of the production of language, specifically fluent and grammatical speech (and signing). Both these changes appear to be present at least by the appearance of *H. erectus* (Holloway, Broadfield & Yuan 2004). The smaller current set of *H. habilis* crania are less well-preserved in some crucial areas, but they are also argued to point to the same features in this earliest of *Homo* species (Tobias 1987).

A third class of brain feature that has been linked to cognitive patterns is bilateral torque. When viewed from above, the human brain can be said to show a sort of twist; at the front, the right hemisphere usually protrudes a little further forward than the left, while at the back, a left protrusion or petalia is seen. While some other hominid species show the back or occipital petalia, the bilateral pattern is argued to be distinctive to hominins (Holloway & De La Coste-Lareymondie 1982) – and it is asymmetry in the frontal region, above all, that has been argued to influence many aspects of our species' unusually high degree of cerebral lateralisation, including handedness (see e.g. Zilles et al. 1996).

Thus far, then, these morphological criteria point to interesting developments in hominin cognitive capacity, in place with the appearance of the genus *Homo*. If existing assumptions about the role of EQ are correct, intelligence by some gross measure took a significant step forward at this time. Furthermore, if the function of specific areas of the modern human brain can be extrapolated back to early *Homo*, then the dawn of our genus may also have seen at least some development of language-related cognitive skills away from the ape and towards the position now represented by ourselves, while species' level handedness may also have typified hominins from that time or even earlier.

2.2 Arguments from behaviour

2.2.1 *Stone tools*

The most enduring record of hominin behaviour is that of stone tools. These (like any isolated technology) give only a minimum definite measure of the technological capacity of their makers, and stone tools in fact may not be a particularly sensitive measure of overall technological level anyway: The stone tools made in some

regions of New Guinea until recently, for instance, are similar to the earliest known from Africa, despite the fact that the makers' cultivation technology, social organisation, etc. are exceedingly complex. However, stone tools are preserved where nothing else is, and as a result, their study and classification, and the attempts to make inferences from them about the kinds of cognitive processes they demanded of their makers, all have a long and ever-developing history in archaeology. The very brief sketch of main points which follows is based in Clark's (1977) classic taxonomy, still used as it is, as a convenient heuristic, in everything from technical articles to standard texts (e.g. Davidson 2002; Lewin & Foley 2004). This will provide a framework against which to set the conflicting views presented further on in the chapter.

In this classic picture, the first incontrovertible tools are stones from which, without any further preparation, a few flakes have been removed – yielding (along with the useable flakes) a sharp edged chopper. This technology appears first at 2.6 million years at much the same time as *H. habilis*, a coincidence that suggests (but does not guarantee) that *H. habilis* was the first to make it; it continued to be used for some 1.5 million years by *H. erectus* in Asia. Elsewhere hominins moved on about 1.5 million years ago to produce a bifacial core flaked on both sides to produce a pleasingly symmetrical teardrop shape long called a handaxe, the type artefact of the Acheulian. This action sequence and resulting symmetrical handaxe, apparently standardised, emerge in the archaeological record and then endure with little change for over a million years.

Classically, the next phase is that of prepared cores. In these, a stone is understood to be first shaped by removing stone, not to produce useable flakes, but solely so that multiple useable flakes can next be repeatedly struck from the core. This produces in effect more centimetres of cutting edge per kilogram of stone, and results in a radial core rather than the bilaterally-symmetrical product. Such prepared cores become standard in many hominin repertoires only by about 300,000 years ago. Fully modern *H. sapiens* emerged in Africa at about 195,000 years (McDougall, Brown & Fleage 2005), and some time after this long, slim blades surge as a percentage of stone tools found; deliberate, regular production of their proportions has been held to necessitate far more careful preparation of cores. This is followed by an explosion of fine and varied "microlith" technology, encompassing, as part of their great variety, stone elements formed for combination with other tool components (such as a spear haft) to create compound tools. This final efflorescence comprehends not only an increase in number of techniques and resulting tool types, but also a greatly accelerating rate of technological change as such, and also importantly, an increase in standardisation and differentiation between identifiable artefact types.

What are the implications for cognition? The traditional interpretation of the classic picture stresses a 2.6 million year history of progression in stone tool techniques and thus in evinced cognitive skill. It also specifies the steps in this progression. The earliest forms represent opportunistic creation of sharp flakes, with little foresight as regards the geometry of the resulting flakes or core (however useable they may transpire to be), or of planned actions in creating them. The Acheulian artefact represents a new ability to hold in the mind a template of the desired form, and to impose the concept onto the raw materials; a sustained distinction between mental and real world affairs arguably attesting to a capacity for symbolic thought. Prepared cores give evidence of a further level of cognitive complexity; the ability to structure actions hierarchically by first producing one desired shape, from which the real target – flakes of a quite different shape – can then be struck. Finally, blades and the compound tools formed by hafting microliths are the mark of the fully modern mind. Their diversity of form, combined with clear standardisation (that is, uniformity with and distinction between types) gives abundant evidence of diverse mental “templates” in the minds of the makers, while the varied techniques between them demonstrate a cognitive ability to mentally retain and to choose between multiple hierarchical strategies. The conjoining of parts to form a compound tool, in which the whole is formally and functionally different from its component elements, is a further step never yet seen in nonhuman animals.

In recent decades, however, both the evidence base and the implications of the classic picture have been questioned. This reassessment has taken two very different forms. One line of argument questions whether the stone tool sequence really reveals significant cognitive abilities in any species prior to our own. Another trend of interpretation, conversely, finds evidence of appreciable cognitive sophistication at the very dawn of stone tool technology.

What one might call the minimalist approach to the archaeological record is typified by Davidson (e.g. in Davidson 2002). This approach focuses closely on the archaeological record alone, holding it to be the only reliable source of information on past cognitive achievements. Within this, it urges archaeologists to adopt an interpretation of past technology that attributes to them the least human-like abilities logically possible for the behaviour seen, in order to avoid imposing on minds now extinct the cognitive patterns of our species today. This departs from the routine patterns of evolutionary argument applied to nonhuman animals.⁵

5. Routine evolutionary argument is not logically water-tight, but it is what a betting person might call the “best bet”. Consider an extant species #1 that exhibits features A and B, both demonstrably tied to cognitive feature X. All other extant species in the lineage lack feature X, and lack also the associated features A and B. One then finds an extinct species of the same lineage yet more closely related to species #1 than to any others, and it is accompanied by evidence of A and B. From viruses to dinosaurs to the alien species of a different planet, all other things being equal,

Thus, among other arguments, Davidson has long held that the symmetry and apparent standardisation of the best examples of the Acheulean handaxe, so intuitively persuasive of a mental template to the modern mind, need be no more than the automatic outcome of more sustained knapping than is found in earlier technology. Even later examples of this phase of technology, striking (to us) in their symmetry, are argued to present no evidence for deliberate production. The flakes removed from them are the only products that we can be sure are intended objects of the makers' actions; thus the handaxe is argued to be misnamed, and cannot be held to have been deliberately planned and created to satisfy practical, ceremonial or aesthetic needs. Combined with similarly cautious interpretations of other aspects of the archaeological record, Davidson and colleagues conclude that even the cognitive capacity for basic symbol use, the very rudiments of language, cannot be attributed to any species prior to our own – or indeed even to our species, prior to the first colonisation across a major water barrier, a mere 60,000 years ago at most (Noble & Davidson 1996).

Where Davidson queries the status of the Acheulian as marking the presence of mental templates, his colleague Moore (2004, 2005) extends his minimalist approach further, calling into question the interpretation of the artefacts of the next classic phase as marking the presence of a new level of hierarchical mental operations. Limiting his own evidence base firmly to stone tool technology, Moore argues that the artefacts that typify this later period can be produced, like earlier ones, by repetitions of only one single, short sequence of physical and cognitive operations, “the basic flake unit”. With Davidson, he points out that this short sequence itself produces a desirable product in the form of a useable flake, and thus its performance logically demands no motivation beyond this. By means of simply extending the number of repetitions of this basic flake unit on a single core, first Oldewan “choppers”, then bifacial “handaxes”, then radial cores automatically appear. Blades, too, are not in themselves necessarily the marker of the modern human mind that they are held to be in the classic view; on the contrary, a certain percentage of them can be the statistically inevitable byproduct of basic flaking. Moore's revolutionary analysis thus argues for no significant or qualitative change in the basic cognitive operations necessary to produce stone tool technology from the Oldewan right up until archaeologically very recent times some 40,000 to 50,000

routine evolutionary argument would infer the presence of cognitive feature X in the new species. No particular significance would be attached to the fact that species #1 may be ourselves, A and B might be stone tool production and changes in brain morphology respectively, and X might comprise the ability to hold a mental image, engage in hierarchical operations, or communicate via symbols. In the minimalist argument, however, the assumption of X is unwarranted for as long as A and B can be explained in any way congruent with the features of earlier species in the lineage.

years ago (Moore pers. com. September 2005); that is, within only the last 20–25% of the lifetime of our own species, *H. sapiens*, let alone hominins as a group.

Minimalist approaches can, however, produce an apparent dissonance between the interpreted record of hominin behaviour and other strands of evidence. For example, if the reconfigured and energy-hungry larger brains of species in the genus *Homo* were not being used for more sophisticated tool production (or resource exploitation in the form of fire use and hunting, or the support of symbolic behaviour in language, as the minimalist strand of archaeological interpretation argues), then what pressures could have supported and enabled the appearance of these costly investments? Other evolutionary drivers for the observed morphological changes must be sought.

Conversely, another new trend of interpretation sees evidence of an unappreciated degree of cognitive sophistication in even the earliest stone tools. Selection, initial reduction and transport of the optimal kind of stone for flaking; an understanding of the appropriate (and quite constrained) geometry in that stone which provides a viable target to strike at in order to produce flakes, albeit not the capacity to actually prepare cores wholesale by creating these angles in the first place; an ability, nevertheless, to somewhat modify existing striking platforms, or correct minor flaws produced by knapping that could affect further flake production; a capacity to order the removal of flakes in a way that allowed such continued knapping (i.e. the removal of a series of flakes) rather than quickly destroying the necessary geometric properties of the striking platform; the selection of optimal size, shape and percussion points of the hammerstone, as well; the understanding (and the manual control) to strike with precision and with a force appropriate to the stone – all these have been attributed to makers of the earliest stone tool technology, at least in some locales (e.g. de la Torre 2004; Delagnes & Roche 2005).

Further, some of the very earliest tools also require a superior grasp of geometric configurations than is currently demonstrated in tool making by extant chimpanzees (Toth et al. 1993; Schick et al. 1999). Toth, Schick and colleagues have shown that, while by no means lacking in ingenuity in the making of flakes, to date even human-enculturated bonobos are apparently unable to recognise and exploit the particular points on the stone's geometry that provide viable targets for flake production. In spite of an enhanced appreciation of the cognitive and communicative abilities of modern apes, therefore, this recent research still emphasises a functionally distinctive further degree of cognitive capacity in the earliest tool makers.

In archaeology, then, the current theoretical position as regards the relation between stone tool technology and cognition is complex and conflicted. Some tenable position, however, may emerge in applications below.

2.2.2 *Hunting of large prey*

The appearance of evidence for hunting of large prey has been used as another behavioural marker of increased cognitive skills, and also of linguistic ones. True, social carnivores such as lions and wolves hunt large prey without the benefit of language, while chimpanzees have also been found to hunt co-operatively. However, social carnivores have many tens of millions of years of evolution in the context of such hunting; they are demonstrably adapted to it physically, and so might equally be assumed to be adapted to it through specially dedicated cognitive mechanisms.

On the other hand some chimpanzee troops actively hunt, while others, at least on the data available to date, do not. This pattern of variation in hunting activity in chimpanzees currently suggests that their hunting skills are supported by general, undedicated cognitive abilities, rather than being the inevitable instinctive result of a hardwired hunting-specific cognitive specialty. However, as yet chimpanzees have been seen to hunt only smaller species than themselves, where capture and dispatch involves pursuing and cornering the animal, then simply grabbing it and either flailing it to death (Stanford 1995) or biting it – sometimes, indeed, commencing to eat the unfortunate animal alive (Boesch & Boesch 1989).

The hunting and killing of mammoth, bison or rhinoceros, in contrast, would seem to be a different and far more demanding matter; nothing comparable has been seen in chimpanzees. First, hunters must prepare hunting tools that, for hominins, replace the teeth and claws of specialist large-prey carnivores. Further, they must also plan and co-ordinate their big game hunts, anticipating the actions of the prey and their co-hunters, and co-operating with the latter. The extended series of planned and integrated actions needed to bring down large prey has been argued to require explicit exchange of information between individuals during planning of the hunt, and even in the education of hunters (e.g. Roebrooks 2001).

In the past, hunting has therefore been modelled as a driving force in the evolution of language and cognition (e.g. Lee & DeVore 1968). More recently and more modestly, if co-operating hunters must discuss hunting and do so in the absence of the animals actually hunted, systematic big-game hunting can be argued to be evidence of linguistic abilities that minimally reach the level firstly of symbols, and secondly of displacement, the use of the symbols in the physical absence of their referent.

Turning to the archaeological record for hunting of large prey, interpretation is clouded by the possibility of mere scavenging, a behaviour known in modern hunter-gatherers. As a result big game hunting, like systematic fire use, is by some authors allowed and by others denied to *H. erectus* (e.g. Wynn 2002). However, finds such as that of sharpened wooden shafts, in a range of lengths up to 2.3 metres yet each with a centre of balance resembling a modern javelin, and found in association

with butchered big game remains (Thieme 1997), point to such hunting by 400,000 years ago, giving, by the big game argument, a new minimum date for symbol use and displacement.

2.2.3 *Planned water crossings*

Another potential behavioural marker of advanced cognitive and linguistic abilities is water crossings. As with hunting, the migration of a viable group of hominins across a substantial water barrier has been argued to demand a high level of forward planning, technical ability and actual discussion (Davidson & Noble 1992). Attribution of deliberate water crossings to a species has its own problematic aspects; the interpretation of the record is complicated not only by the possibility of land bridges (by which what is now an island may, in the past, have been connected to the mainland) but also by the known possibility of rafting.

This is the term used for the accidental transport of animals across water barriers on natural rafts of flotsam such as regularly form, for example, during the flooding of rivers. These rafts can occasionally be quite large and stable, so separating accidental rafting from deliberate water crossings is not as straightforward as might appear; it cannot be done solely by the length of the crossing made. First, is the species a known good rafter, frequently appearing on islands that other species have not reached? (Small rodents, for example, are outstanding rafters; ungulates less so.) If not, is the species accompanied in the island record by others of equally poor rafting ability, indicating that rafting to this island could have been an easy and frequent natural process? Only if the answer is again no does the presence of the species begin to demand the hypothesis of planned human action.

At this stage, then, the earliest clear evidence for planned water crossings is held to stand at some 40,000–60,000 years ago, with the colonisation of Australia by fully modern *H. sapiens* (Davidson & Noble 1992). Note that this is well within the most recent half of our own species' existence.

3 The Flores hominins

Flores is an island which itself forms part of an island chain stretching from the South East Asian mainland in the west, to Papua-New Guinea and Australia in the east and south-east. Recent archaeological work has uncovered evidence for three hominin species on the island.

The first hominins reached Flores by at least 880,000 years ago, as evidenced by stone tools of this age in the Soa Basin (MM pers. obs., Morwood et al. 1998). There is no associated skeletal evidence.

The form of the tools and their date of deposition initially pointed to the maker as most likely *H. erectus* (Morwood et al. 1999), a species known from independent skeletal finds to have been present contemporaneously to the west, on the Asian mainland and also on Java. (Java is now an island, but in the past land bridges periodically connected it to the Asian mainland [Itihara et al. 1985; de Vos et al. 1994] and it thus possesses many mainland species.) However, the species attribution of this early toolmaker on Flores is now being rethought, towards a descendent of an unknown but even earlier species than *H. erectus*. This does not affect the conclusions to be drawn below, except as it will point towards an even earlier leap towards so-called humanlike communication and cognition.

The second hominin species evident on Flores, *Homo floresiensis*, probably evolved *in situ* on the island by endemic dwarfing of the earlier population that occupied the Soa Basin. Excavations at Liang Bua, a cave in the limestone uplands of West Flores, has yielded skeletal remains of at least nine individuals from deposits between 95,000 and 12,000 years ago (Morwood et al. 2004). All are characterised by unique morphological traits. In terms of understanding their place in the hominin family tree, the least significant of these is the fact that adults are about 1 metre in height (hence they are colloquially known as “hobbits”). A massive volcanic eruption about 12,000 years ago, represented by a thick layer of tuffaceous silt in the Liang Bua deposits, seems to have caused the demise of both *H. floresiensis* and *Stegodon* (a primitive type of elephant, and the only other large-bodied mammal to have colonised Flores).

Above the tuff layer, in deposits around 10,500 years old, an abrupt leap in technology together with skeletal evidence marks the arrival of the third hominin species known from Flores – fully modern humans (*H. sapiens*). This coincides with the first evidence at Liang Bua for ochre, ornaments and formal disposal of the dead, as well as more intensive use of plant resources (probably including cultivation of root and tree crops). Later, about 4,000 years ago, the arrival of the Neolithic is evident in the appearance of pottery, ground adzes and cereals, as well as the introduction of pig, dog, monkey, porcupine and civet cat, all of which were transported to the island by people (Morwood et al. 2004). Thus, while *H. sapiens* and *H. floresiensis* must have been broadly concurrent in the region since at least 55,000–35,000 years ago (O’Connell & Allen 2004; O’Connor, Spriggs & Veth 2002; Roberts et al. 1994), it appears that at Liang Bua the two species did not co-exist.

Although *H. floresiensis* was extant as recently as 12,000 years ago, the closest common ancestor between ourselves and the “hobbits” nevertheless appears to lie at an extraordinary time depth. Morphological analysis shows that the species shares many traits with the early *Homo erectus* population evident at the 1.8 million year old site of Dmanisi in Georgia – for instance, the overall shape of the

cranium and endocast (Falk et al. 2005). However, some traits (such as the morphology of the mandibular symphysis and body proportions) hark back to even earlier in the African hominid sequence (Brown et al. 2004; Morwood et al. 2005). It is thus not yet clear whether *H. erectus* was ancestral to *H. floresiensis*, or alternatively whether both are derived from an earlier common ancestral species. In the latter case, standard evolutionary argument would predict that this putative earlier species already possessed the endocast features shared by its two daughter species, so doubts over the species classification of the early toolmaker on Flores will not affect the inferences to be made here.

The recent finds appear to overturn key assumptions regarding the interaction of morphology and behaviour in extinct hominins. In the process they also tend to weaken models that favour a very recent emergence for crucial aspects of the modern human mind, including some that contribute to language.

3.1 The early toolmakers on Flores: Evidence and significance

As noted above, the earliest stone tools known to date on Flores have been attributed to *H. erectus*, on two widely used grounds; skeletal evidence of the contemporaneous presence of the species, albeit to the west on the past or present Asian mainland, and the form of the tools themselves. Whether it is *H. erectus* itself, or the descendent of an even earlier common ancestor of both *erectus* and *floresiensis*, what is remarkable about this species is the place where they were toolmaking, for the island of Flores lies to the east of the well-known Wallace line, a significant water gap across which few species have successfully rafted naturally.

Biologically speaking, the apparent continuity of the island chain from the South East Asian mainland to Australia is an illusion. It spans the geologically-recent collision of two continental shelves, each bearing landmasses on which life has evolved in separate directions since the Mesozoic. This has produced two clear zones that are geographically so close, yet biologically so distinct, that they helped lead Alfred Wallace independently to the idea of evolution by natural selection.

However, in the no-man's land of extremely deep water between the Asian and Australian continental shelves is Wallacea. This is a scattering of islands marked generally with a very depauperate fauna – the few representatives from both biogeographic zones which have managed to fly, swim or raft to the islands. Flores lies here. It has been divided at all times from both landmasses. Specifically, the nearer Asian biogeographic zone has been distant by at least two water crossings; one of these represents the famed Wallace Line itself, and is a channel so fraught with currents that prior to modern human intervention only birds, highly mobile rafters like small rodents, and excellent swimmers such as elephant species succeeded

in migrating eastwards into Flores.⁶ Significantly, no other primate species managed the feat, and this is in spite of the smaller body size and apparently greater rafting potential of the numerous monkey and ape species in the Asian biogeographic zone.

What, then, can account for the presence, in mid-Wallacea and at least 880,000 years ago, of hominins? Routine patterns of argument point to deliberate water crossing, with all the cognitive and linguistic skills this may demand. Against this, though, stand many other strands of argument and evidence. If the species was *H. erectus*, there is the sheer difficulty of accepting that a species which may not even have mastered fire, and which in most of Asia spent its entire existence repeatedly making the simplest of definitely recognisable stone tools, should have been cognitively equipped for such a feat. If the species was the descendant of an even earlier common ancestor of *erectus* and *floresiensis*, its apparent watercrossing achievement is if anything even harder to credit.

H. erectus has never been found in comparable circumstances before. This points to the absence of habitual deliberate water crossings by the species – but equally to the absence of natural flotsam-rafting proclivities. Nor is there any evidence of hominins reaching the Australian mainland at this extraordinarily early date, as might be expected of a genuinely skilled seafaring species – but colonisation of Australia, unlike island-hopping to Flores, entails the further challenge of setting off towards an apparently empty horizon. This is a venture any log-paddling hominin of modest seafaring abilities (and reasonable foresight) must be reluctant to undertake.

The scenario found on Flores is thus unprecedented, and difficult to pronounce upon. Currently the best guess may be a series of freakish events, such as a tsunami washing a group out to sea on flotsam, and sea currents then depositing this group on islands “downstream”. Far-fetched though this may seem, the time-frames involved make statistically vanishingly-rare events a possibility.

Such a scenario would put the geographic origins of the Flores hominins not on the closest islands to the east of Flores but on those to the north, upstream along the strong ocean currents of the region. For this reason excavations are commencing at limestone shelters and Upper Pliocene-Middle Pleistocene fossil sites on Sulawesi, the most likely source area. Further modelling must await the results of multiple excavations there and on other islands in the archipelago.

6. Elephants are surprisingly well-equipped for island hopping, being large, buoyant, strong and frequent swimmers, moving in genetically-viable herds, and being provided with their own natural snorkel.

3.2 *H. floresiensis* on Flores: Evidence and significance

3.2.1 *Arguments from morphology*

Unlike the elusive first hominin on Flores, the second species is represented by well preserved skeletal remains yielding abundant information about its morphology.

In relation to the evolution of language and cognition, the most striking feature of *H. floresiensis* is morphological: For a hominin, the brain of the “hobbit” is smaller than even its small body size would lead us to expect – smaller, too, than the external dimensions of the skull would indicate, due to substantial relative thickening of the bone in some areas. At 380 cm³ by traditional mustard-seed measurement (Brown et al. 2004) and 417 cm³ by calculations based on data from the scanned cranium (Falk et al. 2005), the brain is the smallest known for the genus *Homo*. Importantly, it is strikingly smaller than that of either *H. erectus* or even the cranially-challenged Dmanisi hominins. With a recent increase in the estimated body size of *H. floresiensis* (cf. Brown et al. 2004; Morwood et al. 2005), the “hobbit’s” EQ is also a drop on that of any known potential *Homo* ancestor, being now well in the range for australopithecines, and even chimpanzees. This is in contrast to the pronounced evolutionary trend towards higher EQ seen throughout the rest of the hominin lineage.

The inference from EQ alone would thus be that *H. floresiensis* was virtually ape-like in cognitive and linguistic abilities. Indeed, based on existing patterns, it would be plausible to argue that the species had actually regressed cognitively since its separation from the rest of its genus; that it was a sort of hominin dodo, living on an island Eden in which the relaxation of mainland pressures allowed the loss of previously essential abilities – in short, that the little “hobbits” were cute, but dumb.

Against this, however, stands all remaining morphological and behavioural evidence. As regards morphology, first, Falk et al. (2005) have shown that the endocast resembles that of early *Homo* (the comparison specimens were *H. erectus*, *H. habilis* crania being relatively rare and poorly preserved) rather than those of apes or of earlier hominins assigned to *Australopithecus* and *Parathropus*. This includes not only its overall shape, but also some more specific morphological features.

First among these is the position of the lunate sulcus, which is relatively posterior, suggesting that this typical *Homo* reorganisation of the brain was present in *H. floresiensis*. Second, although Falk et al. (2005) decline to identify the convolutions that make up Broca’s area, they state that the morphology in the relevant region of the *H. floresiensis*’ brain is different from that of apes and that “the general morphology in this area would be consistent with their existence” (Falk et al. 2005: 245). Third, although not diagnostic of the genus *Homo*, it is interesting that *H. floresiensis* displays extremely pronounced bilateral torque, albeit in the opposite direction from the rest of its genus (a pattern sometimes seen today in left-handed

individuals, and more often in women than men [Bear et al. 1986]: note that the type specimen discussed here is judged to be female). These three specific features fit the suite seen in the genus *Homo* from the earliest times, and in the current theoretical context can be assumed to have been inherited from *erectus* or from an earlier shared *Homo* ancestor of the two species.

However, Falk et al. have also shown that *H. floresiensis* also differs in some ways from the *erectus* specimens. First the endocast, highly crenulated as it is, reveals still further degrees of crenulation of the brain surface in an area of the frontal lobe known as Brodmann's area 10. The relative degree of folding and hence by implication the expansion of this area exceeds that found even in modern *H. sapiens*. Falk et al. point out that, in both extant human and ape species, this region of the forebrain is associated with planning ahead (including the resumption of interrupted activities), and with instituting novel activities. Secondly, the temporal lobes (which the reader will recall also support our species' language skills) are described as extremely wide, again in comparison even to *H. sapiens*, let alone *H. erectus*.

The "hobbits" thus appear to have some distinctive "derived" features of the brain relative to *H. erectus*. That is, they possess some features which (on the face of it) evolved within the species at or after the time of their isolation from their ancestral hominin stock, be this erectine or earlier.⁷

What, then, were the realised cognitive and linguistic abilities of this peculiar species; what could demonstrably be achieved by a hominin with virtually chimpanzee-like EQ, yet showing (relative to these apes) derived features of brain configuration typical of *Homo* and (relative to the rest of the genus *Homo*) further derived features that are unique to the species, and seem almost an "exaggeration" of some of our own species' specialisations? To answer this one must turn again to the archaeological record.

3.2.2 Arguments from behaviour

3.2.2.1 *H. floresiensis* and water crossings. First, as yet nothing interesting can be said about *H. floresiensis*' capacity for water crossings. This is because it is not certain whether the species evolved on Flores itself as the descendant of a merely accidentally-rafted earlier *Homo* species, or whether it began its career as a species

7. Derived features of a lineage are those of its features which evolved after its split from some nominated ancestral stock. It is thus a relative term; feathers are a derived feature of birds when birds are compared with reptiles, but a "primitive" feature of parrots when they are compared with other birds. Low EQ has until now always been primitive in hominins, as it typifies the putative earlier forms and in the absence of other evidence would be considered simply a retention of that earlier form. However, if a larger-brained species (such as *H. erectus*) is indeed the immediate ancestor of *H. floresiensis*, then the latter's low EQ represents a derived feature of the species within the genus *Homo*; a departure from its ancestral *Homo* stock.

on the mainland or some other island, with its presence on Flores thus representing a separate series of island-hopping events. As for *H. erectus* above, its abilities here can only be determined by an extensive and long-term program of excavations across the island chain.

Turning to evidence for two other behavioural indicators, though, the current state of play on Flores is substantially different.

3.2.2.2 *H. floresiensis* and big game hunting. Associated with the *H. floresiensis* sequence is abundant evidence of hunting. Bones burnt by fire and scarred with the cut-marks of their butchery indicate the consumption of a variety of animals. Some are relatively small: bats and the endemic giant rats. But there is also abundant evidence at the site for the hunting of two larger and more dangerous animals: *Stegodon* (an endemically dwarfed elephant species about the size of a modern water buffalo) and Komodo dragon (which even today remains a threat to unwary *H. sapiens* in the region). After capture, these prey were evidently transported, at considerable cost of energy and time, to the cave – where clusters and circles of reddened and fire-cracked pebbles point to the presence of hearths (Morwood et al. 2005).

Many of the *Stegodon* were juveniles or neonates, indicating selective targeting of the more vulnerable individuals. Some, however, are adult. This reminds us that the term “pygmy” is relative; an adult pygmy *Stegodon* of the period weighed some 500kg. Even to today’s *H. sapiens* this would count as big game. To *H. floresiensis*, at not much more than half our height and weight, they were big game indeed. The frequency and success with which they were hunted point very strongly indeed to cognitive and linguistic skills well in advance of extant apes; arguably, also, in advance of *H. erectus*, for whom big game hunting and even the widespread use of fire is often discounted (e.g. Wynn 2002).

3.2.2.3 *H. floresiensis* and stone tools. Literally thousands of stone tools have been recovered in the sequence associated with *H. floresiensis*, of which approximately six thousand have been subjected to further study (Moore 2004, 2005). The assemblage includes small flakes struck from radial bipolar and burinated cores, as well as points, perforators, and microblades (Morwood et al. 2004).

From his technological analysis of the assemblage, Moore (2004) questions the degree of differentiation and standardisation in the artefact assemblage at the site – the features, note, that unequivocally indicate the presence of mental templates in the classic model. Moore also finds no evidence of prepared platforms, and hence of hierarchical operations at this level in the stone tool production.

However, Moore (2004) nevertheless emphasises that a co-ordinated mental and physical sequence is required for the consistent, error-free production of the basic flakes, and *H. floresiensis* showed a very high degree of skill and control in

such knapping. Further, though platform preparation may be absent, there are abundant indications of higher hierarchical cognitive patterns in other aspects of the stone tool manufacturing process. *H. floresiensis* gathered and initially processed the raw material for knapping outside the cave: River cobbles were size-selected and reduced, and the resulting segments of no less than 4–5 cm were then transported to the cave, where they would be further reduced by prolonged knapping to generate multiple usable flakes (Moore 2005). Furthermore, as part of this latter process, material would be cycled through no less than four different flaking techniques, and this in a patterned (non-random) fashion. Cognitively hierarchical sequences are evident in the total sequence of planned material testing, procurement, transport and modification.

As regards the presence of mental templates, with their classic association with symbolic thought, certain tools also stand out from the bulk of the assemblage both technically and formally. The standard flake production technique at Liang Bua was to remove just a couple of flakes, then flip the core to address another area, producing multiple usable flakes from one core (Moore 2005). However, to make one of the tools which is formally suitable for use as a perforator, 39 flakes were removed in sequence along the edge of the tool, which thus curves in relatively smooth concave lines to a sharp point. Technological difference coincides with formal difference, converging on a not unreasonable attribution of an intended goal shape or mental template in the mind of the tiny maker.

Finally, many of the flakes systematically produced from tiny bipolar, buri-nated and radial cores were too small for hand-held use. Why then, were they painstakingly produced?

In ourselves, such microliths are produced in order to be hafted. One possible explanation, then, is that they were also hafted by our cousin species; that is, that the technology of *H. floresiensis* indeed included composite tools. During preparation of this chapter, studies of the residues on some of these microblades have added an additional line of evidence to support this position, revealing patterns of resin residues on the tiny blades that are suggestive of hafting (Lentfer pers. com.). (Resin is the common glue used by *Homo sapiens* in many parts of the world when hafting stone blades onto wooden handles or spear shafts.) If so, then taken in the context of the total assemblage, the implications may be great for both classic and minimalist approaches to the significance of given grades of stone tool technology in relation to assigned cognitive skills.

3.2.2.4 The behavioural suite of *H. floresiensis*. What the evidence above indicates about the cognitive capacities of the “hobbit”, and the degree of difference between this species and its commoner closest nonhuman relative, *Pan troglodytes*, can be summarised by visualising the difference between the hunting ventures of the two.

Chimpanzee behaviour differs in different regions, but the hunt may be preceded – at most – by careful gazing at the prey (usually monkeys in the canopy above), followed by a burst of excited and positive social interaction among the hunters (Goodall 1986); then the hunt begins. Among *H. floresiensis* (unless we are to assume that the “hobbits” too flailed young *Stegodon* to death), it appears that any hunting was preceded, long before this stage, by the preparation of stone tools. Furthermore, this preparation itself sometimes involved temporally and geographically separated stages within hierarchically-composed operations.

During the actual hunt, chimpanzees in some regions (though not others) are indeed argued to cooperate (Boesch & Boesch 1989). The degree of co-ordination necessary to bring down the far larger and more dangerous prey of *H. floresiensis* points to an equal or greater capacity for cooperation and role division among the little “hobbits”. After the capture of the prey, the differences between the “hobbits” behaviour and that of common chimpanzees increases further. Among apes, the prey is torn apart and eaten on the spot. However, in an echo of the stone tool manufacturing sequence, among *H. floresiensis* the larger *Stegodon* were apparently butchered to smaller, more portable joints, and next, yet more time and energy was invested in carrying the joints, or the entirety of smaller prey, back to the cave.

Here, still further preparations precede or have preceded a meal. River pebbles are gathered and transported to the cave, where a hearth is then constructed; fuel is also gathered and brought in from outside the cave, in another series of operations necessary to create a fire, and then sustain it. Finally, long after the chimpanzees have wiped the last stains of blood from their chins and dispersed, the “hobbits” already cooked and ate their prey, accompanied by further dissection using tools that have left their marks upon the already charred bones.

This brings us to one more distinction between *H. floresiensis* and chimpanzees. Among the latter, the meat at a kill site is sometimes shared to a small degree: During consumption, successful hunters may respond to begging gestures by donating some of their meat to the begging animal, particularly if it is a political ally (Nishida et al. 1992) or an oestrus female (Stanford 1995). What about *H. floresiensis*? In particular, given that *H. floresiensis* walked upright, it would seem likely that females experienced the same difficulties that upright posture presents to modern humans carrying infants (including, note, even the more precocious ape infants). That is, one arm at least must be given to completely supporting the infant against the now-vertical body (see Savage-Rumbaugh, Shanker & Taylor 1998). Chimpanzee mothers are less handicapped in this way, and on occasion will hunt in some regions (Boesch & Boesch 1989), but for the mother of infant hobbits, haring about the countryside after large, swift and dangerous prey appears out of the question. Did this section of the community – the ones dealing with the greatest nutritional demands – simply go without the meat of all but the smallest and most

easily captured animals? In fact, did “hobbit” mothers face the regular theft of their own prey by males, that is so common an experience among successful female chimpanzee hunters? Though possible, new evidence makes this seem unlikely.

For there is one more find recently unearthed at Liang Bua that points to positively human levels of cooperation. It is the radius of an adult individual which has healed from a break so severe as to leave the final radius bent almost at right angles (Morwood et al. 2005). Clearly, the individual survived the many weeks or months of severe incapacitation that the complete bone severance and the subsequent difficult angle of the healing would have demanded. He or she also went on to survive much longer still, with a spectacular deformation of the arm; one that seems to have restricted its use considerably. Not just the limited sharing of the occasional meat bonanza, but prolonged social provisioning of diverse food types, at a very high level, appears to have been within the capacity of *H. floresiensis*.

In the light of this evidence, it appears reasonable to consider the possibility that the species had a social structure in which groups or individuals foraged as they could, but at least sometimes shared the results around a fire at home base. The unencumbered logically must have provided a greater proportion of the meat of big game; the females with small young may have provided a higher proportion of any plant matter and small prey that made up the balance of the diet.

3.2.3 *Evidence of behaviour: Concluding remarks*

Where one is dealing with isolated fragments of evidence, minimalist interpretation is a rational response. Where multiple lines of evidence repeat the same patterns, however, the concept of what is the most conservative or parsimonious explanation begins to change.

H. floresiensis appears to present not one but several examples of behaviour that entails co-ordination of diverse activities, separated in space and/or time, yet all necessary to and culminating in a final product or advantageous end. Individually, these behaviours have often been taken as evidence of long-term planned action that demands a capacity for extended hierarchical cognitive operations, an ability to hold in the mind a concept distinct from current reality, and to communicate this concept via symbols, albeit not necessarily grammar. Turning to morphology, the reorganised brain of *H. floresiensis*, with its broadened temporal lobes and elaborated Brodmann’s area 10, converges with behaviour that demands just such brain changes. Taken together, then, we suggest that the entire suite of morphology and behaviour in *H. floresiensis* is most parsimoniously accounted for simply by attributing to the species the mastery of these distinctively human-like cognitive and linguistic capacities.

On the other hand, the pervasive association between mammalian brain size and intelligence still stands: Size may not rule supreme, but it still rules in part (see

also below). Note, too, that there is no evidence that the snack-sized “hobbit” brain could produce the diverse, standardised and complex technology that typifies *H. sapiens* from about 40,000 onwards. In combination with their tiny albeit reconfigured brains, this absence of evidence takes on somewhat more weight than it would alone. Just how smart the “hobbits” could have been under ideal conditions, we will never know – but on balance, though well above apes, it is a safe bet that they were also substantially less intelligent than ourselves.

The final outcome, then, is a picture of a cogitating, symbolically-communicating, co-operating species, that nevertheless adapted morphologically to its island conditions. It did so with a sacrifice of body and brain size, and probably of net intelligence too, that is fairly typical of non-human animals on remote islands.

This is a picture that makes us question what it means to be categorically human or ape. However, such tension between the categorical and the continuous in our models of the past is a predictable result of the evolutionary process, part and parcel of standard evolutionary scenarios. It is to these that we now turn.

4 Implications of the Flores hominins for the evolution of cognition and language

4.1 A downgrading of EQ as the sufficient cause of cognitive and linguistic abilities

First the Flores findings demonstrate, not that the relationship between EQ and cognitive and linguistic abilities in hominins is unimportant, but certainly that it is looser than has hitherto been assumed. As part of the total suite of features that make up a hominin species, other aspects must be capable of creating larger changes in realised behaviour than have previously seemed possible at the low-EQ end of the hominin range.

As part of this observation, and for the very first time, the case of *H. floresiensis* also strikes at the root of arguments that enhanced cognitive abilities and/or language are or could be entirely the epiphenominal products of sheer bulk of brain, relative or absolute (e.g. Chomsky 1982; Gibson, Rumbaugh & Beran 2001). This perennial and appealingly parsimonious argument can be rescued now only by replying that there is an important threshold level of brain bulk for such behaviours as are seen in *H. floresiensis*. Chimpanzees would then be far closer to this major threshold than has hitherto been assumed, while *H. floresiensis* would just squeak over to the other and far more gifted side.

Against any such “modified threshold” argument stands, first, the fact that other hominin species of equal or larger cranial gifts, from australopithecines on, show

far less evidence of a spectacular threshold effect. A second, logically more minor point is the fact that subsequent hominin species continued to invest in ever-larger brains and EQs. This is an enormously expensive evolutionary strategy, and its continuance would seem at least a little odd if all of the behaviour that later and larger hominin brains demonstrably supported, up and well into the history of our own species, was already possible at brain size and EQ in the australopithecine range.

4.2 An added emphasis on the significance of brain reorganisation?

Conversely, the case of *H. floresiensis* appears to highlight the significance of brain reorganisation. In the past, the *Homo* reorganisation that includes a relatively posterior lunate sulcus has always been found on relatively and absolutely larger hominin brains. Thus, rather than losing a percentage of the primary visual cortex (PVC) that the lunate sulcus demarcates, the species could be said to be gaining in other regions of the brain, be they modelled as parietal, frontal, or indeed temporal (Deacon 1997; see also Wilkins & Wakefield 1995). The “hobbit”, however, was not in this win/win situation. In absolute terms its brain is comparable in size to that of a modern chimpanzee, and falls below the australopithecine average (Holloway, Broadfield & Yuan 2004), while it is both relatively and absolutely smaller than the brain of *H. habilis* and *H. erectus*. Depending on the position taken regarding australopithecine brain reorganisation, the “hobbit’s” reorganisation would on the face of it appear to represent a genuine sacrifice of PVC relative to other hominins.⁸ This implies either that the species was unique in its lineage in having an especially reduced role for the visual processing functions that the PVC supports – which in the total context seems vanishingly unlikely – or that some further adaptation of the species’ brain permitted the variation.

Here the possibilities are many, and all debateable. The crenulations that typify the “hobbit” may have continued medially, tucked in between the hemispheres, with this development enabling the posterior position of the lunate sulcus on the external surface of the brain (Dean Falk, pers. com. July 2005). If so, some other region in the brain would still have to “give”, but it need not be PVC. Alternatively, the recently-documented difference between PVC in extant humans versus some other extant primates, tentatively linked by its finders to superior visual abilities in humans (Preuss, Qi & Kaas 1999), may be another answer. If the “hobbit’s” PVC was indeed small but efficient by virtue of already possessing the distinctively and arguably superior organisation seen on ourselves, this would also date the newly-discovered differences in PVC to a period preceding the branching of *H. floresiensis* from the

8. Holloway (1975) has long defended the position that australopithecines also possess the feature of posterior location of the lunate sulcus.

rest of the genus *Homo*. However, this line of argument rests at many points on new evidence and inferences as yet unsubstantiated. Alternatively again, and with just the same peculiarity in mind as regards his position on australopithecines (see Note 8 and Jerison 1984), Holloway and colleagues document individual variation in the position of the lunate sulcus in chimpanzees which suggests that, at least in captive specimens, individuals can operate successfully with unusually small PVC areas (Holloway, Broadfield & Yuan 2001). So far, though, the described variation in chimpanzees does not unequivocally extend to posterior location of the lunate sulcus to the degree Holloway proposes for australopithecines. Whether it would encompass the location of the lunate sulcus in *H. floresiensis* is as yet unclear.

What is clear is that there is a problem raised by the “hobbit’s” retention of *Homo*-like brain reorganisation in the face of brain reduction to the size of a chimpanzee. This brain reorganisation has been argued to support cognitive abilities such as advanced spatial skills, multimodal association, and language (e.g. Wilkins & Wakefield 1995; Deacon 1997; Holloway, Broadfield & Yuan 2004; cf. also Stout, Toth & Schick 2000). If, as appears, the reorganised brain of *H. floresiensis* was indeed generating such potentially adaptive behaviours, then any evolutionary contortions that allowed retention of *Homo*-like brain reorganisation in the face of shrinking absolute size would have been backed by powerful selective pressures.

4.3 Frontal lobes, temporal lobes and “human” behaviour

However, it is in any case unlikely that all of the “hobbits” talents can be attributed to its *Homo*-like brain organisation. Indeed, it is unlikely that any or all of the morphological features that *H. floresiensis* shares or may share with *H. erectus* – be they brain reorganisation, cerebral torque, Broca’s area, or all three together – are sufficient to support the “hobbit’s” behaviour. After all, these features are also found together in other early members of the genus *Homo* without unequivocally producing the same degree of demonstrated behavioural effects. In logical terms then, these brain changes may have been necessary to support *H. Floresiensis*’ behaviour, but for as long as species which possess them, such as *H. erectus*, are held to be nevertheless less behaviourally advanced, the archaeological record as a whole suggests that the shared brain features were not in themselves sufficient to support hobbitine behaviour.

The picture is different when one turns to the significance of the derived features of *H. floresiensis*’ brain, regarding Brodmann’s area 10 and the temporal lobes. In both modern human and nonhuman primates, Brodmann’s area 10 supports aspects of cognition that allow long-term planning. This includes, in ourselves, memory retrieval, the planning of adaptive responses to novel conditions, “multi-tasking” and the resumption of interrupted tasks (Allman, Hakeem & Watson

2002; Burgess et al. 2005). The temporal lobe in primates and ourselves supports a variety of functions including (again) aspects of memory, plus audition and comprehension of vocalisation – this latter including, in ourselves, language (Bear, Connors & Paradiso 2001).

The unique derived features of the brain seen in *H. floresiensis* are thus relevant to the behavioural changes to be accounted for in this species: Many of their behaviours point to substantial abilities for long-range planning, including the co-ordination of diverse activities towards a single end over an extended time frame, while the association of symbol-use with hunting at the level seen in *H. floresiensis* is long-standing.

Further, the brain changes themselves seem parsimoniously accounted for in no other way to date. Importantly, if both smaller and larger-EQ species lack the features of relatively expanded Brodmann's 10 and broadened temporal lobes to the degree seen in *H. floresiensis*, then allometric trends cannot encompass the features (i.e. they are not merely passively towed along as the automatic result of a brain getting bigger). In such a situation, one would normally conclude that the features were actively favoured by the classic process of natural selection in the course of the species' evolution.

4.4 Language, cognition and parallel evolution

If *H. floresiensis*' cognitive abilities and ours do indeed both exceed those of our most recent common ancestor (*erectus* or not), then the two of us together represent a rather strong case of parallel evolution. This is a well-attested pattern of evolutionary change in which daughter lineages of a common and relatively recent ancestral species subsequently and independently evolve in the same direction – with the result that the sister lineages are more similar to each other, on some relevant measures of interest, than they are to their common ancestor.⁹ In the case of

9. Thus armour plate reduction in different species of freshwater stickleback, which has occurred independently many times around the world from a marine stickleback starting point, is a case of parallel evolution *par excellence* (Colosimo et al. 2005).

It is relevant here to distinguish parallel from convergent evolution, as the distinction justifies the emphasis on the status of early *Homo* that follows in this chapter. Parallel evolution is the product of similar selective pressures acting on similar genetic material, to produce features that owe their similarity to both environmental and phylogenetic (or “ancestral starting point”) commonalities. (Thus, for example, in the case of freshwater sticklebacks, the very same genetic feature of the marine species has been exploited many times in different locales to produce the same change from a marine ancestral species to a freshwater one; Colosimo et al. 2005.) Convergent evolution, though, is the product of similar selective pressures acting on very different genetic and phenotypic material, to produce only superficially similar solutions each time. The fact that both fish and dolphins have similar external body morphology is a case of convergent evolution.

H. floresiensis and ourselves, the relevant measures of interest are the presence of relative enlargements in temporal and specific frontal lobe areas, together with the behavioural shift to fire use, big game hunting, and (it now seems) compound tool production. As described above, these brain changes appear to be lacking in early *Homo*, and the behavioural changes are known or argued to be lacking as well. This gives two directions for future research.

4.4.1 *A new emphasis on early Homo*

The first important point is that in parallel evolution the similar marked evolutionary careers of the daughter lineages are assumed to be attributable, in part, to qualities of the ancestral species. It is a reasonable suspicion, and employs exactly the same reasoning as one would use in accounting, in part, for the extravagances of the Mitford girls by scrutinising their otherwise undistinguished parents. Few people become famous (or notorious); not just one but most of the Mitford sisters, independently, became famous (or notorious); what, one asks, were their parents like?¹⁰ The inference may seem unjustifiable to those accustomed to the strength of formal logic, but notice the use of a broader pattern than the mere fact of change in one child (i.e. species) or even one group of siblings (i.e. many related species); in the case of the Mitford sisters, one also automatically takes into account the background frequency of fame at that time and place when one points the finger at their parents. Similarly, the background frequency of a novel feature in a particular era and ecosystem is also relied upon to pick out the kind of situation in which suspicion really does, reasonably, fall on ancestral species.

The argument has thus come full circle, from downplaying the significance of the cognitive abilities of early *Homo* to scrutinising them. Even if the behaviours seen in *H. floresiensis* are denied to early *Homo*, the case of *H. floresiensis* still raises this question: What changes, in place at least by our last shared ancestor with *H. floresiensis*, acted as a platform on which the distinctive derived aspects of both human and “hobbit” brain and behaviour could next independently and advantageously be built? Was this all-important foundation the brain reorganisation, for which *H. floresiensis* at least appears to pay so problematically in PVC? Was it

As fish and dolphins are still ultimately related, albeit many hundreds of millions of years in the past, the issue of recency of the common ancestor is important in deciding whether an observed change is the result of convergent or parallel evolution – that is, in practical terms, how much of a facilitating foundation role to attribute to the common ancestor in accounting for the evolutionary trends later seen. The entire hominin lineage has arisen in only 6–7 million years, the flicker of an evolutionary eyelash, so parallel evolution rather than convergent evolution is the standard process to invoke here.

10. The parents themselves deny all responsibility. According to their father, “I am normal, my wife is normal, but my daughters are each more foolish than the other” (Lovell 2002).

perhaps the expansion of Broca's area, with which the hobbit's brain too is "not inconsistent" (Falk et al. 2005)? Was it both of these, and more? It is a question well beyond the scope of this chapter – but it is worth reiterating that the case of the "hobbit" focuses attention on the ancestral species we share with *H. floresiensis*, as an important platform for later changes in brain morphology and evinced behaviour. This holds true whether that earlier common ancestor is finally held to be *H. erectus*, or an even earlier hominin.

4.4.2 *And a welcome source of new constraints on model building*

There is also a second important implication from the model of parallel evolution. An earlier ancestral condition lays a foundation for the subsequent parallel development of other features, but (in the absence of other evidence) it is not usually modelled as compelling it, with the later changes inevitably spilling out merely with the passage of time. Rather, the daughter lineages are normally modelled as independently encountering a new environmental situation that is absent for the parental species. This new environment, acting in each case on a similar ancestral set of features, is held to be equally necessary in making a certain kind of evolutionary change both possible and advantageous.

So what similar pressures, applied to two species of hominin, could have promoted the independent appearance of very similar brain features and cognitively-supported behaviours many millennia later? Was it a change in resource distribution, a change in population density, size or structure, a change in social organisation, in patterns of exchange, or some chain of many of these events – or something else again? The hypotheses about just what triggered the transition from a pre-human to some interestingly human grade could fill volumes (and do). Again, then, this is a question for the future, well outside the scope of this chapter.

Until now, though, there has been only one incontrovertible case study to work from – ourselves. This has made it very difficult to model the evolution of those very features of ourselves that are most distinctive and interesting to us; there have been few means of constraining hypothesis, little chance of cross-checking proposed evolutionary scenarios. Now, standard models of parallel evolution lead to the expectation that, in broad terms, some of what happened to our ancestors in Africa probably also happened to *H. floresiensis* in Southeast Asia. By checking for similar patterns in the history and environment of *H. floresiensis*, and weighing these against controls elsewhere, some scenarios can be supported and others perhaps excluded. As ever, the inference is not watertight; but it is a routine line of modelling that is certainly a great deal better than nothing. *H. floresiensis* may prove extremely useful in providing a greatly needed second case that will help to guide and constrain future modelling in the evolution of language and cognition.

5 Conclusion

As regards the Flores hominins themselves, morphological and behavioural evidence converges on a model of *H. floresiensis* as a thinking, planning, symbolically-communicating species. As regards our understanding of hominin evolution generally, this coherent picture creates a further coherence at another level; it brings our view of our species' cognitive gifts into line with regular evolutionary patterns, normalising our situation vis-à-vis the rest of the animal world. Rather than being a species apart, unmatched and unmatchable in all important respects – a species for which all the usual patterns of argument must be disallowed – the Flores case shows *H. sapiens* as in truth one of many species in a lineage that has long specialised, so to speak, in intelligence, or the individual plasticity in social, communicative and ecological behaviour that goes by this name.

As regards the evolution of language and cognition, so far the Flores finds have overturned a long-standing possibility that enlargement of the brain could overshadow all else in the transition from ape to human-like levels of cognition. The finds have also pointed to the very dawn of *Homo*, the species standing at the branching point of *H. floresiensis* and ourselves, as being of great interest in modelling the evolution of language and cognition. Early *Homo* species' report cards, as delivered by archaeologists, have commonly read little better than "tries hard, may do better next millennium". However, the Flores hominins suggest that by the appearance of even the earliest *Homo* species, key changes were in place that later enabled the emergence of more archaeologically visible and overtly "human" developments in hominin behaviour and inferred cognition. The brain of *H. floresiensis* also supports models that look, for the basis of these subsequent developments, in our species' enlargement of temporal lobe and specific frontal lobe areas. Finally, in combination with standard models in evolutionary theory, the Flores species and the Wallacean region promise in the future to provide much-needed constraints in modelling of the evolution of language and cognition.

For archaeology, the immediate imperative is to expand the known record of hominin evolution in Flores and surrounding islands. What species were present? What was their ecological context? What clues for behaviour, and hence for cognition and language, can be found in association with each species?

For researchers in cognitive and linguistic fields, the Flores hominins may encourage us to pursue the more expanded models for the evolution of language and thought. If the impetus to language and planning had its seeds at or near the very dawn of the genus *Homo*, and if these abilities can then flourish in a hominin with an EQ very little more than that of modern chimpanzees, then models which one way or another stress an extended time frame for language become of greater interest (e.g. Bickerton 1990; Deacon 1997; Pinker 1994).

Input will also be necessary from other disciplines. At the outset of this chapter, we stated that in the biological sciences, and especially in reconstruction of the past, certainty is a luxury that one must learn to live without. We put it that this does not open up a field to unconstrained speculation: One is still bound to observe relative degrees of reasonableness. Together with the use of broad patterns based on extensive comparative data, it is clearly reasonable to call for the maximum of case-specific evidence.

This is where a consistent application of reasonableness becomes severely constraining. For in the dangerous business of moving from the general to the particular – from common patterns of evolutionary change to the particular case of the evolution of a human-like mind – a huge amount of relevant information is possessed by a sweep of disciplines from animal communication to zoogeography. Each discipline offers constraints; patterns of data and of argument that in different ways narrow down what is possible and impossible, likely or less likely, in the evolution of the mind. A solid understanding of the appearance of the human mind on earth will only be constructed, extremely painstakingly, from the aggregation, integration, and mutual adjustment of theory and data from all these many disciplines. Given the many researchers now working on these tasks, and the increasing communication amongst them, it is reasonable to hope that we will eventually converge on a coherent, mutually-acceptable model, and a stronger understanding of the abilities that underpin the current success of ourselves, the only hominin species that now remains.

References

- Allman, John, Atiya Hakeem & Karli Watson. 2002. Two phylogenetic specializations in the human brain. *Neuroscientist* 8. 335–346.
- Bear, D., D. Schiff, J. Saver, M. Greenberg & R. Freeman. 1986. Quantitative analysis of cerebral asymmetries. *Archives of Neurology* 43. 598–603.
- Bear, Mark F., Barry W. Connors & Michael Paradiso. 2001. *Neuroscience: Exploring the brain*. Baltimore, MD: Lippincott Williams & Wilkins.
- Bickerton, Derek. 1990. *Language and species*. Chicago, IL: The University of Chicago Press.
- Boesch, Christophe & Hedwige Boesch. 1989. Hunting behavior of wild chimpanzees in the Tai National Park. *American Journal of Physical Anthropology* 78. 547–573.
- Brown, Peter, Thomas Sutikna, Michael J. Morwood, Rapien P. Soejono, Jatmiko, E. Wahyu Saptomo & Rokus Awe Due. 2004. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431. 1055–1061.
- Burgess, Paul W., Jon S. Simons, Iroise Dumontheil & Sam J. Gilbert. 2005. The gateway hypothesis of rostral prefrontal cortex (area 10) function. In John Duncan, Louise McLeod & Peter Phillips (eds.), *Measuring the mind: Speed, control, and age*, 215–246. Oxford: OUP.

- Jerison, Harry J. 1984. Fossil evidence on the evolution of the neocortex. In Edward Jones and Alan Peters (eds.), *Comparative structure and evolution of the cerebral cortex*, Vol 1. 285–309. New York, NY: Plenum Press.
- Lee, Richard B. & Irvén DeVore. 1968. *Man the hunter*. Chicago, IL: Aldine.
- Lewin, Roger & Robert A. Foley. 2004. *The principles of human evolution*. Oxford: Blackwell.
- Lovell, Mary S. 2002. *The Mitford girls: Biography of an extraordinary family*. London: Abacus.
- McDougall, Ian, Francis H. Brown & John G. Fleage. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433. 733–736.
- McHenry, Henry M. 1982. The pattern of human evolution: Studies on bipedalism, mastication and encephalization. *Annual Review of Anthropology* 11. 151–173.
- Moore, Mark W. 2004. Middle Pleistocene blade technology at Liang Bua Cave, Western Flores. *Australian Archaeological Association Conference: Networks and Narratives*, 13–15 December 2004, Armidale, Australia.
- Moore, Mark W. 2005. The design space of lithic technology. Armidale, NSW: University of New England PhD thesis.
- Morwood, Michael J., Paul B. O'Sullivan, Fachroel Aziz & Asaf Raza. 1998. Fission-track ages of stone tools and fossils on the east Indonesian island of Flores. *Nature* 392. 173–176.
- Morwood, Michael J., Fachroel Aziz, Nasruddin, Douglas R. Hobbs, Paul B. O'Sullivan & Asaf Raza. 1999. Archaeological and palaeontological research in central Flores, east Indonesia: Results of fieldwork, 1997–98. *Antiquity* 73. 273–286.
- Morwood, Michael J., Radian P. Soejono, Richard G. Roberts, Thomas Sutikna, Chris S. M. Turney, Kira E. Westaway, William J. Rink, Jian-Xin Zhao, Gert D. van den Bergh, Rokus Awe Due, Douglas R. Hobbs, Mark W. Moore, Michael I. Bird & L. Keith Fifield. 2004. Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* 431: 1087–1091.
- Morwood, Michael J., Peter Brown, Jatmiko, Thomas Sutikna, E. Wahyu Saptomo, Kira E. Westaway, Rokus Awe Due, Richard G. Roberts, T. Maeda, Sri Wasisto & Tony Djubiantono. 2005. Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* 437: 1012–1017.
- Nishida, Toshisada, Toshikazu Hasegawa, Hitoshige Hayaki, Yukio Takahata & Shigeo Uehara. 1992. Meat-sharing as a coalition strategy by an alpha male chimpanzee. In Toshisada Nishida, William C. McGrew, Peter Marler, Martin Pickford & Frans B. M. de Waal (eds.), *Human origins* (Topics in Primatology 1), 159–174. Tokyo: University of Tokyo Press.
- Noble, William & Iain Davidson. 1996. *Human evolution, language and mind: A psychological and archaeological enquiry*. Cambridge: CUP.
- O'Connell, James F. & James Allen. 2004. Dating the colonisation of Sahul (Pleistocene Australia-New Guinea): A review of recent research. *Journal of Archaeological Science* 31. 835–853.
- O'Connor, Sue, Matthew Spriggs & Peter Veth. 2002. Excavation at Lene Hara cave establishes occupation in East Timor at least 30,000–35,000 years ago. *Antiquity* 76. 45–49.
- Pinker, Stephen. 1994. *The language instinct: The new science of language and mind*. Harmondsworth: Penguin.
- Preuss, Todd, Huixin Qi & Jon H. Kaas. 1999. Distinctive compartmental organization of human primary visual cortex. *National Academy of Sciences, USA* 96. 11601–11606.
- Roberts, Richard G., Rhys Jones, Nigel A. Spooner, M. John Head, Andrew S. Murray & Michael A. Smith. 1994. The human colonisation of Australia: Optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. *Quaternary Science Review* 13. 575–583.

- Roebrooks, Wil. 2001. Hominid behavior and the earliest occupation of Europe: An exploration. *Journal of Human Evolution* 41. 437–461.
- Savage-Rumbaugh, Sue, Stuart G. Shanker & Talbot J. Taylor. 1998. *Apes, language and the human mind*. Oxford: OUP.
- Schick, Kathy, Nicholas Toth, Gary Garufi, E. Sue Savage-Rumbaugh, Duane Rumbaugh & Rose Sevcik. 1999. Continuing investigations into the stone tool-making capabilities of a Bonobo (*Pan paniscus*). *Journal of Archaeological Science* 26(7). 821–832.
- Stanford, Craig B. 1995. Chimpanzee hunting behavior and human evolution. *American Scientist* 83(3). 256–261.
- Stout, Dietrich, Nicholas Toth & Kathy Schick. 2000. Stone tool-making and brain activation: Positive emission tomography (PET) studies. *Journal of Archaeological Science* 27. 1215–1223.
- Thieme, Hartmut. 1997. Lower Paleolithic hunting spears from Germany. *Nature* 385. 807–810.
- Tobias, Phillip V. 1987. The brain of *Homo habilis*: A new level of organization in cerebral evolution. *Journal of Human Evolution* 16. 741–761.
- Torre, Ignacio de la. 2004. Omo revisited: Evaluating the technological skills of Pliocene hominids. *Current Anthropology* 45(4). 439–465.
- Toth, Nicholas, Kathy D. Schick, Sue Savage-Rumbaugh, Rose A. Sevcik & Duane M. Rumbaugh. 1993. *Pan* the tool-maker: Investigations into the stone tool-making and tool-using capabilities of a bonobo (*Pan paniscus*). *Journal of Archaeological Science* 20. 81–91.
- Vos, John de, Paul Yves Sondaar, Gert D. van den Bergh & Fachroel Aziz. 1994. The *Homo* bearing deposits of Java and its ecological context. In Jens Lorenz Franzen (ed.), *100 years of Pithecanthropus: The Homo erectus problem*, 129–140. Frankfurt: Forschungsinstitut Senckenberg.
- Wilkins, Wendy K. & Jennie Wakefield. 1995. Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences* 18. 161–226.
- Wynn, Thomas. 2002. Archaeology and cognitive evolution. *Behavioral and Brain Sciences* 29. 389–438.
- Zilles, Karl, Andreas Dabringhaus, Stefan Geyer, Katrin Amunts, Meishu Qü, Axel Schleicher, Emmanuel Gillissen, Gottfried Schlaug & Helmut Steinmetz. 1996. Structural asymmetries in the human forebrain and the forebrain of non-human primates and rats. *Neuroscience and Biobehavioral Reviews* 20. 593–605.

Evolving artificial minds and brains

Pete Mandik, Mike Collins and Alex Vereschagin

We explicate representational content by addressing how representations that explain intelligent behavior might be acquired through processes of Darwinian evolution. We present the results of computer simulations of evolved neural network controllers and discuss the similarity of the simulations to real-world examples of neural network control of animal behavior. We argue that focusing on the simplest cases of evolved intelligent behavior, in both simulated and real organisms, reveals that evolved representations must carry information about the creature's environments and further can do so only if their neural states are appropriately isomorphic to environmental states. Further, these informational and isomorphism relations are what are tracked by content attributions in folk-psychological and cognitive scientific explanations of these intelligent behaviors.

1 Introduction

Many kinds of explanations of intelligent behavior make reference to mental representations, that is, they explain an organism's ability to behave intelligently in virtue of an organism's having mental representations. The existence of such explanations, "representational explanations" for short, raises many questions, of which two will be the focus of this chapter. The first is the question of whether representational explanations of intelligent behavior are the best explanations of intelligent behavior or if we might instead do better with explanations that make no reference to mental representations. We will argue that we can do no better than representational explanations. The second question that arises is the question of what representational explanations are referring to when they refer to representations. What *are* representations? We demand not just an account of what representations are, but additionally we demand an account that explains how representations can be the sorts of things that help explain intelligent behavior. We will sketch such an account. The goal of this chapter then, is twofold: to argue that we need representations to explain

intelligent behavior and to sketch an account of what sorts of things representations must be if they are to explain intelligent behavior.

Several opponents of representational explanations have built their cases by starting with the simplest examples of intelligent behavior and attempting to demonstrate that in such examples, no representations are to be found and thus, no representations need be referred to in order to explain the behaviors at hand. This is the strategy followed, for example, by roboticists and artificial intelligence researchers such as Brooks (1991) and Beer (1990) in their arguments for the possibility of intelligence without representation. We will employ a similar strategy but we will be drawing different conclusions. We will examine some of the simplest cases of intelligent behavior and demonstrate that in these cases the behavior at hand is best explained in terms of representations. Further, our account of representations will be fully realist and reductive. To say that the account is realist is to say that the attributions aren't purely instrumental ways of speaking *as if* the creatures had representations. It is instead to pick out states of creatures that would be there independently of our speaking of them. To say of our account that it is reductive, we will be identifying representational states in ways that are straightforwardly explicable in terms of states of creatures' nervous systems and relations between their neural states and environmental states.

One way to examine the simplest examples of intelligent behavior is to examine the simplest examples of organisms that behave intelligently. This strategy confers the following advantage. The simpler the creature the easier it will be to keep track of the creature's internal structures, the structures of the creature's environment, and the relations between the two kinds of structure in virtue of which the former count as representations of the latter. Further, dealing with extremely simple cases will allow for tractable computer simulations of creature behavior as well as simulations of the evolutionary forces that contribute to the emergence of such behaviors.

Our motive for caring about the evolutionary background of the simplest cognitive behaviors emerges from the following presumptions. We presume, and are unlikely alone in doing so, that the simplest forms of intelligent behaviors are adaptive. That is, intelligent behaviors, at least of the simplest varieties, provide biological benefits to the organisms that perform them. We presume also that just as there was a time in the history of the universe that there were no biological organisms, there was a time in the history of the universe that there were no organisms performing intelligent behaviors. Since *abiogenesis* is the term referring to the hypothesized emergence of life from non-living matter, we coin the term *apsychogenesis* to refer to the hypothesized emergence of intelligence from non-intelligent systems. When, in the history of the universe did abiogenesis and apsychogenesis occur? No one knows, but we doubt that apsychogenesis preceded abiogenesis. They either coincided or abiogenesis occurred first. However, the latter option

strikes us as the more plausible of the two. Adding to our growing list of presumptions, we further presume that the problem of understanding apsychogenesis is best understood in the context of an evolutionary framework. Thus we are led to ask: What pressures applied to non-intelligent organisms yielded the earliest and simplest forms of intelligence? If mental representations are to underwrite intelligent behavior, then questions of the evolvability of intelligence will be closely related to questions of the evolvability of mental representations.

We will tackle the topics of intelligence, representation, and evolution by examining computer simulations of evolved adaptive behaviors. The simulated organisms, behaviors, and environments will be simple enough to make tractable questions concerning the relations that constitute representation and the roles representations play in adaptive intelligent behaviors.

The structure of the rest of the chapter is as follows. First we will briefly examine a few cases in which representations are invoked to explain the intelligent behaviors of humans and non-human animals. The goal here will not be to extract a definition of representation from these examples but instead to only note a few key features of the roles representations play in such explanations. Formulating a definition of representation is a goal to be achieved (or at least approximated) toward the end of the chapter and not a presupposition to be made at its outset. Following the examination of these sample explanations, we will describe the basic intelligent behavior of positive chemotaxis and highlight the ways in which the problem that chemotaxis poses for organisms can be solved in a variety of ways involving representations. Next we describe mathematical and computer models of positive chemotaxis. The models are informed by neuroanatomical and neurophysiological data from real animals. Finally we discuss what account of representation seems best supported by the models.

2 Mental representations in explanations of intelligent behavior

Let us take a brief look at a folk-psychological explanation of a piece of intelligent behavior. Consider George. George is opening a refrigerator. Why? What explanation is available for this action? A folk-psychological explanation will advert to a collection of psychological states that jointly constitute a cause of George's behavior. An example collection of such states would include a desire, a perception, and a memory. One explanation of George's behavior then would advert to George's *desire* to drink some beer, George's visual *perception* that there is a refrigerator in front of him, and George's memory that he put some beer in the refrigerator the day before.

There are a few useful points to note about this explanation. First, the psychological states are not individually sufficient to cause a behavior, but must act in concert. A belief that there is beer in front of you will contribute to causing you to move toward it if combined with a desire for beer and will contribute to causing you to move away from it if combined with a fear of beer. Similarly, a desire for beer will contribute to causing you to move forward if combined with a belief that beer lies ahead and cause you to move in some other direction if combined with some other belief. In summary, psychological states contribute to the causes of behavior by acting in concert.

A second useful point to note about this sort of explanation is that the psychological states are identified in part by their representational content and in part by what attitude the person is taking toward that content. In the case of George's memory that he put some beer in the refrigerator, the representational content of the memory is that George put some beer in the refrigerator and the attitude is one of remembering. Different types of attitude can be taken toward one and the same content (e.g. *remembering* buying beer; *planning on* buying beer) and one and the same attitude type can be taken toward different contents (e.g. perceiving *that there is a beer in front of me*, perceiving *that there is a slice of pizza in front of me*). In summary, psychological states that are causes of intelligent behaviors admit of a distinction between their representational contents and the attitude that is taken toward those representational contents.

A third useful point to note about these sorts of explanation is that we can make attributions of content without explicit knowledge of what, in general, representational content is. We construct such explanations on the fly without knowing, for example, what the right theory of content is or even having a theory of content in mind. We plan to exploit this in what follows. We will present relatively clear cases of synthetic organisms that behave in ways explainable in terms of representational states and we will do so before offering a definition of what representations are or what representational content is. This leaves open to empirical investigation what the best accounts of representation and content are as opposed to a matter that must be settled *a priori* before such investigations take place.

It is worth noting that the power of representational explanation is not simply some story we tell ourselves and each other sustained by our own (possibly mistaken) views of ourselves. One way to appreciate the power of such explanations is to appreciate them in the context of explaining the behaviors of non-human animals. The literature is filled with such examples. We briefly mention just a few. Consider the impressive feats of maze learning exhibited by rats. A Morris water maze is filled with water rendered opaque to obscure a platform that will offer a rat a chance to rest without having to tread water. When placed in the maze for the first time, a rat will explore the area and eventually find the platform. When the rat

is returned to the starting position, the rat does not repeat the exploratory strategy but instead swims straight to the remembered location of the platform. Apparently, the perceptual inputs gained during the exploration were utilized to compute the straight-line path to the platform. The rat's behavior is thus explicable in terms of psychological states such as perceptions and memories and computations that operate over them. Much more detail can be given, to be sure, but for now our main concern is only to call these sorts of explanation to the reader's attention. Much more detail concerning, for instance, the neural underpinnings of perception, memory, and computation, will be supplied later. Gallistel (1990: 2) describes another such example:

Every day two naturalists go out to a pond where some ducks are overwintering and station themselves about 30 yards apart. Each carries a sack of bread chunks. Each day a randomly chosen one of the naturalists throws a chunk every 5 seconds; the other throws every 10 seconds. After a few days experience with this drill, the ducks divide themselves in proportion to the throwing rates; within 1 minute after the onset of throwing, there are twice as many ducks in front of the naturalist that throws at twice the rate of the other. One day, however, the slower thrower throws chunks twice as big. At first the ducks distribute themselves two to one in favor of the faster thrower, but within 5 minutes they are divided fifty-fifty between the two "foraging patches." ... *Ducks and other foraging animals can represent rates of return, the number of items per unit time multiplied by the average size of an item.* (emphasis ours)

In both the cases of the rats and the ducks, the ultimate explanation called for is going to require mention of some relatively subtle mechanisms inside of the animals that are sensitive to properties of the environment. To get a feel for what might be called for, contrast the way in which we would explain, on the one hand, the movements of the rat toward the platform or the duck toward the bread and, on the other hand, a rock falling toward the earth. The rock's movement is explained by a direct appeal to a fundamental force of nature that constitutes the attraction between the respective masses of the earth and the rock. Such a direct appeal to a fundamental force will not explain the rat's movement to the platform. This is not to say, of course, that something non-physical is transpiring between the rat and the platform. There is of course energy flowing between the two that impacts the rat in ways that ultimately explain its behavior. But unlike the case of the rock, the transference of energy from platform to rat will only have an impact on the rat's behavior insofar as the rat is able to transduce the information carried by that energy into a code that can be utilized by information processing mechanisms in its central nervous system. Such mechanisms will be able to store information in the form of encoded memories and make comparisons between encoded memories and current sensory input to compute a course of action toward

a goal state. Going into further detail of how the nervous system of an animal might encode such information and perform such computations can get quite complicated. Before proceeding it will be useful to turn our attention toward nervous systems much simpler than those of vertebrates.

3 Modeling the simplest forms of intelligence

Chemotaxis – directed movement in response to a chemical stimulus – is one of the simplest forms of organism behavior. It is an adaptive behavior as when, for example, positive chemotaxis is used to move toward a food source or negative chemotaxis is used to move away from a toxin. The underlying mechanisms of chemotaxis are relatively well understood and amenable to modeling and simulation (Mandik 2002, 2003, 2005). Chemotaxis is appropriate to regard as cognitive. As we will argue below, it constitutes what Clark and Toribio (1994) call a “representation hungry” problem. To appreciate the informational demands that chemotaxis places upon an organism, it is useful to consider the problem in the abstract. The central problem that must be solved in chemotaxis is the navigation of a stimulus gradient, and the most abstract characterization would be the same for other taxes such as thermotaxis or phototaxis. To focus on a simplified abstract case of positive phototaxis, imagine a creature traversing a plane and utilizing a pair of light sensors – one on the left and one on the right. Activity in each sensor is a function of how much light is falling on it in such a way that the sensor closer to the source of light will have a greater degree of activation. Thus, the difference in the activity between the two sensors encodes the location of the light source in a two-dimensional egocentric space. Information encoded by the sensors can be relayed to and decoded by motor systems responsible for steering the creature. For example, left and right opposing muscles might have their activity be directly modulated by ipsilateral sensors so that the greater contraction corresponds to the side with the greatest sensor activity, thus steering the creature toward the light.

Consider now the problem of phototaxis as confronted by a creature with only a single sensor. The one-sensor creature will not be in a position to directly perceive the direction of the light since activity in a single sensor does not differentiate from, say, light being three feet to the left or three feet to the right. Of course, the creature might try to exploit the fact that the sensor is moving and make note of changes in sensor activity over time, but such a strategy will be available only to creatures that have some form of memory. Exploiting the change of sensor activity will require a means of comparing the current sensor activity to some past sensor activity.

Note the folk-psychological explanation of how a human would solve the problem of one-sensor taxis. To imagine that you are in a gradient it will do to

imagine that you are literally in a fog so dense that while you can ascertain how dense it is where you are, you cannot ascertain in which direction the fog gets more dense and in which direction it gets less dense. However, after walking for a while you notice that the fog is much less dense than it was previously. By comparing your current perception of a less dense fog to your memory of a more dense fog against the background of your knowledge that you have been walking, it is reasonable for you to infer that you are moving out of the area of greatest concentration. Conversely, if your current perception reveals a greater concentration of fog than remembered, it is reasonable for you to infer that you should turn around if you want to get out of the fog.

There are several points we should get from the above discussion. The first is that the informational demands of one-sensor chemotaxis can be readily appreciated from the point of view of folk-psychological explanation. The same point of view allows us to construct possible solutions to the problem of one-sensor chemotaxis: A creature that is both able to perceive the current concentration and remember the past concentration is thus in the position to make an inference about whether to keep moving ahead or turn in order to reach a desired location in the gradient.

One-sensor chemotaxis is accomplished by natural organisms. One particularly well studied example is the nematode worm *Caenorhabditis Elegans* (*C. Elegans*). Despite having four chemosensors, a pair in the head and a pair in the tail, there are good reasons to believe that the worm effects one-sensor, not four-sensor, chemotaxis (Ferrée & Lockery 1999). First off, the worms are able to effect chemotaxis even when their tail sensors are removed. Second, the two sensors in the head are too close together for there to be an appreciable difference between the activity in each of them in response to local concentration of attractant. Third, when navigating chemical gradients on the effectively two-dimensional surface of a Petri dish, the worms are positioned on their sides with the pair of head sensors orthogonal to the gradient. Fourth, artificial neural network controllers inspired by the neurophysiology of *C. Elegans* with only a single sensor input are able to approximate real chemotaxis behaviors in simulated worms. These simulations are especially interesting to examine in some detail.

We next briefly review work done in simulating *C. Elegans* chemotaxis in Shawn Lockery's lab at the University of Oregon Institute of Neuroscience. In particular we focus here on work reported in Ferrée and Lockery (1999: 263–277) and Dunn et al. (2004). Ferrée and Lockery construct a mathematical model of the control of *C. Elegans* whereby the time-derivative of the chemical concentration is computed and used to modulate the turning rate of the worm in the gradient. One of our purposes in reviewing this work is to point out how it, at best, supplies only a partial explanation of how the actual nervous systems of *C. Elegans* regulates chemotaxis. Ferrée and Lockery begin by constructing a model network that makes

many simplifying assumptions about the neuroanatomy and neurophysiology of the relevant circuits in *C. Elegans*. They hypothesize that the worm must “assess the gradient by computing the temporal derivative of concentration as it moves through the chemical environment” and that the behavioral upshot of this assessment is that the worm “attempts to keep its head pointed up the gradient”. Their model network consists of five neurons whose various states of activation model voltage. The single sensory input has a state of activation that reflects the local concentration of the chemical attractant. Two output neurons model the voltages of dorsal and ventral motor neurons whose relative voltages determine the worm’s neck angle. The remaining three neurons are interneurons. Each of the five neurons is connected to every other neuron by both feed-forward and feedback connections thus making a recurrent network. Ferrée and Lockery optimized network parameters by using a simple simulated-annealing training algorithm to maximize a fitness function defined in terms of the change of chemical concentration. The optimized network resulted in simulated worm behavior similar to that of real worms: “oriented movement up the gradient and persistent dwelling at the peak.” However, Ferrée and Lockery point out that it is not obvious how the networks are effecting these behaviors: “Simple inspection of the parameters ... does not necessarily lead to an intuitive understanding of how the network functions, however, because the neural architecture and optimization procedure often favor a distributed representation of the control algorithm.” To derive “an intuitive mathematical expression for this algorithm” they manipulated the analytic solution to the linear system of equations that comprise their mathematical model. The analytic solution for the linear recurrent network is an equation wherein the rate of turning is equal to the sum of a turning bias and the cumulative effect of chemosensory input on the rate of turning. This equation produces exactly the same response to chemosensory input as the original optimized network. In order to “further improve our intuition about chemotaxis control in this model”, Ferrée and Lockery produce a Taylor expansion of the equation in time-derivatives of the input. The extracted rule for chemotaxis control equates rate of turning with a sum whose first term is a turning bias, the second term is the zeroth time derivative of chemical concentration, the third term is the first order time derivative of chemical concentration, the fourth term is the second order time derivative of chemical concentration, and so on. Next they compared simulated behavior wherein only some of the terms are kept. With just the turning bias and the zeroth order term, the resultant behavior was not chemotaxis but instead just a circular motion around the starting position. Adding the first order term resulted in chemotaxis as did adding the first and second order terms. Likewise adding the first order but omitting the zeroth order term.

Ferrée and Lockery describe their accomplishment as follows: “Using analytical techniques from linear systems theory, we extracted computational rules that

describe how these linear networks control chemotaxis” (Ferrée & Lockery 1999: 276). However, we find the resultant mathematical *descriptions* unsatisfying insofar as they do not constitute *explanations* of how the networks effect chemotaxis. And they do not constitute explanations because too little has yet been said about what the underlying mechanisms are and how it is that they are functioning. When we say that they do not supply a complete account of the mechanism, by “mechanism” we intend it in the sense of Craver (2001: 58): “Mechanisms are collections of entities and activities organized in the production of regular changes from start or set up conditions to finish or termination conditions” (See also Craver & Darden 2001; Machamer, Darden & Craver 2000; Bechtel & Richardson 1993).

To get a feel for what we think is still missing, recall the earlier discussion between the difference between two-sensor chemotaxis and one sensor chemotaxis. In the case of two-sensor chemotaxis, the difference in activity between the left and right sensors can be straightforwardly exploited by a steering mechanism that would guide the animal right up the gradient. For example, left and right steering muscles could be connected to the sensors in such a way that the greater activity in the right sensor will result in a greater contraction in the right steering muscle thus turning the head of the worm toward the direction of the greatest concentration. If the worm’s head is pointed directly in the direction of the greatest concentration then the activity in the left and right sensors will be approximately equal as will be the amount of contraction in the left and right steering muscles, thus keeping the worm on course. In this description of the two-sensor case, we have at least a sketch of what the mechanisms underlying chemotaxis are. We are not in a comparable position yet with Ferrée and Lockery’s mathematical description. The computation rule tells us *that* the time derivative of the concentration is being computed, but we are not yet in a position to see *how* it is being computed. We know enough about the underlying mechanisms to know that there is sufficient information present upon which to compute the time derivative, because we know that the chemical concentration detected by the sensor is changing over time as the worm moves through the environment. However, we need to know more than that the information is *there*. We need to know how the information is *encoded* and subsequently *used* by the organism. As Akins (2001: 381) puts a similar point:

Information that is carried by, but not encoded in, a signal is information that is available only *in theory*. To say that the information is *present* is to say only that there exists a computable function which, if used, would yield the correct result ... It is present, as it were, from the point of view of the universe. But no creature has ever acted upon information that is available only in principle.

Lockery and his colleagues are not blind to this sort of shortcoming. In a subsequent publication Dunn et al. (2004: 138) write “The chemosensory neurons

responsible for the input representation are known ... as are the premotor interneurons for turning behavior ... Much less is known about the interneurons that link chemosensory input to behavioral output". To get a further handle on what the interneurons might be doing, Dunn et al. run simulations of networks optimized for chemotaxis. The networks in these simulations have a single input neuron, one output neuron, and eight interneurons. All of the neurons in each network are connected to each other and have self-connections as well. After optimization and testing, the networks that performed successful chemotaxis were subjected to a pruning procedure whereby unused neurons and connections were eliminated. Dunn et al. report that the pruned yet still-successful networks have only one or two interneurons and they all have inhibitory feedback among all of the neurons. Dunn et al. proposed that the main function of this feedback is "to regulate the latency between sensory input and behavior" but we note that while this latency regulation may indeed be occurring, it certainly does not explain how successful chemotaxis is accomplished. The mere introduction of a delay between input and response surely cannot suffice for successful chemotaxis. We hypothesize that the crucial yet underappreciated mechanism in the successful networks is the existence of recurrent connections. Recurrence has been noted by many authors (e.g. Mandik 2002; Churchland 2002; Lloyd 2003) as a mechanism whereby a network may instantiate a form of short-term or working memory, since activity in the network will not simply reflect the information currently coming into the sensory inputs, but also reflect information feeding back and thus representing past information that came into the sensory inputs. We hypothesize, then, that the recurrence is implementing a form of memory that allows the network to compute the time derivative of the concentration in virtue of both encoding information about the current concentration (in the state of the sensor) and encoding information about past concentration (in the signal propagated along the recurrent connections).

To test this hypothesis, we conducted our own simulations of *C. Elegans*' single-sensor chemotaxis. For our simulations we utilized the Framsticks 3-D Artificial Life software (Komosinski 2000) that allowed for the construction and testing of worms in a simulated physics and the optimization of the worms using simulated Darwinian selection. The morphologies of our synthetic worms are depicted in Figure 1 and their neural network topologies are depicted in Figure 2. Networks are modular. One module constitutes a central pattern generator that regulates forward motion by sending a sinusoidal signal to the chain of muscles that control flagellation. Another module regulates steering with a single sensory neuron, three interneurons, and one output neurons. This five-neuron steering network is recurrent with every neuron in it connected to every other.

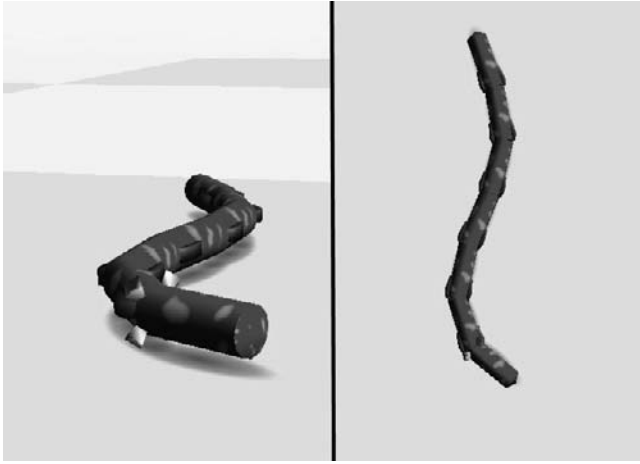


Figure 1. Synthetic C. Elegans. On the left, front view. On the right, top view

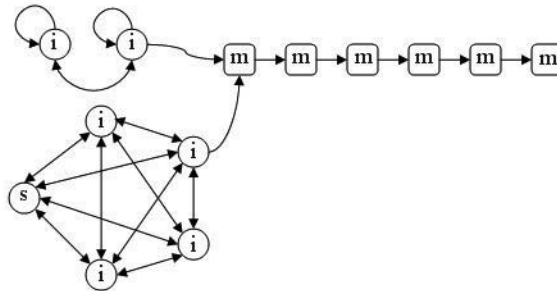


Figure 2. Neural network for the synthetic C. Elegans. Neurons include one sensor (s) and several motor neurons (m) and interneurons (i). Single-headed arrows indicate flow of information from one neuron to the next. A double-headed arrow between two neurons indicates both a feed-forward and a feedback connection between them

In our simulations the initial morphologies and network topologies were set by hand. The connection weights, however, were optimized through a Darwinian process whereby mutations are allowed only for connection weights and not to morphologies or network topologies. Fitness is defined in terms of overall lifetime distance. This forced the worms both to maintain a high velocity and also to extend their lives by replenishing their energy store with found food. We compared the performance of three kinds of orientation networks: fully recurrent networks with sensory input, recurrent networks with no sensory input (“blind” networks),

and strictly feed-forward networks with sensory input. Four populations of each of the three kinds of orientation networks were subjected to the evolutionary simulation for 240 million steps of the running program.

Results are shown in Figure 3 of the lifetime distances averaged over the four populations for each of the three kinds of orientation networks. The performance of the blind networks involved the maximal distance accomplished by worms with maximally optimized velocities but no extension of lifespan through finding food beyond whatever food they collided with accidentally. Worms with sensory inputs and recurrent connections were able to maximize their lifespan through food-finding by chemotaxis. Further, their swimming behaviors were similar to those exhibited by real *C. Elegans*: directed movement up the gradient and dwelling at the peak. Worms without recurrent connections were conferred no advantage by sensory input. Our explanation of this is that without the recurrent connections to constitute a memory, the worms are missing a crucial representation for the computation of the change of the local concentration over time. We turn now to examine the nature of these underlying representations.

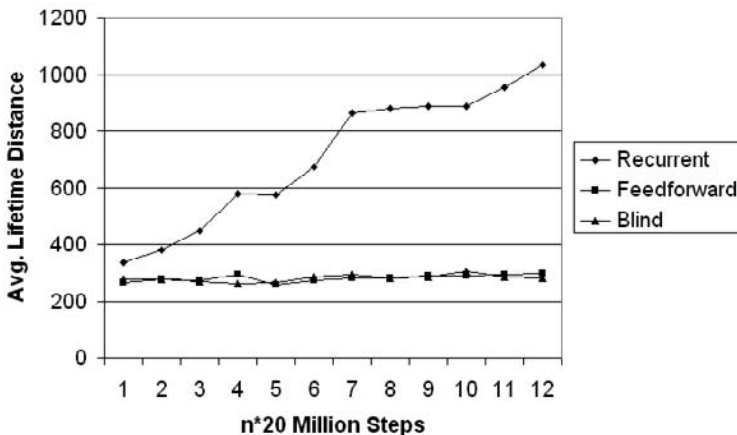


Figure 3. Results of the experiment comparing recurrent, feed-forward, and blind networks in an evolutionary simulation of chemotaxis

4 What the representations are in the models

We admittedly do not yet have a complete explanation of *C. Elegans* chemotaxis, but we do have a pretty good sketch of what is going on: Heading in the gradient is determined by a computation that takes as inputs both a sensory representation

that encodes information about the current local concentration and a memory representation that encodes information about the past local concentration. The existence of a memory mechanism was predicted by the folk psychological explanation and supported by the simulation experiments. Further, we are in a position with respect to these models to make some remarks about what the representations are and what relations obtain that determine the representational contents. In the orientation networks we may discern three types of representations: sensory representations, memory representations, and motor representations. The sensory representations are states of activations in the chemo-sensory input neuron, the memory representations are signals conveyed along recurrent connections, and the motor representations are states of activation in neurons that output to muscles. In each case, the contents of the representations are the things that are represented. In the sensory case, what is represented is current local concentration. In the memory case, what is represented is past local concentration. In the motor case, the representation is a command signal and what is represented is a level of muscular contraction.

The question arises of what the relation between representation and the represented is such that the former is a representation of the latter. Two major sorts of suggestion common in the philosophical literature on representational content seem initially applicable to the case of the chemotaxis networks: informational approaches and isomorphism-based approaches. The first sort of suggestion is that the relations that underwrite representation are causal-informational. On such a suggestion, it is in virtue of being causally correlated with a particular external state that a particular internal state comes to represent it. In the chemotaxis examples, there are indeed relations of causal correlation between the representations and what they represent. In the case of the sensory representation, there is a reliable causal correlation between the sensor state and the current local concentration and in the memory case there is a reliable causal correlation between the recurrent signal and the past local chemical concentration. The informational view must give a slightly different treatment of motor representations since commands are the casual antecedents of their representational targets (Mandik 1999, 2005).

The isomorphism suggestion seems applicable as well, though before discussing its application we need to spell out the relevant notion of isomorphism. An isomorphism is a structure-preserving one-to-one mapping. A structure is a set of elements plus a set of relations defined over those elements. So, for example, a set of temperatures plus the hotter-than relation constitutes a structure as does a set of heights of a mercury column in a thermometer and the taller-than relation. A structure-preserving one-to-one mapping exists between a set of temperatures and a set of heights just in case for any height and the next higher one they are mapped respectively to a temperature and the next hottest one.

Information-based theories of representational content make it a necessary condition on a representation r of a thing c that r carry information about (causally correlate with) c . Isomorphism-based theories of representational content make it a necessary condition on a representation r of a thing c that r and c be elements in structures wherein an isomorphism obtains that maps r to c . Can we adjudicate between the informational and isomorphism suggestions? More specifically, can the way in which attributions of representation in the explanations of the network control of chemotaxis be used to favor information-based theories over isomorphism-based theories or vice versa? We see the respective roles of the notions of representation, information, and isomorphism in this context as follows. The sensory and memory states are able to drive successful chemotaxis in virtue of the informational relationships that they enter into with current and past levels of local chemical concentration, but they are able to enter into those informational relations because of their participation in isomorphisms between structures defined by ensembles of neural states and structures defined by ensembles of environmental states. In brief, in order to have the representational contents that they have, they must carry the information that they do and in order to carry the information that they do they must enter into the isomorphisms that they do. To spell this out a bit further will require spelling out two things: First, why it is that representation requires information and second, why information requires isomorphism.

We begin with the reason why representation requires information. A large part of the reason representation requires information in the example of the chemotaxis networks is because of the sorts of representation that we are talking about, namely sensory and memory representations. It is part of the nature of sensory states that they carry information about the current local situation of an organism and part of the nature of memory states that they carry information about the past. Another way to appreciate the carrying of information is to realize that if the networks didn't encode information about the current and past chemical concentrations then they would not be able to give rise to the successful chemotaxis behavior. Consider the blind worms: they were deprived of the means of encoding information about the present chemical concentration. Consider also the worms with strictly feed-forward networks. Without recurrent connections, they were deprived of the means of encoding the relevant information about the past. It seems that the crucial aspect of attributing sensory and memory representations in explaining successful one-sensor chemotaxis is that such attributions track the information-bearing properties of the states.

To see why isomorphism is important, it helps to begin by considering how hard it would be to not have isomorphism. First off, note that, as Gallistel (1990) has pointed out, a one-to-one mapping can be considered as structure preserving

even if the structures involved are defined only in terms of sets of elements and the identity relation. On such schemes the resultant representations are what Galistel calls “nominal representations”. For example, the set of numbers assigned to players on a sports team is a set of nominal representations in this sense. There is a one-to-one mapping between numbers and players and the only relation between numbers that is mapped onto a relation between players is identity: one and the same number can only be mapped onto one and the same player. Larger numbers, however, need not indicate larger or heavier players. Nonetheless, they still satisfy the requirements for isomorphism, since the mapping is structure preserving. Similarly, even if the information-bearing states of a nervous system constitute a set of nominal representations of environmental states, they would nonetheless satisfy the requirements for isomorphism.

Setting aside identity-based nominal representations as genuine isomorphisms, there is still a serious difficulty the informational theorist faces concerning the alleged dispensability of isomorphism. Even if there were a logically possible scheme that had information without isomorphism, it is incredibly difficult, if not impossible, for such a scheme to be evolved or learned. We can see the point concerning evolution in the context of the synthetic *C. Elegans* in our artificial life simulations. Organisms’ bodies, as well as the environments they are situated in, contain many physical systems that have states that fall into natural ordering relations. Consider, for example, that chemical solutions can be more or less concentrated, or that neural firings can have higher or lower rates or higher or lower voltages. It is hard, if not impossible, to see how there could be a counter-example to the following claim: Any situation in which a particular level of neural activation can be used to carry information about a particular level of chemical concentration is also going to be a situation in which a slightly higher level of neural activation can be used to carry information about a slightly higher level of chemical concentration. In other words, organisms and their environments are rich in structures and it is hard to see how elements in those structures can be evolved to enter into informational relationships without the structures themselves also entering into isomorphism relationships.

While our argument is, to our knowledge, unique, it is worth mentioning certain similarities between our argument, which is specifically about evolution and some other arguments that focus on learning that have appeared in the literature on isomorphism. Cummins (1996) and Churchland (2001) both endorse isomorphism based theories of representational content and both argue that a creature can only be in a position to have states that carry information about external states if the creature’s internal states are embedded in a network of internal states that may be regarded as constituting knowledge of or a theory of the target domain. As Cummins (1997: 356–537) puts the point:

Distal properties generally cannot be directly transduced. Instead, the detection of distal properties must be mediated by what we might as well call a theory about that property. To detect cats (an instantiation of catness) requires a theory that says, in effect, what sorts of proximal stimuli are reliable indicators of catness. To detect cats visually, you have to know how cats look. The same goes for colors, shapes, and sizes: for these to be reliably detected visually under changes in perspective, lighting, and distance requires knowledge of such facts as that retinal image size varies as the inverse square of the distance to the object. Much of the knowledge that mediates the detection of distal properties must be acquired: we are, perhaps, born with a tacit knowledge of Emmert's Law, but we are not born knowing how cats look, or with the ability to distinguish edges from shadows.

(For an argument similar to Cummins' see also Churchland 2001: 131–132; for an argument that a creature can extract information from a perceptual representation only if certain isomorphisms obtain between states of perception and what is perceived, see Kulvicki 2004.)

Based on the above sorts of arguments, we draw the following conclusions about the nature of representation, at least as it applies to the simplest cases of creatures behaving intelligently in virtue of possessing mental representations. Attributions of representations to organisms are not simply heuristics to be abandoned later when better ways of explaining behavior are discovered. They are attributions of real properties of organisms and real relations between organisms and their environments. The representations attributed are states of the nervous systems of the creatures that represent environmental (and bodily) states in virtue of carrying information about those and a requirement on the acquisition by the organism of such states is that the states enter into isomorphism relations between neural and other structures.

One sort of objection that we've encountered in various personal communications is that the notion of isomorphism employed above should instead be replaced with the notion of homomorphism where, in brief, the main difference between the two is that where isomorphisms involve one-to-one mappings, homomorphisms involve mapping one structure into (not onto) another. Homomorphism comes up in the literature on non-mental representations such as scientific representation (Suarez 2003) and the representations pertinent to measurement theory (Matthews 1994; Suppes & Zinnes 1965) but we think that isomorphism is more appropriate for mental representation. Trying to utilize the notion of homomorphism for mental representation would involve the idea that structure A represents B if and only if B is homomorphic to A which involves B mapping into (not onto) A. This allegedly allows for, among other things, A to be a partially accurate model of B. We might think of these mappings as, for example, a mapping of

physical objects or an empirical system into the real numbers allowing us to say that numbers represent physical objects.

One problem with the above homomorphism based suggestion is that we don't simply want to establish a relation between two sets: the representations and the represented. We want instead to establish a set of relations, more specifically, a set of relations that will allow us to say, for example, of each height of the mercury column, whether it represents a temperature and, if so, which one. Similarly, we want to say of each temperature, if it is represented by a height of the mercury column and, if so, which one. We especially want to avoid attributing multiple contents to one and the same representation, as in, saying of a height of the mercury column that it represents multiple temperatures.

Attributing representations to an organism must involve partitioning the state-space of the organism and the state-space of its environment such that there is a one-to-one mapping between the two sets of regions. Thus there is a certain supervenience guaranteed between mental contents and neural vehicles: there should be no mental (content) differences without physical (vehicular) differences. We do not want to attribute multiple contents if the organism is not capable of distinguishing them. This is analogous to the case of the representation of the past in our experiment. The chemosensory input carries information about both the present and the past, but the feed-forward networks are incapable of distinguishing present from past. The attribution of contents to an organism is an attempt to portray the world as it is carved up by the creature's point of view: Elements of the world that the creature cannot distinguish cannot make a difference discernable from the creature's point of view.

We close by briefly mentioning what the above account of representation in very simple systems might possibly say about the philosophically vexing problem of the representation of inexistent objects. The problem of representations of things that do not exist – gold mountains, square circles, etc. – constitutes one of the largest problems that inspire philosophers to worry about representation. It might even be framed as an objection to our view: Our account of representation, couched in terms of information and isomorphism, cannot account for the representation of inexistent objects, since if something doesn't exist something else can neither carry information about it nor be isomorphic to it. We have several brief remarks to make on this topic.

The first remark is that our primary concern is to give an account of how representations might underwrite certain kinds of explanation and it is unclear that representations of inexistent objects play roles in explanations in virtue of their contents and not simply in virtue of their vehicular properties. An underappreciated point is that we may very well be wrong when we think something has content.

It may very well be that any appearances to the contrary from the first person point of view are explicable in terms of the first person indistinguishability of vehicles that have content and vehicles that do not. We suggest, then, that certain seeming representations, namely, so-called representations of things that do not exist, actually have no content. If the content of a representation is identical to the thing it represents, then a representation of a thing that does not exist is a representation with a content that does not exist. If nothing exists that is identical to the representation's content, then the representation has no content. Whatever role such representations play in explanations of behavior must, then, be due to their vehicular properties.

This line of thought derives a nihilism about content from a relatively dismissive view of nonexistent objects. Some philosophers, however, may be much more tolerant of things such as non-actual possible worlds. Our response to such philosophers is that insofar as there may be a sense in which things that do not exist in the actual world nonetheless exist, then there may be a sense in which things in non-actual worlds are able to enter into the requisite information and isomorphism relations with neural representations in the actual world. (See, for example, Swoyer 1991 for a discussion of isomorphism-based representations and non-actual possible worlds.)

We close, then, with a final remark on the topic of the representation of nonexistent objects. Whatever the status of the representation of nonexistent objects, it is safe to say that such representations are irrelevant in the explanations of the simplest cases of cognition. Certainly, the representation of a thing that does not exist cannot be the most basic case of representation. Restricting our attention to the simplest cases of representation, we see that the simplest cases of things that can be represented are things that not only actually exist, but also actually enter into relations of information and isomorphism with neural states.

Acknowledgements

This work was supported in part by grants to Pete Mandik from the National Endowment for the Humanities and James S. McDonnell Foundation Project in Philosophy and the Neurosciences. Pete Mandik is grateful for feedback from members of audiences of presentations of this material at the 2004 International Language and Cognition Conference in Coffs Harbour, Australia; the City University of New York Graduate Center Cognitive Science Symposium and Discussion Group; and the Neurophilosophy: The State of the Art conference at Caltech.

Mike Collins is grateful to audiences at the City University of New York Graduate Center Cognitive Science Symposium and Discussion Group and the Fall 2004 meeting of the New Jersey Regional Philosophical Association at Felician College.

References

- Akins, Kathleen. 2001. Of sensory systems and the 'aboutness' of mental states. In William Bechtel, Pete Mandik, Jennifer Mundale & Robert S. Stufflebeam (eds.), *Philosophy and the neurosciences: A reader*, 369–394. Oxford: Blackwell.
- Bechtel, William & Robert Richardson. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.
- Beer, Randall. 1990. *Intelligence as adaptive behavior*. San Diego, CA: Academic Press.
- Brooks, Rodney. 1991. Intelligence without representation. *Artificial Intelligence* 47. 139–159.
- Churchland, Paul. 2001. Neurosemantics: On the mapping of minds and the portrayal of worlds. In K. E. White (ed.), *The emergence of mind*, 117–147. Milano: Fondazione Carlo Elba.
- Churchland, Paul. 2002. Catching consciousness in a recurrent net. In Andrew Brook & Don Ross (eds.), *Daniel Dennett: Contemporary philosophy in focus*, 64–80. Cambridge: CUP.
- Clark, Andy & Josefa Toribio. 1994. Doing without representing? *Synthese* 101. 401–431.
- Craver, Carl. 2001. Role functions, mechanisms and hierarchy. *Philosophy of Science* 68. 31–55.
- Craver, Carl & Lindley Darden. 2001. Discovering mechanisms in neurobiology: The case of spatial memory. In Peter K. Machamer, Rick Grush & Peter McLaughlin (eds.), *Theory and method in neuroscience*, 112–137. Pittsburgh, PA: University of Pittsburgh Press.
- Cummins, Robert. 1996. *Representations, targets, and attitudes*. Cambridge, MA: The MIT Press.
- Cummins, Robert. 1997. The LOT of the causal theory of mental content. *Journal of Philosophy* 94. 535–542.
- Dunn, Nathan A., Shawn R. Lockery, Jonathan T. Pierce-Shimomura & John S. Conery. 2004. A neural network model of chemotaxis predicts functions of synaptic connections in the nematode *Caenorhabditis Elegans*. *Journal of Computational Neuroscience* 17(2). 137–147.
- Ferrée, Thomas C. & Shawn R. Lockery. 1999. Computational rules for chemotaxis in the nematode *C. Elegans*. *Journal of Computational Neuroscience* 6. 263–277.
- Gallistel, Charles R. 1990. *The organization of learning*. Cambridge, MA: The MIT Press.
- Komosinski, Maciej. 2000. The world of Framsticks: Simulation, evolution, interaction. *International Conference on Virtual Worlds* 2. 214–224.
- Kulvicki, John. 2004. Isomorphism in information carrying systems. *Pacific Philosophical Quarterly* 85. 380–395.
- Lloyd, Dan. 2003. *Radiant cool: A novel theory of consciousness*. Cambridge, MA: The MIT Press.
- Machamer, Peter, Lindley Darden & Carl Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67. 1–25.
- Mandik, Pete. 1999. Qualia, space, and control. *Philosophical Psychology* 12(1). 47–60.
- Mandik, Pete. 2002. Synthetic neuroethology. *Metaphilosophy* 33(1–2). 11–29.
- Mandik, Pete. 2003. Varieties of representation in evolved and embodied neural networks. *Biology and Philosophy* 18(1). 95–130.
- Mandik, Pete. 2005. Action oriented representation. In Andrew Brook & Kathleen Akins (eds.), *Cognition and the brain: The philosophy and neuroscience movement*, 284–308. Cambridge: CUP.
- Matthews, Robert. 1994. The measure of mind. *Mind* 103. 131–146.
- Suarez, Mauricio. 2003. Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science* 17(3). 225–244.

Suppes, Patrick & Joseph L. Zinnes. 1965. Basic measurement theory. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of mathematical psychology*, 1–76. New York, NY: John Wiley and Sons.

Swoyer, Chris. 1991. Structural representation and surrogative reasoning. *Synthese* 87. 449–508.

Multi-agent communication, planning, and collaboration based on perceptions, conceptions, and simulations

Peter Gärdenfors and Mary-Anne Williams

This study seeks to enhance current understanding of cognitive capabilities from perception to collaboration by developing an evolution inspired framework infrastructure designed to support planning and meaningful communication between agents as a means to effective collaboration. We have implemented parts of the framework in the robot soccer domain where agents interact in a complex and dynamic environment to illustrate our approach. The framework is based on a cognitive scaffold of cued and detached representations based on conceptual spaces that can be used to develop and communicate plans, and pursue shared goals. We show how grounded representations can be used to explain agent behaviour and to construct agents that can share information about their inner world for the purpose of collaboration.

1 Introduction

Current understanding of cognition is still rudimentary, and as a result explaining the behaviour of biological agents and constructing new artificial agents that can communicate and collaborate to achieve shared goals in complex and dynamic environments remains a difficult and challenging task. A richer and deeper understanding of the interplay among the various cognitive processes that take place in an agent's inner world and the generation of cognitive capabilities such as perception, planning, communication and collaboration is necessary.

This study seeks to enhance current understanding of cognitive capabilities developing an evolution inspired framework and accompanying infrastructure designed to support planning and meaningful communication between agents as a means to effective collaboration in complex and dynamic environments. We have implemented parts of the framework in the robot soccer domain where agents

interact in a complex and dynamic environment to illustrate our approach. The framework is based on a cognitive scaffold of grounded cued and detached representations, and can be used to achieve meaningful communication among a group of agents such as robots. It allows them to exchange information, to construct and share plans, and to pursue shared goals individually and collaboratively, but perhaps more importantly it provides a flexible and agile infrastructure that captures the meaning of representations.

We show how grounded representations (Gärdenfors & Williams 2003; Williams et al. 2005) allow agents to share information represented in their inner world for the purpose of collaboration. An agent's inner world is built on representations from which symbolic and linguistic descriptions can be derived and shared.

Section 2 describes a useful classification of the different kinds of representations which underpin cognitive capabilities. Section 3 briefly describes the architecture and properties of agents, and outlines some cognitive capabilities that agents typically possess. It describes the components of an agent's inner world and highlights the role of representations, and the interrelationships between sensation, perception, conception, simulation, and planning. In Section 4 we illustrate our framework using examples from the robot soccer domain where robots based on our framework exhibit high levels of cognitive capability as they set about playing soccer autonomously.

2 Cued and detached representations

Entities in representations can be objects, properties, relationships, actions, events, processes, etc. They can range from low level perceptual representations all the way up to high level logic and linguistic expressions. There is no need to be prescriptive about representations, since for the purpose of understanding cognitive capabilities like planning and collaboration we can make a distinction between two types of important classes of representations – *cued* and *detached* (Gärdenfors 1996, 2003). In cued representations percepts and/or concepts stand for something that is *present* in the surrounding situation. A detached representation, on the other hand, is a simulation or abstract concept: something the individual can utilise *regardless* of whether what it represents is present or not. A detached representation can even stand for something that does not exist at all.

Being able to use and manipulate a detached representation requires that one can *suppress* the sensations one has for the moment; otherwise they will come into conflict with the representation. Detached representations are necessary for planning and for all other higher cognitive functions. Figure 1 summarises the classification of the four kinds of entities: sensations, perceptions, conceptions, and

simulations. Conceptions and simulations can be detached, i.e. no direct connection with the surrounding world is required, and can be created and manipulated independently of sensations whilst perceptions are cued, i.e. coupled to sensory impressions.

An important property of representations is their *groundedness* (Williams et al. 2005), i.e. the degree to which the entities in the representation correspond to the entities they represent. Cognitive processes must ensure that representations are appropriately grounded. Clearly, grounding plays an important role in both cued and detached representations and the quality of planning and communication will be related to the quality of the grounding of the underlying representations. We introduce conceptual spaces in the next section and show how they can be used to provide the infrastructure for a grounded bridge from sensations through perceptions to simulations and plans.

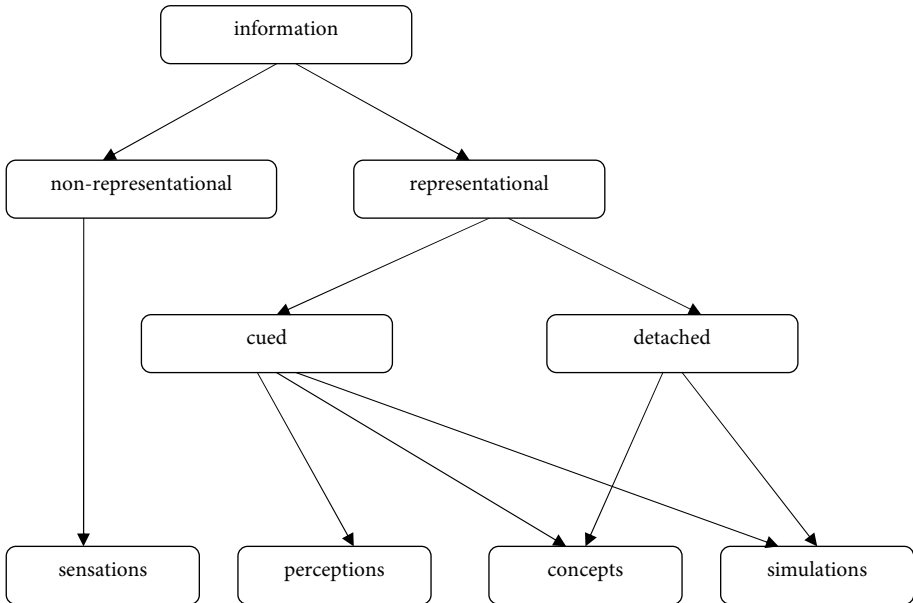


Figure 1. Interrelationships between cued and detached representations

3 Cognitive capabilities for agents

Mobile robots, infobots, humans, apes, and other animals are examples of agents. An agent is a self-contained *autonomous* system whose inner world interacts with an outer world (the environment). This interaction involves the transfer of input

(e.g. sensorimotor data) and output (e.g. information or actions). Within the inner world of an agent, information gained via interaction is fused and transformed during perception, conception, simulation, planning processes.

In this study we are mainly concerned with the formation and management of representations for the purposes of communication and collaboration. The idea being that grounded representations not only provide the basis for perception, the formation and evolution of concepts and simulated world models, but they provide the essential infrastructure for problem solving, decision making, planning, communication, and collaboration.

In order to understand cognition and communication better and to design agents that possess higher forms of cognition, we must give an account of how agents represent, and subsequently exploit, their outer worlds.

Our cognitive framework is based on the following interrelated representations that inhabit an agent's inner world:

- *Sensations* can be *externally-oriented* or *internally-oriented*. Externally-oriented sensations mediate sensory impressions or data from the outside world, whilst internally-oriented sensations are generated internally to the system often for regulatory or feedback purposes.¹
- *Perceptions* are interpreted/processed sensory impressions.
- *Conceptions* are concise representations of classes of entities and/or individuals which may exist in either the inner or the outer world.
- *Simulations* are not directly governed by sensory impressions.² They can be thought of as perceptions taken from memory or combined from cued and detached perceptions and conceptions.
- *Plans* are recipes for action.

Even simple organisms have sensations – sensory impressions that report what is happening within the body. More advanced animals have perceptions that consist of interpretations of sensations. In particular, perceptions can provide knowledge about what is going on in an agent's immediate surroundings. In the animal world it is probably only mammals (and perhaps also birds) that have simulations – in their *inner world* they can direct their thinking towards something that is not immediately present in the surrounding environment. Conceptions are not only built from perceptions and simulations, but they influence them as well. Conceptions are particularly important when developing artificial agents with active

1. In humans the externally-oriented sensations are sight, smell, taste, hearing, and touch. Some internally-oriented sensations include the sensory feedback mechanisms for motor control and posture, and the vestibular system of birds and mammals which provides feedback mechanisms for balance by detecting gravity.

2. Simulations were called “imaginations” in Gärdenfors (2003).

communication and collaboration capabilities. Perceptions, conceptions, simulations and plans are the basic elements from which the inner world is constructed.

3.1 Sensations and perceptions

Our basic assumption is that all components in animal cognition have a biological function. So if sensations are fundamental, what is their function? Humphrey (1993) claims that the biological role of sensations is to tell about what is *happening right now* to the organism.

The sensations often involve an (unconscious) *evaluation* of whether what is happening is good or bad. This is sometimes referred to as “raw feels.” Sensations are often closely aligned with basic choice of action – eat, fight, flee or mate. This applies especially to more primitive organisms.

As a preparation for presenting perceptions, let us consider what happens in the brain when you throw something. The throw begins in the cerebrum and is guided by the cerebellum. Naturally, the brain gets feedback from the sensors in the muscles of the hand and arm about the direction they are taking during the throw. The catch is that the signals that come back from the muscles via the nerve fibres are too *slow*. It has been calculated that the loop from the signals that go from the brain to the muscles and back to the brain takes something between 200 and 450 milliseconds (Denier van der Gon 1988). Unfortunately, that is much too long for the brain to be able to check that the arm is moving in the right direction during a throw.

It turns out that a kind of *simulator* has been created in the brain that quickly estimates what the *anticipated* result of the signals to the muscles will be (Grush 1997, 1998, 2004; Wolpert & Kawato 1998; Wolpert, Doya & Kawato 2003). The signals that leave the motor part of the cortex are sent by the cerebellum to both the arm and the simulator. A calculation is made in the brain of what is about to happen in the arm and the result is sent back to the cerebellum, which adjusts the arm’s continuing movement (Ito 1984).

The calculation loop in the simulator is faster than the physical loop that goes via the body’s muscles. It has been estimated that the brain can actually correct a signal it has sent to the muscles within 70 milliseconds. The simulator *bluffs* in the sense that it has not received any real feedback from the body. But with careful training it can learn to bluff sufficiently correctly. Without such a simulator we would never be able to solve the control problem involved in the art of aiming.

Grush (1997, 1998) proposes that *Kalman filters* (Kalman 1960; Kalman & Bucy 1961) form the best way of modelling the simulators. The core idea behind this powerful construction from control theory is that the filter maintains an optimal estimate of the state of the real process. The Kalman filter compares its

expectation of what the result of an action should be to what it actually is. On the basis of the mismatch it adjusts its estimate of the real process. In Section 4, we will present an example of how Kalman filters are used in the soccer playing robots.

An agent that can not only find out what is happening with its own body via sensations, but also receives signals about *what is going on out there in the world* has better opportunities to foresee the future and thus survive in an inhospitable world. This is the role of the *perceptions*. First and foremost, perceptions provide information about the spatial structure of the world and what objects are in it.

Our proposal is that perceptions build on the type of simulator described above. You become much better at aiming if there is a mechanism that more quickly *fills in* the result of the signals sent to the muscles. We assume that there are similar complementing mechanisms for many other types of sensations (Gärdenfors 2003; Grush 1997, 1998; Hesslow 2002). Perceptions are sensations that are *reinforced* with simulations. In control theoretic terms such simulators are *forward predictors*.

One important property of a simulator is that it does not need to rely exclusively on the signals coming from sense organs: It can also *add on* new types of information that can be useful in simulating. It does not matter much if this information has no direct counterpart in the surrounding world as long as the simulations produce the right result, i.e. lead to appropriate actions. This way the simulator's output will be richer than the influx of sensations (Grush 1997, 1998). For example, much of our reasoning about *causes* depends on this kind of addition.

The brain is full of mechanisms that contribute new information. In particular, there are many well-studied examples of the visual process. They show that we have plenty of simulators that *complement* the signals provided by our senses. Such complementations create the *representations* with which thinking works, since what we experience is not only that which is presented by our sensory receptors but also that which is recreated, i.e. *represented*, by our simulators.³ The weaker or the more incomplete the sensations are, the more important it is for the brain to create its own perceptions.

One role of perceptions is illustrated by the fact that a cat chasing a mouse that runs in behind a curtain can predict that it will come out the other side. Such behaviour presumes the cognitive ability called *object permanence* by Piaget. This implies that the cat retains some kind of *representation* of the mouse even when its sensory impressions of the mouse are gone. The cat has expectations about the

3. For a more thoroughgoing discussion of the concept of representation and a comparison with other definitions, see Gärdenfors (2003) who regards representations as *theoretical quantities* from the way they are treated in scientific theory.

mouse. Object permanence can be seen as a form of simulation of the same kind as that involved in controlling a throw effectively.

The example of the cat's hunting method shows that it is important for the brain to *coordinate* input from the different senses when it is going to create useable representations of objects. After all, the brain handles impressions from the different receptors in rather well separated centres, at least in the initial stages of the processing. Sometimes we speak of a *modal* treatment of impressions. But for many tasks, vision, hearing and the sense of touch must be able to work together. In order to be able to meet such challenges of coordination, the human brain has developed representations of its surroundings, objects, actions and so on that are *independent* of sensory modalities. The literature refers to such representations as being *cross-modal*.

Sensations are modal, i.e. they are tied to a certain sensory domain: visual, auditory, etc. In contrast, perceptions are cross-modal. The explanation is probably that because the perceptions deal with what exists in the world, they are more useful to an agent if they are accessible for all thinking processes, and not bound to some special sensory modality.

3.2 Conception

Conception *builds on* and *influences* perception, simulations and planning. In this section we describe conception in detail and show how it can be used to develop artificial agents with high level cognitive capabilities.

3.2.1 Building concepts via categorisation in conceptual spaces

Agents classify and categorise information into classes as a way of simplifying and making sense of all the information impinging on their sensors and/or input devices. Concepts can be created by perception and they also form some of the building blocks for the simulations that the animal or the agent uses in its planning activities.

Categorisation involves partitioning objects into concepts which are cognitively useful groups or categories. The categorisation mechanisms *fill in* the sensations provided by the sensors of an animal or an agent. Once formed concepts can be used to build cued and detached representations and simulations, e.g. perceptible concepts like *colour* and *shape* can simplify visual processing, and more detached concepts like *action* or *task* can enable complex cognitive processes such as planning. Such concepts can also be converted into symbolic or linguistic entities for the purpose of high level reasoning, problem solving, decision making, planning, and communication.

Concepts help agents reduce the complexity of the information they need to manage. For example, without concepts agents such as robots would not be capable

of representing visual information beyond the pixel level, and as a result would not develop a world model that could support even simple forms of object recognition and reasoning. The ability to form, combine and manipulate concepts enhances an agents' capacity for problem solving, communication, collaboration etc. as they forage around information rich and dynamic environments. The ability to categorise new sensorimotor information and to ground it to objects and entities perceived in the outer world allows agents to exhibit appropriate behaviour in previously unencountered situations.

Few concepts can be specified using necessary and sufficient conditions: Mathematical entities like *triangles* can, but almost all everyday objects, like *chairs* for example, defy explicit definition. In the domain of robot soccer the way around the problem is simplified by colour-coding objects of interest, i.e. ball, goals, beacons, other robots, etc. For the conceptual spaces that we will use in our framework, the prototype theory of concepts better catches the structure of concepts. However, the theory of conceptual spaces puts the prototype theory in a mathematical model that can also explain how concepts are learned in the first place and how they develop over time.

It is well known that the process of categorisation is complex. It depends on the specific task at hand, the context, as well as the background knowledge and the goals of the agent (Gärdenfors 2000). The conceptual space framework adopts a similarity-based approach to categorisation. Historically, there have been two approaches to understanding categorisation: the similarity-based methods and the causally-based methods. Similarity-based approaches tend to justify themselves using the idea that objects are categorised according to how similar they are to a prototype or (cluster of) exemplar(s). For example, *yellow* is more similar to *green* than it is to *blue* (Hahn & Ramscar 2001). Other approaches tend to use an agent's background knowledge and causal understanding to explain the process of categorisation. Recently, however, a more unified picture has emerged, namely, for the vast majority of concepts the similarity-based approaches can be extended to accommodate theoretical knowledge (see Gärdenfors 2000).

Hampton (1993) argues that similarity-based approaches can capture conceptual structure and deep knowledge as well as perceptual similarity. Others convincingly demonstrate experimentally how causal knowledge can be integrated into similarity-based approaches by giving more weight to dimensions that represent causes than those that represent effects (Gärdenfors & Williams 2001).

3.2.2 *The basic structure of conceptual spaces*

In this section, we describe how conceptual spaces provide a powerful substrate upon which to build effective concept and knowledge management capabilities that integrate information from multiple sensory and symbolic sources. Using the conceptual

spaces framework agents can build rich and grounded world models from a wide variety of internal and external knowledge (re)sources, e.g. sensors, ontologies, databases, knowledge bases, the Semantic Web, web services, other agents etc.

Conceptual spaces (Gärdenfors 2000) represent information by *geometric* structures rather than by symbols. Information is represented by points (standing for individuals or objects) and regions (standing for properties and relations) in dimensional spaces. Many semantic structures, for example similarity relations, can be modelled in a natural way by exploiting *distances* in the space.

A conceptual space consists of a number of *quality dimensions*. Examples of such dimensions are: colour, pitch, temperature, time, weight, size and the three ordinary spatial dimensions. These examples are closely connected to what is produced by our sensory receptors. However, there are also quality dimensions that are of an abstract non-sensory character. For example Gärdenfors (2004) extends the analysis to include functional concepts.

The primary role of the dimensions is to represent various “qualities” of objects in different *domains*. Since the notion of a domain is central to our analysis, it should be given a more precise meaning. To do this, we rely on the notions of separable and integral dimensions taken from cognitive psychology (Garner 1974; Maddox 1992; Melara 1992). Certain quality dimensions are *integral* in the sense that one cannot assign an object a value on one dimension without giving it a value on the other. For example, an object cannot be given a hue, without also giving it a brightness value. Or the pitch of a sound always goes along with a loudness. Dimensions that are not integral are said to be *separable*, as for example the size and hue dimensions. Using this distinction, the notion of a *domain* can now be defined as a set of integral dimensions that are separable from all other dimensions.

The domains form the framework used to assign *properties* to objects and to specify *relations* between them. The dimensions are taken to be independent of symbolic representations in the sense that we can represent the qualities of objects, for example by vectors, without presuming an explicit language in which these qualities are expressed.

The notion of a dimension should be understood literally. It is assumed that each of the quality dimensions is endowed with certain *topological* or *geometric* structures. As a first example, take the dimension of *time*. In science, time is a one-dimensional structure that is isomorphic to the line of real numbers. If *now* is seen as the zero point on the line, the future corresponds to the infinite positive real line and the past to the infinite negative line.

If it is assumed that the dimensions have a metric, one can talk about distances in the conceptual space. Such distances represent degrees of *similarity* between the objects represented in the space. Hence, conceptual spaces are suitable for representing different kinds of similarity relations. Similarity for us can be perceptual or

abstract. Perceptual similarity is derived from the perceptual elements of objects (e.g. colour), whilst abstract similarity can be determined by other characteristics (e.g. number of team-mates in attacking positions). Using perceptual similarity alone is typically not sufficient for many categorisation tasks. For example, a large *orange* could be incorrectly categorised as a *robot soccer ball* on the basis of perceptual similarity since both are regions of orange pixels in a classified image.

A paradigmatic example of a domain involves *colour*. In brief, our cognitive representation of colour can be described by three dimensions. The first dimension is *hue*, which is represented by the familiar *colour circle*. The topological structure of this dimension is thus different from the quality dimensions representing time or weight which are isomorphic to the real line.

The second psychological dimension of colour is *saturation*, which ranges from grey (zero colour intensity) to increasingly greater intensities. This dimension is isomorphic to an interval of the real line. The third dimension is *brightness*, which varies from white to black and is thus a linear dimension with end points. Together, these three dimensions, one with circular structure and two with linear, constitute the colour domain which is a subspace of our perceptual conceptual space.

This domain is often illustrated by the so-called *colour spindle* (see Figure 2). Brightness is shown on the vertical axis. Saturation is represented as the distance from the centre of the spindle. Hue, finally, is represented by the positions along the perimeter of the central circle.

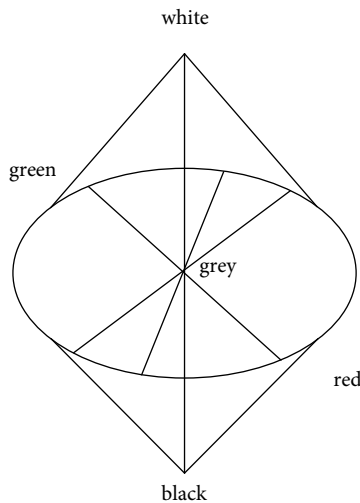


Figure 2. The colour spindle

When it comes to implementing a knowledge representation system based on conceptual spaces, it is of course the architect who decides the domains and their underlying structures, at least for the basic domains. This involves selecting the dimensions of the domain; specifying the topology or geometry of the domain; and, in the case of a metric space, specifying its metric. A rule of thumb for the metric is that integral dimensions combine by the Euclidean metric, while separable dimensions combine by the city-block (Minkowski) metric.

3.2.3 *Properties and concepts*

The theory of conceptual spaces will first be used to provide a definition of a *property*. The following criterion (Gärdenfors 2000, 2004) is central:

Criterion P: A property is a convex region in some domain.

Criterion P presumes that the notion of *betweenness* is meaningful for the relevant quality dimensions. This is, however, a rather weak assumption which demands very little of the underlying geometric structure. The criterion does not presume that one can identify sharp borders between properties; it can be applied also to *fuzzy* properties or properties that are defined by *probabilistic* criteria. What convexity requires is that if two object locations v_1 and v_2 both satisfy a certain membership criterion, for example, has a certain degree or probability of membership, then all objects between v_1 and v_2 also satisfy the criterion.

What is the cognitive motivation for Criterion P? Gärdenfors (2000) argues that the assumption that properties are convex makes concept *learning* more efficient. Another kind of argument has been proposed by Jäger and van Rooij (2006) who show that if two agents, humans or robots, communicate about the properties of objects that can be described in a metric conceptual space using a fixed set of words, then the optimal communication strategy is to let each word represent a region that is determined by a set of prototypes for the words and the Voronoi tessellation generated by these prototypes. It is well-known that if the metric is Euclidean, Voronoi tessellations always result in a partitioning of the space into convex sets.

Most properties expressed by simple words in natural languages seem to be natural properties in the sense specified here. For instance, *colour terms* in natural languages express natural properties with respect to the psychological representation of the three colour dimensions. It is well-known that different languages carve up the colour circle in different ways (Gärdenfors 2000), but all carvings seem to be done in terms of convex sets (Berlin & Kay 1969; Taft & Sivik 1997).

Properties, as defined by criterion P, form a special case of *concepts*. We define this distinction by saying that a property is based on a *single* domain, while a concept may be based on *several* domains. The distinction between properties and concepts

has been obliterated in the symbolic as well as connectionist representations that have dominated the discussion in the cognitive sciences. In particular, both properties and concepts have been represented by *predicates* in first order languages.

When several domains are involved in a representation, some principle for how the different domains are to be *weighed* together must be assumed. These weights influence the distances in the conceptual space. The relative weights of the domains depend on the *context* in which the concept is used. Hence, we assume that in addition to the regions associated with each domain, the concept representation contains information about the *prominence* of the different domains. The prominence values of different domains determine which *associations* can be made and thus which *inferences* can be triggered by a particular use of a concept. The prominence values can change with the context, and with the knowledge and interests of the user. For example, if you are eating an orange, its taste will be more prominent than if you are using an orange as a ball when playing with a robot soccer player, which would make the shape domain particularly prominent.

Concepts are not just bundles of properties. The proposed representation for a concept also includes an account of the *correlations* between the regions from different domains that are associated with the concept. For example, there is a very strong (positive) correlation between the colour of objects and their function in a crafted colour-coded robot soccer environment.

These considerations of prominence and correlations motivate the following definition of concept representation:

Criterion C: A concept is represented as a set of convex regions in a number of domains together with a prominence assignment to the domains and information about how the regions in different domains are correlated.

Conceptual spaces provide a bridge between sensorimotor information and detached representations. Sensorimotor data can be mapped to a conceptual space and cued and detached representations can then be derived in various ways for a multitude of purposes. Concept and knowledge management involves the formation and evolution of concepts and categories. Robots, for example, need to categorise both physical and abstract entities of interest in their environment (Chella, Frixione & Gaglio 1998). Developing mechanisms for constructing categories from sensory and symbolic information is becoming an increasingly important area in robotics.

Conceptual spaces accommodate context-sensitive categorisation via a geometrical representation that facilitates concept formation and concept management. Standard and tractable computational geometry algorithms can be employed to implement conceptual spaces. In Gärdenfors and Williams (2001) we demonstrated the feasibility of using geometric algorithms to build and manage categories

in conceptual spaces, and furthermore we showed how it was possible to reason about categories and other conceptual regions.

3.3 Simulation

The notion of simulation was introduced in our discussion about perception in Section 3.1. In order to consider it in its own right, imagine a general stands in front of a big map of the combat zone with markers for his own and the enemies' troops. The map is not the real war, but it provides a *picture* of the war. The reports coming in from outside make the picture change bit by bit. When the general moves his markers on the map, his troops do not similarly jump around on the battlefield. Through his rearrangements, the general can *test* various developments of the battle and choose to give the order that he thinks best in relationship to his picture of what will happen. Of course, there is no guarantee that the reality will be as he foresees it, but he has no other options.

The general does not reach his decision on the basis of direct knowledge of the world, but by using a *representation* of the world, namely his map. His interaction with the map and its symbols corresponds to the kind of simulator discussed above. Note that the general uses the markers on the map *both* as presentations of how the battlefield looks *and* as representations when he tries various alternatives. So the representations are tools that can suitably be used to *anticipate* the consequences of different possible actions. Above all, it is much easier and quicker to move the troop markers around on the map than to move the troops in reality.

The point of the metaphor is that certain animals' brains, especially humans, utilise representations of this kind, what we call *simulations* – things that we can move around in our heads before trying to move them in reality. Naturally, we do not literally have maps and markers in our heads, but there is something there that functions in the same way.

There is nothing that says there has to be any feedback from the body to the perceptual simulators. Actually, it is only a little extra step for the evolutionary process to let the simulator work without any signals being sent to the body at all – the simulator is *detached* from the sensory and motor systems. This is what happens when the organism creates for itself a *representation* of what will happen if a certain action is undertaken. We submit that the power of simulation is completely decisive for the development of all higher thought processes.

3.4 Planning

Our framework is based on an agent's inner world being made up of representations: perceptions, conceptions, and simulations, which are the basic building

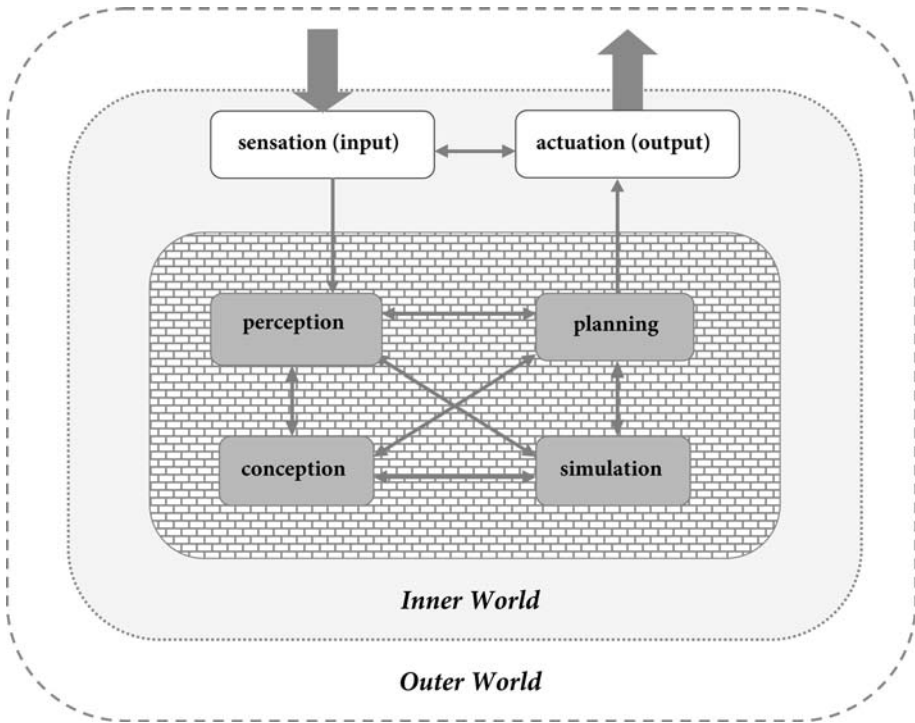


Figure 3. Open agent system architecture

blocks for *plans*. For our purposes we consider that problem solving, reasoning, and decision making can be viewed as special types of planning.

3.4.1 *Planning requires perceptions and simulations*

When an agent can represent *different* actions and their consequences in its inner world, i.e. different approaches to reaching a goal, then *choice* enters the picture. Representations of alternative actions must be detached, and the capacity for a conscious choice therefore presupposes an inner world, i.e. perceptions and simulations. An animal or agent dedicated to trial-and-error does not choose – the action carried out is generated unconsciously, without first being represented in the inner world, and to some extent haphazardly.

Planning involves the development of a plan where a plan is a set of actions that needs to be completed (serially and/or in parallel) in order to achieve an objective. Plans can include redundancy and backup actions and action sequences. Agents plan in their inner world by using representations of relevant entities upon which they can simulate pertinent actions and events. Simulations can generate

different sequences of action possibilities by executing different choice patterns. As noted in the previous section representations can be cued or detached.

Another way of describing planning is that the agent allows the simulator that guides its body to run, but without having the real behaviour engaged. The motor of thought is declutched from the wheels of the body.

One suitable definition of what planning implies is this (Gulz 1991: 46): *An individual plans its actions if it has a representation of a goal and of its present situation, and can generate representations of a number of actions that can lead forward to the goal.*

From this definition it follows that the representations of actions must be detached; otherwise a stimulus-response chain might just as well decide the behaviour chosen by the individual.

In artificial agent settings, conceptual spaces are valuable for representing the state of the world before and after the planned actions. Since the information has been divided into separable domains and since the representational format is geometrical and not symbolic, the traditional frame problem (McCarthy & Hayes 1969) does not arise (Gärdenfors 2000). This is one of the advantages of using conceptual spaces as a framework. In addition conceptual spaces can play an important role in representing plans themselves particularly for the purpose of textual and/or linguistic communication.

3.4.2 *Anticipatory planning with detached representations of goals*

Ethologists, who study animal behaviour, appear to be largely in agreement that certain animal species can plan in the sense defined here. Yet all examples of planning among animals available in the ethological literature concern planning for *current needs using current resources*. Apes and other animals start planning because they are hungry or thirsty, tired or frightened. Their motivation comes from the present state of the body.

Man seems to be the only animal that can *plan for future needs*. We can *foresee* that we will be hungry tomorrow and save some of our food; we realise that it will be cold and windy in the winter, so we build a shelter in plenty of time. Suddendorf and Corballis (1997) refer to mental time travel and Gulz (1991) calls the capacity to plan for the future *anticipatory planning*.

The reason why it is more difficult to plan for future needs than for current ones has to do with the different representations that are required for the two types of planning. When planning to satisfy current needs, one must be able to represent in one's inner world the possible actions and their consequences, and to determine the value of the consequences in relation to the needs one has *at that moment*. But no separate representation of that need is required. To plan for future needs, on the

other hand, one must also be able to *represent these potential needs* (and to understand that some of them will arise).

When it comes to planning in robots, the distinction between cued and detached planning boils down to how the *goal* of the robot is handled by the system. If the goal is externally set, like the desired temperature in a thermostat or the target of a missile, then even if the system has detached representations of the environment that it can use in its planning, the result will be cued planning. On the other hand, if the goal is one of the detached representations of the robot that can be modified in its planning or negotiated in communication with other robots, then the robot will exhibit anticipatory planning. This form of planning will be more flexible, but also more complicated to handle computationally because the search space is much larger to allow detached representations of entities to feature in plans. Allowing only cued representations puts a limit on the number of entities to be manipulated in the planning scenario simulations, and if the entities are restricted even further to cued representations of physical objects then the number of entities will be finite, and for many planning problems small in number.

3.5 Communication and collaboration

Communication between agents, biological and artificial, is highly diverse and can occur across a wide range of media and channels. From a specific agents perspective communication with the outer world is a form of interaction and information exchange with the outer world. The information conveyed in a communication ranges from highly structured, e.g. language, to unstructured information, e.g. another agents intentions. Agents can communicate via *interpreted actions* such as human gestures, robot ball chasing behaviour, or *interpreted information* derived from artifacts like written text, chemicals signs, sounds, physical objects, etc. Communication by its nature involves multiple agents, and for the purposes of collaboration it can take place between two active agents, or between an active and a passive agent. An example of the former would be a dialog between two people using *Google Talk*,⁴ and an example of the latter would be the communication between mating dragonflies via visual cues inherent in the female's body size, shape and colour patterns.

The quality of the communication will depend on the degree to which the agents involved share the meaning of the underlying entities represented in the message. In our framework conceptual spaces capture meaning and as a result they play a crucial role in perception, simulation, planning, communication and collaboration.

4. Google Talk can be used to make calls or send instant messages over the Internet.

An important issue which we do not have space to address fully which has a major impact on the quality of communication and collaboration between agents is the ability of agents to model the capabilities of other agents, e.g. mind, emotions, intentions, etc.

4 Cognitive capabilities for playing soccer robots

4.1 Robot soccer domain

The Robot Soccer World Cup competition, RoboCup (www.robocup.org), is an international research initiative whose vision is to build a team of robots capable of beating the human FIFA World Cup team by 2050. There are several leagues at RoboCup and our example involves the Four-Legged League where Sony AIBO robots play soccer.

The AIBO robot (www.aibo.com) is an entertainment robot produced by Sony, see Figure 4. It is 80 (w) x 278 (h) x 319 (d) mm in size, and has a 576 MHz 64bit RISC CPU, 32-Mb of main memory, and 16Mb of external memory on a Sony memory stick.



Figure 4. The Sony AIBO Robot: Main externally-oriented sensors are camera; infrared distance sensor; microphone; tactile sensors on head, paws, back and chin. Actuators include speaker, joint motors, and LED's on its face and back

The AIBO vision system is driven by a 350K-pixel CMOS camera. The AIBO has an infrared range finder, a stereo microphone, an acceleration sensor, vibration sensor, and several touch sensors located on the head, paws, back, and chin. The AIBO's hind limbs have knee joints, and the forelimbs elbow joints. The head and limbs each have three degrees of freedom. AIBOs also have a wireless LAN capability which can be utilised for querying the robot's inner world representations by humans and for communicating with other agents. The Sony's OPEN-R Software Development Kit (www.open-r.org), a set of Open Source C++ interface libraries, is used to develop applications for controlling the AIBO's hardware.

The Four-Legged League competition involves teams of four robots that play a modified form of soccer on a field 5.4m x 4m in size depicted in Figure 5. The rules resemble human soccer but cater to the capabilities of the four-legged AIBO robot. A game consists of two 10 minute halves. Penalty shoot-outs can be used in the case where a game is a draw. During a match each robot is in one of the following states: initial, ready, set, playing, and penalised. Robots are penalised for a range of reasons including holding the ball for more than 3 seconds, for deliberately obstructing the ball, and for charging the goalie.

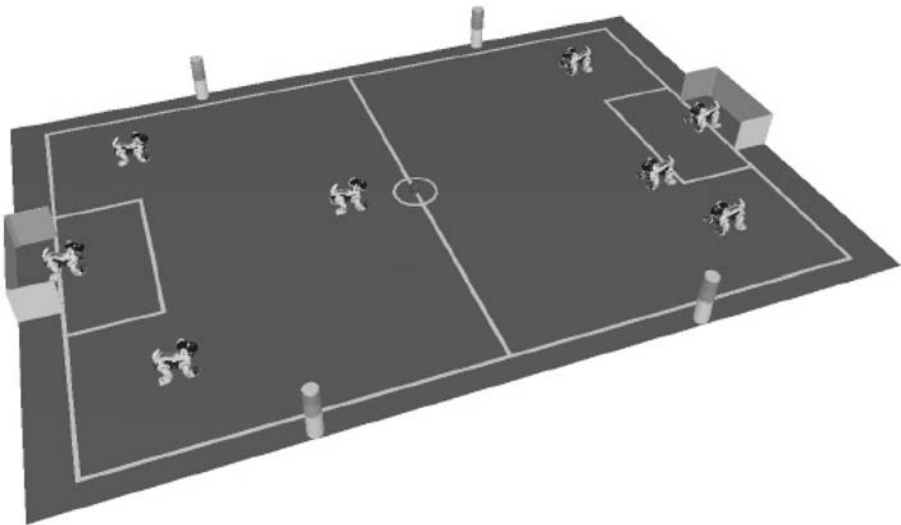


Figure 5. The Outer World in the Four-Legged League at RoboCup 2005. RoboCup Four-Legged League Rules: <http://www.tzi.de/4legged/pub/Website/Downloads/Rules2005.pdf>

4.2 Robot sensations, perceptions, and conceptions

Sensations in robots are the raw sensorimotor signals it can detect from its hardware. Perceptions in robots are derived from the hardware signals from the robots sensors (camera, microphone, accelerometer, etc.), and actuators (head motors, limb motors, speakers, etc.).

We illustrate several kinds of representations based on a Robot Soccer System (Chang et al. 2004) in Figure 6 which are constructed from raw robot camera sensor data. Figure 6(a) is a 2D visualisation of the *raw* camera data, and Figure 6(b) is a processed version of Figure 6(a) where specific colours (YUV values) of pixels are used to determine if they *belong* to specific objects of interest – a ball, a beacon and a goal are clearly identified. The information represented in Figure 6(b) can be used to find the distance, heading and elevation, from the robot’s camera to the various objects of interest which in turn can be used to calculate the pose of the robot in a global reference frame. Information represented in Figure 6(b) can be combined with a conceptual representation of robot location, i.e. $\text{robot}(\text{id}, x, y, \varphi)$,⁵ to create a conceptual representation of the location of objects on the field, i.e. $\text{object}(\text{o}, r, \varphi, \psi)$.⁶ The set of object concepts can be visualised for ease of human understanding as soccer objects such as goals, robots, ball in specific locations on a simple 2D representation of soccer field – see Figure 6.

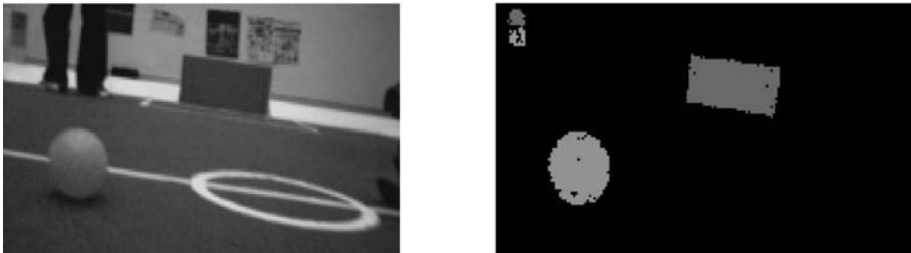


Figure 6. Perceptual representations of a robot soccer player’s the outer world (a) digital image derived from a robots camera, (b) perceptual representation of the image after processing to identify objects of interest

5. id is a robot’s identifier, x and y are coordinates of the robot, φ is the heading of the robot in a predefined world coordinate system.

6. o can be one of [*ball*, *beacon*, *goal*, *team-mate*, *opposition-robot*, *obstacle*], r is the distance from the camera of the robot to the object, φ is the heading to the object and ψ is the elevation to the object relative to the robot’s camera system.

A number of entities are conceptualised by the robots including colour, soccer objects such as ball, beacon, goal, team-mate, and opposition-robot. So too are more abstract concepts such as actions, e.g. chase and kick, skills e.g. dodge, player positions e.g. goalie and forward. Conceptual spaces are used to describe some of the concepts in particular colour and player locations. Concepts are used to describe the world based on sensations and perceptions, e.g. robot(Beckham,1,2,40degrees) and object(ball,15cm,30degrees,20degrees).

4.3 Robot simulations and planning

The perceptual representations in Figure 6 are cued, whilst concept representations like object and robot can be cued or detached. Detached representations are extremely powerful. They can be manipulated independently of the external world and can be conceived without necessarily being perceived. Some examples of detached representations are absent objects, as well as past and potential future world states. More specifically our robots are able to conceive of objects and states of the world that they may never have perceived in the past or will ever encounter in the future, e.g. even if they have never previously observed object(ball,15cm,30degrees,20degrees), they can not only represent it, but they can manipulate it, communicate it, and use it in simulations and plans.

For our purposes the inner worlds of robot soccer players are constructed from background knowledge about objects of interest (e.g. physical size of the ball, beacons, and field), direct sensorimotor information (e.g. raw data from the robot camera), perceptual information (e.g. a processed camera image), conceptual information (e.g. orange is a colour), and simulations (e.g. path of the ball after a kick).

The robot soccer players use Kalman filters to build simulations of their outer world, Figure 7 illustrates a representation of the robot soccer player's outer world, often referred to as the world model.

The world model is a detached representation of the current state of the outer world. Robots use their world model to decide on the most appropriate action to perform next and plan ahead. The quality of the robots decision making is proportional to the quality of the groundedness of the world model. The outer world changes continuously, however the robots receive a new vision frame 30 times per second.⁷ Updating a robot's world model is a nontrivial exercise and it has been found that the most effective way to maintain a world model over time is to use simulations based on Kalman filters.

7. See Figure 6(a) for an example of a vision frame.

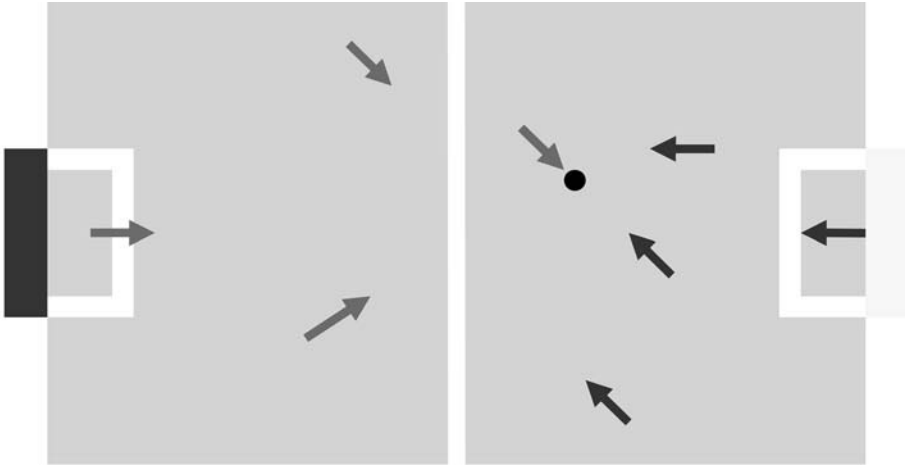


Figure 7. A robot's representation of the outer world. The field has a halfway line and two penalty regions in front of the goal areas. Robots are represented by arrows which indicate their heading direction. Teams are made up of four robots. The robots can perceive the ball, goal, and other robots but not the white lines on the field, however they do conceive of them

Kalman filters update a robot's world model using perceptions derived from a vision frame and other sensorimotor information obtained from the robots sensors and actuators together with an estimate of a future state and some background knowledge regarding the objects and their behaviour in the outer world.

A robot's world model is updated using a Kalman filter which simulates the behaviour of objects on the soccer field. The Kalman filter involves a process model that fuses perceptual information and background information with a simulated estimate to yield a best guess of the state of the field, i.e. the location of robots and the ball on the field. A Kalman filter processes perceptual information by maintaining an estimate of the state of the field as a simulation. The simulator continually provides a prediction of the future state of the outer world by evolving its most recent estimate, i.e. its expectation. This prediction is then combined with perceptual information derived from the sensors in order to reach a new estimate based both on the expectation and the sensor information available (Grush 2004).

A Kalman filter simulates the evolution of an agent's outer world via feedback control: The filter estimates the outer world state at some time and then obtains feedback via actual sensorimotor measurements (Kawato 1990). There are two main equations that drive Kalman filters: time update equations and measurement update equations. The time update equations project the current state and error covariance estimates forward in time to obtain the a priori estimates for the next time step. The measurement update equations provide feedback by incorporating

a new measurement into the a priori estimate to obtain an improved a posteriori estimate. The time update equations act as prediction simulations and the measurement update equations act as feedback corrections. The final estimation algorithm resembles a two step cycle: prediction based on *simulation* followed by a reality-check via *perception*.

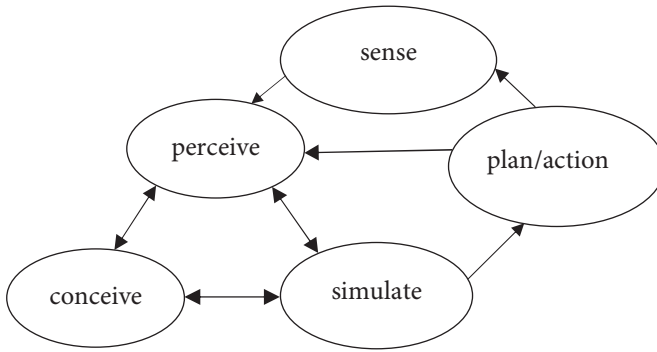


Figure 8. Individual robot processing pathways

4.4 Robot communication and collaboration

AIBOs possess a wireless network card and a wireless LAN can be established that allows the robots to send and receive messages. In this section we provide two examples of how our framework has been used to build high level cognitive capabilities such as communication and collaboration. The first example involves the development of collaborative game play plans which allow robots to collaboratively adopt attacking and defensive behaviours, and the second example demonstrates how detached representations, namely simulations, can be used by robots to share information concerning the whereabouts of the ball. Collaborative ball tracking allows a robot, who is too far away from the ball to perceive it directly. In real robot soccer matches the ball is often occluded by other robots and so the ability to collaboratively track the ball is a powerful team capability that can be exploited for significant competitive gain.

4.4.1 Collaborative planning based on conceptual spaces

The key ingredient for successful robot soccer collaboration is effective communication. As discussed in Section 3.5 effective communication can only be achieved if the parties involved share the meaning of the underlying communicated concepts. Conceptual spaces can not only model concepts generated from perceptual information, but they are sufficiently versatile to also model abstract concepts, e.g.

kick-off. In order to develop collaborative plans for soccer we developed a conceptual space of game play plans that cover important and prototypical situations (Karol et al. 2004). We identified a set of quality dimensions and an underlying similarity measure can be used to build plans given any state of the outer world. Some examples of dimensions used are: position of the ball, number of team-mates in attacking positions, the current score.

Each robot maintains his own world model using a Kalman filter and broadcasts it to his teammates so that each robot can also use world models from teammates,⁸ or even combine world models from several teammates, e.g. those close to the ball. The robot's decisions can then be based on his own world model, a specific teammate's world model or a combination of teammate's world models. The processing pathways illustrated in Figure 8 commence with sensing actions which are then used to generate a set of perceptions which is combined with background information and concepts, and various estimates to produce an updated world model using a Kalman filter based simulation.

In order for robots to execute collaborative game play plans they must share detailed information about their outer world models. For example, passing is an important, albeit crucial, skill in soccer, and to affect a successful deliberate pass two robots need to coordinate their behaviour. They do so by sharing rich information about the objects they perceive, as well as their location, heading, and intentions. This information is represented in the world model and in the robots perceptions and represented as concepts using relations like robot and object described in Section 4.2 for communication across the wireless network.

In Karol et al. (2004) we showed how conceptual spaces can be used to represent *prototypical situations* and *important situations* that are encountered during a robot soccer match. Example situations are: kick off, a single attacking player in the goal penalty area, a player with the ball in a goal-end corner, or a player with the ball on the field border. A collection of situation cases can generate a conceptual space for the purpose of providing strategic decision making assistance to robot soccer players. Each case consists of a set: $\text{Case} = \{\text{StateOfField}, \text{DegreeOfPossession}, \text{ActionsToBeTaken}\}$ where the *StateOfField* is a description of the locations of all robots and the ball, *DegreeOfPossession* is a measure of the degree to which a team controls the ball. It is high when a team member is holding the ball and is in the clear and low when the opposition has control of the ball. The *ActionsToBeTaken* constitutes a collaborative plan; it describes the actions that the robot soccer players should adopt.

Using such representations of cases, we can apply the techniques of case-based reasoning. Gilboa and Schmeidler (2001) have shown that giving a limited number

8. For example the team mate who is most certain of his location and closest to the ball.

of cases and an evaluation of their outcomes, under rather general conditions, there exists a *similarity measure* and a *utility function* such that the value of a given new case can be calculated as the weighted values of the previous cases, where the weights are proportional to how similar the new case is to the old cases. For many scenarios the similarity measure can be represented as a metric over a conceptual space.

The benefit of case-based reasoning is that the value of a new case, with a particular set of ActionsToBeTaken, can quite easily be calculated from a set of specially selected cases, in particular cases where the ActionsToBeTaken have lead to particularly clear positive or negative outcomes, that is, cases to copy or to avoid. If the averaging technique described above leads to a disastrous outcome, then this case should be added as a case to avoid. And if it leads to an outcome that is much better than expected, the case should also be added. In this way, the robot soccer players will be able to learn from their successes and failures, without ending up in an exponential growth of computations.

In addition to the object level features we also use several meta-level features (or global parameters) that can aid in the selection of the appropriate strategy. For example, these meta-level features can help us identify a subset of the possible cases that should be considered in a given situation. Furthermore, they can also be used to resolve conflicts when two or more cases are “equally” similar.

4.4.2 Collaborative tracking the ball based on simulations

The two step cycle of prediction based on *simulation* and *perception* that the robot uses to determine their location on the field is fairly straightforward because the simulation is based on concepts of distance, heading and elevation of a single robots camera, however determining a shared view of the ball position across several robots is more difficult because the ball kinematics are not known. In order to determine the position of the ball using information from several team members multiple observations from different perspectives need to be combined in a meaningful fashion.

Initially a robot localises itself using the Kalman filter simulations as described in Section 4.3, and then each robot broadcasts its determination of the ball position and velocity as well as other internally-oriented information regarding its own pose, its neck tilt and pan angles. When a robot receives such a message he calculates the camera locations of each robot and updates the ball position and velocity using a Kalman filter simulation where information about ball position and velocity is fused. How robots go about sharing information is a complex task and the following issues become apparent: (i) should a robot broadcast the information he has about the ball even if it is highly uncertain, (ii) sharing information takes time in which case the balls position and velocity in the outer world may have changed by the time a robot receives information from a team mate, (iii) how

is the accumulated uncertainty among the robots estimations and simulations combined in a meaningful fashion, and (iv) if a robot is about to kick the ball then should he wait until he has done so before broadcasting so that his team mates will look for the ball in the area he is kicking to rather than where it currently is.

5 Discussion

Constructing multi-agent systems that can communicate and collaborate to achieve shared goals in complex and dynamic environments is a difficult and challenging task. In this study we developed a much needed framework and infrastructure that facilitates modelling the interplay among the various cognitive processes that take place in an agent's inner world from which cognitive capabilities such as communication and collaboration are generated. We showed how the framework and accompanying infrastructure can be used to support meaningful communication and effective collaboration within a team of mobile robots.

The framework allows agents to share perceptions, conceptions, simulations, plans, and intentions for the purposes of achieving meaningful communication and effective collaboration. For example, agents can share cued and detached representations of physical and abstract objects, plans, intentions, and as a result they can collaboratively pursue shared objectives.

We showed how grounded cued and detached representations and simulations provide the right kind of cognitive infrastructure to generate sophisticated cognitive capabilities. An agent's inner world is built on representations from which concepts, plans and linguistic descriptions can be derived and shared. In order to achieve the degree of meaningful communication required for a team of robots to effect flexible collaborative behaviour, e.g. passing, and team game play plans, the robots must have a rich and grounded conceptual structure from which to build their communication, and that effective communication and collaboration among intelligent agents requires the sharing of the meaning of concepts and inner world representations. By rich we mean able to capture similarity judgments among concepts e.g. configurations of the field, and by grounded we mean that representations are meaningful in relation to the external world.

References

- Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.

- Chang, Mario, Gamini Dissanayake, Dinesh Gurram, Fedja Hadzic, Gerry Healy, Alankar Karol, Christopher Stanton, Michael Trieu, Mary-Anne Williams & Astrid Zeman. 2004. *Robot world cup soccer 2004: The magic of UTS unleashed!* <http://www.unleashed.it.uts.edu.au/TeamReports>.
- Chella, Antonio, Marcello Frixione & Salvatore Gaglio. 1998. An architecture for autonomous agents exploiting conceptual representations. *Robotics and Autonomous Systems* 25(3–4). 231–240.
- Denier van der Gon, J. J. 1988. Motor control: Aspects of its organization, control signals and properties. *Congress of the International Electrophysiological Society* 7. Amsterdam: Elsevier.
- Gärdenfors, Peter. 1996. Cued and detached representations in animal cognition. *Behavioural Processes* 36. 263–273.
- Gärdenfors, Peter. 2000. *Conceptual spaces*. Cambridge, MA: Bradford Books.
- Gärdenfors, Peter. 2003. *How Homo became Sapiens: On the evolution of thinking*. Oxford: OUP.
- Gärdenfors, Peter. 2004. Emulators as sources of hidden cognitive variables. *Behavioral and Brain Sciences* 27. 403.
- Gärdenfors, Peter & Mary-Anne Williams. 2001. Reasoning about categories in conceptual spaces. *International Joint Conference on Artificial Intelligence (IJCAI)*, 385–392. San Francisco, CA: Morgan Kaufmann.
- Gärdenfors, Peter & Mary-Anne Williams. 2003. Building rich and grounded robot world models from sensors and knowledge resources: A conceptual spaces approach. *International Symposium on Autonomous Minirobots for Research and Edutainment (AMIRE)*.
- Garner, Wendell R. 1974. *The processing of information and structure*. Potomac, MD: Erlbaum.
- Gilboa, Itzhak & David Schmeidler. 2001. *A theory of case-based decisions*. Cambridge: CUP.
- Grush, Rick 1997. The architecture of representation. *Philosophical Psychology* 10. 5–23.
- Grush, Rick 1998. Wahrnehmung, Vorstellung und die sensomotorische Schleife. In Frank Esken & Heinz-Dieter Heckmann (eds.), *Bewußtsein und Repräsentation*, 169–192. Paderborn: Ferdinand Schöningh.
- Grush, Rick 2004. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27. 377–396.
- Gulz, Agnetha 1991. *The planning of action as a cognitive and biological phenomenon*. Lund: Lund University Cognitive Studies 2.
- Hampton, James. 1993. Prototype models of concept representation. In Iven van Mechelen, James Hampton, Ryszard S. Michalski & Peter Theuns (eds.), *Categories and concepts: Theoretical views and inductive data analysis*, 67–95. London: Academic Press.
- Hahn, Ulrike & Michael Ramscar. 2001. *Similarity and categorization*. Oxford: OUP.
- Hesslow, Germund. 2002. Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences* 6(6). 242–247.
- Humphrey, Nicholas K. 1993. *A history of the mind*. London: Vintage Books.
- Ito, Masao 1984. *The cerebellum and neural control*. New York, NY: Raven Press.
- Jäger, Gerhard & Robert van Rooij. 2006. Language structure: Psychological and social constraints. *Synthese*. Published online 26 August 2006. <http://www.springerlink.com/content/67786v01n02m20pt/?p=c2f40d855edf443fa7d9fee6f90b9218&pi=0>.
- Kalman, Rudolf E. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82(d). 35–45.
- Kalman, Rudolf E. & Richard S. Bucy. 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering* 83(d). 95–108.

- Karol, Alankar, Bernhard Nebel, Christopher Stanton & Mary-Anne Williams. 2004. Case-based game play in the RoboCup Four-Legged League: Part I, the theoretical model. *RoboCup 2003: Robot Soccer World Cup VII (Lecture Notes in Computer Science 3020)*, 739–747. Berlin: Springer.
- Kawato, Mitsuo 1990. Feedback-error-learning neural network for supervised learning. In Rolf Eckmiller (ed.), *Advanced neural computers*, 365–372. Amsterdam: North-Holland.
- Maddox, W. Todd 1992. Perceptual and decisional separability. In F. Gregory Ashby (ed.), *Multi-dimensional models of perception and cognition*, 147–180. Hillsdale, NJ: Lawrence Erlbaum.
- McCarthy, John & Patrick J. Hayes. 1969. Some philosophical problems from the standpoint of artificial intelligence. In Donald Michie (ed.), *Machine intelligence* 4, 463–502. New York, NY: Elsevier.
- Melara, Robert D. 1992. The concept of perceptual similarity: From psychophysics to cognitive psychology. In Daniel Algom (ed.), *Psychophysical approaches to cognition*, 303–388. Amsterdam: Elsevier.
- Suddendorf, Thomas & Michael C. Corballis. 1997. Mental time travel and the evolution of human mind. *Genetic, Social and General Psychology Monographs* 123. 133–167.
- Taft, Charles & Lars Sivik. 1997. Salient color terms in four languages. *Scandinavian Journal of Psychology* 38. 26–31.
- Williams, Mary-Anne, Peter Gärdenfors, John McCarthy, Alankar Karol & Christopher Stanton. 2005. A framework for evaluating groundedness of representations in systems: From brains in vats to mobile robots. *International Joint Conference on Artificial Intelligence (IJCAI) Workshop on Agents in Real-Time and Dynamic Environments*. AAAI Press.
- Wolpert, Daniel M. & Mitsuo Kawato. 1998. Multiple paired forward inverse models for motor control. *Neural Networks* 11. 1317–1329.
- Wolpert, Daniel M., Kenji Doya & Mitsuo Kawato. 2003. A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 358(1431). 593–602.

The modal-logical interpretation of the causation of bodily actions

Hiroyuki Nishina

We explore the semantics of action verbs based on how a model skeleton propagates its motions. A bodily action is dissected into intervals encompassing motions which are expressed in terms of two kinds of vertices: endpoints and joints. Rotation/turn functions formulate the motions of vertices and the causation of these motions by their respective supporting vertices. The causation relation between vertices is the basis for the accessibility relation of our possible worlds. The proposition that a vertex moves in a world is valued by the rotation/turn functions at each interval. Rearranging the valid formulae in each world leads to a propagation structure. We suggest that this structure belongs to the local path in a BY-clause modifying the event in the conceptual structure for the verb.

1 Introduction

There hardly seems room to argue against the necessity of visual information for comprehending the meaning of bodily action verbs. We will introduce a strategy for building up the structures of actions from the constituent rotations/turns defined over the skeleton of their performers, and propose an algorithm for abstracting propagation structures from these structures. We will also discuss the possibility that the causation of actions underlies their cognition and that it provides a key to the successful understanding of the meanings of action verbs.

2 Conceptual structure and modal formulae

2.1 Conceptual structure and 3D model

Jackendoff argues that the difference between the meanings of *duck* and *goose* is difficult to describe in terms of the usual binary features using e.g. [$\pm$ long neck] (Jackendoff 1990: 32–34). With regard to this, he advocates that Marr's (1982) 3D

model structure with enrichments such as those proposed in Biederman (1987) shows promise and can be interfaced with the theoretical structure of his conceptual semantics via a set of rules.¹ This correspondence effects a translation of visual information into linguistic format, enabling one to describe what can be seen. Marr's approach is interesting, Jackendoff argues, because the 3D model encodes the decomposition of objects into parts, the geometric systems of spatial axes around which objects are organized, and the relations among the parts. The 3D model is valuable for the analysis of the semantics of verbs of action. Its natural extension can be deployed to encode action patterns such as throwing or saluting. If the lexical entries for the corresponding verbs *throw* and *salute* contain a 3D model representation of the relevant actions, Jackendoff suggests, no distinction at all needs to be made in conceptual structure. Then, verbs such as *run*, *jog*, and *lope* will be treated equally as "verbs of locomotion", and verbs such as *throw*, *toss* and *lob* will be treated equally as "verbs of propulsion", and there will be no need for features as they are insufficient for description and an inadequate semantic apparatus.

One of the reasons that a distinctive semantic feature system is invoked in linguistics is its compatibility with semantic compositionality. When words are composed into phrases and finally into a whole sentence by syntactic rules such as phrase structure rules, categorical rules etc., the semantic or interpretive rules corresponding to the syntactic rules apply to the composed parts to yield the full semantic representation or interpretation. Features, assigned to lexical items or syntactic categories, and combinations of these, work consistently in a compositional way. This means that any semantic distinction that results in a syntactic difference must be encoded in conceptual structure, which in turn implies that differences appearing only in 3D model structure can have no syntactic effects. Therefore, Jackendoff argues, the members of each of the above verb sets ("verbs of locomotion" and "verbs of propulsion") can be identical in their conceptual structure and differ only in the associated 3D model. It is important to note that the syntactically irrelevant description of meaning is missing from the conceptual structure.

2.2 Modal formulae vs. conceptual structure

In the following, we will show how bodily actions denoted by verbs of action are expressed in existential movement formulae that are modally valuated with respect to the rotation/turn functions calculated from the movements in the motions made by its performers. By citing examples in which the inherent and specific

1. Biederman (1987) proposed the Recognition-by-Components Theory, in which the representation of surface curvature is a consequence of features representing axis properties and properties of the contour that is swept along an axis in order to generate an object part.

information of the action verbs' meanings is duplicated as optional manner-adverbial expressions subcategorizing the action verbs, we will point out a possibility that the modal logic formulae denoting that part of meaning can be expressed in the conceptual structure. In Section 3, we will consider a way to describe the motions denoted by bodily action verbs in terms of rotation/turn functions and point out the causative relation involved in the rotations/turns described by these functions. In Section 4, we will see that accessibility holds among causative relations and we will propose a way to value an existential movement formula with respect to the rotation/turn functions involved in an example as model. In Section 5, we will discuss the conceptual structure for action verbs. In Section 6, we will consider a way to simplify the modal formulae derived from action tokens by proposing the propagation structure as semantic representation of an action verb. After these investigations are done, in Section 7, we will return to the problem posed at the outset, which we restate as: Where should we put the modal formulae for actions in the conceptual structure for the relevant verbs?

3 Action as a set of rotations

3.1 Primacy of visual information in the cognition of action

Before we can decide whether we refer to objects in the singular or plural, we normally face them by sight, even if their singularity/plurality is a mere grammatical, not natural, phenomenon. Pointing this out might not be necessary. We, however, put emphasis on this fact, mentioning some examples from Mam, a language in western Guatemala. Mam has no general word for 'lie', as in *lie down*, *lie on the ground*, *lie in bed* (Aitchison 1996: 27). Instead, one has a choice of numerous different words. To verbalize different 'lying' actions properly, a Mam speaker has to know not only who or what is lying – that is, whether it is a human, an animal or a thing – but also the position adopted, as shown in the partial selection of terms for a human male in (1).

(1) *Verbs of action in Mam*

- a. *mutsl* 'he is lying on his stomach'
- b. *pak'l* 'he is lying on his back'
- c. *tʃaltʃ* 'he is lying on his side'
- d. *qinl* 'he is lying outstretched'
- e. *leql* 'he is lying sprawled (probably intoxicated)'
- f. *kutʃl* 'he is lying alone in a house (probably sick)'

This strongly reminds us that we normally discern our object of description by sight.

3.2 Modeling of the human skeleton

A bodily action is a successive change of posture done for a purpose by moving parts of the performer's body. Naturally, these movements are highly dependent upon the physical construction of the parts of its performer. Biomechanical analyses have been applied to actions of humans and animals, shedding light on the kinematic mechanism of human motions (Badler, Phillips & Webber 1992).² Our ultimate interest is in how the cognition of action is related to linguistic semantic (and syntactic) structures. To probe into this relation, we introduced the description of bodily actions via rotations/turns around the joints of the human skeleton (Nishina 1999).

When modeling the physiological joint structure into a model skeleton, we borrow a limited number of actual joints and endpoints and subsume them under the "content vertices" of the model skeleton. With respect to an arbitrary individual, its total content vertex set (C) is the union of its simplified joints (J) and endpoints (E):

- (2) *Constituency of an individual*
 $C = J \cup E$, where $J \cap E = \emptyset$

We choose some pairs of vertices from this set and connect them into a "support set", as will be discussed in Section 3.4, because rotation and turn presuppose support.

3.3 Rotations and turns around joints

The joints of a real human are generally subcategorized into five groups or more by their modes of rotation (Zatsiorsky 1997). We, however, radically simplify the categorization into two types of rotation: *rotation* and *turn*. *Rotation* (R) is the type of rotation in which the rotating vertex, the center vertex of this rotation and the vertex supporting the center belong to the same plane, and *turn* (T) is the rotation in which the rotating vertex belongs to a plane perpendicular to the edge including the axis (edge) of the turn. We simulate the real rotations involved in actions using any or both of these two simplified rotations. In the following part-list, each endpoint/joint part of the body corresponds to a vertex or a subset of vertices.

2. Simulation of humans has a comparatively long and rich tradition of research. As a superb example of the products from agent technology, *Jack* has been developed by Badler, Phillips and Webber (1992). The default polyhedral human figure in *Jack* is composed of 69 segments, 68 joints, 136 degrees of freedom, and 1183 polygons (including cap and glasses). The reason that we use our own more primitive and discrete skeleton and rotation/turn functions in this chapter is that the purpose of our research lies in the abstract representation of movement propagation on the skeleton.

(3) *Simplified skeleton*

{Head: endpoint; Neck: joint [R/T]; Shoulders: joints [R/T]; Elbows: joints [R/T]; Wrists: joints [R]; Fingers: endpoints; Hips I (Sacro-iliac): joint(s) [R/T]; Hips II (Pelvis): joints [R/T]; Knees: joints [R]; Ankles joints[R/T]; Toes: endpoints}

The joints act as the centers of rotation (R) and as the axis of turn (T).

3.4 Mapping bodily actions into support sets

From a physical point of view, when an individual carries out a motion, one or more parts remain stable and “support” the moving part(s). Thus, when a human walks, the foot and leg pushed forward on one side are supported by the foot and leg on the other side, which, in turn, are grounded on an outer object, such as the ground or its equivalent. This shows that support underlies motion, acting as the basis for transmitting movements to the point of operation where external causation has an effect. Note that in a case of rotation, a vertex acting as the center of the rotation of another vertex also supports the latter, even if the latter stops rotating. Thus, each motion presupposes a structure in which more than one of its content vertices support others. When a human assumes a still standing posture, making no motions, its skeleton forms a set of “support” relations:

(4) *Support relation*

c_i “supports” c_j iff $\langle c_i, c_j \rangle \in C \times C$, where $i \neq j$ and C is the total content vertex set

Seen as one type of content vertex set, a support set is the set of content vertices, each member of which is the endpoint of some support relation. Support relations may hold within one individual or between two or more individuals, with the latter case being beyond the scope of this chapter. As a unit of movement, we define a “direct” support relation:

(5) *Direct support set*

$\{\langle c_i, c_j \rangle \mid c_i \text{ “supports” } c_j, \text{ and there is no } c_k \text{ such that } c_i \text{ “supports” } c_k \text{ and } c_k \text{ “supports” } c_j\}$

Let us suppose that support is a transitive relation. We can then get an indirect support set by connecting pairs transitively from the direct support set. The union of “direct” and “indirect” support sets is the “total” support set, resulting in a graph from which pairs of vertices are selected to define “rotate” pairs and “turn” pairs.

(6) *Total support set*

Direct support set $\cup$ Indirect support set

3.5 Rotation and turn functions

As the total support set is a default structure for action, rotations are defined on this set. When a rotates around b on a model skeleton, we can interpret this as a causative event where the center b causes a to *rotate* around it. Likewise, we can say the axis $\langle c, b \rangle$ causes a to *turn* around it. We will see how the two types of rotation – *rotation* and *turn* – are mapped into a support set.

Let us define a function returning the movement caused by a content vertex rotating around another vertex as the movement's presence or absence:

(7) *Rotation function*

$$\pm R^i(\langle c_j, \sigma/\mu \rangle, c_k) = \sigma/\mu, \text{ where } c_j \text{ supports } c_k$$

The functor “R” shows that the relevant rotation is *rotation*, not *turn*, its polarity “ $\pm$ ” distinguishes the closing/opening rotation, and the superscript “i” denotes the number of the interval during which the relevant rotation continues.³ The content vertices involved are denoted as “c’s”, and “ σ ” and “ μ ” distinguish “stay” and “movement” states. This function, with c_j as the center of *rotation* in its stay or movement state, takes c_k as its input and returns as its output the state μ if c_k has been moving during the interval i and the state σ if it has not. The number of the intervals must be minimal in the sense that each interval is distinct from the others in its rotation/turn membership.

Turn is a bit different. The “center” of a turn is expressed as an edge, not as a vertex as has been seen in *rotation*. A “double-headed” type of function is invoked for this purpose:

(8) *Turn function*

$$\pm T^i(\{\langle c_j, \sigma/\mu \rangle, \langle c_k, \sigma/\mu \rangle\}, c_l) = \sigma/\mu, \text{ where } c_j \text{ and } c_k \text{ support } c_l \text{ respectively, and } c_j \text{ directly supports } c_k$$

The polarity “ $\pm$ ” distinguishes the clockwise/counter-clockwise turn seen from the center of the skeleton. This function, with the edge $\langle c_j, c_k \rangle$, each of whose member is in the state σ or μ , as the axis of *turn*, takes c_l as its input and returns as its output μ if c_l has been moving during the interval i and σ if it has not. For a turn to be successful, it is necessary that each vertex of the axis pair supports the turning vertex. In terms of support, the one vertex of the pair nearest to the turning vertex

3. The polarity attached to the rotation function shows only the direction of the relevant rotation, not the specific angle of rotation, which is specifiable if necessary. Our main topic in this chapter, however, is the valuation based on rotation/turn functions, so we omit the specification of the angle from the functions throughout this chapter. The same convention applies to the turn function.

is directly supported by the other, which means that there is no intervening vertex in between.⁴

3.6 Causative analysis of rotation and turn

Generative semantics proposed a causative analysis which decomposes verb meaning in terms of semantic primitives, opening up a possibility of analyzing the semantic contents packed within a lexical verb (Lakoff 1970; McCawley 1970). Thus, the lexicalized form of the causative verb *kill* is obtained by a predicate raising from the embedded structure of semantic primitive predicates (Postal 1972):

- (9) [CAUSE X [BECOME [NOT [ALIVE Y]]]]
 ⇒ [CAUSE X [BECOME [(NOT ALIVE) Y]]]
 ⇒ [CAUSE X [(BECOME NOT ALIVE) Y]]
 ⇒ [(CAUSE BECOME NOT ALIVE) X Y]
 ⇒ [*kill* X Y]

In this tradition, Talmy (1976) gives a detailed analysis of the causation types of *break*. He, however, concludes that the use of the single deep verb CAUSE is insufficient, proposing a subcategorization of the causative *break* into five types: resulting-event causative (basic causative), causing-event causative, instrument causative, author causative and agentive causative. Using the third of these types of causative, instrument causative, we are able to describe the causative structure implied in the two types of rotation:

- (10) a. *b* “rotates” *a* by acting as the center of *a*’s rotation.
 b. *<c, b>* “turns” *a* by acting as the axis of *a*’s turn.

Here, “rotate” and “turn” are used as causative transitive verbs. Both sentences are well-formed without any problem. According to Talmy’s subcategorization of causation, they seem to fall under “instrument causation”, because they allow the *by*-phrases playing the semantic role of “instrument”. Talmy (1976: 69) proposes that this type of causative is derivable from the underlying structure whose main predicate is an INSTRUMENT causative and is modified with a WITH phrase (for *with*, *in* and *by*).

4. Note that if embedding by substitution applies, indirect causativity will result from this direct support relation between the axis pair. This is an unwelcome fact, because in the case of turn, each of the vertices of its axis directly causes the relevant vertices to turn independently. We can evade this problem by decomposing the turn function into the conjunction of two functions. See the discussion of (36) in Section 6.2.

Before going into details, let us briefly review the semantic analysis of verbs of motion. In his elaborate framework of conceptual semantics, Jackendoff (1983, 1990) proposes a semantic analysis by decomposing causative verbs into a binary CAUSE function, taking two arguments, with an entity as its first argument and an event as its second.⁵

$$(11) \quad [_{\text{Event}} \text{CAUSE} ([_{\text{Thing}} x], [_{\text{Event}} y])]$$

This roughly means that an entity (x) acts physically and/or mentally upon another entity so that an event (y) involving the latter entity will come into existence. Using a causative predicate, Jackendoff (1983: 175) argues that some action verbs have the following conceptual structures:

- (12) a. *The wind pushed Sim into the room.*
 b. $[_{\text{Event}} \text{CAUSE} ([_{\text{Thing}} \text{WIND}], [_{\text{Event}} \text{GO} ([_{\text{Thing}} \text{SIM}], [_{\text{Path}} \text{INTO} ([_{\text{Thing}} \text{ROOM}])])])]$
- (13) a. *Beth threw the ball out the window.*
 b. $[_{\text{Event}} \text{CAUSE} ([_{\text{Thing}} \text{BETH}], [_{\text{Event}} \text{GO} ([_{\text{Thing}} \text{BALL}], [_{\text{Path}} \text{OUT} ([_{\text{Thing}} \text{WINDOW}])])])]$

The “Event” in the second argument position of the CAUSE function in each structure consists of a functor “GO” and its arguments “Thing” and “Path”.⁶ The GO

5. Causative analyses in generative semantics such as the one by Postal have a thing not an event (or a proposition) as its first argument, but there are recent studies in which the first argument is an event (a proposition), not a thing, like in Parsons (1990: 105–126). This originates in Davidson’s analysis (e.g. 1967a, 1967b). The reason for this is that greater generality is achieved by having an event as the first argument than a thing as the first: $[_{\text{Event}} \text{CAUSE} ([_{\text{Event}} \text{DO}([_{\text{Thing}} x], [_{\text{Action}} z]), [_{\text{Event}} y])]$. We can see this from the fact that the analysis with the event as its first argument is adequate for accommodating the sentential event of “John’s blowing bubbles made us laugh”. But Jackendoff argues that if we consider a sentence like “John’s blowing bubbles made us laugh by making us realize how drunk we all were”, the means (instrument) expression, *by*-phrase, cannot be taken to fill the variable z in the structure. As a restrictive modifier, this expression is attached to the causative clause with an event as its first argument. Thus, Jackendoff allows an event-causer type as well as a thing-causer type. Note that a causative analysis with a thing as its first argument is more suitable for our framework in that the causer needs the status of a vertex/either of a vertex pair, not of an edge, in the cases of *rotation/turn*. In our framework, a rotating/turning joint acts as the causer by itself. Even if we were to adopt the Davidsonian type of causative, however, we could naturally take an edge where one vertex supports another, as its causer, and another pair whose first member is the second of the causer pair, as its causee, in which case both causer and causee would be events.

6. The “localistic” theory by Gruber (1976) and others claims that many verbs and prepositions appear in two or more semantic spaces and form intuitively related paradigms. Specifically, the relative semantic relations between *go* (or *change*, in connection to the prepositions *from* and/or

function denotes a physical/abstract translation of an entity/property on a route specifying its course of translation. The meanings of both sentences, however, are different in that the former lacks the willful agency of the subject, the causer, while the latter has one. This difference is not specified in these simplified representations.

3.7 Interpretation of rotation and turn functions

By adopting the causative analysis in Jackendoff's conceptual semantics and applying it to our cases, we can integrate (10a) and (10b) into the interpretations of rotation and turn functions:

- (14) *Interpretation of rotation and turn functions*
- When *a* rotates around *b*, the rotation function approximately says "*b* CAUSES *a* to MOVE by *b*'s acting as a center".
 - When *a* turns around $\langle b, c \rangle$, the turn function approximately says "*b/c* CAUSES *a* to MOVE by *b/c*'s acting as an axis".
 - When there is no rotation/turn, the rotation/turn function shows no causation, returning STAY.

Hence we will use MOVE as a semantic primitive instead of the semantic primitive GO, as is used in Jackendoff's framework.

One might think it more realistic from a physical point of view to admit that these centers/axes themselves undergo some motion when acting as centers/axes, but from the theoretical viewpoint taken here we do not count them as motions.

4 Sample action

4.1 Simplified action

We are ready to see how an action by a part is to be described in our rotational framework. Suppose that the model skeleton shown in the middle of Figure 1,

to), *be* and *keep* are kept in different semantic fields: spatial location and motion, possession, ascription of properties and scheduling of activities. Jackendoff (1990: 26) accounted for the paradigms by claiming that they are each realizations of the basic conceptual functions as follows:

- $GO ([], [_{Path} FROM ([]), TO ([])])$
 - $BE ([], [_{Place}])$
 - $STAY ([], [_{Place}])$

These conceptual structures analyze the meaning of *go* (or *change*), *be* and *keep* respectively. Jackendoff argued that the four semantic fields have parallel conceptual structure, claiming that one of the three types of structure underlies the (partial) semantic structure of each verb.

whose content vertex set is $\{a, b, c, d\}$, has been performing the motions as shown in the left and right of Figure 1 over a period of time. This could be taken as an abstract example or the realistic case of folding an index finger or the like. Note that there is no *turn* here.

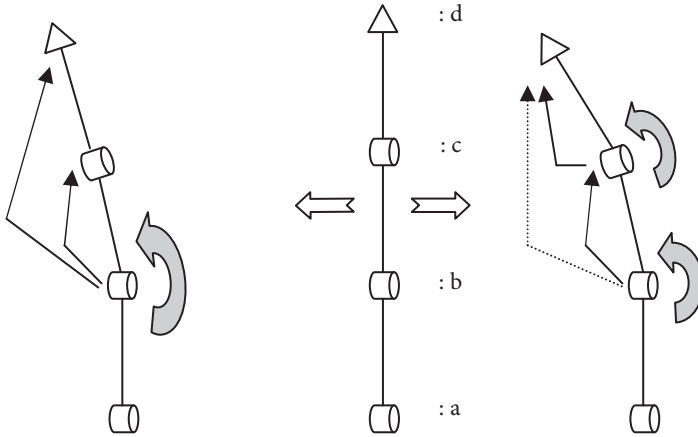


Figure 1. Direct rotations (left) and direct and indirect rotations (right)

The middle of Figure 1 is a starting posture, showing a support structure with no motion. It has the following two types of set:

- (15) *Direct support set in Figure 1* $\{ \langle a, b \rangle, \langle b, c \rangle, \langle c, d \rangle \}$
- (16) *Indirect Support Set in Figure 1* $\{ \langle a, c \rangle, \langle a, d \rangle, \langle b, d \rangle \}$

Every single movement made by this skeleton is to be captured in terms of rotation and/or turn functions defined on the total support set, which is the union of the direct and the indirect support sets:

- (17) *Total support set for the skeleton in Figure 1*
 $\{ \langle a, b \rangle, \langle b, c \rangle, \langle c, d \rangle, \langle a, c \rangle, \langle a, d \rangle, \langle b, d \rangle \}$

Suppose that each of the left and right postures in Figure 1 has been directly assumed from the middle posture. The directions of rotation are shown with thick curved arrows. In the motion indicated by the change from the middle posture to the left, the vertices c and d have moved simultaneously because of the rotation around the center b . The two arrows in the left figure show these causations. These two cases of rotation are identified as “direct rotations”. In the motion from the middle posture to the right, the vertex d undergoes a double rotation, where d not only

directly rotates around c but also “indirectly rotates” around b (d directly rotates around c , which directly rotates around b , therefore d indirectly rotates around b).

The motion depicted by the middle and the right skeletons is described in terms of the causative analysis discussed in the above as:

(18) *Motion completed by the right posture in Figure 1*

b causes c to move by b 's closing rotation, while c causes d to move by c 's closing rotation simultaneously. b indirectly causes d to move via c by b 's closing rotation.

In the right figure, the two arrows show direct causations of movement, and a dotted arrow shows an indirect causation.

4.2 Causations in rotations

Let us now apply the rotation function to each potential center of rotation on the skeleton whose posture has shifted from the middle one to the right one shown in Figure 1. Here, there is only one interval. As was specified in the definition (7), rotation relations presuppose support relations. The rotations counted on the right skeleton are:

(19) a. $1ab: \pm R^1(<c_a, \sigma>, c_{b/c/d}) = \sigma$

During the 1st interval, b , c and d are undergoing neither closing nor opening rotation around a , which is staying.

b. $1bc: +R^1(<c_b, \sigma>, c_c) = \mu$

During the same interval, c is undergoing a closing rotation around b , which is staying.

c. $1cd: +R^1(<c_c, \mu>, c_d) = \mu$

During the same interval, d is undergoing a closing rotation around c , which is moving.

Among these rotations, we find a transitive pair of causative relations of movement between content vertices: b causes c to move in $1bc$, and c causes d to move in $1cd$. Connecting these causative relations leads to b 's indirect causation of d to move (by way of c). Indirect causations such as this are attained by substitution, via substituting the left-hand side expression of $1bc$ for the μ in the pair of the left-hand side expression of $1cd$. Under the identity of that pair with the pair of the input c_c and the output μ of the function $1bc$, we have $1bd$:

(20) *Function embedding by substitution*

$$a. \quad 1bc: +R^1(<\underline{c_b}, \sigma>, \underline{c_c}) = \mu \quad 1cd: +R^1(<c_c, \mu>, c_d) = \mu \Rightarrow$$

$$b. \quad 1bd: +R^1(<c_c, +R^1(<c_b, \sigma>, c_c)>, c_d) = \mu$$

Looking at the function (20b) from inward to outward, we can see that b , staying as a center, is causing c to move by rotation, while c , making such a movement as a center, is causing d to move by rotation. Thus, an indirect causation of movement is described explicitly as a formula with one rotation function embedded in another.

4.3 Model for modal predicate logic

In modal predicate logic, if M is a model with W as its set of possible worlds, R as its accessibility relation, and V as its valuation, the truth-value of a formula in w given M is definable (Gamut 1991; Hughes & Cresswell 1996). Nishina (1999) proposes that the propagation of the movements of an action is described by modal logic based on rotation functions. If we assign vertices to worlds, causation to accessibility and rotation/turn functions to valuation, in the model, we can value the formula that there is a moving vertex in a world:

- (21) *Existential movement formula*
 $\exists x \in C [\text{MOVE}(x)]$ in w , where $w \in W$

With this model, we are able to represent the propagation of the movements of an action. In the following sections, we will discuss the main constituents in this model, first worlds and accessibility and second valuation.

4.4 Accessibility from worlds

We saw in Section 4.1 that the direct support set involved in the bending action in Figure 1 is $\{ \langle a, b \rangle, \langle b, c \rangle, \langle c, d \rangle \}$. While the non-moving part $\langle a, b \rangle$ does not directly participate in the causation of action, it constitutes the remnant part of the skeleton, acting as the basis supporting the movements involved. In the case of a human skeleton in its normal standing posture, the earth first supports its feet, which then support the knees, which in turn support the pelvis, which supports the backbone, and so on up to the head. For these relations to be successfully connected, it is necessary to allow transitivity between the support relations over the skeleton. Recall that all the causative relations of movement possible on the skeleton in motion are defined over its total support set of content vertices. From the two thick arrows in the right skeleton of Figure 1, we can see that its direct causation set is: $\{ \langle b, c \rangle, \langle c, d \rangle \}$. Additionally, we have an indirect causation set as: $\{ \langle b, d \rangle \}$, which has been attained by function embedding. The union of these sets is the total causation set $\{ \langle b, c \rangle, \langle b, d \rangle, \langle c, d \rangle \}$.

Now let us make the following assumption:⁷

(22) *Content vertex to world correspondence*

Each of the content vertices constituting a skeleton corresponds to the world co-indexed with that vertex, where that vertex is MOVING.

We will also assume the following condition:

(23) *Accessibility condition*

A world is “accessible” from another world if and only if the content vertex in the latter *directly or indirectly* causes the content vertex in the former to MOVE.

The worlds defined for valuation are connected with the causative relations holding between them. To allow the full extensibility of the propagation of the movements caused by rotation and turn, we must utilize direct and indirect causation as the accessibility relation for the valuation. Applying the vertex-world correspondence to the motion described by the middle and right postures in Figure 1, we find that each element of the content-vertex set $\{a, b, c, d\}$ corresponds to its counterpart of the world set $\{w_a, w_b, w_c, w_d\}$, respectively. Based on the total causation set of the motion made by this skeleton, we define the accessibility relation between worlds as an ordered pair of a world before another world, in which the content vertex in the latter is “accessible from” that vertex in the former:

(24) *Accessibility relation set for the right motion in Figure 1.*

$\{ \langle w_b, w_c \rangle, \langle w_b, w_d \rangle, \langle w_c, w_d \rangle \}$

4.5 Valuation of formulae on rotation functions

We utilize the value of a rotation/turn function as the condition for valuation, V . For the rotation function, if and only if a function denoting the movement of content vertex c_k by rotation returns μ (moving) as its value, the existential movement

7. The assumption of vertex to world correspondence raises an interesting problem. According to (22), in every world, there is a vertex moving or not moving. The representation of this sentence invokes an existential quantifier in its formula. Thus far, the situation leads us to choose intensional predicate logic for the construction of a model. Note, however, that (22) also says that there is only one vertex of that kind. This means that each world has a set of one member, a singleton set. Selecting one member from a singleton set causes no plural possibilities. It is impossible that one world is different from another in domain. So regarding the movement formula as a kind of proposition, despite its quantifier, we have chosen intensional propositional logic. The valuation of quantified formulae between the worlds whose domains can be different would increase complexity. We will leave this to further studies.

formula is valuated as true in the world relevant to that vertex. Otherwise, the formula is valuated as false:

(25) *Valuation of rotation function*

$$\begin{aligned} V_{W_k} (\exists x \in C [\text{MOVE}(x)]) &= 1 \text{ iff } \pm R^i (\{ \langle c_j, \sigma/\mu \rangle, \langle c_k, \sigma/\mu \rangle \}, c_l) = \mu, \\ \text{otherwise } V_{W_k} (\exists x \in C [\text{MOVE}(x)]) &= 0 \end{aligned}$$

For turn, if and only if the function denoting the movement of content vertex c_l by turn returns μ as its value, the existential movement formula is valuated as true in the relevant world. Otherwise, the formula is valuated as false:⁸

(26) *Valuation of turn function*

$$\begin{aligned} V_{W_l} (\exists x \in C [\text{MOVE}(x)]) &= 1 \text{ iff } \pm T^i (\{ \langle c_j, \sigma/\mu \rangle, \langle c_k, \sigma/\mu \rangle \}, c_l) = \mu, \\ \text{otherwise } V_{W_l} (\exists x \in C [\text{MOVE}(x)]) &= 0 \end{aligned}$$

With all these, we are able to define the model M with W as a set of worlds, R_{acc} as accessibility relation and V as valuation. In our discussion to follow, we will omit the valuation function and write merely the true formula without the functor V or the truth value, unless they are necessary.

4.6 From rotation to movement formula

In the model we have been discussing, we valuate the truth-value of the existential movement formula (21) in each world as follows:

(27) *Valuation with respect to each world for the right motion in Figure 1:*

- a. $\neg \exists x \in C [\text{MOVE}(x)]$ in w_a : In world w_a , there is no moving vertex caused by any center.
- b. $\neg \exists x \in C [\text{MOVE}(x)]$ in w_b : From $+R^1(\langle c_a, \sigma \rangle, c_b) = \sigma$, where there is no moving vertex in world w_b , since there is no other vertex working as the center causing such a rotation.
- c. $\exists x \in C [\text{MOVE}(x)]$ in w_c : From $+R^1(\langle c_b, \sigma \rangle, c_c) = \mu$, where, in world w_c , there is a moving vertex caused by a staying center in world w_b .
- d. $\exists x \in C [\text{MOVE}(x)]$ in w_d : From $+R^1(\langle c_c, \mu \rangle, c_d) = \mu$ where, in world w_d , there is a moving vertex directly caused by a moving center in world w_c .
- e. $\exists x \in C [\text{MOVE}(x)]$ in w_d : From $+R^1(\langle c_c, +R^1(\langle c_b, \sigma \rangle, c_c) \rangle, c_d) = \mu$ where, in world w_d , there is a moving vertex indirectly caused by a staying center in world w_b .

8. In the case of valuation the turn function must be decomposed into the conjunction of two functions.

Note that the function in (27e) is doubly valuated in world w_d , through (27c) and (27d).

4.7 Modal formulae

Based on the causation relations denoted by the rotation functions, accessibility relations are calculated between the worlds where a content vertex is moving. First, to get an idea of accessibility, see Figure 2:

	w_a	w_b	w_c	w_d
w_a				
w_b				
w_c		A		
w_d		A	A	

Figure 2. Accessibility relations from worlds

The accessibility relations (24) are depicted in Figure 2. Each A-mark indicates that the world of its row is accessible from the world of its column. Thus the light grey cells show that worlds w_c and w_d are accessible from world w_b , and the dark grey cell that world w_d is accessible from world w_c . From Figure 2, we can see the following:

(28) *Accessibility relations from world w_b :*

- Since w_c and w_d are the only worlds accessible from w_b and the existential movement formula holds true in these worlds, this formula is necessary in w_b .
- $\Box[\exists x \in C [\text{MOVE}(x)]]$ in w_b

This accessibility relation is shown with thick arrow 1 in Figure 2. Likewise, we can also see the following:

(29) *Accessibility relation from world w_c :*

- w_d , where the existential movement formula holds, is the only world accessible from w_c .

- b. $\Box[\exists x \in C [\text{MOVE}(x)]]$ in w_c

In Figure 2, this accessibility is shown with thick arrow 2. Connecting the light grey and the dark grey cells in the diagram, we have the following:

- (30) *Double accessibility relations from world w_b* :
- a. The necessity formula seen in (29b) holds only in w_c of the two worlds w_c and w_d , accessible from w_b . Thus, this formula is possible in w_b .
 - b. $\Diamond[\Box[\exists x \in C [\text{MOVE}(x)]]]$ in w_b

From the definition of its valuation, it is important to note that necessary formulae hold in the world corresponding to an endpoint. We, however, neglect the accessibility from endpoints, which do not act as causers of motion, in the denotation of the agent of the verb *per se*.

Each of the attained modal formulae checks the presence or absence of the exhaustiveness of the vertices whose movements are caused by the rotation of another vertex as their center, from which accessibility is measured in terms of direct and indirect causation.

5 Conceptual structure for *kick*

5.1 Inalienable adverbial instrument

Several verbs of physical action show an interesting pattern: They take a prepositional phrase acting as the point of action (seen from a physical point of view).

- (31) a. *Sue kicked Fred with her right foot/*a stick.*
 b. *Sue kicked Fred.*

The prepositional phrase (PP) *with her right foot* in (31a) is an adverbial, having an “instrument” role in thematic relations. As the unacceptability of *with a stick* in (31a) shows, the instrument must be a body part of the “agent” – that is in particular, being inalienably possessed by the “agent”. The sentence means that Sue, by using her right foot, caused her right foot to come into contact with Fred with some impact. This implies that there is information about the path through which the energy generated in Sue reached her right foot, the point of action on the target, from which it is passed on to Fred. The meaning of the verb, *kick*, in (31b) is less specific as to whether the left foot or the right one was involved in the action. Thus, the meaning denoted by this type of verb requires as instrument some body part the name of which is associated with the respective verb. Aside from *kick*, we have *shoulder*, *grasp*, *grip*, *elbow*, *flick*, *snap* etc. They form a natural category: “verbs of action with inalienably

possessed instrument”. In the conceptual structure of this verb type, the instrument must be an inalienable part of the agent.

5.2 Japanese verbs of action

We find verbs of this category in a variety of languages. One such example is Japanese: *atama-de kangaeru* ‘consider with head’, *kata-de ninau* ‘carry with shoulders’, *ude-de burasagaru* ‘hang down with arm’, *te-de nigiru/nigirisimeru* ‘grasp/grip with hand’, *hizi-de tuku* ‘nudge with arm’, *yubi-de haziku* ‘flick with thumb or fingers’, *asi-de aruku* ‘walk with feet/legs’, *asi-de keru/humu* ‘kick/step on with feet’.

- (32) a. *Taro ga Jiroro o migi asi de ket-ta.*
 Taro NOM Jiro ACC right foot INST kick-PAST
 ‘Taro kicked Jiro with his right foot.’
- b. *Sono koto o zibun no atama de yoku kangae-nasai*
 that matter ACC self GEN head INST well consider-IMP
 ‘Consider that matter well with (your) mind.’
- c. *Kare wa raketto o ryoote de nigirisime-ta.*
 he TOP racket ACC both hands INST grip-PAST
 ‘He gripped a racket with both (his) hands.’

In each of these sentences, the expression containing the *-de* case marker of instrumental (alternatively, postposition) as optional adverbial specifies the point of operation explicitly. The NP-*de* expressions in fact redundantly specify the point of operation as precisely the relevant verbs do themselves.

5.3 Conceptual structure for *kick*

According to Jackendoff’s categorization of verbs, *kick* is decomposed as the causation by an agent of an inchoation of a state:

- (33) *Sue kicked Fred with her right foot.*

$$\left[\left[\begin{array}{l} \text{CS}([SUE_i], [\text{INCH}[\text{BE}([HER_i \text{ RIGHT FOOT}], [\text{AT}[\text{FRED}]]))]) \\ \text{AFF}([SUE_i], [\text{FRED}]) \end{array} \right] \right]$$

$$\left[\left[\begin{array}{l} \text{BY} \left[\begin{array}{l} \text{CS}([SUE_i], [\text{AFF}([HER_i \text{ RIGHT FOOT}], [\text{FRED}]])) \\ \text{AFF}([SUE_i], [HER_i \text{ RIGHT FOOT}]) \end{array} \right] \right] \right]$$

Note that the function BE takes a thing as its first argument and a place as its second argument, denoting the state where the former is in the location of the second. The

inchoative function *INCH* takes a state as its argument, denoting the change into that state. The causative function *CS* takes a thing and an event as arguments, denoting the latter's causation by the former. *AFF* designates its first argument as agent and its second as patient. The instrumental meaning expressed by the PP of the sentence is mapped in terms of the function *BY*, the means expression, whose argument is a restrictive modifier on the main conceptual clause. The function *BY* only specifies that its first argument causes the first argument of *AFF*, as agent, to act on the second of *AFF*, as patient, while the first argument of *CS* is agent and the first of *AFF* patient. Thus, the means expression in (33) says that Sue, acting on her right foot, caused her right foot to act on Fred, without saying any more as to what the particular actions were.

We have discussed the conceptual structure of *kick*. We will see how its denotation is to be described in our rotation framework.

6 Rotations and turns in kicking action

6.1 Action of kicking

When describing bodily action verbs, we face some cases where only a few of the motions involved in actions are sufficient for description and other cases where even all the motions are not sufficient. The kicking action belongs to the former case, its partial description provides sufficiently distinctive information to define the meaning of the corresponding verb. So, we explicitly utilize only the sufficient part of the direct support set for defining the movements involved in the action, by limiting it to the lower half of the skeleton:

(34) *Direct support set for kicking*

$$\{ \langle c_S, c_N \rangle, \langle c_S, c_H \rangle, \langle c_S, c_{H'} \rangle, \langle c_H, c_K \rangle, \langle c_{H'}, c_{K'} \rangle, \langle c_K, c_A \rangle, \langle c_{K'}, c_{A'} \rangle, \langle c_A, c_T \rangle, \langle c_{A'}, c_{T'} \rangle \}$$

Here, c_N denotes a joint simplified from the neck and c_S a joint simplified by putting the pair of sacro-iliac joints together. c_H and $c_{H'}$ are simplified from the pair of right and left hip joints respectively. c_K is from the right knee joint, c_A from the right ankle joint, and c_T the endpoint of the right foot simplified from the right toes, with $c_{K'}$, $c_{A'}$ and $c_{T'}$ their left counterparts.

The following postures in Figures 3 and 4 show the rotations/turns at each interval.

Note that it should not be counted as a part of the meaning whether the moving parts are left or right, which need not be specified in the definition. The figure at $t_{\underline{0}}$

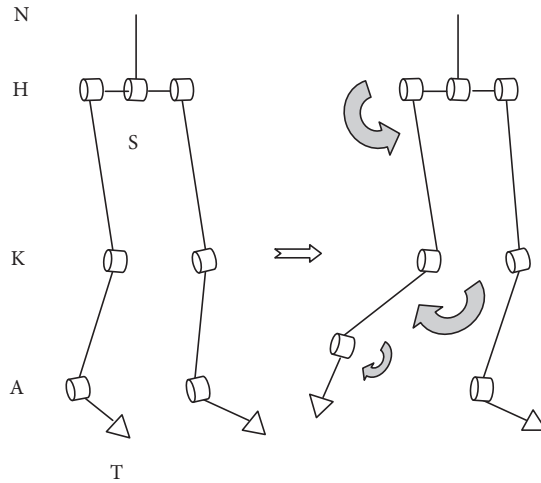


Figure 3. Kicking: Interval 1 (t_0-t_1) and interval 2 (t_1-)

is a normal standing posture at default. At interval 1, the right foot is being pulled up backward, while the right knee is being raised. At interval 2, the right leg and foot are being pushed forward, as the right thigh is being pushed forward. At interval 3, the right thigh and leg continue to be pushed forward, with the right foot kept in a fixed position relative to the right leg.

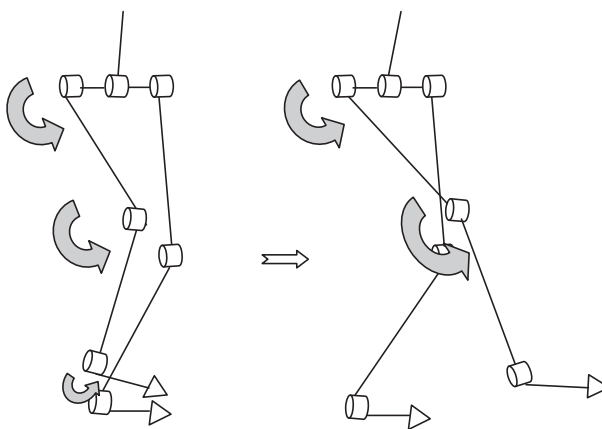


Figure 4. Kicking: Interval 2 (t_1-t_2) and interval 3 (t_2-t_3)

6.2 Rotation/turn functions for kicking

We will investigate how rotations and turns are involved in each of the motions that form the action of kicking. Calculating the direct rotation/turns from the movements involved in these motions, we have:

$$\begin{aligned}
 (35) \quad & \text{Direct turn and rotations at interval 1} \\
 (1SHK): & +T^1(\{<c_s, \sigma>, <c_H, \sigma>\}, c_K) = \mu \\
 (1KA): & +R^1(<c_K, \mu>, c_A) = \mu \\
 (1AT): & -R^1(<c_A, \mu>, c_T) = \mu
 \end{aligned}$$

Embedding a direct function into another by substitution leads to an indirect function. Note that when we substitute the left-hand side of (1KA) for μ of (1AT), we have:

$$(1KT): -R^1(<c_A, +R^1(<c_K, \mu>, c_A)>, c_T) = \mu,$$

where one rotation function is embedded in another rotation function. It is important to note that (1SHK) is a case of turn, where the axis is not a single vertex but an edge, which is shown as the set of pairs of vertex and movement/stay in the function. When the left edge of a turn function is used for substitution, we must divide it by each pair. So, (1SHK) is decomposed into the conjunction of two functions (1SK) and (1HK):

$$\begin{aligned}
 (36) \quad & \text{Decomposition of turn function} \\
 (1SHK) & \Rightarrow (1SK) \ \& \ (1HK) \\
 +T^1(\{<c_s, \sigma>, <c_H, \sigma>\}, c_K) & = \mu \Rightarrow +T^1(<c_s, \sigma>, c_K) = \mu \\
 & \ \& +T^1(<c_H, \sigma>, c_K) = \mu
 \end{aligned}$$

In the following discussion, we will use the conjunction of functions as equivalent to a turn function. When substituting the left-hand side of (1SHK) for μ of (1KA), we utilize the conjunction of the divided turn functions, (1SK) & (1HK), in place of (1SHK). Here, we can see hybrid cases of substitution where a rotation and a turn coexist in the motions involved. Thus, by substituting the left-hand side of (1SK) for μ of (1KA) and that of (1HK) for μ of (1KA), we have (1SA) and (1HA) respectively:

$$\begin{aligned}
 (37) \quad & \text{Indirect rotations at interval 1} \\
 (1SA): & +R^1(<c_K, +T^1(<c_s, \sigma>, c_K)>, c_A) = \mu \ \& \\
 (1HA): & +R^1(<c_K, +T^1(<c_H, \sigma>, c_K)>, c_A) = \mu
 \end{aligned}$$

We can attain doubly indirect cases of rotation. Substituting the left-hand side of (1HA) for μ of (1AT) and the left-hand side of (1SA) for μ of (1AT), respectively, we have the following results:

(38) *Doubly indirect rotations at interval 1*

$$(1HT): -R^1(<c_A, +R^1(<c_K, +T^1(<c_H, \sigma>, c_K)>, c_A)>, c_T) = \mu$$

$$(1ST): -R^1(<c_A, +R^1(<c_K, +T^1(<c_S, \sigma>, c_K)>, c_A)>, c_T) = \mu$$

At the second interval, the joints acting as the centers/axes of rotation/turn are the same as those at the first interval, though the directions of the rotations around knee and ankle at the second interval, are opposite to those at the first.

(39) *Direct turns/rotations at interval 2*

$$(2SHK): +T^2(\{<c_S, \sigma>, <c_H, \sigma>\}, c_K) = \mu$$

$$\Rightarrow (2SK): +T^2(<c_S, \sigma>, c_K) = \mu \ \& \ (2HK): +T^2(<c_H, \sigma>, c_K) = \mu$$

$$(2KA): -R^2(<c_K, \mu>, c_A) = \mu,$$

$$(2AT): +R^2(<c_A, \mu>, c_T) = \mu$$

By substitutions, we have the following indirect functions:

(40) *Indirect rotations at interval 2*

$$(2HA): -R^2(<c_K, +T^2(<c_H, \sigma>, c_K)>, c_A) = \mu$$

$$(2SA): -R^2(<c_K, +T^2(<c_S, \sigma>, c_K)>, c_A) = \mu$$

$$(2KT): +R^2(<c_A, -R^2(<c_K, \mu>, c_A)>, c_T) = \mu$$

Substituting twice leads to the following:

(41) *Doubly indirect rotations at interval 2*

$$(2HT): +R^2(<c_A, -R^2(<c_K, +T^2(<c_H, \sigma>, c_K)>, c_A)>, c_T) = \mu$$

$$(2ST): +R^2(<c_A, -R^2(<c_K, +T^2(<c_S, \sigma>, c_K)>, c_A)>, c_T) = \mu$$

Let us turn to the third interval, at which the hip and the knee act as the same axis/center of turn/rotation again, in the same direction, as those seen at the second interval, while there is no more rotation around the ankle at this interval:

(42) *Direct turns/rotations at interval 3*

$$(3SHK): +T^3(\{<c_S, \sigma>, <c_H, \sigma>\}, c_K) = \mu$$

$$\Rightarrow (3SK): +T^3(<c_S, \sigma>, c_K) = \mu$$

$$\& \ (3HK): +T^3(<c_H, \sigma>, c_K) = \mu$$

$$(3KA): -R^3(<c_K, \mu>, c_A) = \mu$$

$$(3KT): -R^3(<c_K, \mu>, c_T) = \mu$$

By substitutions, we have the following indirect functions:

(43) *Indirect rotations at interval 3:*

$$(3HA): -R^3(<c_K, +T^3(<c_H, \sigma>, c_K)>, c_A) = \mu$$

$$(3SA): -R^3(<c_K, +T^3(<c_S, \sigma>, c_K)>, c_A) = \mu$$

$$(3HT): \text{Compatible with (1HT)}$$

$$(3ST): \text{Compatible with (1ST)}$$

Here we can see no (3AT), which holds at the second interval. This is due to the lack of the rotation around the ankle. There are, however, two cases of double application, (3HT) and (3ST), by way of (3KT).

6.3 Valuation at interval 1

Based on the causative relation denoted by rotation/turn functions, we attain the accessibility at the first interval as follows:

- (44) *Accessibility relation at interval 1*
- $$\{ \langle w_s, w_k \rangle, \langle w_h, w_k \rangle, \langle w_k, w_a \rangle, \langle w_k, w_t \rangle, \langle w_a, w_t \rangle, \langle w_s, w_a \rangle, \langle w_s, w_t \rangle, \langle w_h, w_a \rangle, \langle w_h, w_t \rangle \}$$

We value the truth-value of the existential movement formula in each world. Comments on the causative relations are omitted due to lack of space:

- (45) *Valuation in each world at interval 1*
- $\neg \exists x \in C [\text{MOVE}(x)]$ in w_s and w_h
 - $\exists x \in C [\text{MOVE}(x)]$ in w_k : from (1SK) and (1HK)
 - $\exists x \in C [\text{MOVE}(x)]$ in w_a : from (1KA), (1HA) and (1SA)
 - $\exists x \in C [\text{MOVE}(x)]$ in w_t : from (1AT), (1KT), (1HT), and (1ST)

Let us now turn to the valuation between the worlds. Based on the above formulae and comments on the causative relations, we value each modal formula, following the valuation conditions (the comments as given in (28a), (29a) and (30a) are omitted).

- (46) *Valuation between worlds at interval 1*
- $\Box [\exists x \in C [\text{MOVE}(x)]]$ in w_s, w_h, w_k and w_a
 - $\Diamond [\Box [\exists x \in C [\text{MOVE}(x)]]]$ in w_s, w_h , and w_k
 - $\Diamond [\Diamond [\Box [\exists x \in C [\text{MOVE}(x)]]]]$ in w_s and w_h

6.4 Valuation at interval 2

The same turn and two rotations that were observed at the first interval hold at the second interval as well, except that at the second, the two rotations have reversed their directions. Since the distribution of the causation relations on the support set remain unchanged over the first and second intervals, the accessibility at the second is the same as that at the first:

- (47) *Accessibility relation at interval 2: = (44)*

Since the polarity differences in rotation functions at the second interval do not reflect on the valuation of the formulae, we have the same array of formulae as those seen at the first interval:

- (48) *Valuation in each world at interval 2:* = (45)
 (49) *Valuation between worlds at interval 2:* = (46)

6.5 Valuation at interval 3

We now go on to the third interval. It is important to note that one rotation, $+R^2(<c_A, \mu>) = \mu$, the one holding at the second interval, is missing here. There is no more rotation around the ankle at this interval. The accessibility, different from that common between the first and second intervals, is as follows:

- (50) *Accessibility relation at interval 3*
 $\{<w_s, w_k>, <w_h, w_k>, <w_k, w_a>, <w_k, w_t>, <w_s, w_a>, <w_s, w_t>, <w_h, w_a>, <w_h, w_t>\}$

Accordingly, the rotation/turn functions calculated from the motions at the third interval are reduced to (3SHK) = (3SK) & (3HK), (3KA), (3KT), (3HA), (3SA), as was shown in Section 6.2. On these we value formulae as follows:

- (51) *Valuation in each world at interval 3*
 a. $\neg\exists x \in C [\text{MOVE}(x)]$ in w_s and w_h
 b. $\exists x \in C [\text{MOVE}(x)]$ in w_k, w_a and w_t

Furthermore we can attain the following modal formulae:

- (52) *Valuation between worlds at interval 3*
 a. $\Box[\exists x \in C [\text{MOVE}(x)]]$ in w_s, w_h and w_k
 b. $\Diamond[\Box[\exists x \in C [\text{MOVE}(x)]]]$ in w_s and w_h

At this interval, there is found no formula with a triple modal operator, since there is no rotation/turn around the right angle.

6.6 Propagation structure of action

The set of formulae valued at each interval shows the propagation of causation of rotation/turn. What should we do with these sets to enable them to partially contribute to the semantic representation of action verbs? For this, let us consider a way to convert the above formulae into a simpler representation. Consider the following algorithm:

(53) *Simplification algorithm*

- (i) Simplify modal operators $\Diamond$, $\Box$, and the existential formula as P, N and m , respectively.
- (ii) Sort all the formulae by the world to which they belong.
- (iii) Select the longest formula for each world.
- (iv) Enclose all the parts of that formula with brackets labeled with the intervals to which the formulae belong in that world.

Applying this algorithm to the above formulae leads to more transparent representations:

(54) *Modal structure for movement propagation*

- a. at w_K : $[_{1/2} P[_{1/2/3} N[_{1/2/3} m]]]$
- b. at w_A : $[_{1/2} N[_{1/2/3} m]]]$
- c. at w_T : $[_{1/2/3} m]$
- d. at w_S and w_H : $[_{1/2} P[_{1/2/3} P[_{1/2/3} Nm]]]$

The different indices to the same bracketing show the recurrence of the same propagation over the relevant intervals. A bracket may contain another bracket with the same index, in which case they are propagation varieties, that is, the formula containing its embedded one and solely the embedded one, at the same interval. This structure shows the coverage and remoteness of propagation from each world (each vertex with an endpoint/joint name) at consecutive intervals. In it, N and P denote the total and the partial exhaustiveness of movement propagation on the causative path from the world (vertex), respectively, and the dimensionality of the operators denotes the extendibility of the propagation. If we intend to avoid the extreme specificity of subcategorizing action tokens of the same type into more than one type, we could reduce the number of valuations by limiting the accessibility distance (the dimensionality of formulae) from which modal formulae are valued. This constitutes an argument for our proposal.

7 Microstructure of causation of action

7.1 Causation in rotation/turn revisited

From a cognitive point of view, it is interesting that the existential formula used in our analysis is a formula with a predicate MOVE. This is compatible with the semantic primitive GO in terms of Jackendoff's binary causative function. While an action is proceeding on its performer, in its microstructure a vertex/pair of vertices as the center/axis of rotation/turn causes another vertex to move from one locus to

another, in which case the causer acts as Agent and the mover as Theme in terms of thematic relations. Assigning thematic roles to the arguments of the CAUSE and GO functions, we can see that THEME, a moving object in the metaphorical sense implied in the thematic theory by Gruber (1976), falls on the first argument of GO in the embedded event.

(55) *Causation of movement*

$$\begin{array}{l}
 \text{[}_{\text{Event}} \text{ CAUSE} \\
 \quad (\text{[}_{\text{Thing}} \text{ a content vertex}], \\
 \qquad \qquad \qquad \text{(Agent)} \\
 \\
 \text{[}_{\text{Event}} \text{ GO ([}_{\text{Thing}} \text{ another content vertex}], \\
 \qquad \qquad \qquad \text{(Theme)} \\
 \\
 \text{[}_{\text{Path}} \text{ FROM a place point TO another place point}]))] \\
 \qquad \qquad \qquad \text{(Source)} \qquad \qquad \text{(Goal)}
 \end{array}$$

The event in the second argument of CAUSE denotes a translation of a content vertex from one place point to another, where GO acts as a predicate with a “Thing” as its first argument and a “Path”, as its second. It is interesting to note that the center vertex or (each vertex of) the axis pair in the first argument of the causative is assigned “Agent”. The reason for this is that the center/axis plays a decisive role in propagating the transitivity of causation on the support structure. The second argument of the causative is an event, in which a “Thing” GOES along a path consisting of a SOURCE and a GOAL. This path begins with a location and ends with another location.

Suppose that the first vertex (in an agentive role), as a center/(part of) axis, causes the second (in a thematic role) to GO, and if the second, as an independent center/(part of) axis, causes the third to GO, note that, on the latter of these edges, the second vertex bears a new role of agent. Thus, on a sequence of connected causation pairs of vertices, such as $\{ \langle a, b \rangle, \langle b, c \rangle, \langle c, d \rangle, \dots \}$, where a causes b , which, in turn, causes c , ..., a vertex – bearing Theme as a going (moving) object as the second member of an pair – switches its role to Agent as a causer of a rotation/turn in the next pair.

7.2 Linguistic evidence for causation of action

Using the framework of conceptual structure discussed in Section 5, we are able to attribute the role of transmitting the “kinetic force” of action to CAUSE, and assume that the Theme carries this force and passes it to another vertex and so on. If this successive transmission of force from one content vertex to another on the skeleton finally reaches the point of operation for the action, the (partial) action within the skeleton (body) of its performer completes the denotation of the

intransitive or transitive action verb. In a way, the process can be viewed as a successive sequence of causative relations of action at the micro level, starting from the supporting vertex to the point of operation. The force that has just been passed on up to the point of action will be acting on the target of the relevant action, that is, the denotation of the grammatical object.⁹ This process is represented as a causative relation of the conceptual structure, at the normal granularity of description. It is important to remember that in Section 5, we discussed that the inalienable instrument expressions, *with-NP* phrases in English and *NP-de* phrases in Japanese, co-occur with verbs of bodily action, showing the point of the relevant actions. Based on this fact, we can conclude that the modal structure for movement propagation, which was attained in the preceding section, represents the mode of transmission of energy from the supporting point in the performer up to the point of action, in terms of thematic relations.

We still have a question, however: Where should we put the propagation description in the conceptual structure? We will consider this question in the following section.

7.3 Global path vs. local path

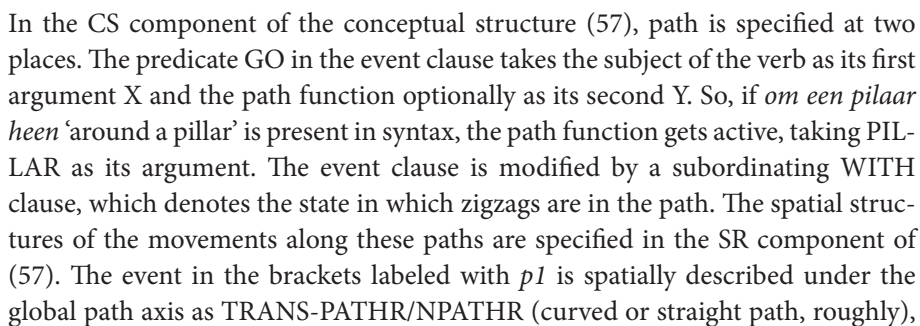
The studies on the semantic representation of the curvature inherent in the figures and the motions of objects have begun to attract researchers' attention. Among them, van der Zee is eminent. Conceptual structure (CS) is assumed to encode our knowledge of the categorical status of auditory, visual, olfactory and haptic information in a neutral or a-modal format. Spatial structure (SR) is assumed to

9. Croft (1991: 159–172) proposes, based on Lakoff's analysis, that each individual event is the semantic entity corresponding to an ordinary simple verb. He argued that the following causal sentence is decomposed into the "atomic events" as:

(ii) *John broke the boulder with a hammer.*

John → hand → hammer → boulder → (boulder) → (boulder)
 Vol Grasp Contact Change State Result State

The atomic events are: physical, volitional, affective and inductive causations. In the sequence below the sentence, each arrow belongs to a causation event in which the endpoint-entities connected via the arrow are the participants of the same causation event. Of the five causation events above, Grasp, Contact and Change State causations are physical causations and processes, where there is a transmission of force, which is one criterion for an atomic event, Croft argues. In Section 7.2 in the text, we have seen that the force that has been passed on to the point of action will be acting on the target of the relevant action. The point of action corresponds to "hand" in (ii). It is very interesting to see that the volitional causation between *John* and *hand* is the very path that should be analyzed by our propagation structure, while the connected event-parts to its right correspond to the target of the action.



which specifies the whole course the subject takes by ‘zigzagging’. The subordinating clause labeled with *p2* is spatially described under the local path axis as TRANS-PATHR (curved path, roughly), which denotes how the subject moves by ‘zigzagging’. Thus, note that the local path part, in the SR, which in turn gives a 3D description, specifies the detailed description of the change in the figure of the verb’s denotation, not the change in its translation.

7.4 Right place for propagation structure

We discussed the instrument phrases that modify Japanese verbs of bodily action in Section 5, where we pointed out that the *NP-de* instrument phrases express the point of operation for the relevant bodily actions. We also saw that in the conceptual structure (33), two types of causative function, CS, contain the point of operation HER RIGHT FOOT in the event clause and in the BY-subordinating clause, which modifies the event clause. Remember that in Sections 7.1 and 7.2, our propagation structure for kicking describes the sequence of causative relations on the skeleton up to the point of operation. These facts lead us to conclude that the entry for the local path, in SR, that describes the movement of BY-clause (instrument clause) modifying the event in CS might be a candidate position for the propagation structure.¹⁰

8 Concluding remarks

To capture the ordering and kinds of the parts alternating in the movements involved in each action, we have proposed a propagation structure as the modal logical representation for bodily actions. We have also shown that it is derivable from our framework that the MOVE predicate of the formula is a unit force transmittable over the network of transitive causative movements between all the

10. Van der Zee (2000) proposes as the conceptual structure for the transitive verbs *turn* and *rotate* the event where the grammatical subject bears the Agent, which CAUSES the Theme (born by the object), to GO AROUND AXIS OF the Theme. Then, one might ask if this analysis could apply to the description of the microstructure of causation inherent to the bodily action. Its application would, however, be difficult in two points. First, even if the application should be successful, the same causative analysis would have been used twice, at the macro-level of meaning description and at the micro-level of connection of causation relations, which might lead to a tautology. Second, that micro-level use would require a transitive set of relations on which causation is successively definable, which is missing in the SR attached to this CS. An independent mechanism of a skeleton-like structure would be needed, but such a representation would be inappropriate here as a normal macro-level of meaning description.

endpoints and joints forming the skeleton of the performer of an action. This might possibly suggest that human cognition of actions is dependent on the transitivity of causation of movement, thus constituting a base for understanding the meaning of action verbs. For future research, we intend to check the validity of our framework by developing a substantial inventory of propagation structures for verbs of bodily actions.¹¹

References

- Aitchison, Jean. 1996. *The seeds of speech: Language origin and evolution*. Cambridge: CUP.
 Ann, Jean. 1993. Physiological constraints on two-finger handshapes. *Chicago Linguistics Society (CLS)* 28(1). 1–14.
 Badler, Norman I., Cary B. Phillips & Bonnie L. Webber. 1992. *Simulating humans: Computer graphics, animation, and control*. Oxford: OUP.
 Biederman, Irving. 1987. Recognition by components: A theory of human image understanding. *Psychological Review* 94(2). 115–147.

11. For this purpose, a test was carried out by describing a hand sign (Nishina 2003). In American Sign Language (ASL), ‘to ask’ (= to inquire or to question) has more than 6 variant forms (Stokoe, Casterline & Croneberg 1976). One of the typical expressions for ‘I ask you’ is signed in the following manner: As the hand moves out smoothly to the side, its extended index finger is crooked. Although we have left unanalyzed the intricate linguistic structure the sentence has, we have merely applied our tool for description to this sign as a typical sample of action. The period during which the hand shape is being made can be divided into three intervals:

(iii) *Handshape for ‘I ask you’*

- Interval 1: While the elbow is being opened, the proximal interphalangeal joint (PIP) is being lightly closed
 Interval 2: With the two movements in the first interval maintained, the distal interphalangeal joint (DIP) is being closed.
 Interval 3: With all the three movements still ongoing, the distal metacarpophalangeal joint (MCP) is being closed.

We base our analysis on the hand structure proposed by Ann (1993) and Köpf-Maier (1997). The support set for the relevant part of the skeleton for this action is $\{<c_s, c_e>, <c_e, c_c>, <c_c, c_m>, <c_m, c_p>, <c_p, c_d>, <c_d, c_t>\}$, where, aside from the bold-face capitalized joints, S stands for shoulder, E elbow, C carpometacarpal joint (CMCJ), and T tip of finger. Skipping over to the conclusion, the propagation structure for the sign of ‘I ask you’ is:

(iv) *Propagation structure for ‘I ask you’*

- a. at w_E : $[_3 P[_2 P[_1 P[_1 N[_{1/2/3} m]]]]]$
 b. at w_M : $[_{1/2} P[_{1/2} P[_{1/2} N[_{1/2/3} m]]]]]$
 c. at w_P : $[_2 P[_1 N[_{1/2/3} m]]]$
 d. at w_D : $[_2 N[_{1/2/3} m]]]$

Remember that rotation/turn function returns one vertex’s state after the relevant rotation/turn over an interval in terms of moving or staying as output.

- Croft, William. 1991. *Syntactic categories and grammatical relations*. Chicago IL: Chicago University Press.
- Davidson, Donald. 1967a. The logical form of action sentences. In Davidson (ed.), 105–122.
- Davidson, Donald. 1967b. Causal Relations. In Davidson (ed.), 149–162.
- Davidson, Donald (ed.). 1967. *Essays on actions and events*. Oxford: OUP.
- Gamut, L. T. F. 1991. *Intensional logic and logical grammar* [Logic, Language and Meaning 2]. Chicago, IL: The University of Chicago Press.
- Gruber, Geoffrey. 1976. *Lexical structures in syntax and semantics*. Amsterdam: North Holland.
- Hughes, George E. & Max J. Cresswell. 1966. *A new introduction to modal logic*. Oxford: OUP.
- Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge, MA: The MIT Press.
- Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: The MIT Press.
- Köpf-Maier, Petra (ed.). 1997. *Wolf-Heidegger's atlas of human anatomy I: Systemic anatomy, body walls, upper and lower limbs*. Basel: S. Karger AG.
- Lakoff, George. 1970. *Irregularity in syntax*. New York, NY: Holt, Rinehart and Winston.
- Marr, David. 1982. *Vision*. New York, NY: W. H. Freeman and Company.
- McCawley, James D. 1970. English as a VSO language. *Language* 46. 286–299.
- Nishina, Hiroyuki. 1999. Extracting semantic features from spatiotemporal information on action using modal logic. Paper delivered at the *Colloquium on Dynamic Logic*. Czech Academy of Science, Prague.
- Nishina, Hiroyuki. 2003. How humans identify actions? One proposal for extracting semantic contents from the spatiotemporal information of action. In Eva Hajičová, Anna Kotěšovcová & Jaromír Mírovský (eds.), *International Congress of Linguists XVII*. CD-ROM. Matfyzpress, MFF UK.
- Parsons, Terence. 1990. *Events in the semantics of English*. Cambridge, MA: The MIT Press.
- Postal, Paul M. 1972. On the surface verb “REMIND”. In Charles J. Fillmore & D. Terrence Langendoen (eds.), *Studies in linguistic semantics*, 180–270. New York, NY: Holt Rinehart and Winston.
- Stokoe, William C., Dorothy C. Casterline & Carl G. Croneberg. 1976. *A dictionary of American Sign Language on linguistic principles*. (New edition.) Burtonsville, MD: Linstok Press.
- Talmy, Leonard. 1976. Semantic causative types. In Masayoshi Shibatani (ed.), *The grammar of causative constructions* [Syntax and Semantics 6], 43–116. New York, NY: Academic Press.
- Zatsiorsky, Vladimir M. 1997. *Kinematics of human motion*. Champaign, IL: Human Kinetics Publishers.
- Zee, Emile van der. 2000. Why we can talk about bulging barrels and spinning spirals: Curvature representation in the lexical interface. In Emile van der Zee & Urpo Nikanne (eds.), *Cognitive interfaces: Constraints on linking cognitive information*, 143–182. Oxford: OUP.

Do we access object manipulability while we categorize?

Evidence from reaction time studies

Anna M. Borghi, Claudia Bonfiglioli, Paola Ricciardelli, Sandro Rubichi and Roberto Nicoletti

In two experiments we investigate whether different decision tasks were influenced by object manipulability. In Experiment 1, participants had to categorize objects (drawings or words) into artifacts or natural kinds. In Experiment 2, manipulability was made relevant to the task; participants were to decide whether or not objects could be picked up and put inside a backpack. Overall natural objects received faster responses than artifacts. In Experiment 2 an effect of manipulability was also found, but only with natural kinds, probably due to the fact that they convey information associated both with action (“how”) and function (“what for”). The same pattern of results found with drawings and words suggests that words also activate motor information on how to grasp objects.

1 Introduction

How we manipulate and interact with objects has been a long-standing research issue that is still far from being settled. Not only do we need information on how to grasp objects but also on how to use them. In other words, we learn *how* to hold a knife and *what* it is for. It is still a matter of debate whether these two different notions are represented separately or are embodied together in object concepts (for reviews see Borghi 2005; Carlson & van der Zee 2005; Coventry & Garrod 2004).

Studies on conceptual organization have shown that categories differ in the weight they put on different types of knowledge: The recognition of artifacts depends more on functional features (e.g. cup – used to drink) than the recognition of natural objects (e.g. dog) (Warrington & Shallice 1984). However, while the notion of “visual” features is fairly clear (i.e. shape, color, etc.), the notion of

“functional” feature is underdefined, as it may include either action or function information or both (Borghi 2004). Recent neuropsychological studies suggest that action (“how”) and function (“what for”) information, i.e. information on how to manipulate and interact with an object and information on how to use it properly, might differ. Buxbaum, Veramonti and Schwartz (2000) report cases of apraxic patients with impaired manipulation knowledge but intact function knowledge. These cases double dissociate from a case of an agnosic patient who was able to determine how to manipulate certain objects, but was not able to define their function or the context in which they would be utilized (Sirigu, Duhamel & Poncet 1991). Brain activation results showed that the response of the left ventral premotor cortex and the left middle temporal gyrus was stronger for manipulable than for non-manipulable objects, whereas no regions of the cortex were more activated by function relative to action judgements about artifacts (Kellenbach, Brett & Patterson 2003). These results indicate that the brain responds preferentially to how we interact with objects, rather than to what they are used for, and suggest that action and function knowledge do not overlap.

Other functional neuroimaging studies have shown that action information is automatically activated by viewing objects and pictures, and that the same areas are involved when forming motor imagery and when activating information on tools. In a PET study Martin et al. (1996) found that naming tools, compared to naming animals, differentially activated the left middle temporal gyrus – an area nearly identical to the area activated by action generation tasks – and the left premotor cortex, an area generally activated when participants imagine themselves grasping objects with their dominant hand. Grafton et al. (1997) found that retrieval of actions associated with tools produced activation in the left premotor cortex. More recently, Chao and Martin (2000) carried out a fMRI study showing that the left premotor cortex responds selectively to photographs of tools but not to other objects such as animals, faces, and houses. Consider that this different activation pattern cannot be due to the distinction between artifacts and natural objects, nor can it be due to the fact that only tools are characterized by functional information – houses, too, are artifacts and have a specific function. Alternatively, the selective response for tools might be due to the fact that, differently from the other considered categories, tools are manipulable objects. In a PET study Gerlach, Law and Paulson (2002) showed that in a categorization task the left ventral premotor cortex was activated with both artifact and natural manipulable objects – more specifically, it was activated during categorization of fruit/vegetables and clothing, relative to animals and non-manipulable artifacts.

Behavioral evidence on manipulability is sparser. Bub, Masson and Bukach (2003) showed photographs of hand gestures followed by the photographs of objects associated with the gestures (for example, pinch: match and needle).

Participants had to depress a key until they started to gesture in response to the object color. No advantage of pairs in which the gesture associated with the color and the gesture typically associated to the object corresponded was found. The performance was better when the color and the gesture associated with the object corresponded only when a cue indicated whether to gesture to the color or to the object. This suggests that passively viewing an object does not necessarily imply the activation of gestural knowledge; this knowledge is recruited only when competing sources of gestural representations are activated.

Behavioral studies with compatibility paradigms, i.e. paradigms implying some kind of correspondence between stimuli and responses, indicate that the vision of objects elicits motor information, related in particular to reaching and grasping movements (Tucker & Ellis 1998, 2001, 2004). Ellis and Tucker (2000) and Tucker and Ellis (2001) asked participants either to respond to a high or a low tone or to categorize objects of different size located behind a screen as natural or artifact, using either a power grip or a precision grip. A compatibility effect between the kind of grasp and the object's size was found. A similar compatibility effect was found between the direction of the wrist rotation and the kind of grasp required by the object: For example, bottles facilitated responses with a clockwise wrist rotation, toothbrushes responses with a counter-clockwise wrist rotation. These compatibility effects were found regardless of whether the objects were natural or artifacts – they only depended on the way objects could be reached and grasped.

The studies reported leave some questions open. First, they demonstrate an automatic activation of information on grasping only when the motor system is pre-activated – for example, the response consists of different ways to grasp a device rather than of a simple key pressure. To our knowledge the only exception is the study by Saffran, Kossler and Keener (2003) who in a word association task found a higher proportion of verbs produced with pictures than with words and with manipulable than with non-manipulable objects. However, to our knowledge no study has investigated the effect of manipulability on the speed of processing of different concepts. Finding an effect of manipulability on response times would be important in order to verify whether information on how to interact with objects is part of our conceptual representation and directly affects behavior.

Second, these studies do not allow to clearly distinguish what kind of motor information is automatically activated – whether information on how to interact with objects or on how to use them.

Third, they leave open the question of whether or not compatibility results as those described imply access to conceptual knowledge. According to an influential account, two different routes to action exist: a direct vision-to-action route, mediated by on-line dorsal system processes, and a mediated vision-to-semantics-to-action ventral route (Castiello & Jeannerod 1991; Fagg & Arbib 1998; Rumati &

Humphreys 1998). From this account it follows the prediction that pictures activate manipulation information more directly than words (Phillips et al. 2002). Consider that the fact that pictures activate motor information more directly than words does not necessarily imply that words do not activate motor information. Recent behavioral evidence suggests that words also activate motor information (Barsalou 1999; Glenberg & Kaschak 2002; Klatzky et al. 1989; Pecher & Zwaan 2005; Pulvermüller 2003). Compatibility effects between object size and kind of grasp were found also when names instead of images of objects are presented (Tucker & Ellis 2004). Borghi, Glenberg and Kaschak (2004) found with a part verification task that responding by pressing a button in a direction compatible with the part location (e.g. responding upward to verify that the name “head” designates a part of a horse) was faster than responding in a direction incompatible with the part location. There is also evidence that the semantic meaning of words affects the grasping and reaching kinematics (Gentilucci 2003). The meaning of words such as *large* or *small* had an effect on the grip aperture in the initial grasp kinematics (Gentilucci & Gangitano 1998; Glover & Dixon 2002).

Overall, these studies prove that the semantic meaning of words affects the motor system. However, the tasks used always required either a motor preparation or a specific activation of the motor system – for example participants had to prepare a specific motor response or had to perform a reaching or a grasping behavior to answer. Let us make an example of what we mean by “motor preparation” preceding the experiment. Klatzky et al. (1989) assessed whether priming a hand shape facilitated judging whether actions performed with objects were sensible. They found compatibility effects between different hand postures and action sentences – for example, the sentence “aim with a dart” was processed faster when preceded by a pinch posture than by other postures. Crucially, in Klatzky et al.’s study participants were submitted a phase of “motor preparation”, i.e. they learned to associate the prime, which could be presented either visually or verbally, to a specific gesture they had to perform.

Whereas some experiments are preceded by a motor preparation, others imply a “specific activation of the motor system”, i.e. the responses required involve the same kind of motor actions the stimulus is supposed to elicit. As an example of what we intend with “specific activation” of the motor system consider the study by Tucker and Ellis (2004) who found a compatibility effect between the size of object referred to by names and the kind of motor response required by the task. In their experiment object size was not relevant to the task, which consisted of deciding whether the objects were artifacts or natural kinds, but the motor action used to provide the response (a kind of grasping response that mimicked either grasping objects with a precision or with a power grip) is the same we use while grasping

objects. Consider, instead, classic RTs experiments in which the motor response consists simply of pressing a button on the computer keyboard. If we provide evidence of activation of motor information also in the last case, this would show that object concepts directly incorporate information on how to manipulate them, and that this information is immediately accessible to be used for acting.

The first aim of the present study is to verify whether object manipulability influences response times in a categorization task that does not require pre-activation of the motor system. As discussed later, finding an effect of manipulability in a categorization task would be very relevant to embodied theories of cognition, according to which concepts are grounded in sensorimotor experiences with their referents. The second aim of this study is to assess whether the activation of motor-related information differs depending on how the visual input is presented, i.e. as a word or as a drawing. Finding activation of motor information also with words would suggest that it does not depend only on the direct vision-to-action route, but that access to conceptual knowledge contributes in explaining it. In fact, similar results obtained in the same task with pictures and words, demonstrating that motor information is activated by both pictures and words, cannot be explained solely on the basis of the activation of a direct vision-to-action route. Rather, they would demonstrate that conceptual knowledge is accessed, and that the ventral system is probably involved. Further, we aim to disambiguate the effects of action and function. For this reason we used manipulable and non-manipulable artifacts and natural objects. Finally, in Experiment 1 we asked participants to categorize objects into natural or artifacts, in Experiment 2 we made manipulability relevant to the task by asking participants whether the objects could be put inside a backpack or not. With artifacts we predict an interference between functional information and the motor program necessary to accomplish the task in both experiments. We also predict that the activation of action information leads to an interference for manipulable objects in Experiment 1 and to a facilitation in Experiment 2, when manipulability was relevant to the task.

2 Experiment 1

If object concepts automatically activate action (“how”) information, this information could affect the motor program necessary to accomplish the task leading to process differently manipulable and non-manipulable objects, and to respond differently to manipulable objects with the dominant hand than with the other one. Indeed, prehension movements are more associated with the more skilled dominant hand (see also Handy et al. 2003).

Concepts referring to artifacts surely contain more functional attributes than natural objects, therefore it can be assumed that the activation of function interferes more with artifacts than with natural kinds.

2.1 Method

Participants and design

Twenty-two right-handed students of the University of Bologna took part in the experiment. They either volunteered their participation or received course credit for their time.

Materials

Sixty-four items controlled for familiarity, number of syllables, word frequency, visual complexity were selected from the Lotto, Dall'Acqua and Job (2001) database of Italian words and black-and-white drawings of common objects. A pre-test was performed in order to select, from the original set of 64 items, clearly manipulable and clearly non-manipulable items. 12 participants were presented with one word at a time and had to decide whether the object it referred to was manipulable or not. In this way 40 items were selected: 10 manipulable artifacts (knife), 10 manipulable natural objects (carrot), 10 non-manipulable artifacts (boat) and 10 non-manipulable natural objects (palm). While selecting the items we made sure that within each group of items (manipulable natural kinds, non-manipulable natural kinds, manipulable artifacts, non-manipulable artifacts) there were members of different categories. Within each group there were items of at least 3 categories, and there were at least 2 members of each category. For example, non-manipulable artifacts included 3 members of the category "vehicle", 5 of the category "building", and 2 of the category "furniture". Even though the items in each condition were not many, this does not prevent the generalization of the results, as the selected items were quite typical everyday concepts of different categories, among the most frequently studied in the literature. Each of the selected items was presented 4 times, in 2 different modalities: as a word and as a drawing. The pictorial stimuli were presented centrally on the screen. The left-right orientation of the pictorial stimuli with a protruding part (e.g. hammer) on the screen was balanced.

Procedure

Participants sat in front of a computer monitor. Each trial began with a fixation point (+) displayed for 1000 ms. Immediately after the fixation point disappeared, depending on the experimental block either a word or a picture appeared, remaining on the screen until the participant's response. Participants had to decide whether the presented word or picture referred to an artifact (e.g. knife) or to a

natural object (e.g. palm). Half of them pressed the right button in response to artifacts and the left button in response to natural objects. The other half had the opposite mapping. Participants received feedback on reaction time (RT) after pressing the right key (the reaction times were displayed), as well as after pressing the wrong key ("ERROR") or after taking 3000 ms to respond ("You have not answered"). The next trial began after the feedback disappeared. The experiment consisted of 2 practice blocks of 24 practice trials each and of 2 experimental blocks, one for words and one for drawings. The presentation order of the experimental blocks varied for each participant. In each experimental block the 40 critical trials were presented twice, in a different random order for each participant. After each block, participants could take a brief break. Overall the experiment lasted about 20 minutes.

2.2 Data analysis and results

RTs more than 2 standard deviations from each participant's means were excluded from the analysis. Correct RTs were entered into two mixed ANOVAs, one on the participants' data and one on the materials. In the ANOVA with participants as random factor Kind of Concept (artifact, natural kind), Manipulability (manipulable, non-manipulable objects) and Presentation Mode (drawing, word) were manipulated within participants, while Mapping (artifact-left hand/natural-right hand vs. the opposite) was manipulated between-participants. In the ANOVA with materials as random factor, Kind of concept and Manipulability were between materials factors, whereas Presentation Mode and Hand of Response (right, left) were within materials factors.

The two main effects of Kind of Concept and Presentation Mode were significant both in the analysis on participants (indicated by F^1) and on materials (indicated by F^2) (respectively $F^1(1, 20) = 17.39$, $MSe = 1283.4$, $p < .001$; $F^2(1, 36) = 6.19$, $MSe = 2439.5$, $p < .05$; $F^1(1, 20) = 4.17$, $MSe = 2063.8$, $p < .0001$; $F^2(1, 36) = 625.45$, $MSe = 765.7$, $p < .0001$). Responses to natural kinds were, on average, 23 ms faster than responses to artifacts, and drawings were elaborated 109 ms faster than words. The interaction between Kind of Concept and Presentation Mode, $F^1(1, 20) = 5.06$, $MSe = 513.2$, $p < .05$, was significant due to the fact that with artifacts pictures were elaborated 30 ms faster than words, whereas with natural objects the difference between pictures and words was reduced to 15 ms.

The 3-way interaction between Kind of Concept, Manipulability and Presentation Mode was also significant, $F^1(1, 20) = 5.18$, $MSe = 476.1$, $p < .05$; $F^2(1, 36) = 9.71$, $MSe = 765.7$, $p < .01$. As can be seen in Figure 1, pictures were processed always faster than words and natural kinds always produced faster responses than artifacts. The only exception were natural manipulable pictures, which were slower

than natural non-manipulable pictures (Newman-Keuls, $F^1 p = .06$; $F^2 p < .05$), and did not differ from pictures of both manipulable and non-manipulable artifacts.

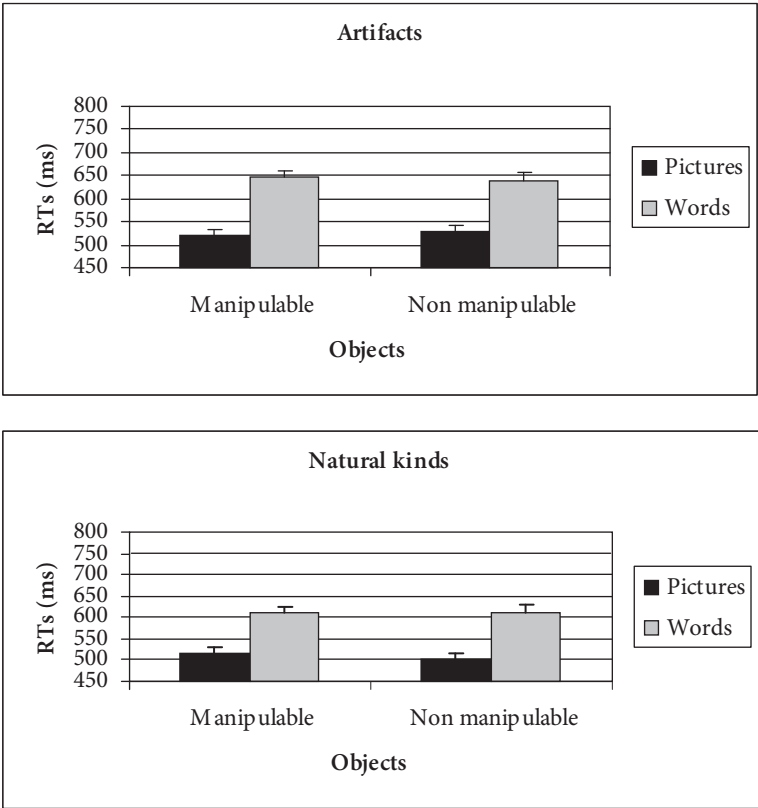


Figure 1. Experiment 1: 3-way interaction between Manipulability, Kind of Concept and Presentation Mode

In the analysis on materials Kind of Concept and Hand of Response interacted, $F^2 (1, 36) = 12.79, p < .01$, due to the fact that responses to natural kinds with the right hand were the fastest (Newman-Keuls, $p < .001$). However, a closer look at the data showed that this was mainly due to non-manipulable objects.

2.3 Discussion

The results suggest that both action and function information are activated. The finding that pictures of manipulable natural kinds were responded to slower than pictures of non-manipulable ones may be due to the fact that the activated action

information causes an interference with the motor program necessary to accomplish the task. Even if the motor program was a simple key press and the required motor response was not the one we typically use for interacting with objects, object processing activated the motor system. This interpretation is supported by the similarity between our results and those by Gerlach, Law and Paulson (2002). They found that specific neural areas were activated with manipulable objects; crucially, in response times this activation corresponded to a disadvantage in processing manipulable fruit and vegetables compared with non-manipulable animals. The advantage of natural kinds over artifacts, given that it was not due to word frequency, familiarity, and visual complexity, may depend on an interference between the functional information artifacts elicited and the motor program necessary to accomplish the task. This interference effect was weaker with pictures than with words. This is compatible with studies showing that pictures activate affordances more directly than words. However, another possible explanation of the result may be advanced. Compared to artifacts, natural kinds have a higher degree of within category similarity. In the literature, it has been proposed that tasks requiring separate responses to individual category members, such as naming, are negatively affected by perceptual similarity, whereas tasks requiring common responses to different members, such as categorization, are positively affected by perceptual similarity (Humphreys, Riddoch & Forde 2002). According to this explanation, natural kinds, being perceptually more similar than artifacts, should be processed faster than artifacts in a categorization task.

3 Experiment 2

In Experiment 2 participants were asked to categorize items based on whether it was possible or not to pick them up and put them inside a backpack. This task aimed to increase the relevance of manipulability, but again without activating the motor system for response production. It should be noted that in this way only the role of manipulability, but not of function, was made more relevant. Accordingly, we predict no interference but a facilitation of manipulable objects over non-manipulable ones. Moreover, given that with natural objects action and function information correspond, responses to manipulable natural kinds should be faster than responses to manipulable artifacts.

As to function information, there should be again an interference of function information occurring with artifacts. If the advantage of natural kinds over artifacts is replicated, the explanation according to which the difference depends on within category similarity can be ruled out.

3.1 Method

Participants

Twenty right-handed students of the University of Bologna volunteered their participation.

Materials

The materials were the same as those in Experiment 1.

Procedure

The same procedure was used as in Experiment 1, but now participants were shown an open backpack and were asked to press a different key to decide whether the object could be picked up and put inside the backpack (e.g. knife) or not (e.g. palm).

3.2 Data analysis and results

As in the previous experiment, RTs more than 2 standard deviations from each participant's means were excluded from the analysis. Correct RTs were entered into two mixed ANOVAs, one on the participants' data and one on the materials, as in Experiment 1. In F^1 Kind of Concept, Manipulability and Presentation Mode were manipulated within participants, whereas Mapping (manipulable-left hand response key/non-manipulable-right hand response key vs. the opposite) was manipulated between-participants. In F^2 , Kind of Concept and Manipulability were between materials factors, whereas Presentation Mode and Hand of Response were within materials factors.

Natural objects were responded 11 ms faster than artifacts, $F^1(1, 18) = 5.54$, $MSe = 843.6$, $p < .05$, and drawings were processed 112 ms faster than words, $F^1(1, 18) = 77.5$, $MSe = 6520.1$, $p < .0001$; $F^2(1, 36) = 454.82$, $MSe = 1115.4$, $p < .0001$. However, the interaction between Kind of Concept and Presentation Mode, $F^1(1, 18) = 5.68$, $MSe = 520$, $p < .05$, showed that the difference between artifacts and natural kinds was mainly due to the fact that with words artifacts were processed 19 ms slower than natural kinds, whereas with pictures RTs did not differ with artifacts and natural kinds (respectively ms 585, ms 583).

Crucially, Kind of Concept and Manipulability interacted due to the fact that manipulable natural kinds were the fastest to be processed, $F^1(1, 18) = 6.51$, $MSe = 1150.3$, $p < .05$, $F^2(1, 36) = 3.28$, $MSe = 1115.4$, $p = .08$.

The 3-way interaction between Kind of Concept, Manipulability, and Presentation Mode, $F^1(1, 18) = 12.63$, $MSe = 531$, $p < .01$; $F^2(1, 36) = 7.33$, $MSe = 1115.4$, $p < .01$,

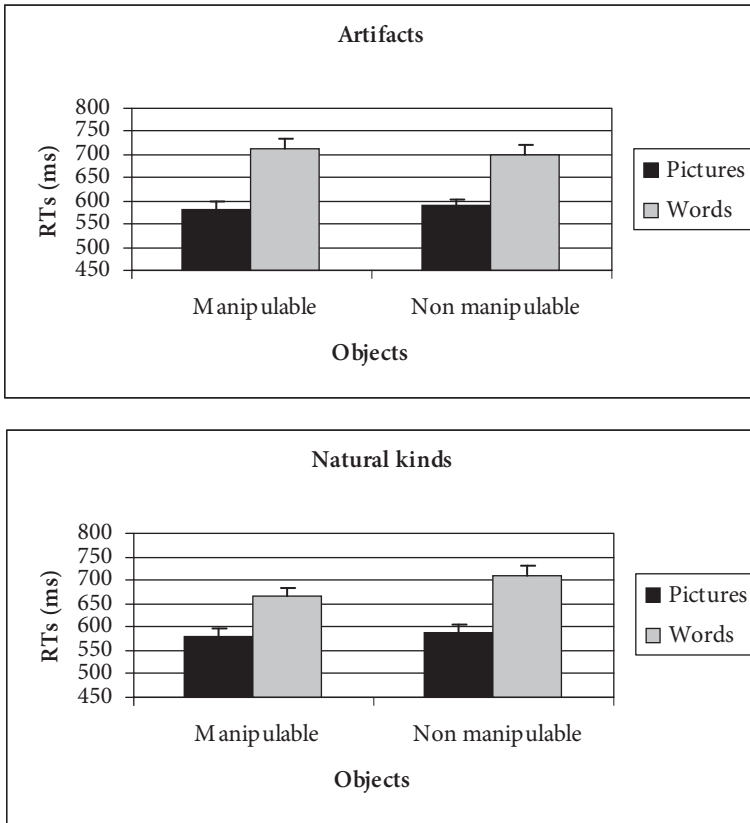


Figure 2. Experiment 2: 3-way interaction between Manipulability, Kind of Concept and Presentation Mode

showed that manipulable natural kinds words were the fastest to be processed among words (Newman-Keuls, $p < .01$), even though pictures were always processed faster than words ($p < .001$) (see Figure 2). Interestingly, pictures of manipulable artifacts were processed faster than pictures of non-manipulable ones, and responses to pictures of manipulable natural objects were faster than those to pictures of non-manipulable ones. However, these differences did not reach significance in the pairwise comparison.

In the analysis on materials the 3-way interaction between Kind of Concept, Presentation Mode and Hand of Response, $F^2(1, 36) = 6.65$, $MSe = 494.6$, $p < .014$ was significant (see Figure 3). Post-hoc Newman-Keuls showed that right hand responses to manipulable natural kinds were significantly faster than all other responses ($p < .02$) but to left hand responses to manipulable natural kinds; the latter were faster than non-manipulable right hand responses ($p < .01$) and faster than

responses to non-manipulable artifacts, but the differences did not reach significance ($p = .07$; $p = .09$). Thus with natural kinds, but not with artifacts, there was a clear advantage of manipulable objects, particularly when they were responded to with the right hand.

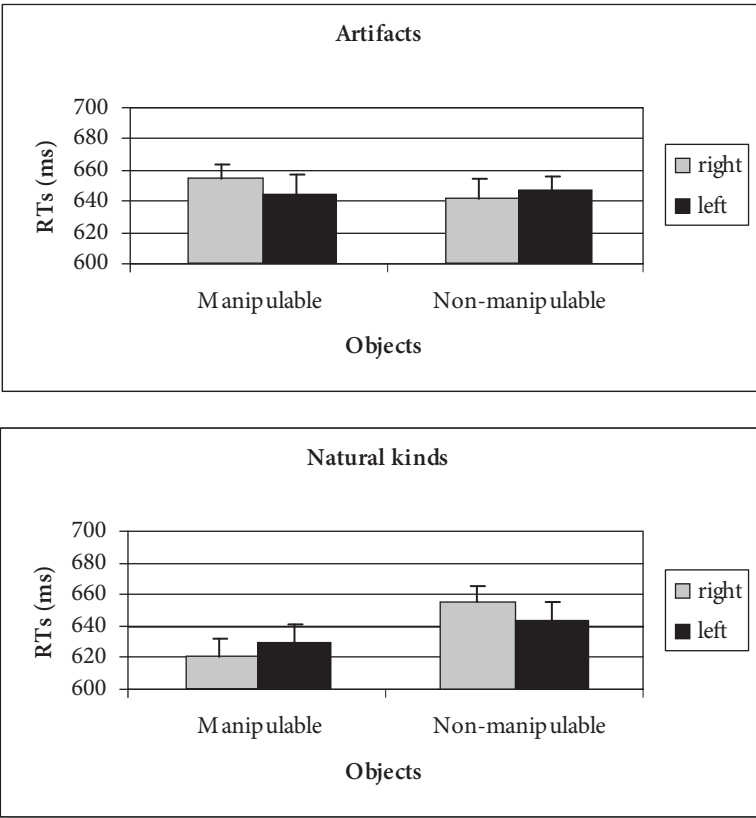


Figure 3. Experiment 2: 3-way interaction between Manipulability, Kind of Concept and Hand of Response

3.3 Discussion

The most significant results were the effect of manipulability and the advantage of natural kinds over artifacts. The effect of manipulability, which was restricted to natural objects, was complemented by the advantage of right hand responses with manipulable natural kinds. The advantage of the right hand with manipulable objects is in line with PET studies showing that imagined right hand movement activate the same area of the brain as manipulable tools (Grafton et al. 1996).

Therefore, it suggests that object names retrieved the right-hand movements associated with them. Consistently with this, Wohlschläger and Bekkering (2003) showed that children prefer to use the dominant hand when imitating the grasping of objects, while no such hand preference is observed when they imitated pointing towards objects.

The facilitation for manipulable natural kinds and not for artifacts is probably due to the fact that the first represent a case in which action and function information overlap. This is not true for manipulable artifacts, for which extracting function may have slowed down response times. The advantage of natural kinds over artifacts suggests that automatic activation of function slowed down RTs with artifacts. The interference of function took place with words, not with pictures, possibly because pictures activate function more directly than words. As the task was not to categorize objects into natural or artifacts, it is very unlikely that participants accessed the category before providing an answer, thus the explanation according to which the advantage of natural kinds over artifacts is due to the higher within-category similarity of the former can be ruled out.

4 General discussion

The present study aimed to investigate whether object concepts conveyed either by drawing or by words automatically activated motor information on how to interact with objects. If that were the case, automatic activation of such information should have an effect on the response times even when the task performed did not imply a motor preparation or a specific activation of the motor system. In Experiment 1 the automatic activation of action information lead to longer RTs while processing pictures of manipulable compared to pictures of non-manipulable natural objects. This interference effect disappeared in Experiment 2, when manipulability was made relevant for the task: the activation of action information facilitated responses to manipulable natural kinds and right hand responses to manipulable natural kinds. This suggests that participants might imagine grasping the objects with their dominant hand. This occurred with manipulable natural kinds and not with manipulable artifacts, likely because with manipulable natural kinds action and function information overlap. The results suggest that both action and function information are automatically activated in decision tasks. When this information was not relevant to the task, its activation leads to interference with the motor program necessary to accomplish it.

In both experiments the processing disadvantage of artifacts over natural kinds clearly suggests that function information is automatically activated, leading to interference with the motor program necessary to accomplish the task.

As to the relationships between Presentation Mode and Manipulability, the similarity of the result pattern obtained with words and pictures indicates that words also activate action information. In fact we found no interaction between Manipulability and Presentation Mode and manipulable words referring to natural kinds were processed faster than all other words. The striking similarity of the results obtained with drawings and with words suggests that the activation of motor information is not directly evoked by the visual stimuli but it is mediated by conceptual knowledge, i.e. by the ventral system.

A last point is worthy of note. When manipulability was relevant for the task, the advantage of natural kinds over artifacts clearly emerged with words. This raises the possibility that words interfere with processing when functional information has to be extracted. This is consistent with preliminary evidence by Bub, Masson and Bukach (2003) who found that naming was not affected by function but that affordances driven by object shape, rather than functional knowledge, are involved in object naming.

5 Conclusion and implication for cognitive science

The results suggest that seeing pictures and processing words referring to manipulable objects activates the motor system. The effect found also with words argues for the involvement not only of the dorsal but also of the ventral system and of long-term visuomotor associations between objects and actions in generating affordances. This is consistent with evidence showing that in order to perform gestures appropriate to objects it is necessary to combine conceptual knowledge with affordances derived from objects (Creem & Proffitt 2001; Buxbaum et al. 2003).

More generally, our results support theories claiming for a strict interaction between language, concepts, and the motor system. Namely, the finding that both drawings and words referring to objects activate information on manipulability is consistent with the so-called “embodied” theories of cognition. Embodied views oppose to standard cognitivist theories according to which the mind is a symbols manipulation device and concepts are given by abstract, amodal and arbitrary symbols (Pylyshyn 1986; Fodor 1987; for more recent formulations see Landauer & Dumais 1997). In the last years much evidence has been collected showing that, in antithesis with theories that posit perception and action as separate spheres (Pylyshyn 1999; Sternberg 1969), perception and action are strictly interwoven (Berthoz 1997). In addition, as mentioned before, recent studies have provided evidence of the tight interrelations between conceptual organization, language and the motor system: More specifically, it has been shown that language recruits the same systems used for perception and action. According to embodied views, con-

cepts are grounded in sensorimotor processes (Barsalou 1999; Gallese & Lakoff 2005; Glenberg 1997). More specifically, they are recordings of the neural activation that arises during perception and action arranged as distributed systems or “simulators” (Barsalou 1999). Thus, conceptual information is distributed over modality specific domains (visual, tactile, auditory, etc.) (Boronat et al. 2005; Martin et al. 1996), and the activation of the specific modalities depends on their relevance during knowledge acquisition and for the current situation (Pecher, Zeelenberg & Barsalou 2003). The difference we found between manipulable artifacts and natural objects is perfectly accounted for by the view that information is distributed over modality specific domains: Namely, with artifacts both action and function are activated, whereas for natural objects either function information is not activated, or action and function coincide, as for example while picking up vs. eating a cherry.

Proponents of a standard cognitivist approach might argue that a theory conceiving of concepts as arbitrary symbols might be able to account for our data. However, there are several reasons why we believe that an embodied explanation is the more appropriate one. First of all, differently from embodied theories, cognitivist theories might eventually provide post-hoc explanation for our results, but would hardly predict them. Further, our behavioral data are perfectly in line with results of neuroscience. Finally, the reliance of standard theories on pre-stored knowledge would lead to a problem of combinatorial explosion. Consider the different effects of Manipulability we found in Experiment 1 and 2: There are so many different actions and so many potential contexts in regard to both manipulable and non-manipulable objects that it would be very difficult to have encoded all of the different action and context links, as well as all the procedures to activate depending on the selected action and context. Amodal views have no principled way to account for the flexibility and contextual sensitivity of our cognitive system. The last, more general reason why we believe that embodied accounts are more convincing than standard ones is that to our knowledge no clear empirical support for amodal symbols has been provided so far (see Barsalou 1999; Barsalou et al. 2003 for a similar argument). For all these reasons we believe that embodied theories provide more simple and convincing explanation of our data, and that they represent a most fruitful and promising approach in cognitive science. Clearly, many outstanding issues remain concerning how action information is activated by object concepts. This chapter represents only a step trying to show that object concepts automatically activate information on how to grasp and manipulate them. The automatic activation of manipulability information allows us to respond quickly and efficiently to objects in our environment. At the same time, as the differences between the two experiments indicate, our cognitive system is highly flexible and perfectly able to activate different manipulability information depend-

ing on the context. More detailed empirical evidence is required in order to verify whether, besides general information distinguishing manipulable and non-manipulable objects, seeing objects and hearing or reading their name lead to automatic activation of specific motor programs in order to interact properly with objects.

Acknowledgements

We would like to thank Luisa Lugli and Antonello Pellicano for discussion and help in preparing the experiments. Thanks also to Cristina Iani, Elena Gherri, Lucia Riggio, Alice C. Roy and Alesia Tessari for helpful suggestions and discussions. Part of the data reported in this chapter were presented by the first author at the Third Annual Summer Interdisciplinary Conference (ASIC) in Cavalese, Italy, 27 June-2 July 2004.

References

- Barsalou, Lawrence W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22. 577–609.
- Barsalou, Lawrence W., Kyle W. Simmons, Aron K. Barbey & Christine D. Wilson. 2003. Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences* 7. 84–91.
- Berthoz, Alain. 1997. *Le sens du mouvement*. Paris: Odile Jacob.
- Borghi, Anna M. 2004. Object concepts and action: Extracting affordances from objects' parts. *Acta Psychologica* 115(1). 69–96.
- Borghi, Anna M. 2005. Object concepts and action. In Diane Pecher & Rolf Zwaan (eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking*, 8–34. Cambridge: CUP.
- Borghi, Anna M., Arthur M. Glenberg & Michael P. Kaschak. 2004. Putting words in perspective. *Memory & Cognition* 32. 863–873.
- Boronat, Consuelo B., Laurel J. Buxbaum, H. Branch Coslett, Kathy Tang, Eleanor M. Saffran, Daniel Y. Kimberg & John A. Detre. 2005. Distinction between manipulation and function knowledge of objects: Evidence from functional magnetic resonance imaging. *Cognitive Brain Research* 23. 361–373.
- Bub, Daniel N., Michael E. J. Masson & Cindy M. Bukach. 2003. Gesturing and naming: The use of functional knowledge in object identification. *Psychological Science* 14. 467–472.
- Buxbaum, Laurel J., Angela Sirigu, Myrna F. Schwartz & Roberta Klatzky. 2003. Cognitive representations of hand posture in ideomotor apraxia. *Neuropsychologia* 41. 1091–1113.
- Buxbaum, Laurel J., Tracy Veramonti & Myrna F. Schwartz. 2000. Function and manipulation tool knowledge in apraxia: Knowing 'what for' but not 'how'. *Neurocase* 6. 83–97.
- Carlson, Laura & Emile van der Zee (eds.). 2005. *Functional features in language and space*. Oxford: OUP.
- Castiello, Umberto & Marc Jeannerod. 1991. Measuring time to awareness. *Neuroreport* 2. 797–800.
- Chao, Linda L. & Alex Martin. 2000. Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12. 478–484.

- Coventry, Kenny R. & Simon C. Garrod. 2004. *Saying, seeing and action: The psychological semantics of spatial prepositions*. Hove: Psychology Press.
- Creem, Sarah H. & Dennis R. Proffitt. 2001. Grasping objects by their handles: A necessary interaction between perception and action. *Journal of Experimental Psychology: Human Perception and Performance* 27. 218–228.
- Ellis, Rob & Mike Tucker. 2000. Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology* 91. 451–471.
- Fagg, Andrew H. & Michael A. Arbib. 1998. Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks* 11. 1277–1303.
- Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, MA: The MIT Press.
- Gallese, Vittorio & George Lakoff. 2005. The brain's concepts: The role of the sensorimotor system in conceptual knowledge. *Cognitive Neuropsychology* 21. 455–479.
- Gentilucci, Maurizio. 2003. Object motor representation and language. *Experimental Brain Research* 153. 260–265.
- Gentilucci, Maurizio & Massimo Gangitano. 1998. Influence of automatic word reading on motor control. *European Journal of Neuroscience* 10. 752–756.
- Gerlach, Christian, Ian Law & Olaf B. Paulson. 2002. When action turns into words: Activation of motor-based knowledge during categorization of manipulable objects. *Journal of Cognitive Neuroscience* 14. 1230–1239.
- Glenberg, Arthur M. 1997. What memory is for. *Behavioral and Brain Sciences* 20. 1–55.
- Glenberg, Arthur M. & Michael P. Kaschak. 2002. Grounding language in action. *Psychonomic Bulletin and Review* 9. 558–565.
- Glover, Scott & Peter Dixon. 2002. Semantics affect the planning but not control of grasping. *Experimental Brain Research* 146. 383–387.
- Grafton, Scott T., Michael A. Arbib, Luciano Fadiga & Giacomo Rizzolatti. 1996. Location of grasp representations in humans by position emission tomography: 2. Observation compared to imagination. *Experimental Brain Research* 112. 103–111.
- Grafton, Scott T., Luciano Fadiga, Michael A. Arbib & Giacomo Rizzolatti. 1997. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage* 6. 231–236.
- Handy, Todd C., Scott T. Grafton, Neha M. Shroff, Sarah Ketay & Michael S. Gazzaniga. 2003. Graspable objects grab attention when the potential for action is recognized. *Nature Neuroscience* 6. 421–427.
- Humphreys, Glyn W., M. Jane Riddoch & Emer M. E. Forde. 2002. The principle of target-competitor differentiation in object recognition and naming (and its role in category effects in normality and pathology). In Emer M. E. Forde & Glyn H. Humphreys (eds.), *Category specificity in brain and mind*, 51–84. Hove: Psychology Press.
- Kellenbach, Marion L., Matthew Brett & Karalyn Patterson. 2003. The importance of manipulability and action in tool representation. *Journal of Cognitive Neuroscience* 15. 30–46.
- Klatzky, Roberta L., James W. Pellegrino, B. P. McCloskey & Sally Doherty. 1989. Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language* 28. 56–77.
- Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104. 211–240.
- Lotto, Lorella, Roberto Dell'Acqua & Remo Job. 2001. Le figure PD/DPSS. Misure di accordo sul nome, tipicità, familiarità, età di acquisizione e tempi di denominazione per 266 figure. *Giornale Italiano di Psicologia* 28. 193–207.

- Martin, Alex, Cheri L. Wiggs, Leslie G. Ungerleider & James V. Haxby. 1996. Neural correlates of category specific knowledge. *Nature* 379. 649–652.
- Pecher, Diane & Rolf A. Zwaan. 2005. *The grounding of cognition: The role of perception and action in memory, language, and thinking*. Cambridge: CUP.
- Pecher, Diane, Ren  Zeelenberg & Lawrence W. Barsalou 2003. Verifying properties from different modalities for concepts produces switching costs. *Psychological Science* 14. 119–124.
- Phillips, Jacqueline A., Glyn W. Humphreys, Uta Noppeney & Kathy J. Price. 2002. The neural substrates of action retrieval: An examination of semantic and visual routes to action. *Visual Cognition* 9. 662–684.
- Pulverm ller, Friedemann. 2003. *The neuroscience of language: On brain circuits of words and serial order*. Cambridge: CUP.
- Pylyshyn, Zenon W. 1986. *Computational cognition: Toward a foundation for cognitive science*. Cambridge, MA: The MIT Press.
- Pylyshyn, Zenon W. 1999. Is vision contiguous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22. 341–423.
- Rumiati, Raffaella I. & Glyn W. Humphreys. 1998. Recognition by action: Dissociating visual and semantic routes to action in normal observers. *Journal of Experimental Psychology: Human Perception and Performance* 24(2). 631–647.
- Saffran, Eleanor M., H. Branch Kossler & Matthew T. Keener. 2003. Differences in word associations to pictures and words. *Neuropsychologia* 41. 1541–1546.
- Sirigu, Angela, Jean Ren  Duhamel & Michel Poncet. 1991. The role of sensorimotor experience in object recognition: A case of multimodal agnosia. *Brain* 114. 2555–2573.
- Sternberg, Saul. 1969. The discovery of processing stages: Extensions of Donder's method. In Wilhelm G. Koster (ed.), *Attention and performance II*, 34–78. Amsterdam: North-Holland.
- Tucker, Mike & Rob Ellis. 1998. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* 24(3). 830–846.
- Tucker, Mike & Rob Ellis. 2001. The potentiation of grasp types during visual object categorization. *Visual Cognition* 8. 769–800.
- Tucker, Mike & Rob Ellis. 2004. Action priming by briefly presented objects. *Acta Psychologica* 116. 185–203.
- Warrington, Elizabeth K. & Tim Shallice. 1984. Category-specific semantic impairment. *Brain* 107. 829–854.
- Wohlschl ger, Andreas & Harold Bekkering. 2003. The role of objects in imitation. In Maxim I. Stamenov & Vittorio Gallese (eds.), *Mirror neurons and the evolution of brain and language*, 101–114. Amsterdam: John Benjamins.

Speaking without the cerebellum

Language skills in a young adult with near total cerebellar agenesis

Alessandro Tavano, Franco Fabbro and Renato Borgatti

Patients with acquired cerebellar lesions show evidence of expressive morphosyntactic deficits (mild agrammatism) and phonologically-guided lexical search impairments. However, recent research documented receptive morphosyntactic and semantically-guided word generation deficits. We investigated language skills in a young adult born without the cerebellum (near total cerebellar agenesis) who remarkably acquired language skills sufficient for normal communication in a 20-year time course. We provide evidence of selective persistent morphosyntactic comprehension and production deficits and of an impaired phonologically-guided word fluency task. We show that language learning without the cerebellum is possible to a significant extent, although normal language acquisition requires procedural learning abilities likely to be sustained by the normal cerebellar activity.

1 Preliminary note

The investigation of the neurobiological and neuropsychological basis of first language acquisition in children is geared to understanding *how* children learn language. Recently, theoretical and experimental insights have been gained by distinguishing between procedural and declarative processes in language learning. Procedural processes pertain to those aspects of language knowledge, such as phonology and morphosyntax, which are acquired effortlessly and without thinking. Declarative processes pertain to explicit linguistic knowledge such as the lexicon. Such distinction interacts in interesting ways with the different contributions of cortical versus subcortical structures and the cerebellum to language processing. It appears that key subcortical structures, such as the basal ganglia, and the cerebellum are especially important in determining the very first phases of language acquisition. Thus, it is important to investigate the type and extent of language

acquisition when the cerebellum and/or the basal ganglia are congenitally damaged and cannot contribute to language acquisition.

2 Introduction

2.1 Neural correlates of language acquisition

In the last 30 years, research on the neural correlates of language acquisition has concentrated on understanding the degree of specialization for language processing of neocortical speech and language areas such as the left posterior frontal cortex (Broca's area, sustaining language production) and the left posterior, superior and middle temporal cortex (Wernicke's area, sustaining language comprehension) (Bates, Thal & Janowsky 1992; Bates, Vicari & Trauner 1999; Chilosi et al. 2001). However, from a neurobiological viewpoint, the distinction between different cortical regions in the cerebral hemispheres is just one of the main functional differences shaping the developing brain. A far less investigated difference pertains to the influence of neocortical versus subcortical (that is, lying underneath the brain's neocortex) and cerebellar structures on the acquisition of higher-order cognitive abilities, and, in particular, of language (Johnson 1997). This is a major topic given that the cerebellum, and many subcortical structures and cortical areas, are functionally organized into networks with great computational capacity (Cappa & Vallar 1992). However, the influence of subcortical structures and the cerebellum on first language acquisition and development is quite difficult to study, since contrary to specialized neocortical areas, they are normally involved in the processing of many different types of information and therefore it is hard to selectively study their impact on the acquisition of single functions.

2.2 Basal ganglia, cerebellum and language

Recent advances in the neurolinguistics of subcortical aphasia have highlighted the involvement of multimodal highly integrative subcortical structures such as the basal ganglia in language abilities (Fabbro 1997). This has been taken to suggest that the basal ganglia take part in the modulation of linguistic chunks already processed at neocortical level (Nadeau & Crosson 1997). The basal ganglia constitute a large system that receives inputs from and feeds outputs to neocortical areas through the thalamus, thereby instantiating a cortico-subcortical network (DeLong 2000). Generally, they have been mainly involved in the control of voluntary fine-grained movement. However, they have been shown to play a role in cognition and mood modulation, and possibly in the processing of structural aspects of language

such as phonology and syntax (Pickett et al. 1998; Tettamanti et al. 2005). The cerebellum, a complex structure which occupies most of the posterior cranial fossa at the lower back of the skull (Altman, Naidich & Braffman 1992), is concerned with movement coordination, smooth movement execution, sequencing and timing of motor acts. It is particularly important for implicit, procedural motor learning (Leiner, Leiner & Dow 1991; Ghez & Thach 2000). However, functional neuroanatomical studies have also shown patterns of cerebellar activation associated with the activation of Broca's area in word and rhyme generation tasks (Petersen et al. 1989; Petersen & Fiez 1993). Clinical studies highlight an association between cerebellar lesions and mild expressive impairment (mild agrammatism, see Schmahmann & Sherman 1998), more evident at the phonological and morpho-syntactic levels (Justus 2004; Justus et al. 2005). A major question at this point is whether the contribution of the basal ganglia and the cerebellum to language is specific or not (Fabbro et al. 2004). In other words: Are the basal ganglia and the cerebellum involved in language processing (specific contribution), or do they simply sustain language processing by providing the network with greater computational capacity (e.g. by modulating the processing of neocortically formed chunks of linguistic information)?

2.3 The cerebellum: Basic structure and functions

Within the cerebellum different functions have been attributed to anatomically distinct regions (see Figure 1). From a dorsal view it is possible to distinguish three main lobes: (1) the anterior lobe; (2) the posterior lobe, further divided into three functional/anatomical regions, the central vermis (mediating visual, auditory, somatic information and affect, as well as posture, locomotion and gaze) and the lateral and intermediate zones of each hemisphere (mediating executive functions such as planning, visuospatial abilities and language); and (3) the flocculonodular lobe, phylogenetically more primitive than the other lobes, controlling balance and eye movement (Ghez & Thach 2000).

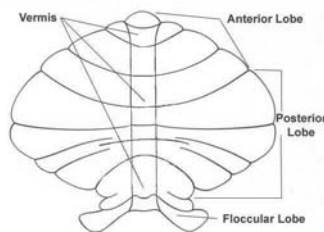


Figure 1. Main cerebellar regions

The cerebellum has major connections with both the brain stem (basic regulative and motor functions) and the supratentorial structures, in particular with temporal and frontal lobes (constituting the “Frontal Lobe System”, which includes Broca’s area and Supplementary Motor Area) (Leiner, Leiner & Dow 1991). Such connections allow the cerebellum to mediate the coordination of motor functions and motor skill learning (Dow & Moruzzi 1958; Schmahmann 1997). As Dow and Moruzzi (1958) showed, the cerebellum is not essential for the initiation of single limb movements, but it governs the acquisition and skilled coordination of balance, posture and spatio-temporal accuracy in limb movement. Disorders of these abilities following cerebellar damage are termed *ataxias*.

Cerebellar *dysarthria* is a characteristic speech-motor disorder within the range of ataxic disorders, which consists of a disruption of voice production and sound shaping due to defective coordination of the respiratory, laryngeal and orofacial subsystems of speech production (Ackermann et al. 1992). This implies imprecise consonant and vowel production, irregular articulatory breakdown, and scanned speech. Fast repetition of syllables is impaired. A reduction in stress contrast and intonational contours induces dysprosodia. Transient mutism may occur after removal of posterior fossa tumors.

Cerebellar non-motor functions include non-motor learning, attention coordination, planning, memory and affect modulation (Leiner, Leiner & Dow 1991; Fiez et al. 1992; Schmahmann & Sherman 1998; Schmahmann & Caplan 2006). The cerebellum allows visuospatial integration of complex figures and is involved in working memory (both verbal and spatial) by operating as an internal clock for the computation of durational parameters in all cognitive tasks. As for affect modulation, damage to the cerebellar vermis may induce aggressive behaviour, disinhibition, and autistic-like behaviours.

2.4 Types of language disorders associated to cerebellar lesions

Given its structural and functional complexity, the cerebellar contribution to language acquisition and breakdown is necessarily multifaceted (Petersen et al. 1989; Silveri, Leggio & Molinari 1994; Schmahmann & Sherman 1998; Ullman 2001; Fabbro et al. 2004).

Clinical studies of acquired lesions provide evidence of language breakdown. Schmahmann and Sherman (1998) suggested that patients with acquired cerebellar lesions may suffer from *Cerebellar Cognitive-Affective Syndrome*, characterized by impaired planning and anticipation, disrupted spatial cognition for complex images, affect changes, expressive speech/language deficits (agrammatism, mild anomia, dysprosodia). Levisohn, Cronin-Golomb and Schmahmann (2000), studying 19 children with excision of cerebellar tumor (which accounts for over 50%

of all brain tumors, see Cohen, Duffner & Tebbi 1982; Bearzotti et al. 2004) confirmed such a clinical picture for a paediatric population, with younger children being the least likely to show deficits in language, cognition or affect, and recovery being attributed to brain plasticity factors. Riva and Giorgi (2000) studied 26 children with excision of posterior fossa tumours and found that lesions to the vermis could yield two different results: postsurgical mutism, which later on in some cases evolved into a form of agrammatism; disturbances of communication, from irritability to autistic-like behaviours. Cerebellar hemispheric lesions were the most likely to induce selective neuropsychological disturbances. Excision of right cerebellar hemisphere tumours more often induced disturbances of auditory sequential memory and expressive language disorders (e.g. lower than normal Mean Length of Utterances, MLU; mild agrammatism, reduced verbal fluency). More often excision of left cerebellar hemisphere tumours was marked by impairments in lexical access and visuospatial abilities.

The issue of the lateralization of neuropsychological functions in the cerebellum has been raised with special reference to language. Projections from the cerebellum to the cerebral cortex are contralateral: The right cerebellar hemisphere projects to the left cerebral hemisphere, and the left cerebellar hemisphere projects to the right cerebral hemisphere (Leiner, Leiner & Dow 1991). Mild agrammatism should thus occur in the presence of right cerebellar hemisphere damage (Mariën et al. 2001; Scott et al. 2001; Silveri, Leggio & Molinari 1994; Silveri et al. 1998; Zettin et al. 1997). However, some authors suggest that there is mixed evidence with respect to this issue (Fabbro 2000; Fabbro et al. 2004; Tavano, Fabbro & Borghatti 2007; Tavano et al. *In press*).

Regarding the characteristics of language disorders associated to cerebellar lesions, the literature provides solid evidence for two types of language disorder: (1) impaired verbal fluency; (2) mild agrammatism. It is worth noting that verbal fluency deficits do not depend on dysarthric speech interferences: Cerebellar patients with impaired speech output (cerebellar atrophy) perform better than cerebellar patients with more preserved speech (localized cerebellar lesion) on verbal fluency tasks. Actually, cerebellar patients seem to have particular difficulties with phonemic clustering, while semantic clustering is preserved, a finding which has led some researchers to posit a disruption of phonemic word-retrieval strategies as the underlying deficit (Molinari, Leggio & Silveri 1997). However, imaging findings suggest a role for the cerebellum, and especially the right cerebellar hemisphere, in semantically-guided word generation (Petersen et al. 1989; Petersen & Fiez 1993; Herholz et al. 1996). These findings have been confirmed at a clinical level: Patients with cerebellar lesions show marked impairments on specific tasks requiring semantically-guided word generation (such as synonym generation, verb from noun generation and the like, Fiez et al. 1992; Fabbro et al. 2004). The

nature of word finding problems in cerebellar patients remains therefore open to investigation.

With regard to agrammatism, consistent findings indicate the presence of morphosyntactic deficits, which are usually transient, subtle (Zettin et al. 1997; Fabbro et al. 2004) and more evident in production than in comprehension (Justus 2004). As is the case with aphasia, language typology influences the type of agrammatism displayed. For example, in relation to English, Justus (2004) found that cerebellar patients tended to omit closed class words such as articles while word order was preserved, in line with the fact that the word order in English conveys the major semantic and syntactic relations among phrases. English-speaking cerebellar patients also have problems discriminating grammatical from ungrammatical sentences when required to focus on subject-verb agreement. In relation to Italian, Silveri, Leggio and Molinari (1994) discussed the case of a patient who, besides showing dysarthric speech, tended to omit free grammatical morphemes such as auxiliaries and clitic pronouns, and to substitute verbal inflections with infinitive endings. They attributed these findings to a failure of the cerebellum in timing (modulating) the integration of neocortically formed morphosyntactic features at the level of verbal short-term memory before production. Fabbro et al. (2004) confirmed Silveri, Leggio and Molinari's (1994) findings in two non-dysarthric young adult twins but also found instances of bound grammatical morpheme substitution, leading to subject-verb and gender agreement errors. Such errors may have a major impact on the intelligibility of Italian sentences, where the word order is freer than in English and bound morphology conveys the main morphosyntactic relations among phrases. In recent work on language development in Italian patients with cerebellar malformations, we found that, contrary to acquired cerebellar lesions, congenital lesions may induce permanent language disorders and affect comprehension to a greater extent. We also found errors in free and bound morphology in both production and comprehension (Borgatti et al. 2004; Tavano, Fabbro & Borgatti 2007). These findings allow to draw two main inferences: (1) They underline the possibility that the cerebellum is involved not only in language production but also in language comprehension; (2) by the same token, they open a window into the procedural nature of cerebellar non-motor learning, which would thus be implicated in sustaining the acquisition and *processing* of procedural aspects of language, such as phonology and morphosyntax. An impairment in language processing would explain the presence of deficits in both comprehension and production (Molinari et al. 1997; Ullman 2001).

In this work we present and discuss the course of language acquisition and the actual language skills of a young Italian adult with near total cerebellar agenesis; part of these data was already discussed in Borgatti et al. (2004). Our main aims are the following: 1) to further document that speaking without the cerebellum is

possible; 2) to investigate the grammatical and lexical limits of acquiring language without the support of procedural cerebellar learning. In this way we might be able to provide further indications for future research on the more general issue of the type (processing vs. modulation) of cerebellar contribution to language.

3 Materials and method

3.1 Participant's characteristics

GR is a right-handed male patient aged 32,01 at the time of assessment. He was referred to the our Institute at age 4,06, where he remained until he was able to work autonomously.

3.2 Neurological assessment procedure

Standard neurological examinations followed the lines of the "International Cooperative Ataxia Rating Scale" (Trouillas et al. 1997). The patient received a standard psychiatric assessment (according to the DSM-IV), EEG monitoring and MRI scan.

3.3 Neuropsychological and neurolinguistic assessment procedures

Quantitative and qualitative data were obtained for the following neuropsychological functions:

- *IQ*: Wechsler Adult Intelligence Scale-Revised (WAIS-R, Wechsler 1997), Griffiths Scale (Griffiths 1984);
- *Orientation*: Temporal Spatial Orientation Test (Spinnler & Tognoni 1987);
- *Visuospatial abilities*: Constructive Apraxia Test (Spinnler & Tognoni 1987), Rey Figure "A" (Rey 1967);
- *Executive abilities*: Wisconsin Card Sorting Test (Heaton et al. 1993); Tower of London Test (Shallice 1982);
- *Verbal Memory*: Word span, forward (Spinnler & Tognoni 1987); Verbal Long-Term Memory Test (Spinnler & Tognoni 1987);
- *Procedural memory*: Mirror Star Tracing Task (manual);
- *Language*: GR was administered the Bilingual Aphasia Test (BAT), Part B (Paradis 1987, 1999). The BAT, Part B consists of 32 subtests (427 items) and allows an in-depth investigation of linguistic levels and linguistic skills. We compared GR's performance to the values obtained for a control group of 19 participants (12 males, 7 females) with high educational level (equivalent to MSc or A level degree), mean age 35.6 (age range 31–42), described in detail in Fabbro et al. (2004);

- *Narrative*: The Bird Nest Story, a picture story consisting of six pre-arranged figures which the patient has to describe (Paradis 1987; Tavano, De Fabritiis & Fabbro 2005). GR's performance was compared to the values obtained for a control group of 10 participants extracted from the pool of 19 who were administered the BAT;
- *Reading*: "Prove di lettura MT" (Cornoldi, Colpo & Gruppo 1998) consisting of age-appropriate standardized reading tasks for the Italian school population.

3.4 Statistical analysis

For the narrative measures an analysis was performed using the procedures and software for single-case investigations presented in Crawford and Garthwaite (2002).

4 Results

4.1 Neurological and neuroradiological features

The standard neurological examination (Trouillas et al. 1997) yielded a score of 29/100, corresponding to a diagnosis of cerebellar ataxia (Postural and gait disturbances = 8/34; Limb ataxia = 16/52; Speech apraxia = 3/8; Oculomotor disorders = 2/6). The MRI showed a near-total absence of cerebellum, with a minimal residue of the vermal region, and a flat ventral pons (see Figure 2). This type of malformation was classified as a Type I cerebellar malformation according with Altman, Naidich and Braffman's (1992) criteria (see Figure 2). Our case is similar to the very few ones described by Gardner et al. (2001) under the term "Near-total absence of cerebellum with flat ventral pons and relatively mild clinical affection". The etiopathogenesis of the "Near-total absence of the cerebellum" syndrome is not known: Anatomical considerations on the remaining tissue suggest a very early destruction of normally developing cerebellar tissue as a likely cause (Gardner et al. 2001). Recently a very similar case has been described by Titomanlio, Romano and Del Giudice (2005).

4.2 Development

It is remarkable that, notwithstanding the lesion size, this syndrome affects cognition and behaviour only to a limited extent. However, this is true of the end state of a highly deviant developmental course, which initially shows a marked delay in all functions but gradually recovers over time up to a sufficient extent (see Table 1).

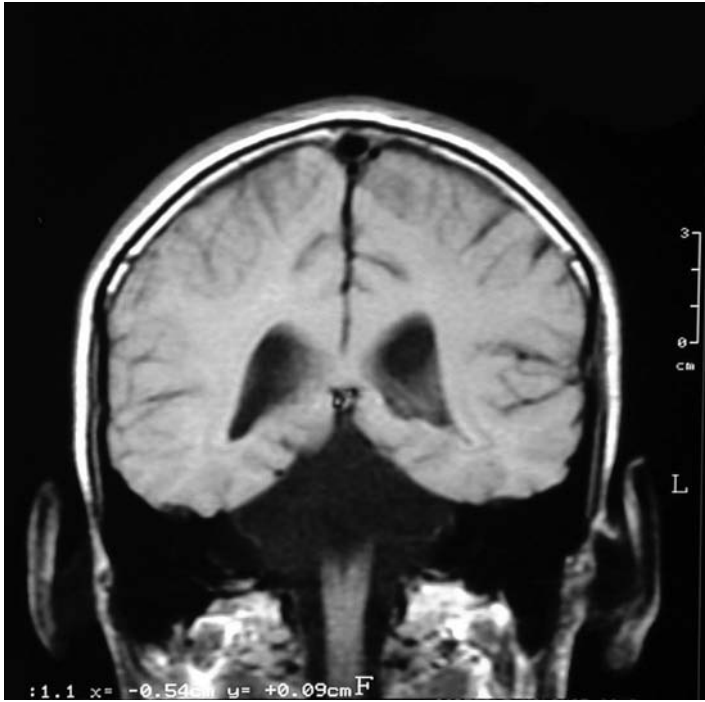


Figure 2. Near-total absence of the cerebellum; coronal section, T1-weighted

Language development was extremely delayed up to 7 years of age, since he showed good phonological skills in terms of single phoneme pronunciation, but coarticulation was still effortful and thus word production was often incorrect. Basic communicative functionality (simple sentences) was in place at around 12 years of age.

4.2.1 Neuropsychological assessment at present

IQ. Administration of the WAIS-R (Wechsler 1997) showed a harmonious profile of mild retardation (VIQ = 72, PIQ = 67, FIQ = 68), with major problems in Arithmetics (standardized score = 3), Coding (ss = 3), Similarities (ss = 4), and Picture Arrangement (ss = 4).

Orientation. The patient was oriented in time and space.

Visuospatial functions. Rey Complex Figure = marked impairment, age equivalent 6.06 years (Copy); 7.06 years (Memory).

Constructional Apraxia = normal (raw score = 13; Mean = 13.2 for the 40–49 age group, SD = 1.48).

Table 1. Main developmental milestones

DOMAINS	GR	CONTROLS*
<u>MOTOR</u>		
Head movement control	5 months	1–3 months
Crawling	5 years	11 months
Walking unsupported	22 years	14 months
<u>LANGUAGE</u>		
Produces at least 2 clear single words	24 months	11 months
Produces 2-word or longer sentences	7 years	24 months
Reading and writing	10 years	6 years
<u>SOCIAL</u>		
Smiles	3 months	2 months
Turns head to speaking adult	7 months	5 months
Plays with other children	9 years	4 years

* based on the Griffiths Scale normative data.

Executive functions. Tower of London Test = marked impairment in attention (raw score = 8.62; equivalent score = 0 for the age group of 40 years).

Wisconsin Card Sorting Test = no impairment (Total errors = 18%, Mean and SD = 24.32 ± 15.11; Total categories completed = 6, Mean and SD = 5 ± 1.6).

Executive functions. Tower of London Test = marked impairment in attention (raw score = 8.62; equivalent score = 0 for the age group of 40 years).

Wisconsin Card Sorting Test = no impairment (Total errors = 18%, Mean and SD = 24.32 ± 15.11; Total categories completed = 6, Mean and SD = 5 ± 1.6).

Verbal memory. Word Span Test = normal (raw score = 4; Mean = 4.75 for the 40–49 age group, SD = 0.86).

Long-term Memory Test = mildly defective performance (raw score = 8.5; Mean = 13.32 for the 40–49 age group, SD = 2.65).

Procedural memory. Mirror Star Tracing Task = normal patterns of improvement through practice, scores are comparable to subjects with the same mental age.

4.2.2 *Neurolinguistic assessment at present*

Voice. Hypophonic, scanned, nasalized (non-occluding velum). Irregular patterns of articulatory breakdown suggest a diagnosis of ataxic dysarthria (*cenò* instead of *sennò*, ‘otherwise’; /zi/ instead of the Italian affirmative particle /si/).

Formal language test. The patient was administered the Italian adaptation of the Bilingual Aphasia Test, Part B (Paradis 1987, 1999, see Table 2).

Table 2. Neurolinguistic assessment

	BAT	GR
Language skills	Comprehension	78 [‡]
	Repetition	100
	Grammatical judgment	98
	Lexical access	92 [†]
	Sentence construction	86 [‡]
	Reading	89 [‡]
	Writing	60 [‡]
Language levels	Phonology	96 [†]
	Morphology	95 [*]
	Syntax	72 [‡]
	Lexicon	93 [†]
	Semantics	85 [‡]

* Mild deficit; † Moderate deficit; ‡ Severe deficit.

Quantitative results highlight a severe impairment of syntax and semantics, which impacts on comprehension, sentence construction and writing. On closer examination of each subtest, the worst performances include Mental arithmetic (13% correct), Sentence dictation (20%), Auditory comprehension of a short story (40%), Reading comprehension of sentences (50%), Syntactic comprehension (70%) and Antonyms (80%).

Phonemic fluency was markedly impaired as well: GR was able to say only 4 words beginning with /t/, where he also made a morphological gender mistake [*tavolo* 'table'; *tamburo* 'drum'; *tino* 'barrel'; *tazzo* [MASC] instead of *tazza* [FEM] 'cup'; *telefono* 'phone']; 3 words beginning with /m/ [*mamma* 'mummy'; *mammifero* 'mammal'; *mammuto* 'mammoth']; 6 words beginning with /p/ [*pipa* 'pipe'; *pupu* childish word for 'excrement'; *pero* 'pear tree'; *polpastrello* 'finger tip'; *pipistrello* 'bat'; *pipì* childish word for 'urine'].

A qualitative investigation of the Sentence construction subtest (cf. Figure 3) shows that the patient had marked difficulties in elicited sentence generation ([*IL*] = omitted words; [MASC] = agreement uncertainties; HOUSE = words provided by the examiner).

Remarkable problems in sentence programming are evident also in the spontaneous 5-minute conversation. Further, a total of 4 morphosyntactic errors (omissions

Sentence Construction Test**(Italian)**

Examiner: bravo senti adesso io ti dico delle parole e tu mi fai una frasetta per esempio se ti dico porta aprire infermiera tu mi puoi dire l'infermiera apre la porta.

Ok? CASA GATTO

GR: il gatto va a casa

E: bene. POLTRONA / DOTTORE / SEDERSI

GR: [IL] dottore si siede nella poltrona

E: bene. SCRIVANIA / APRIRE / CASSETTO

GR: il [MASC] calzett la [FEM]@ il calzetto è nella scrivania

E: bene. ALBERO / VERDE / FOGLIA / VEDERE

GR: a@

E: una frasetta con questi quattro parole ALBERO / VERDE / FOGLIA / VEDERE

GR: l'albero è nel verde cio vedere la foglia

E: me la puoi ripetere?

GR: l'albero è@ # l'albero si vede nel nel verde si vede una foglia

E: va bene l'ultima MATITA / SCRIVERE / BLU / FOGLIO

GR: la matita scrive nel blu fo nel foglio blu

E: va bene

Sentence Construction Test**(approximate English translation)**

F: good now listen I'll say some words and you will use them to build up a simple sentence for instance if I say door open nurse you can say the nurse opens the door.

Ok? HOUSE CAT

G: the cat goes home

F: fine. ARMCHAIR / PHYSICIAN / TO SIT

G: physician sits in the armchair

F: fine. DESK / OPEN / DRAWER

G: the sock the @ the sock is in the desk

F: fine. TREE / GREEN / LEAF / SEE

G: a@

F: a simple sentence with these four words TREE / GREEN / LEAF / SEE

G: the tree is in the green cio to see the leaf

F: can you repeat it?

G: the tree is @ the tree one can see in the in the green one can see a leaf

F: all right the last one PENCIL / WRITE / BLUE / PAPER

G: the pencil writes on the pap blue on the blu paper

F: all right

Figure 3. Qualitative investigation of sentence construction subtest

of free grammatical morphemes, definite articles and direct and indirect clitic objects) were present in spontaneous speech:

- (1) *[MI] piace la formula 1* '[TO ME] like formula 1'
- (2) *[I] genitori sono morti* '[THE] parents are dead'
- (3) *un altro fa [L] elettricista* 'another works as [THE] electrician'
- (4) *mia sorella [LA] casalinga* 'my sister [THE] housewife'

Narrative. On the Bird Nest Story GR produced a story which is comparable to that of control subjects but presents with many morphosyntactic and word finding errors (see Table 3, Crawford & Garthwaite 2002).

Table 3. The bird nest story

Parameters	GR	Controls*	Analysis
Total words	86	123.6 (44.57)	p=0.22
Narrative time	49	58 (21.72)	p=0.35
Fluency (words per minute)	105,3	120.27 (25.96)	p=0.29
Narrative words	73	114.6 (40.03)	p=0.17
MLU	10,42	11.04 (4.63)	p=0.45
TOKENS	29	50.2 (18.21)	p=0.14
TYPES	21	33.8 (10.1)	p=0.12
TY/TO RATIO	0,72	0.7 (0.11)	p=0.43
Omission of free grammatical morphemes (GM)	–	–	–
Substitution of FREE GM	1	–	–
Addition of FREE GM	1	–	–
Substitution of BOUND GM	1	–	–
PHON paraphasias	–	–	–
SEM paraphasias	1	–	–
ANOMIAS & CONTENT WORD ERRORS	2	–	–

* Means (standard deviations) across 10 control subjects

The patient did not understand the gist of the story.

Reading. Administration of the MT reading tests evidenced a score average for a 5th grader as far as correctness, rapidity and comprehension are concerned.

5 Discussion

We presented the case of a young adult male with near-total cerebellar agenesis who, through an exceptionally long time course, managed to acquire language and

generally higher-order cognitive abilities to a remarkable extent (Borgatti et al. 2004). At present the patient is able to use language effectively for his everyday activities, including familiar and work relationships. However, he shows clinically evident neuropsychological deficits, marked permanent speech disturbances and impairments of both receptive and expressive language.

The time course of functional acquisition initially presents with a severe diffuse motor deficiency. While head control is gained within the first year of life, developmental motor milestones such as crawling are far more delayed and independent walking is attained at 22 years of age. It appears that the near-total absence of the cerebellum especially affected the coordination and timing of sequences of limb movements necessary to deambulate, such that the more complex the dynamic interplay of motor segments required by each developmental milestone, the later its acquisition.

Although most evident for motor skills, such difficulties may be seen also in other domains. The results of both formal and naturalistic speech and language tasks suggest that a disruption in the timing of sequencing skills may help explain problems in both the coordination of articulatory muscles (e.g. coarticulation of speech sounds) and the sequencing of phrases (syntax). Social abilities do not seem to be as delayed as motor and linguistic skills, notwithstanding the fact that up to age 10 GR's personality was extremely passive, and active interaction with peers was not sought. At present no psychopathological traits can be distinguished. Therefore, the overall exceptionally delayed developmental trend seems to highlight a distinction between social abilities, which were generally preserved, and motor and speech/language abilities, which were markedly impaired, possibly because of a general underlying deficit in the timing of sequences of coordinated actions.

The neuropsychological assessment documents an end-state harmonious profile of mild retardation (VIQ = 72, PIQ = 67, FIQ = 68); subtests requiring procedural motor efficiency (Coding), procedural memory of operations (Arithmetics), the extraction of higher-order semantic inferences (Similarities) and visuospatial abilities (Picture arrangement) are the most impaired. As for visuospatial abilities, no signs of constructional apraxia were present. Thus, the highly severe impairment on the Rey Complex Figure Test cannot be preeminently attributed to peripheral graphomotor disturbances. Actually, as shown in Schmahmann and Sherman (1998) and Borgatti et al. (2004), this finding is likely to represent a major inability to appreciate the organizing structure of the figure. Alternatively, such findings could be attributed to executive (planning and attention) difficulties, given the evidence that the cerebellum participates in sustaining the frontal lobe functions (Courchesne et al. 1994). The Tower of London Test results are in agreement with this hypothesis. However, the fact that the patient performed within normal limits on the Wisconsin Card Sorting Test shows that GR does not suffer

from frontal symptoms, such as perseveration. Therefore, GR seems to be impaired on tasks which require smooth handling of complex material, whether of a visuospatial or linguistic nature, although visuospatial information seems harder to process. Finally, he shows a normal learning pattern on the Mirror Star Tracing Test, and the overall performance is in line with his mental age.

A qualitative voice assessment showed a mild degree of ataxic dysarthria, which is not sufficient to interfere with language abilities as they result from formal tests, and spontaneous and narrative speech. As for language, the main findings can be summed up as follows:

1. GR presents with major impairments in syntactic comprehension, which is likely to affect also sentence construction, and in general he presents with both receptive and productive impairments.
2. Somewhat milder lexical-semantic deficits can be evidenced, but phonemically-guided word retrieval (word fluency task) seems far more impaired than semantically-guided lexical search, given that among the lexical-semantic subtests included in the BAT the patient scored lower than normal only on the Antonyms subtest, while he scores at ceiling on both Synonyms and Semantic Categories subtests.

These findings are in agreement with clinical studies of patients with acquired cerebellar lesions. Indeed, acquired lesion symptoms are subtle and tend to regress rather rapidly over time, while in our case they persisted over the highly delayed developmental time course (see Fabbro et al. 2004; Borgatti et al. 2004). The patient shows expressive morphosyntactic difficulties in his spontaneous and narrative speech such as closed-class word omission, which is typical of agrammatic speech as far as Italian is concerned, notwithstanding the fact that his overall narrative performance is within normal limits (see Table 3). GR also presents with some clear instances of gender agreement errors (substitution of inflectional morphemes), a fact which distinguishes his case from the ones presented in Fabbro et al. (2004), and generally from the typical cases of agrammatism in Italian.

Silveri, Leggio and Molinari (1994) and Zettin et al. (1997) argue that, in the case of mild agrammatism following cerebellar focal lesions no actual language impairment is to be envisaged but these symptoms can be attributed to a deficit in the timing of the interplay of morphological stems and endings at the level of verbal working memory, an ability which depends on the cerebellum. Therefore, according to these authors the cerebellum has a modulating role in language. In our case such difficulties are present also on the receptive side, since the patient scored 61 out of 87 correct items on the Syntactic Comprehension subtest of BAT. The presence of morphosyntactic difficulties on both receptive and expressive sides suggests the likelihood of an underlying impairment in language *processing*, and

this very fact would argue for a specific contribution of the cerebellum to language processing (Justus 2004). In our case, support for this last view comes also from the fact that assessment of formal repetition abilities, which involve verbal working memory as well, shows no impairment and, further, word span measures are within normal limits. Furthermore, the patient's mean length of utterance on the narrative task was within normal limits, a fact which argues against a memory limitation in online handling of verbal information. It follows that the verbal working memory deficit hypothesis cannot fully explain our findings. Moreover, GR shows both omissions of free grammatical morphemes (function words) and substitutions of bound (inflectional) grammatical morphemes; taken together these data argue for a defective acquisition of grammatical knowledge. Such perspective can be projected on other findings as well. For example, the phonemically-related impairment in word search could be understood as a manifestation of a general procedural phonemic (initially articulatory) learning insufficiency which determined a less than optimal lexical organization on phonemic basis (Ullman 2001). Notice that this would contribute to explaining why generally the patient's lexicon seems poorly organized and he shows some problems in semantic inferencing. However, this issue needs further investigation in future studies.

In conclusion, even if plasticity factors may have relocated cerebellar functioning at a cerebral level as it seems to be the case in some instances of near-total absence of the cerebellum syndrome (Gardner et al. 2001), in GR's case such relocation seems far from complete. From this very exceptional case we can conclude that speaking without the cerebellum is possible but that normal language acquisition may not be an exclusively or even – at the first stages – mainly neocortical matter. It actively requires the involvement of the cerebellum to support procedural processing of specific linguistic features (Molinari et al. 1997, Ullman 2001).

References

- Ackermann, Hermann, Matthias Vogel, Dirk Petersen & Michael Poremba. 1992. Speech deficits in ischaemic cerebellar lesions. *Journal of Neurology* 239. 223–227.
- Altman, Nolan R., Thomas P. Naidich & Bruce H. Braffman. 1992. Posterior fossa malformations. *American Journal of Neuroradiology* 13. 691–724.
- Bates, Elisabeth S., Donna Thal & Jeri S. Janowsky. 1992. Early language development and its neural correlates. In Isabel Rapin & Sidney J. Segalowitz (eds.), *Handbook of neuropsychology*, Vol. 7, 69–110. Amsterdam: Elsevier.
- Bates, Elisabeth S., Stefano Vicari & Doris Trauner. 1999. Neural mediation of language development: Perspectives from lesion studies of infants and children. In Helen Tager-Flusberg (ed.), *Neurodevelopmental disorders*, 533–581. Cambridge, MA: The MIT Press.

- Bearzotti, Federica, Alessandro Tavano, Paolo Pecile, Miran Skrap & Franco Fabbro. 2004. Neuropsychological and language disorders in 12 adolescents with brain tumor. *Saggi: Child Development and Disabilities* 30. 69–84.
- Borgatti, Renato, Alessandro Tavano, Guido Cristofori & Franco Fabbro. 2004. Language development in children with cerebellar malformations. In Franco Fabbro (ed.), *Neurogenic language disorders in children*, 127–148. Amsterdam: Elsevier.
- Cappa, Stefano F. & Giuseppe Vallar. 1992. Neuropsychological disorders after subcortical lesions: Implications for neural models of language and spatial attention. In Giuseppe Vallar, Stefano F. Cappa & Claus W. Wallesch (eds.), *Neuropsychological disorders associated with subcortical lesions*, 7–41. Oxford: OUP.
- Chilosi, Anna Maria, Paola Cipriani, Barbara Bertuccelli, Lucia Pfanner & Giovanni Cioni. 2001. Early cognitive and communication development in children with focal brain lesions. *Journal of Child Neurology* 16(5). 309–316.
- Cohen, Michael E., Patricia K. Duffner & Cameron K. Tebbi. 1982. Brain tumors in children: Diagnosis and management. In Cameron K. Tebbi (ed.), *Major topics in pediatric and adolescent oncology*, 241–289. Boston MA: GK Hall.
- Cornoldi, Cesare, Giovanni Colpo & M. T. Gruppo. 1998. *Prove di lettura MT per la scuola elementare* 2. Firenze: Organizzazioni Speciali.
- Courchesne, Eric, Jeanne Townsend, Natacha A. Akshoomoff, Saitoh Osamu, Rachel Yeung-Courchesne, Alan J. Lincoln, Hector E. James, Richard H. Haas, Laura Schreibman & Lily Lau. 1994. Impairment in shifting attention in autistic and cerebellar patients. *Behavioural Neuroscience* 108. 848–865.
- Crawford, John R. & Paul H. Garthwaite. 2002. Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40. 1196–1208.
- DeLong, Mahlon R. 2000. The basal ganglia. In Eric R. Kandel, James H. Schwartz & Thomas M. Jessel (eds.), *Principles of neural science*, 853–867. New York, NY: McGraw-Hill.
- Dow, Robert S. & Giuseppe Moruzzi. 1958. *The physiology and pathology of the cerebellum*. Minneapolis, MN: The University of Minnesota Press.
- Fabbro, Franco. 1997. Introduction: Subcortical aphasia. *Journal of Neurolinguistics* 10. 251–253.
- Fabbro, Franco. 2000. Introduction to language and cerebellum. *Journal of Neurolinguistics* 13. 83–94.
- Fabbro, Franco, Alessandro Tavano, Stefano Corti, Nereo Bresolin, Paola De Fabritiis & Renato Borgatti. 2004. Long-term neuropsychological deficits after cerebellar infarctions in two young twins. *Neuropsychologia* 42. 536–545.
- Fiez, Julie A., Steven E. Petersen, Marshall K. Cheney & Marcus E. Raichle. 1992. Impaired non-motor learning and error detection associated with cerebellar damage. *Brain* 115. 155–178.
- Gardner, Robert J., Lee T. Coleman, Leslie A. Mitchell, Lindsay J. Smith, Anthony S. Harvey, Ingrid E. Scheffer, Elsdon Storey, Michael J. Nowotny, Robert A. Sloane & Lionel Lubitz. 2001. Near-total absence of the cerebellum. *Neuropediatrics* 32. 62–68.
- Ghez, Claude & William T. Thach. 2000. The cerebellum. In Eric R. Kandel, James H. Schwartz & Thomas M. Jessel (eds.), *Principles of neural science*, 832–852. New York, NY: McGraw-Hill.
- Griffiths, Ruth. 1984. *The abilities of young children*. High Wycombe: The Test Agency.
- Heaton, Robert K., Gordon J. Chelune, Jack L. Talley, Gary G. Kay & Glenn Curtiss. 1993. *The Wisconsin card sorting test*. Odessa: Psychological Assessment Resources.

- Herholz, Karl, Alexander Thiel, Klaus Wienhard, Uwe Pietrzyk, Hans-Martin von Stockhausen, Hans Karbe, Josef Kessler, Thomas Bruckbauer, Marco Halber & Wolf-Dieter Heiss. 1996. Individual functional anatomy of verb generation. *Neuroimage* 3. 185–194.
- Johnson, Mark H. 1997. *Developmental cognitive neuroscience*. London: Blackwell.
- Justus, Timothy. 2004. The cerebellum and English grammatical morphology: Evidence from production, comprehension, and grammaticality judgements. *Journal of Cognitive Neuroscience* 16. 1115–1130.
- Justus, Timothy, Susan M. Ravizza, Julie A. Fiez & Richard B. Ivry. 2005. Reduced phonological similarity effects in patients with damage to the cerebellum. *Brain and Language* 95. 304–318.
- Leiner, Henrietta C., Alan L. Leiner & Robert S. Dow. 1991. The human cerebro-cerebellar system: Its computing, cognitive and language skills. *Behavioural Brain Research* 44. 113–128.
- Levisohn, Lisi, Alice Cronin-Golomb & Jeremy D. Schmahmann. 2000. Neuropsychological consequences of cerebellar tumor resection in children. Cerebellar cognitive affective syndrome in a paediatric population. *Brain* 123. 1041–1050.
- Mariën, Peter, Sebastiaan Engelborghs, Franco Fabbro & Peter P. de Deyn. 2001. The lateralized linguistic cerebellum: A review and a new hypothesis. *Brain and Language* 79. 580–600.
- Molinari, Marco, Maria G. Leggio, Maria C. Silveri. 1997. Verbal fluency and agrammatism. *International Review of Neurobiology* 41. 325–339.
- Molinari, Marco, Maria G. Leggio, Alessandra Solida, Roberto Corra, Sandro Misciagna, Maria C. Silveri & Laura Petrosini. 1997. Cerebellum and procedural learning: Evidence from focal cerebellar lesions. *Brain* 120. 1753–1762.
- Nadeau, Stephen E. & Bruce Crosson. 1997. Subcortical aphasia. *Brain and Language* 58. 355–402.
- Paradis, Michel. 1987. *The assessment of bilingual aphasia*. Hillsdale, NJ: Lawrence Erlbaum.
- Paradis, Michel. 1999. *Valutazione dell'afasia bilingue*. Bologna: EMS.
- Petersen, Steven E. & Julie A. Fiez. 1993. The processing of single words studied with positron emission tomography. *Annual Review of Neuroscience* 16. 509–530.
- Petersen, Steven E., Peter T. Fox, Michael I. Posner, Mark Mintun & Marcus E. Raichle. 1989. Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience* 1. 153–170.
- Pickett, Emily R., Erin Kuniholm, Athanassios Protopapas, Joseph Friedman & Philip Lieberman. 1998. Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: A case study. *Neuropsychologia* 36. 173–188.
- Rey, André. 1967. *REY Reattivo della figura complessa*. Firenze: Organizzazioni Speciali.
- Riva, Daria & Cesare Giorgi. 2000. The cerebellum contributes to higher functions during development. Evidence from a series of children surgically treated for posterior fossa tumors. *Brain* 123. 1051–1061.
- Schmahmann, Jeremy D. 1997. Rediscovery of an early concept. *International Review of Neurobiology* 41. 3–27.
- Schmahmann, Jeremy D. & David Caplan. 2006. Cognition, emotion and the cerebellum. *Brain* 129. 290–292.
- Schmahmann, Jeremy D. & Janet C. Sherman. 1998. The cerebellar cognitive affective syndrome. *Brain* 121. 561–579.
- Scott, Richard B., Catherine J. Stoodley, Philip Anslow, Caroline Paul, John F. Stein, Elaine M. Sugden & Christopher D. Mitchell. 2001. Lateralized cognitive deficits in children following cerebellar lesions. *Developmental Medicine and Child Neurology* 43. 685–691.

- Shallice, Tim. 1982. Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Series B* 298. 199–209.
- Silveri, Maria Caterina, Maria G. Leggio & Marco Molinari. 1994. The cerebellum contributes to linguistic production: A case of agrammatism of speech following right cerebellar lesion. *Neurology* 44. 2047–2050.
- Silveri, Maria Caterina, Anna M. Di Betta, Valeria Filippini, Maria G. Leggio & Marco Molinari. 1998. Verbal short-term store-rehearsal system and the cerebellum. Evidence from a patient with a right cerebellar lesion. *Brain* 121. 2175–2187.
- Spinnler, Hans & Gianni Tognoni. 1987. Standardizzazione e taratura italiana di test neuropsicologici. *The Italian Journal of Neurological Sciences* Suppl 8. n. 6. Milano: Masson.
- Tavano, Alessandro, Paola de Fabritiis & Franco Fabbro. 2005. Contributo alla valutazione standardizzata dell'eloquio narrativo nei bambini. *Giornale di Neuropsichiatria dell'Età Evolutiva* 25(1). 48–64.
- Tavano, Alessandro, Franco Fabbro & Renato Borgatti. 2007. Language and social communication in children with cerebellar dysgenesis. *Folia Phoniatrica et Logopaedica* 59(4); 201–209.
- Tavano, Alessandro, Rita Grasso, Chiara Gagliardi, Fabio Triulzi, Nereo Bresolin, Franco Fabbro & Renato Borgatti. In press. Disorders of cognitive and affective development in cerebellar malformations. *Brain* 130.
- Tettamanti, Marco, Andrea Moro, Cristina Messa, Rosa M. Moresco, Giovanna Rizzo, Assunta Carpinelli, Mario Matarrese, Ferruccio Fazio & Daniela Perani. 2005. Basal ganglia and language: Phonology modulates dopaminergic release. *NeuroReport* 16. 397–401.
- Titomanlio, Luigi, Alfonso Romano & Ennio Del Giudice. 2005. Cerebellar agenesis. *Neurology* 64(6). E21.
- Trouillas, Paul, Toshio Takayanagi, Mark Hallett, Robert D. Currier, S. H. Subramony, Karl Wessel, Alan Bryer, Hans C. Diener, Steve Massaquoi, Christopher M. Gomez, Paula Coutinho, Mongi Ben Hamida, Giuseppe Campanella, Alessandro Filla, Lawrence Schut, Dagmara Timann, Jérôme Honnorat, Norbert Nighoghossian & Bala Manyam. 1997. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. *Journal of the Neurological Sciences* 145. 205–211.
- Ullman, Michael T. 2001. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition* 4. 105–122.
- Wechsler, David. 1997. *WAIS-R. Wechsler Adult Intelligence Scale – Revised*. Firenze: Organizzazioni Speciali.
- Zettin, Marina, Stefano F. Cappa, Adele D'Amico, Roberto Rago, Claudio Perino, Daniela Perani & Ferruccio Fazio. 1997. Agrammatic speech production after a right cerebellar haemorrhage. *Neurocase* 3. 375–380.

Ontologies as a cue for the metaphorical meaning of technical concepts

Helmar Gust, Kai-Uwe Kühnberger and Ute Schmid

Due to the rapid development of new technical devices, there is the need to find concise names for such tools. In order to give an explanation for the chosen names, we propose to interpret a large class of them as metaphorical expressions. Good examples are concepts taken from information technology. In order to compute the metaphorical meaning of such concepts we will use Heuristic-Driven Theory Projection (HDTP) extended by ontological knowledge guiding the corresponding computation. HDTP is a theory to compute metaphoric relations by generalizations and a productive transfer of knowledge from the source to the target domain. Our analysis shows that the productivity of metaphors can be modeled by the expansion of the ontological basis of HDTP.

1 Introduction

It is generally accepted that the modeling of metaphorical expressions of natural language in a formal system is a hard problem. Up to now there is no convincing automatic procedure that computes the meaning of metaphorical expressions in a broad variety of domains. On the other hand the close relationship between metaphors and analogies is striking (cf. Gentner et al. 2001). Modeling analogies was extensively examined in cognitive science and artificial intelligence at least during the last twenty-five years. Recently published, the monograph Gentner, Holyoak and Kokinov (2001) can be seen as a summary of important approaches towards a modeling of analogies. The classical analysis of analogical reasoning assumes a mapping of a source domain to a target domain by associating certain properties and relations between the two domains or the involved concepts. By loosening the idea of a mapping slightly, it is possible to conceptualize an analogy as the establishment of an analogical relation associating source and target. A natural hypothesis is that the mechanisms that can explain analogical reasoning are strongly related to mechanisms that allow the interpretation of metaphorical expressions.

In technical and scientific disciplines, there is the need to find concise names for newly invented or developed tools and concepts. For example, in IT glossaries, one can find many concepts which share the property of being lexicalized metaphors. Examples of such lexicalized metaphors are *computer*, *mouse*, *folder*, *backbone*, *menu*, *browser*, *buffer*, *virus*, *worm*, *daemon*, *printer* etc. The metaphorical meaning of several of these concepts is immediately transparent: In these relatively simple cases, the meaning can be established by a (more or less) straightforward transfer of knowledge from the source, i.e. the conventional meaning of the concept, to the target. Examples for such straightforwardly interpretable concepts are *mouse* or *printer*. It is intuitively clear that first, the computer mouse has a (more or less) similar shape compared to a real mouse, particularly because of the cable that can be associated with the tail of a real mouse and the sizes of the two entities and, second, because the degree of movability of a computer mouse and of a real mouse is high. Slightly more difficult are examples like *computer* or *daemon*.¹ In the latter examples, the amount of background knowledge that needs to be assumed is significantly higher in order to give an explanation of the usages of concepts like *computer* and *daemon* in the IT world.

Clearly, an analysis of the metaphorical meaning of technical concepts covers only a small part of the ongoing research concerned with metaphorical expressions. Nevertheless non-trivial insights can be gained with respect to the nature of metaphors. More generally, it is possible to explain in more detail the important aspects of metaphors in creative processes of human cognition, as well as the strong connection between natural language and cognitive abilities. By focusing on a controlled domain, like the domain of technical concepts, human creativity can better be analyzed in comparison to a very general analysis of arbitrary metaphors occurring in natural language. A further cognitive aspect that can be emphasized is the connection of the presented topic to categorization: It turns out that explaining the semantics of metaphors is strongly correlated to the task of categorizing unknown concepts into well-known classes.

In this study, we will use so-called Heuristic-Driven Theory Projection (HDTP) to generate an abstraction process that yields generalizations of given source and target domains (cf. Gust, Kühnberger & Schmid 2006). HDTP is a formal theory designed to model analogical reasoning and to compute metaphoric expressions without referring to similarity measures, structural descriptions represented in a meta-language, or a separated learning process. Rather structural descriptions are constructed and represented in the object language. Furthermore, learning processes occur as a side-effect of the computation. In order to make such computations less complex, the introduction of background knowledge using ontologies is

1. A daemon is a program running in the background and becoming active on specific events.

a standard technique quite often used in artificial intelligence. The task in this study is therefore to give ideas of how an integration of ontological knowledge into HDTP is possible. The single steps that will be developed in this chapter are summarized in the following list:²

- As input a conceptualization of the source and the target domain is given formulated in a many-sorted first-order language.
- An ontology O is used to order the involved concepts in a hierarchy and to code certain additional properties of such concepts that are important for the establishment of the metaphorical relation.
- A heuristic-driven generalization process applied to the source and the target domains yields a structural description (generalization) of source and target by associating facts and laws of the target domain with facts and laws of the source domain. This process is algorithmically guided by heuristics and the underlying ontology O .
- Remaining facts and laws of the source are projected (transferred) to the target and tested using an oracle (experiment) where a positive test yields acceptance and a negative test rejection.

We will model the emergence of the lexicalized metaphorical meaning of new concepts in technical domains by an analogical transfer that is extended by ontological knowledge. This ontological knowledge is guiding the appropriate computation. Some cognitive consequences of our modeling are: The possibility of metaphors to be used in a productive way is explained by the associations of concepts in hierarchical structures representing the meaning of the concepts. Second, the recognition of an appropriate interpretation of a metaphorical expression may require certain abstraction processes. Third, the lexicalization of the metaphorical construction of a meaning may lead to a state where the reconstruction is no longer necessary. Therefore the metaphorical relation may vanish (like in the case of *computer*).

This chapter has the following structure: First, we will roughly discuss some of the major difficulties of modeling metaphorical expressions in natural language. Then an overview of the formal framework of Heuristic-Driven Theory Projection (HDTP) extended by ontological knowledge will be given, which will be applied to compute metaphorical relations. The focus of applying HDTP will be on concepts taken from technical domains, in particular, we will use the metaphoric relation between *mouse* and *computer mouse* as a running example. Finally some concluding remarks will finish this chapter.

2. Representations of examples can be found in Gust and Kühnberger (2006), Gust, Kühnberger and Schmid (2006), Schmid et al. (2003a), and Gust, Kühnberger and Schmid (2003b). A Prolog implementation (without an explicit ontology) can be found in Gust, Kühnberger and Schmid (2003a).

2 The problem of metaphorical meaning

The focus of this chapter will be to examine the metaphorical meaning of technical concepts. Nevertheless some general remarks concerning interpretations of metaphorical expressions occurring in natural language are worth discussing briefly. Consider the following examples:

- (1) *A cold warrior.*
- (2) *Lawyers are sharks.*
- (3) *Gills are the lungs of fish.*
- (4) *His eyes are cold as burning fire.*
- (5) *Big apple* [in order to denote New York].

If metaphorical expressions like Examples (1)-(5) are used, different types of metaphors are instantiated. Whereas Example (1) is a metaphor that uses an adjective in order to express a property of a *warrior* (although the property to be *cold* cannot be (literally) applied to a *warrior*), a classical identification of two incomparable concepts is represented in Example (2). Example (3) is similar to a proportional analogy (cf. Indurkha 1992; Schmid et al. 2003) and Example (4) represents a property assigned to an opposing concept. Finally, Example (5) is a lexicalized metaphor (idiom). The given examples show that metaphors can have quite different forms, not only with respect to the syntactic structure, but also with respect to source and target concepts that are associated with one another by the metaphorical transfer.

Certain types of expressions are strongly related to metaphors. For example, idioms are usually considered as lexicalized and non-productive metaphors, i.e. they are already conventionally interpreted. There are many examples of idioms, because many types of colloquial expressions can be subsumed under this concept:

- (6) *It rains cats and dogs.*
- (7) *Jim brachte seinen Chef auf die Palme.*
Jim bring-PAST his boss on the palm
'Jim drove his boss crazy.'

Clearly, metaphors used in technical domains like *computer* or *mouse* are strongly related to idioms, because they are lexicalized and therefore conventionally interpreted. Other language expressions that are connected to metaphors are examples of irony and similes mostly because of their non-conventional interpretation. Furthermore, certain concepts are unconventionally used in natural language and are sometimes related to metaphorical expressions. For example, if Tom says to Jim:

My fish is in the living room – referring to the wooden fish he bought in Singapore – then this non-conventional usage of the concept fish requires a non-conventional interpretation connected with a certain transfer of properties from real fish to the wooden fish, e.g. the shape of fish.

A further aspect concerns the domain of discourse. As a matter of fact, metaphors seem to occur in nearly every domain of natural language discourse. A general account for an analysis of metaphors would therefore require an enormous knowledge base in which large parts of world knowledge need to be coded. This is different if we restrict our attention to technical concepts. In most cases, we are able to significantly restrict the possible domain of discourse, because usually the corresponding domains are most often clearly specified and clearly distinguishable from non-relevant fields.

It is usually assumed that metaphors are directed, i.e. a sentence like *Lawyers are sharks* may be used as a metaphor, but *Sharks are lawyers* cannot be used as a metaphor in a similar way.³ In this example, there are identifiable source and target concepts which therefore seem to play different roles in the analogy making process. This in turn prefers a modeling in which the source and target domains are clearly separated. Some authors question this idea (cf. Fauconnier & Turner 1998) and try to explain metaphoric interpretations as a cognitive blending process and not a directed mapping (or a directed relation) from the source to the target. The blending process does not seem to be sufficiently defined and is highly dependent on less clear concepts like “context” and “blending space”. Therefore we use the classical idea of establishing a metaphorical relation between source and target in order to obtain an analysis of metaphorical expressions.

In total, a number of problems occur if we try to model metaphors in natural language in a formal system. Here is a summary of some of the important problems:

- Metaphors of natural language are highly context-sensitive.
- Metaphors do not occur in clearly and precisely defined domains, but can associate and connect rather different domains.
- Metaphors do not have a general form but can vary significantly with respect to their specific form.
- Metaphors are related to idioms, similes, non-conventional usage of concepts etc. and there is no clear distinction between these categories.

3. It seems to be less clear whether the directedness condition holds for all types of metaphors in a strong sense. Simply because of the fact that in an appropriate context nearly every grammatical natural language sentence can be interpreted, the direction can be inverted by changing the context appropriately. A sentence like *Sharks are lawyers* can have a reasonable interpretation in a fable where the characters are animals who are assigned certain professions taken from the real world.

By restricting metaphoric expressions to the domain of technical concepts the mentioned problems do not play a similarly important role. One reason for this is that metaphors taken from technical fields are less complex with respect to their contextual embedding in comparison to metaphors occurring in, for example, poetic texts. If they occur newly in a technical domain, then the domain is relatively precisely determined in contrast to the domains of metaphors occurring in poetic contexts.

Metaphors show in another respect an important cognitive feature. They can be used for learning tasks. Imagine a situation where a teacher is teaching elementary physics to high-school students. A classically used metaphor to give students a basic understanding of electrons and their behavior is the following sentence:

(8) *Electrons are the planets of the atom.*

Usually most students immediately understand (8) without any pre-knowledge of the atom model: Electrons revolve around the nucleus, showing a quite similar behavior to planets of a solar system which revolve around the sun. The important fact is that metaphors can be used to produce an intuitive understanding of situations or facts of the world that were previously not conceptualized. Therefore, learning is potentially mediated by the understanding of metaphors.⁴ It is no exaggeration to call analogical and metaphorical reasoning one of the most important cognitive mechanisms for learning new situations and new concepts, in particular, if there is no or only a very weak form of pre-knowledge available. Notice that classical learning theories usually require a more or less large number of examples, exemplars, or instances that can be used for the establishment of a general law (inductive reasoning, case-based reasoning etc.). But in many situations of human learning, such a large number of examples is not available. Therefore an explanation of learning needs to take analogies and metaphors seriously into account. This can be seen as one reason why metaphors and analogies play a crucial role in understanding human cognition. Taking additionally into account the relation between metaphors and categorization and the relation between natural language and cognitive abilities in general, the importance of a theory of metaphors for human cognition cannot be underestimated.

4. Clearly, the idea that electrons revolve around the nucleus is far from being the correct description of the situation; at least it is far from being in accordance to the current physical theories. In our context it is just an example for how metaphors can be used for learning tasks.

3 Heuristic-Driven Theory Projection and ontological knowledge

3.1 Heuristic-Driven Theory Projection

Heuristic-Driven Theory Projection (HDTP) is an algorithmic approach to model analogical reasoning.⁵ In Gust, Kühnberger and Schmid (2006) the syntactic, semantic, and algorithmic foundations were developed. The framework was applied to different domains like classical analogies from naïve qualitative physics or analogies occurring in comparisons of different political systems. In the following we will sketch the basic ideas of HDTP.

Let us first consider two examples of metaphors:

- (9) *Heat flows from a warm medium to a cold medium.*
- (10) *Current is the water in an electric circuit.*

In Figure 1, two graphical representations of the metaphoric expressions (9) and (10) are given. The left diagram represents the correspondence between water-flow from a beaker to a vial and heat-flow from a cup of warm coffee to a Beryllium cube. Notice that *heat* is not treated as a predefined concept in the target domain, but is introduced as an abstract concept by a sentence like (9) corresponding to something that can flow in the source domain (e.g. a fluid like water). The right diagram depicts an analogy between water-flow in a water-pipe system and the flow of current in an electric circuit.

Syntactically, HDTP is defined on the basis of a many-sorted first-order language. First-order logic is used in order to guarantee the necessary expressive power of the account. An important assumption is that analogical reasoning (and therefore the analysis of metaphorical expressions) crucially contains a generalization (or abstraction) process. In other words, the identification of common properties or relations is represented by a generalization of the input of source and target. Formally this can be modeled by an extension of the so-called theory of anti-unification (cf. Plotkin 1970), a mathematically sound account describing the possibility of generalizing terms of a given language using substitutions. More precisely, an anti-unification of two terms t_1 and t_2 can be interpreted as finding a generalization (or structural description) t (of t_1 and t_2) together with two substitutions Θ_1 and Θ_2 , such that $t\Theta_1 = t_1$ and $t\Theta_2 = t_2$. Because there are usually many possible generalizations, anti-unification tries to find the most specific one. An example should make this idea clear. Assume two terms $t_1 = f(x, b, c)$ and $t_2 = f(a, y, c)$

5. The account stands in relation to other computational accounts towards modeling analogies and metaphors like Gentner (1983) or Forbus, Gentner and Law (1995).

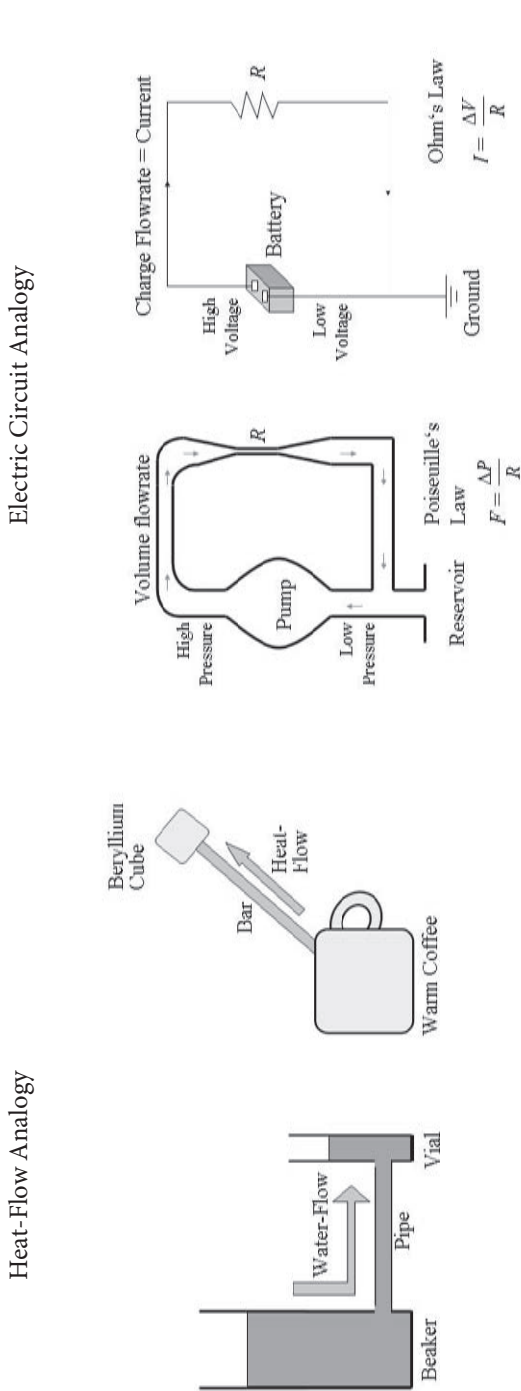


Figure 1. Representation of the famous heat-flow analogy on the left side and the analogy between water-flow and the flow of current in an electric circuit on the right side. The electric circuit analogy is a simplified version of a representation found on the *HyperPhysics* webpage of Georgia State University: <http://hyperphysics.phy-astr.gsu.edu/hphys.html>

are given.⁶ Generalizations are, for example, the terms $t = f(x,y,c)$ and $t' = f(x,y,z)$ together with their corresponding substitutions. But t is more specific than t' , because the substitution $\Theta = \{z \leftarrow c\}$ applied to t' yields: $\Theta t' = t$. Most specific generalizations of two terms are commonly called anti-instances.

Table 1. A conceptualization of the source and target domains of the heat-flow example

Source	Target
Sorts: <i>real, massterm, object, time</i>	Sorts: <i>real, massterm, object, time</i>
Entities: <i>vial: object</i> <i>beaker: object</i> <i>water: massterm</i> <i>pipe: object</i>	Entities: <i>coffee: massterm</i> <i>cup: object</i> <i>b_cube: object</i> <i>bar: object</i>
Functions: <i>observable height: object $\times$ time $\rightarrow$ real $\times$ {cm}</i> <i>observable footprint: object $\times$ time $\rightarrow$ real $\times$ {cm²}</i> <i>operator in: massterm $\times$ object $\rightarrow$ object</i> <i>theoretical volume: object $\times$ time $\rightarrow$ real $\times$ {cm³}</i>	Functions: <i>observable temperature: object $\times$ time $\rightarrow$ real $\times$ {C}</i> <i>operator in: massterm $\times$ object $\rightarrow$ object</i>
Functional definitions: <i>volume(x,t) := footprint(x,t) $\cdot$ height(x,t)</i>	Facts: <i>connected(coffee in cup, b_cube, bar)</i> <i>liquid(coffee)</i>
Facts: <i>connected(beaker, vial, pipe)</i> <i>liquid(water)</i> $\forall t1: \text{time}, t2: \text{time}$: <i>footprint(beaker, t1) > footprint(vial, t1) $\wedge$</i> <i>footprint(beaker, t1) = footprint(beaker, t2) $\wedge$</i> <i>footprint(vial, t1) = footprint(vial, t2)</i>	Experiment: $\forall t1: \text{time}, t2: \text{time}: t2 > t1 \wedge$ <i>temperature(coffee in cup, t1) ></i> <i>temperature(b_cube, t1)</i> $\rightarrow$ <i>temperature(coffee in cup, t2) <</i> <i>temperature(coffee in cup, t1) $\wedge$</i> <i>temperature(b_cube, t2) ></i> <i>temperature(b_cube, t1)</i>
Laws: $\forall t1: \text{time}, t2: \text{time}: t2 > t1 \wedge$ <i>height(water in beaker, t1) > height(water in vial, t1)</i> $\rightarrow$ <i>height(water in beaker, t1) > height(water in beaker, t2) $\wedge$</i> <i>volume(water in beaker, t1) - volume(water in beaker, t2)</i> <i>= volume(water in vial, t2) - volume(water in vial, t1)</i>	

In order to use anti-unification for the analysis of metaphorical expressions, an extension of anti-unification is needed, because not only terms but also complex

6. As usual we assume that variables are denoted by x, y, z and constants by a, b, c .

formulas, in fact whole theories of domains need to be generalized. This becomes clear by considering a possible input that is modeling the heat-flow situation of Figure 1. Such a possible input is depicted in Table 1. The input of the source domain is given by a set of axioms, specifying facts and rules of the source domain, i.e. the water-flow domain. The same holds for the input of the target domain, i.e. the domain where heat flows from the cup of warm coffee to the Beryllium cube. Notice that the proposed conceptualization in Table 1 is just a possible one. In different situations and with respect to different background knowledge, the facts and rules of the two domains can and must be different. In our case, we assume that rather spelled-out knowledge about the source domain is available.⁷

Usually it is assumed that the source domain corresponds to a rich conceptualization whereas the target domain is rather coarsely specified. This is motivated by the fact that knowledge about properties and relations in the source should be mapped to the target and not vice versa, in other words, if the target domain were well-established, it would not be necessary to make a transfer from the source to the target, because the necessary knowledge is already covered. For example, in the heat-flow metaphor (9) the concept *heat* is not present in the target domain, since it is not directly observable. Rather *heat* is projected to the target domain as a counterpart of water from the source domain.

3.2 The algorithm HDTP-A

HDTP is an algorithmic account that produces a generalization of the two given domain theories. This generalization can be interpreted as a structural description of the input theories. The algorithm HDTP-A is intuitively described in Figure 2. A precise specification of HDTP-A is developed in Gust, Kühnberger and Schmid (2006). An implementation of the algorithm can be found in Gust, Kühnberger and Schmid (2003a).

The input is given by a theory Th_s specifying the source domain and a theory Th_t specifying the target domain. Both domains are described in many-sorted first-order logical languages. The algorithm chooses an axiom A_{t1} from the target according to a given heuristics h (for example, a heuristics that orders axioms with respect to their complexity) and searches for an axiom A_{s1} from the source domain that can be generalized with A_{t1} . Again a heuristics h' is used to find an optimal

7. Clearly, it is unrealistic to assume that a state of the art system can handle comprehensive background knowledge in a way humans are able to do. Our modeling here is *a priori* restricted to the knowledge we think is necessary to establish the metaphorical relations. We do not claim to solve the general problem of selecting the relevant parts from world knowledge for understanding arbitrary metaphors.

generalization together with the corresponding substitutions.⁸ Inductively the generalized axiom A_{G_i} is added to the generalized theory Th_G . This process is iterated until all axioms in the target domain are generalized. Finally remaining axioms from the source are projected to the target and tested using an oracle, i.e. a module that contains knowledge about observable data of the world. Oracles are used to model experiments of, for example, a physicist. If the projected axiom is logically consistent and consistent with the observable data in the domain, it is added to the target. If it is either logically inconsistent with the given theory or observable data is contradicting the transferred axiom, the axiom is rejected.

Input: A theory Th_s of the source domain and a theory Th_t of the target domain represented in a many-sorted predicate logic language.

Output: A generalized theory Th_G such that the input theories Th_s and Th_t can be reestablished by substitutions.

Selection and generalization of fact and rules

Select an axiom from the target domain (according to a heuristics h)

Select an axiom from the source domain and construct a generalization
(together with corresponding substitutions)

Optimize the generalization with respect to a given heuristics h'

Update the generalized theory with respect to the result of this process

Transfer (project) facts of the source domain to the target domain provided they are not generalized yet

Test (using the oracle) whether the transfer is consistent with the target theory

Figure 2. An intuitive description of the Algorithm HDTP-A omitting formal details. A precise specification of this algorithm can be found in Gust, Kühnberger and Schmid (2006)

By applying the algorithm HDTP-A a generalized theory can be established. This theory connects convincingly corresponding concepts like the connection between beaker and vial and the connection between the cup of coffee and the beryllium cube (cf. Figure 1). The difficult aspect in this example is the introduction of a concept *heat*: Whereas the flow of water is observable in the source domain, this is not true for the corresponding flow of heat in the target domain. Only the temperature of the warm coffee and the temperature of the beryllium cube can be measured but not the heat-flow. Table 2 shows a simplified solution to the heat-flow example, summarizing the simple associations of the example:

8. There are many possibilities to define appropriate heuristics for the generalization process. A discussion of a variety of heuristics and the advantages and disadvantages of the different choices can be found in Gust, Kühnberger and Schmid (2006).

Table 2. A solution of the heat-flow example

Source	Target	A
(1) connected(beaker,vial,pipe)	connected(coffee in cup,b_cube,bar)	connected(A,B,C)
(2) height(water in beaker,t1) > height(water in vial,t1)	temperature(coffee in cup,t1) > temperature(b_cube,t1)	T(A, t1) > T(B, t1)
(3) height(water in beaker,t1) > height(water in beaker,t2)	temperature(coffee in cup,t1) > temperature(coffee in cup,t2)	T(A,t1) > T(A,t2)
(4) height(water in vial,t2) > height(water in vial,t1)	temperature(b_cube,t2) > temperature(b_cube,t1)	T(B,t2) > T(B,t1)

Notice that the algorithm correctly identifies the beaker with the coffee in the cup and the vial with the beryllium cube. Furthermore, the heights of the water levels in the source domain are correctly associated with the temperature of the coffee and the temperature of beryllium cube, respectively. We summarize the corresponding substitutions Θ_1 and Θ_2 in the following list:

$A \rightarrow \text{beaker} / \text{coffee in cup}$
 $B \rightarrow \text{vial} / \text{b_cube}$
 $C \rightarrow \text{pipe} / \text{bar}$
 $D \rightarrow \text{water} / \text{coffee}$
 $T \rightarrow \lambda x, t: \text{height}(\text{water in } x, t) / \text{temperature}$

The example – although seemingly simple – has a relatively complicated aspect: The system associates an abstract property $\lambda x, t: \text{height}(\text{water in } x, t)$ with *temperature*. The concept *heat* must be introduced as a counterpart of *water* in the target domain by projecting the structure of the λ -term above to the target domain by the following equation:

$$\text{temperature}(x, t) = \text{height}(\text{heat in } x, t)$$

3.3 Integrating ontological knowledge

In AI, there is an increasing interest in examining ontologies for a variety of applications ranging from the classification of documents and improvements of

efficiency of inference processes to the representation of knowledge in knowledge bases.⁹ A very important motivation for the introduction of ontologies is the fact that inference processes can be made more efficient. For example, a given ontology with a subsumption relation based on a many-sorted logic allows to restrict inferences to those rules that are in accordance to the sortal restrictions. A significant reduction of the computational load is possible in such cases. Classical subsumption architectures can have different forms ranging from lattice-like structures (Ganter & Wille 1996; Peylo 2002) to relatively unrestricted semantic networks (Peters & Shrobe 2003).

It turns out that the expressive power of many-sorted first-order logic, i.e. a logic that is augmented by ontological knowledge, is not increased in comparison to ordinary first-order logic (cf. Ebbinghaus & Flum 1999). The reason for this is that ontological knowledge can be reformulated by ordinary first-order logic. For example, if we have a relation *car is_a vehicle* between the concepts *car* and *vehicle*, then we can represent this knowledge by the (one-sorted) first-order formula $\forall x(car(x) \rightarrow vehicle(x))$. The question arises why it is reasonable to make such an extension at all. A reason for augmenting a conceptualization by additional ontological knowledge is given below.

A further remark concerns the way ontological knowledge can be introduced to a system. In principal, there are two possibilities. The first possibility is to translate ontological knowledge to a (many-sorted) first-order language, i.e. representing the ontological knowledge directly on the logical level. There is a standard way to represent this knowledge in a first-order language. The second possibility is to represent the important associations between concepts together with relations between concepts on a different level, namely on the ontological level. An increase of efficiency is reached in both accounts due to sortal restrictions of possible inferences in the first account and additionally inheritance properties of classes in the second. In the second account, there is a further distinction between the logical level, where, for example, the laws of logic are represented, and an ontological level where knowledge about the world is coded.

We need to make the concepts we used so far intuitively more precise. Concerning ontologies there is no generally accepted definition. Usually it is assumed that an ontology consists of a set of concepts *C* which are hierarchically ordered induced by a subsumption relation and certain additional relations between these concepts. We will define ontologies very closely to this idea (cf. Stumme & Maedche 2001):

9. A list of the huge variety of different knowledge-based systems and ontology projects can be found on the internet at <http://www.cs.utexas.edu/users/mfkb/related.html>.

Definition 1: An ontology is a 4-tuple $O = \langle C, is_a, R, \sigma \rangle$ where C is a set of concepts, $is_a \subseteq C \times C$ is a partial order relation on C , R is a set of relations, and σ is a function that assigns to each relation an arity.

In the definition above, the is_a relation corresponds to the subsumption relation mentioned in the text above. Other relations are necessary, in particular, if we want to code knowledge of a whole theory on the ontological level: We do not only need concepts but also attributes, values of attributes, a $part_of$ relation (in order to describe entities) and other relations. Clearly the precise choice of relations occurring in an ontology is dependent on the domain and the problem that is modeled.

A prototypical example for the reduction of the computational load of a problem by introducing ontological knowledge is the steamroller problem, well-known as a benchmark problem for theorem proving (cf. Stickel 1986). We remind the reader of the original steamroller problem which can be represented in a nice natural language paraphrase.

- *Wolves, foxes, birds, caterpillars, and snails are animals, and there are some of each of them.*
- *Also there are some grains, and grains are plants.*
- *Every animal either likes to eat all plants or all animals much smaller than itself that like to eat some plants.*
- *Caterpillars and snails are much smaller than birds, which are much smaller than foxes, which in turn are much smaller than wolves.*
- *Wolves do not like to eat foxes or grains, while birds like to eat caterpillars but not snails.*
- *Caterpillars and snails like to eat some plants.*
- *Prove: There is an animal that likes to eat a grain eating animal.*

Although the paraphrase seems to be rather simple, a flat first-order logical representation is fairly complex: It contains approximately 21 axioms. Moreover it is not obvious how to represent the problem in Horn logic, i.e. a sublogic of first-order predicate logic where every rule has the form $p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow q$ (for p_i and q being literals). This is the reason that standard theorem provers have problems in proving the claim. A significant simplification can be achieved by introducing an ontology separating knowledge about the relation between the occurring sorts from the rest of the domain knowledge. Therefore the ontological knowledge can be used to restrict the resolution process (cf. Walter 1985). A straightforward ontological structure for the steamroller is depicted in Figure 3. Notice that the underlying structure is a lattice, i.e. every pair of sorts can be generalized and specialized. For meaningful generalizations and specializations not all nodes of the lattice are good candidates: Because the extension of the bottom element is empty,

specialization of this node makes no sense. On the other hand, generalizations of the top node are inappropriate since this node cannot be used to restrict possible arguments occurring in predicates. Interesting generalizations are generalizations where certain animal-concepts like *fox* and *snail* are generalized to *animal*. In order to restrict possible generalizations we need to identify an upper structure that makes arbitrary generalizations impossible. A natural idea is to block generalizations that do not have a common superconcept within a set of maximal generalizable concepts (the leaves of the upper structure).

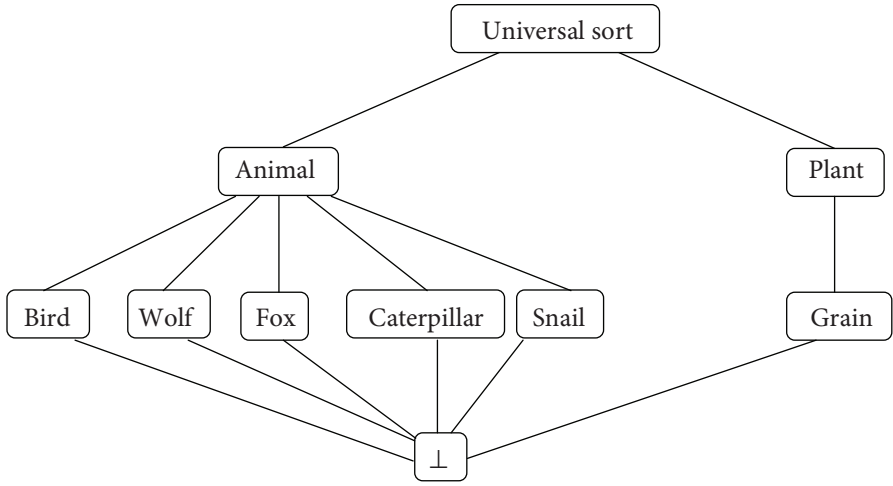


Figure 3. The steamroller ontology specifying the concepts involved in the problem and their corresponding relations with respect to the subsumption relation

We need to make the idea of an upper structure of an ontology precise that can be used for restricting possible generalizations.

Definition 2: Assume an ontology $O = \langle C, is_a, R, \sigma \rangle$ and a set $C_{gen} \subseteq C$ are given. The set of pairs of generalizable expressions O' relative to a function S mapping terms to sorts is defined as follows

- (i) If t_1 and t_2 are terms, then $\langle t_1, t_2 \rangle \in O'$ iff there exists $c \in C_{gen}$ with $S(t_1)$ is_a c and $S(t_2)$ is_a c .
- (ii) If ϕ and ψ are complex, then $\langle \phi, \psi \rangle \in O'$ iff all corresponding subterms can be generalized according to the first condition.

The intended interpretation of the generalizable expressions C_{gen} is a set of leaves of an upper structure. Definition 2 distinguishes between terms of a logical language, specified in condition (i), and complex formulas, specified in condition (ii).

The reason is that in HDTP generalizations of whole theories and not just terms play an important role in order to model metaphorical relations. Hence, we need to take into account both, complex rules as well as terms.

As a matter of fact, it turns out that the complexity of the steamroller problem can significantly be simplified by adding ontological knowledge in form of sortal restrictions as described above. Adding ontological knowledge reduces the number of axioms and the computational load significantly (cf. Walter 1985). This can then be used for efficient theorem provers. In a similar fashion an enrichment of HDTP with ontological knowledge can reduce the complexity of the computation of metaphorical relations. Furthermore, ontological knowledge provides a cognitive cue for analyzing the metaphorical relation. This idea will be developed in the following section.

4 The application of HDTP to technical concepts

4.1 Adding ontological knowledge to HDTP: Mice and mice

We want to start our considerations concerning the metaphorical meaning of technical concepts with a relatively simple example: The metaphor *mouse* in order to denote a computer mouse. Obviously there are certain similarities between a real mouse and a computer mouse:

- Both have a quite similar shape.
- Both have a relatively similar size.
- Both can move, respectively, can be moved very rapidly.
- Etc.

Nevertheless, most properties of the real mouse and the computer mouse are different: Whereas a real mouse is a biological organism, a computer mouse is a technical device designed for a special purpose, namely to be able to point at buttons, fields etc. on a screen. Furthermore, a real mouse has many negative connotations, e.g. that it can carry viruses and bacteria etc., whereas a computer mouse cannot transport biological viruses. In interpreting a metaphor like *computer mouse* the interpreter associates certain properties of a real mouse with (or transfers certain properties to) a computer mouse. Table 3 summarizes a possible conceptualization of the situation.

As usual the conceptualization distinguishes between sorts, objects, observables, facts, and rules. The mouse has parts like tail, body, eyes etc., whereas the computer mouse is made of a cable, a body, button(s) etc. All these parts stand in the *part_of* relation to the whole object *mouse*. Even the properties are quite similar:

Table 3. A possible conceptualization of a mouse and a (computer) mouse

Source: Mouse	Target: Computer Mouse (cm)
Sorts: <i>object, time, shape</i> , etc.	Sorts: <i>object, time, shape</i> , etc.
Objects: <i>tail, body, eyes, ears, head</i> , etc.	Objects: <i>cable, body, button</i> , etc.
Observables: $size_a$, $shape_a$, $movability_a$, etc.	Observables: $size_a$, $shape_a$, $movability_a$, etc.
Facts: $movability_a(mouse) = high$, $shape_a(tail) = longish$, $part_of(tail, mouse)$, $size_a(mouse) = small$, etc.	Facts: $movability_a(cm) = high$, $handy(cm)$, $part_of(cable, cm)$, $shape_a(cable) = longish$, etc.
Rules: –	Rules: $handy(x) \rightarrow size_a(x) = small$

Both, a mouse as well as a computer mouse are highly movable and longish in shape. From the rule $handy(x) \rightarrow size_a(x) = small$ it can be deduced that computer mice are small-sized. This in turn allows the association of $size_a(mouse) = small$ and $size_a(cm) = small$.¹⁰ Clearly further properties can become important dependent on the task that should be modeled by the account.

A conceptualization such as the one represented in Table 3 makes certain assumptions with respect to the domain of discourse. It should be noted that pure facts and rules cannot be sufficient to identify a unique model.¹¹ For example, consider as a domain the natural numbers and the attribute values *longish* and *small*: Then the number 1 clearly satisfies the property to be small, whereas a large natural number can be interpreted as longish. Therefore it is necessary to associate ontological knowledge with domains in order to restrict the possible objects that can occur in the reasoning process. Hence further constraints are imposed on the available background knowledge necessary for an appropriate generalization process.

Now we add ontological knowledge to the description of the source and target domains. Figure 4 depicts a hierarchical dependency of certain concepts that may play a role in a reasoning process. As a matter of fact the conceptualization is somehow arbitrarily chosen and alternatives, for example, a more fine-grained distinction of the *is_a* relation with respect to artifacts and mammals or a different treatment of abstract concepts are conceivable. The interesting feature of coding the ontological level directly in a hierarchical structure is the possibility to use the ontological relations in the matching process. In other words, the association of, for example, *body* (as a part of the mouse) and *body* (as a part of the computer

10. $size_a$ and $shape_a$ refer to the attributes which have values in *size* and *shape*, respectively. We will not discuss the specific problems of the semantics of dimensional adjectives here (cf. Bierwisch & Lang 1989), but rather use a naïve representation for them.

11. This observation corresponds to the fact that every logically specified theory has infinitely many different models.

mouse) can be processed directly on the ontology, without the need of explicit logical inferences. This makes the generalization process more visible in comparison to the ontology-free representation. Several kinds of arrows and relations are used in the representation of the ontological knowledge: Ordinary lines denote the *is_a* relation, double lines the *part_of* relation, double arrows denote the association generated by the abstraction process. Furthermore, ordinary arrows denote the *has_property* relation and dashed lines a value of an abstract (observable) concept. As an example we mention that the abstract property *size* can have the values *small* and *medium*. Notice that the node “???” is intended to represent the technical device of a computer mouse.

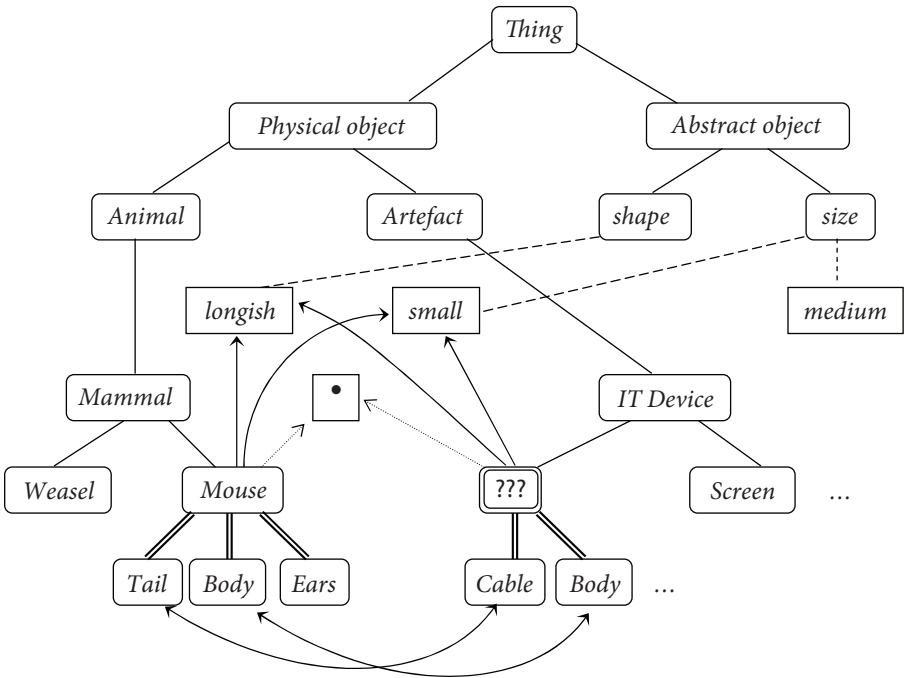


Figure 4. A possible ontology specifying the hierarchical dependencies of mouse and computer mouse. The ordinary line denotes the *is_a* relation, the double line the *part_of* relation, and the double arrows at the bottom denote associations that can be derived syntactically by the generalization process. The meaning of the ordinary arrows is *has_property* and the meaning of the dotted arrows is the attribute-value relation

How can an understanding of the interpretation of a technical concept like *computer mouse* look like? The relevant features are the values *small* for *size_a* and *longish*

for $shape_a$. If a system can match the values *small* and *longish* to both, the *mouse* and the *computer mouse*, the association of these two concepts is straightforward. Then, further associations, for example, the one matching of *tail* and *cable*, are a simple extension of the already existing association. In this way, a metaphorical relation can be established and an understanding of the technical concept is possible.

Attribute values like *longish* and *small* seem to be crucial in order to establish the association of the two concepts *mouse* and *computer mouse*. Further properties that are mentioned in Table 3 – for example, $handy(x)$ or $movability_a(x) = high$ – are not represented in Figure 4 due to complexity reasons in this diagrammatic representation. But obviously these missing concepts can also be straightforwardly introduced into the graph. These properties can clearly be dependent on the context. Interpreting metaphorically an IT device as a mouse requires different properties than interpreting a particular soft toy as a mouse. Therefore functionality aspects like in the case of a computer mouse do not play such a role in the soft toy example. Here most important are aspects that focus on the mouse-like shape.

In the diagram, there is also a node “•” connected with the nodes *mouse* and “???”. Such a node can be used to represent an abstract property *mouse-like* and can function as a complex property containing several different properties like $size_a(x) = small$, $shape_a(x) = longish$, $handy(x)$ etc. as well as more spelled-out information concerning the shape of a mouse. A natural way to interpret this node is to associate it with a generalization of the respective properties.

It seems to be the case that the representation of theory-like knowledge in an ontology is a rather complicated issue, because not only hierarchical aspects like the subsumption relation need to be modeled, but also relatively arbitrary relations that come together with the axiomatization of the source and the target domain. Although this is not surprising, there is a great advantage of this representation: The dependencies and the relations between the involved concepts become immediately visible and can be read off from the diagrammatic representation. Clearly this also has its limits: In a hierarchical representation such dependencies reach rapidly a point where the complexity of the graph prevents a simple and straightforward interpretation and extraction of knowledge.

In Definition 2 we specified an upper structure that is intended to restrict possible generalizations in a hierarchical ontology. In the structure depicted in Figure 4, it seems to be obvious that generalizations of the node *physical object* and *abstract object* should be prevented by the account, because they do not result in concepts that can be reasonably used in the algorithm. Other generalizations are not restricted by the upper structure; hence we do not restrict other possible generalizations.

4.2 Some consequences for cognitive science

We think that ontological knowledge is *the* important cue for being able to productively interpret metaphors that were never heard before, as well as learning new situations by metaphoric expressions. These cues can be seen as a possibility to explain the highly developed adaptive behavior of humans. If analogical and metaphorical inferences can be easily performed, then the adaptation of an agent to a rapidly changing environment is not surprising as it seems to be on first sight. Using the establishment of metaphorical relations a small number of available conceptualizations is sufficient for a productive transfer to new situations. Such a transfer can provide a hint for the agent to act appropriately in new and unfamiliar situations.

There are many examples for metaphorical (and therefore concise) names of technical concepts: Names for IT devices are a rich field for such examples. But there are other less obvious examples. Consider, for example, the name *Beluga* for the transport airplane of Airbus Industries. Without prior knowledge there is no chance to give this name any reasonable meaning. But if we know how the animal *Beluga Whale* looks, then the name for the airplane is striking. In the *Beluga* case, the metaphor is mainly based on the shape of the two objects. For the expert, it may be the case that more properties are involved that can be used to establish a meaningful metaphorical relation.

We think that our considerations have shown that the following hypothesis can be established with respect to technical concepts:

If a concise name for a new technical concept must be found, a natural possibility to find such a name is to establish a metaphor and a metaphorical relation. Furthermore, ontological knowledge about sortal restrictions, hierarchical dependencies of involved concepts, and further relations that hold between concepts, seem to be crucial in order to be able to interpret such names of technical concepts.

From the analysis above it seems to be essential that some specific properties are constitutive for establishing metaphorical relations. In order to generate these specific properties an abstraction process is necessary for which those properties are generic. Because the connections and relevant properties may be initially only in the mind of the inventor of the IT device, it is important to detect these properties over and over again in an interpretation process. Clearly, the interpretation is strongly dependent on the intention of the inventor. For example, it can be the case that the functional aspect of the relation between source and target is stressed or that a shape-like property play an important role. An example for the functional relation between source and target would be a concept like *folder* used in current operating systems. Conventionally a folder is a physical object in which files, pa-

pers, certificates etc. can be stored and ordered. In the development of operating systems it was the case that the functional role of folders was implemented in a software program: Folders of an operating system are used to store and order data, i.e. files with arbitrary features. From a function perspective, this is the same role a conventional folder in an administration office has. Nevertheless there is no similarity between the conventional folder and a (computer) folder, at least at the physical or structural level.

5 Concluding remarks and further research

In this chapter, we showed how Heuristic-Driven Theory Projection (HDTP) can be extended by ontological knowledge in order to give an explanation of metaphorical expressions used for naming technical concepts. The mechanism codes the relevant parts of the background knowledge in an ontological structure. In this ontological structure the association between the source and the target domain is immediately visible. Furthermore the ontological level is separated from the ordinary reasoning process. Because of the flexibility of HDTP it is straightforward to extend this approach by ontological background knowledge.

From a general cognitive perspective concise names for technical concepts and devices are naturally explained by a metaphorical relation between the standard usage of the concepts and the technical concepts. Being confronted with a technical concept that was never conceptualized before, it seems to be the only possibility to assume that a metaphorical transfer from the conventional to the non-conventional domain takes place. This in turn seems just to be a special case for a more general cognitive principle where metaphorical and analogical relations are *the* mechanism for learning and adapting to new situations. In particular, this mechanism is not only limited to natural language expressions, but can be applied to other cognitive abilities as well. For example, certain aspects of perception, action, learning, and memory seem also to be based on analogical reasoning.

Future research will focus on the implementation of the dynamical change of ontological knowledge. Connected with this aspect are learning aspects which are important in establishing metaphorical relations. As a matter of fact appropriate theories about learning aspects and dynamic changes of ontologies are rather rare. In particular, it is important to find a way to learn and to generate ontological knowledge – a problem that arises in fields like document management, knowledge management, or semantic web applications as well.

References

- Bierwisch, Manfred & Ewald Lang. 1989. *Dimensional adjectives: Grammatical structure and conceptual interpretation* [Springer Series Language and Communication 26]. Berlin: Springer.
- Ebbinghaus, Heinz-Dieter & Jörg Flum. 1999. *Finite model theory*. Berlin: Springer.
- Fauconnier, Gilles & Mark Turner. 1998. Conceptual integration networks. *Cognitive Science* 22(2). 133–187.
- Forbus, Kenneth D., Dedre Gentner & Keith Law. 1995. MAC/FAC: A model of similarity-based retrieval. *Cognitive Science* 19(2). 141–205.
- Ganter, Bernhard & Rudolf Wille. 1996. *Formale Begriffsanalyse. Mathematische Grundlagen*. Berlin: Springer.
- Gentner, Dedre. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7. 155–170.
- Gentner, Dedre, Keith J. Holyoak & Boicho N. Kokinov. 2001. *The analogical mind. Perspectives from cognitive science*, Cambridge, MA: The MIT Press.
- Gentner, Dedre, Brian F. Bowdle, Phillip Wolff & Consuelo Boronat. 2001. Metaphor is like analogy. In Dedre Gentner, Keith J. Holyoak & Boicho N. Kokinov (eds.), 199–253.
- Gust, Helmar & Kai-Uwe Kühnberger. 2006. Explaining effective learning by analogical reasoning. In Ron Sun & Naomi Miyake (eds.), *28th Annual Conference of the Cognitive Science Society in cooperation with the 5th International Conference of the Cognitive Science (in the Asian-Pacific region)*, 1417–1422. Mahwah, NJ: Lawrence Erlbaum. <http://www.cogsci.rpi.edu/CSJarchive/proceedings/2006/docs/p1417.pdf>.
- Gust, Helmar, Kai-Uwe Kühnberger & Ute Schmid. 2003a. Anti-unification of axiomatic systems. Available at: <http://www.cogsci.uni-osnabrueck.de/~helmar/analogy/>.
- Gust, Helmar, Kai-Uwe Kühnberger & Ute Schmid. 2003b. Metaphors and anti-unification. *Twente Workshop On Language Technology: Algebraic Methods in Language Processing (AMILP'03)* 21. 111–123.
- Gust, Helmar, Kai-Uwe Kühnberger & Ute Schmid. 2006. Metaphors and Heuristic-Driven Theory Projection (HOTP). *Theoretical Computer Science* 354(1). 98–117.
- Indurkha, Bipin. 1992. *Metaphor and cognition*. Dordrecht: Kluwer.
- Peters, Stephen & Howard E. Shrobe. 2003. Using semantic networks for knowledge representation in an intelligent environment. *IEEE International Conference on Pervasive Computing and Communications (PerCom'03)* 1. 323–337.
- Peylo, Christoph. 2002. *Wissen und Wissensvermittlung im Kontext von internetbasierten intelligenten Lehr- und Lernumgebungen*. Berlin: Akademische Verlagsgesellschaft Aka GmbH.
- Plotkin, Gordon D. 1970. A note on inductive generalization. *Machine Intelligence* 5. 153–163.
- Schmid, Ute, Helmar Gust, Kai-Uwe Kühnberger & Jochen Burghardt. 2003. An algebraic framework for solving proportional and predictive analogies. In Franz Schmalhofer, Richard M. Young & Graham Katz (eds.), *European Conference on Cognitive Science (EuroCogSci03)*, 295–300. Mahwah, NJ: Lawrence Erlbaum.
- Stickel, Mark E. 1986. Schubert's steamroller problem: Formulation and solutions. *Journal of Automated Reasoning* 2(1). 89–101.
- Stumme, Gerd & Alexander Maedche. 2001. FCA-Merge: Bottom-up merging of ontologies. In Bernhard Nebel (ed.), *International Joint Conference on Artificial Intelligence (IJCAI 2001)* 17, 225–230. Seattle, WA: Morgan Kaufmann.
- Walter, Christoph. 1985. A mechanical solution of Schubert's Steamroller by many-sorted resolution. *Artificial Intelligence* 26. 217–224.

Anti-realist assumptions and challenges in philosophy of mind

Drew Khlentzos

The plausibility of Naturalism – the view that the mind is a purely natural phenomenon that can be explained scientifically (if explained at all) – is a hotly contested topic in cognitive science, as every philosopher and cognitive scientist knows. It is not widely recognised, however, that some of the more popular arguments against naturalism rest upon anti-realist metaphysical assumptions. This is a problem since the most plausible defences of naturalism presuppose a realist metaphysics. In this chapter, I shall chart one of these anti-realist assumptions and show how it features as a crucial premise in a leading anti-naturalist argument, the Knowledge Argument.

1 Introduction

Materialism is the thesis that all facts are physical facts. In particular, facts about conscious awareness are physical: No non-physical properties are involved in seeing colours or feeling pains, for example. Some think that materialism is flatly inconsistent with the data of conscious experience – we know simply by seeing colours or experiencing pains that the phenomenal properties directly presented to us in these conscious episodes could not possibly be physical, anti-materialists aver. For no physical property is essentially subjective or possessed of a characteristic appearance or feel as colour sensations or pain sensations are. Hence, even if we knew all the physical facts we still would not know what it is like to see red if we had never experienced any colour sensations before. This is the Knowledge Argument against materialism. Further, it seems as if there could be a complete physical duplicate of me who thought and acted exactly as I do but who lacked any sensory experience. This is the Zombie Argument against materialism. In this chapter I shall discuss the Knowledge Argument, leaving the Zombie Argument for another occasion.

Materialism seems most plausible to those who think of the human mind as a complex product of evolution: a marvellously intricate physical system produced by natural forces. If humans are a part of a physical world whose nature is only partially accessible to them then the world is, in an important sense, mind-independent. This is the view of realism that, in naturalistic form, provides the natural background metaphysics for materialism.¹ So much so that one might wonder whether anti-materialist arguments are not, at least to some extent, driven by anti-realist commitments? I think that this is indeed the case and it is the purpose of this study to demonstrate a highly influential anti-materialist argument, the Knowledge Argument, does in fact rest upon a crucial unacknowledged anti-realist assumption. The anti-realist assumption I have in mind is this:

Reality is circumscribed by our knowledge of it such that all truths are knowable.

I shall refer to this assumption as the “K=T Conjecture”. My aim is to show that as the Knowledge Argument rests upon the K=T Conjecture and there are good reasons to reject this conjecture, there are good reasons for materialists who are naturalistic realists to reject the Knowledge Argument. The question then becomes whether a more restricted version of K=T can be defended for the Knowledge Argument. I argue that there are reasons why the appropriately restricted form of K=T also fails. The failure is attributable to the character of our ordinary phenomenal concepts.

Our first task, though, is to clarify the relation between materialism and its anti-materialist critiques. This can be done by attending to some features that are common to anti-materialist arguments against materialism and anti-realist arguments against realism.

2 Challenge or refutation?

It is important to distinguish two questions that are easy to conflate:

1. *Does the Knowledge Argument pose a significant challenge to materialism?*
2. *Does the Knowledge Argument refute materialism?*

It is important to distinguish (1) and (2) because it is consistent to maintain that whilst materialists have not yet answered the serious challenge the Knowledge Argument poses for their theory, the argument does not yet refute materialism. The view for which I shall argue in this chapter is that not only is this a consistent position, it is very plausibly true.

1. Naturalism is the view that what the world contains is to be determined by our best theory of nature.

The *challenge* to materialism presented by the Knowledge Argument is to provide a physical explanation of sensory consciousness. How, if seeing red is nothing over and above a physical process, could a person blind from birth know all the physical facts yet not know what it is like to see red? I take it as plain fact that no one has produced a physical explanation of colour vision that answers this question.

Challenging a theory is one thing, *refuting* it quite another, however. To *refute* materialism one needs to present a sound argument against it – one proceeding from true premises to a conclusion logically entailed by those premises. To *challenge* a theory one need only present phenomena the theory needs to but cannot, explain. Yet whilst a theory that is unable to explain certain phenomena may be incomplete, it may nonetheless be true. A theory inconsistent with known facts, on the other hand, cannot be true.

Of course there is a close link between a theory's inability to respond to central challenges and our subsequent confidence in its truth: We might feel that belief in a theory that is unable to meet certain crucial challenges is unwarranted, that the theory is unlikely to be true in light of these shortcomings. That is reasonable enough, but how much credence one should place in a theory often critically depends upon the credibility of its competitors. If the problems besetting the alternatives are just as bad or worse than those facing the original theory or if the alternative theories are unconvincing, it may be rational to maintain confidence in the original. For it may be that we simply lack the wit to develop the theory in a way that answers its crucial challenges or, again, we may not realise or accept that the theory already has the resources to meet these challenges.

A comparison of anti-materialist critiques of materialism with anti-realist critiques of realism is apposite at this juncture. Realists contend that our thoughts about the world are true or false independently of whether we can recognise that they are, *purely by virtue of how the world is*. Thus we suppose that it is either true or false that Julius Caesar's heart skipped a beat before he crossed the Rubicon even though we have no way of ever telling which of these alternatives holds. The challenge posed by anti-realists to naturalistic realism is to explain how there could be a *physical* mechanism of mental representation that links agents to states of the world to which they may be entirely oblivious: How could Nature set up such a mapping in the first place?

The anti-realist challenge to naturalistic realism is a serious one. In spite of this, the hypothesis that the world is independent of human conceptions of it might still be the best hypothesis, all things considered, to account for the observed features and nature of the world and of humanity's place within it. It might also be that the anti-realist's alternative hypothesis that the world is circumscribed by our conception of it, is demonstrably inferior in accounting for all the relevant

data or else that it suffers from debilitating internal problems concerning its interpretation and logical coherence, or both.

My own view, for which I have previously argued, is that this is indeed the case: In spite of the severity and aptness of the anti-realist challenge, when all is said and done, the hypothesis that the world is independent of human conceptions of it is the only hypothesis deserving of our credence since the anti-realist alternative does indeed suffer from internal problems of coherence in addition to an inability to adequately account for all the evidence (Khlentzos 2004).

There is more than a structural similarity between anti-materialist critiques of materialism and anti-realist critiques of realism, however. For the most plausible version of materialism construes it as a form of naturalistic realism, in my view. The anti-realist challenge to provide a naturalistic account of mental representation and the anti-materialist challenge to provide a naturalistic account of phenomenal consciousness subsequently emerge as two sides of the same coin: a shared opposition to naturalistic realism. It should be unsurprising therefore if the anti-materialist's Knowledge Argument turned out to tacitly rest upon anti-realist assumptions. I turn now to the task of demonstrating that this argument does indeed rest upon such an assumption, indicating how I think that assumption should be resisted.

3 The $K=T$ Conjecture

If mind is part of nature, as the materialist avers, then its workings must somehow be fixed by its material microstructure. This is because the materialist sees all macrophenomena as determined by or supervenient on the microphysical. But it has proved surprisingly hard to explain this notion of supervenience qua asymmetric dependence of macro-facts on micro-facts.

Frank Jackson holds that the supervenience of macrophysical facts on microphysical facts is to be explained in terms of possible worlds (Jackson 1998):

(MPD) Every minimal physical duplicate of our world is a duplicate simpliciter

where a "minimal physical duplicate" of our world is one that duplicates all and only the microphysical facts that determine the nature of our world. Armed with the MPD understanding of "supervenience", Jackson goes on to argue that the microphysical facts *entail* all other facts in every minimal physical duplicate of our world.

But we might well wonder what *grounds* this entailment? In response, David Chalmers, Frank Jackson and David Lewis have each contributed to an influential model tying supervenience and the associated entailment relation to conceptual analysis (Chalmers 1996; Jackson 1998; Lewis 1994). In brief, someone who knew all the microphysical facts and knew all the functional roles our folk concepts

associate with various macrophysical phenomena should be able, in principle, to deduce all the macrophysical truths *a priori* from the complete set of microphysical truths, Φ . I shall refer to this view as *Analytic Materialism*.²

Analytic Materialism is the thesis that if T is a fact of any sort, the conditional $\Phi \rightarrow T$ is necessary *a priori*. Thus analytic materialists deem the very types of facts many have cited as supporting *non-reductive* materialism by virtue of being necessary *a posteriori* – in particular, semantic facts and mental facts – to all be *a priori* deducible from the microphysical facts and thus to support *reductive* materialism. Yet how can paradigmatically necessary *a posteriori* identities such as those that hold between water and H_2O , or, let us suppose, between visual awareness and pyramidal cell activity in the V4 region of the cortex, possibly be the consequents of *a priori* conditionals?

The key lies in a demonstration that such identities can be derived *a priori* from Φ using a doubly-indexed semantics. We do not need to review the details of this approach to appreciate that there is a problem with it. Whilst the alethic modalities “necessary” and “contingent” partition the class of *all* propositions, the epistemic modalities “*a priori*” and “*a posteriori*” partition the class of *knowable* propositions only. Since only truths are knowable, to secure the conclusion that *all* truths are *a priori* deducible from Φ one must first assume that *all truths are knowable*: In other words, that the class of *knowable* propositions, K coincides with the class of *true* propositions, T . This is the $K=T$ Conjecture. It may help to represent the situation diagrammatically, as in Figure 1.

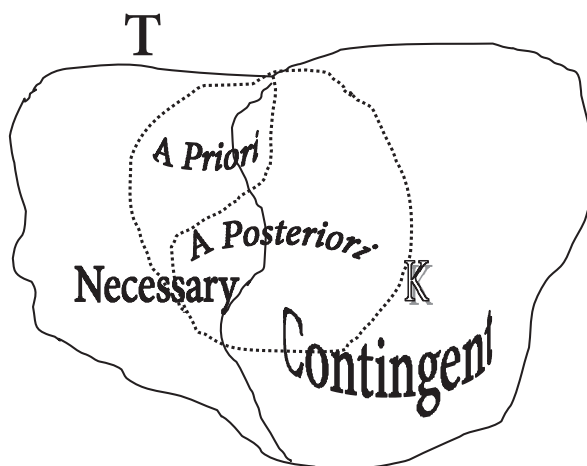


Figure 1. Kripke's model

2. Note that even though Chalmers is a dualist rather than an analytic materialist, he does believe that the fundamental physical facts together with the phenomenal facts entail all other facts.

If the $K=T$ Conjecture holds then there are no unknowable necessary or contingent truths as illustrated in Figure 2.

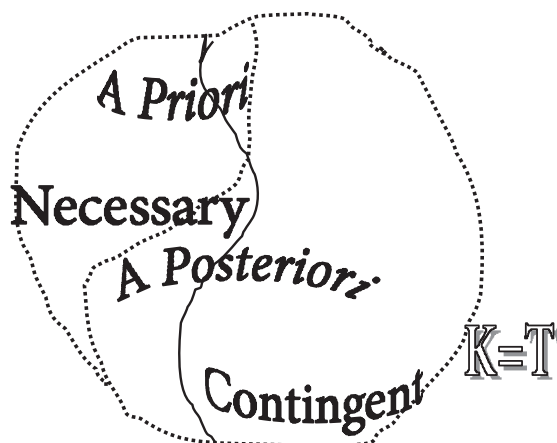


Figure 2. Analytical Materialist model

However, $K=T$ is the definitive thesis of *anti-realism*. What business has materialism with a view that denies the world is mind-independent? Yet the connection between analytic materialism and anti-realism should have been evident since the thesis that all facts are in principle *a priori* deducible from the microphysical truths entails that nothing could *be* a fact unless it is in principle so deducible and no interpretation of “in principle”, however generous, can mask the fact that this thesis puts a significant epistemic constraint on truth.

There is, in fact, a neat and compelling argument due to Frederick Fitch that appears to refute the $K=T$ Conjecture (Fitch 1963).³ Fitch’s Argument can be formalised in a very weak modal logic $S0.5$ extended by adding a one-place connective K such that if q is a well-formed formula then so is Kq . The resulting system, $KS0.5$ say, can be axiomatised:

$$(\Box KT) \Box (q \rightarrow \Diamond Kq), (\Box FK) \Box (Kq \rightarrow q), (\Box K\&) \Box (K(q \& r) \rightarrow Kq \& Kr)$$

$(\Box FK)$ says that knowledge is veridical – necessarily, if we know q then q is true. $(\Box K\&)$ says that knowledge distributes over conjunction – there is no knowing a conjunction is true without knowing both its conjuncts are true. The anti-realist’s

3. There has been extensive discussion of Fitch’s Argument in recent years. For a good sample see van Bentham 2004; Dummett 2001; Edgington 1985; Tennant 2002; Williamson 2002.

distinctive thesis is, of course, ($\Box$ KT) the thesis that necessarily *all truths are knowable*. ($\Box$ KT) thus says it is possible to know every truth (' $\Diamond Kq$ ' says it is *possible* ($\Diamond$) to *know* (K) q)

It is easy to show classically that these three axioms jointly lead to a modal collapse of the Weak Verificationism of ($\Box$ KT) into a Strong Verificationism that holds that *all truths will eventually be known*:

- | | |
|---|----------------------------|
| (1) $p \ \& \ \neg Kp$ | Hypothesis |
| (2) $\Box[(p \ \& \ \neg Kp) \rightarrow \Diamond K(p \ \& \ \neg Kp)]$ | $\Box$ KT |
| (3) $(p \ \& \ \neg Kp) \rightarrow \Diamond K(p \ \& \ \neg Kp)$ | $2\Box E$ |
| (4) $\Diamond K(p \ \& \ \neg Kp)$ | 1, 3 Modus Ponens |
| (5) $K(p \ \& \ \neg Kp)$ | Hypothesis |
| (6) $\Box[K(p \ \& \ \neg Kp) \rightarrow (Kp \ \& \ K\neg Kp)]$ | $\Box K\&$ |
| (7) $K(p \ \& \ \neg Kp) \rightarrow (Kp \ \& \ K\neg Kp)$ | $6\Box E$ |
| (8) $Kp \ \& \ K\neg Kp$ | 5, 7 Modus Ponens |
| (9) $\Box(K\neg Kp \rightarrow \neg Kp)$ | $\Box FK$ |
| (10) $K\neg Kp \rightarrow \neg Kp$ | $9\Box E$ |
| (11) $K\neg Kp$ | $8\&E$ |
| (12) $\neg Kp$ | 10, 11 Modus Ponens |
| (13) Kp | $8\&E$ |
| (14) $\perp$ | 12, 13 &I |
| (15) $\neg K(p \ \& \ \neg Kp)$ | 5, 14 Reductio Ad Absurdum |
| (16) $\Box\neg K(p \ \& \ \neg Kp)$ | $15 \Box I$ |
| (17) $\neg\Diamond K(p \ \& \ \neg Kp)$ | $16 Df \Box$ |
| (18) $\perp$ | 4, 17 & I |
| (19) $\neg(p \ \& \ \neg Kp)$ | 1, 18 Reductio Ad Absurdum |
| (20) $p \rightarrow Kp$ | 19 Equivalence |

The conclusion at (7) is quite literally unbelievable. It implies that it could only be true that Socrates sneezed in his sleep the night before he took the hemlock if someone eventually discovers he did. As the principles ($\Box FK$) and ($\Box K\&$) appear to be beyond reproach and the conclusion is absurd, we are forced to reject ($\Box$ KT). In other words, Fitch's Argument presents us with a *prima facie* refutation of the $K=T$ Conjecture.

4 Application of K=T: The Knowledge Argument

Frank Jackson asks us to imagine an omniscient colour scientist Mary possessed of perfect logical acumen who has never experienced colour vision. Perhaps, as Jackson suggests, she has been interred all her life in a black and white room and only ever seen monochrome shades or perhaps she suffers from achromatopsia. In spite of her chromatic ignorance, whatever its source, Mary knows all there is to know about the physical facts governing colour experience. Yet she does not know what it is like to see red. If she were to be released from her black and white prison (or else have her colour vision restored) then upon seeing a ripe tomato for the first time, she would come to know something she did not previously know: namely, what it is like to see red.

We thus have a straightforward argument against materialism:

1. *Mary knows all the physical facts about colour experience.*
2. *Upon release, Mary learns a new fact about colour experience.*
3. *∴ Not all facts are physical facts.*

I think and shall argue that when suitably clarified this argument is *sound*. Does it not then refute materialism, the thesis that all facts are physical facts? Indeed it does, in my view, *provided* that by “materialism” one means, as Jackson means, *analytic* materialism.

The argument above appears straightforward. What clarification does it require? It requires a more neutral formulation of the first premise before it can be acceptable to materialists. The point in dispute between materialists and anti-materialists is whether the colour facts are physical. That is precisely what the Knowledge Argument sets out to decide. So to avoid an intractable dispute about whether achromatopsic Mary knows *all* the physical facts about colour experience, a dispute that would make the argument a non-starter, the first premise needs to be reformulated. Fortunately a simple reformulation of the first premise is at hand, in terms of supervenience, namely that Mary knows all the physical facts that *determine* colour experience.

Then the Knowledge Argument becomes:

- 1'. *Mary knows all the physical facts that determine colour experience.*
- 2'. *Upon release, Mary learns a new fact about colour experience.*
- 3'. *∴ Not all facts are physical facts.*

Stated thus, before the argument can threaten materialism we need to assume that all facts are *a priori* deducible from the (fundamental) physical facts, as Jackson himself acknowledges.

Materialism is a doctrine in metaphysics. It is a claim about what there is and what it is like. The knowledge argument turns on an epistemological claim, namely, that no story about our world told purely in physical terms – the kind of terms that appear in the materialists' or physicalists' preferred account of the world and its nature – could enable one to deduce the phenomenal nature of psychological states. How is a doctrine in metaphysics supposed to be threatened by a doctrine about the impossibility of a certain sort of deduction? (Jackson 1995: 184)

There is another important reason why the assumption of *a priori* deducibility is required for the argument to be an effective one against materialism. It is that without this assumption the materialist would seem to have a natural and plausible explanation of why Mary is unable to figure out how red things look: The relevant colour facts are not discursively representable, perhaps because the content of colour experiences is non-conceptual or otherwise ineffable. However if all physical facts are in principle *a priori* deducible from the fundamental physical facts this reply is undercut – even if colour vision in imperfectly rational humans ignorant of the relevant colour-determining facts consists of non-conceptual or analogue forms of representation this in no way shows that a perfectly rational human apprised of all the colour-determining facts could not *deduce* how red things looked.

Mary's omniscience with respect to *all* the physical facts, including, if indeed it is a physical fact, how red things look, is *derived* from more fundamental assumptions. The first is the *Supervenience* of all facts on the fundamental (micro) physical Φ -facts. The second we might call the *Completeness of Conceptual Consequence* – the thesis that Φ entails p if and only if p is *a priori* deducible from Φ . The third assumption involved is Mary's *Perfect Logical Acumen*. This assumption is needed to ensure that Mary is able to derive every single logical consequence of the fundamental physical facts in Φ . Without this assumption there would be a natural explanation for why Mary might not know how red things looked: She may not have possessed the logical nous for seeing that the characteristic appearance of red things is also determined by the fundamental facts in Φ .

It will help to have a perspicuous representation of the Knowledge Argument. Let ' p ' be the proposition that Mary, colour-vision restored, would come to learn upon seeing her first ripe tomato – viz. *<red things look like this>*. Let ' $\Phi \models p$ ' be read as asserting that the complete set of microphysical facts Φ semantically entails p and let ' $\Phi \vdash p$ ' assert that p is a conceptual consequence of Φ in the sense that p is *a priori* deducible from Φ :

- | | | |
|---|---|---|
| (1) $\forall p [\text{Tr}(p) \leftrightarrow \Phi \models p]$ | <i>Supervenience</i> | 1 |
| (2) $\forall p [\Phi \models p \leftrightarrow \Phi \vdash p]$ | <i>Completeness of Conceptual Consequence</i> | 2 |
| (3) $\forall p [\Phi \vdash p \ \& \ \text{Km}\Phi \rightarrow \text{Kmp}]$ | <i>Mary's Perfect Logical Acumen</i> | 3 |

(4)	$Km\Phi$	Mary's Φ -Omniscience	4
(5)	$\neg Kmp$	Mary's Chromatic Ignorance	5
(6)	$Tr(p)$	Factuality of p	6
(7)	Kmp	From 1, 2, 3, 4 and 6	1, 2, 3, 4, 6
(8)	$\perp$	From 5 and 7	1, 2, 3, 4, 5, 6

Informally, (1) says that every truth p is made true by ($\models$) the fundamental facts in Φ . (2) says that every truth made true by is also a priori deducible from Φ ; (3) says if some proposition p is derivable from Φ and Mary knows Φ then Mary thereby knows p ; (4) assures us that Mary does know all the fundamental facts in Φ ; (5) says that in spite of all her vast knowledge, Mary does not know the proposition p that describes how red things look; yet p is indeed true (7), so that, feeding this latter information into (1), we get that p is made true by the facts in Φ and thus by (2) derivable from Φ so that Mary *should* know p (using (4)) – a conclusion that is flatly inconsistent with the fact in (5) that she does not know p at all.

We thus have a *reductio* of some one of the five assumptions on which the argument rests, but which one? It cannot be (1), since (1) is a non-negotiable commitment of materialism. Neither can it be (3) or (4) since they comprise essential parts of the idealisation – without assuming Mary has knowledge of all the relevant facts and perfect logical acumen the materialist would have a simple explanation of why she cannot deduce how red things look. We can safely dismiss (6) as a candidate for *reductio* as well. For the argument is only intended to refute materialists who are not eliminativists about conscious experience. That leaves (2) and (5).

The overwhelming consensus in recent responses to the Knowledge Argument is that materialists cannot afford to concede that Mary learns anything new – in other words, that (5) is false. Thus some materialists allege that Mary merely acquires new knowledge of a fact she already knew by merely acquiring new ways of conceiving of the familiar old fact. Others follow David Lewis in holding that she acquires no new knowledge upon seeing her first ripe tomato, merely a new range of cognitive abilities – to visualise red things, recall how they look etc.

Although it lies outside the scope of this chapter to fully assess all such responses, a few remarks are in order. Waiving the massive implausibility of the denial that Mary learns a new fact, it seems to me that these responses fail on their own terms. Let's take the New Knowledge/Old Fact response first:

The idea is that Mary acquires no new fact when her colour vision is restored but only new ways of conceiving of a fact (or facts) she already knows. The fact in question may be that the phenomenal character of visual experiences of red is some function f of physical parameters σ_i say. So Mary already knows the identity $PC[VE_{RED}] = f(\sigma_i)$ under a physical description but **not** under a mentalistic one

according to the view in question. The new thing that she learns when, colour vision restored, she sees her first ripe tomato is the correct application of a mentalistic description, possibly conveyed by a demonstrative such as 'this' as in "Aha! Red things look like *this*!"

Firstly, there clearly *is* a fact that Mary does not know on this account – namely that the very physical property that is $f(\sigma_i)$ is the one that looks like *this*. Secondly, if materialism is true at all, activating any new concept is a physical event and must consequently involve physical changes – for example, the tokening of some physical inscription τ of the concept in the brain. So if Mary already knows all there is to know about the workings of the brain she will already know all there is to know about the physical and functional properties of the neural correlate τ of the new concept achromatopsics acquire once their colour vision is restored and they see red things for the first time. Whence, it is, as before, a mystery how she can learn anything new when she already knows all the facts about the physical and functional role of τ in the brain, including the inferences that involve it and its deductive links with other concepts mirrored in the way the neural tokens of those concepts are activated in the brain.

More importantly, materialists are *not* entitled to assume that the phenomenal character of visual experiences is a *physical* property since this is the very point in dispute! Yet to the extent that it construes the fact Mary discovers, $PC[VE_{RED}] = f(\sigma_i)$, as one she *already* knows under a physical description this is precisely what the Old Fact/New Knowledge response *does* assume. Until such times as they can produce hard evidence for the physical nature of the phenomenal character of visual experiences of red, materialists cannot take the physicality of such properties for granted.

A worrying feature of the contemporary debate is that many materialists *do* seem to think they are entitled to assume, in the context of their debate with anti-materialists, that colour properties are physical. They then construe their task, in the case of the Knowledge Argument, to be one of explaining away the appearance that physically omniscient Mary learns any new fact when she first sees red. In this vein, John Perry outlines a view he calls Antecedent Physicalism.

One sees compelling reasons for physicalism and will not give it up without seeing some clear reason to do so ... We grant that there are subjective characters – so long as qualia or subjective characters are not *defined* as non-physical. Indeed, we insist on the importance of the subjective character of experience. We then construct the best account, or at least a reasonably natural account, of subjective characters on the assumption that they are physical. *Then and only then*, do we look at the neo-dualist arguments to see if they point out some inadequacy or hidden contradiction in our account. (Perry 2001: 27; Italics mine)

Perry's "antecedent physicalist" assumes physicalism or materialism throughout his debate with Jackson. The task is to defend it against Jackson's attempted *reductio* in the Knowledge Argument. Indeed it is, but no materialist, "antecedent" or otherwise, can afford to *continue* to assume materialism when the task turns to saying what is wrong with a purported *reductio* of it!

If there is to be any progress in the debate between materialists and anti-materialists both sides must shelve their ideological commitments when it comes to the evaluation of the evidence. Jackson, for his part, does *not* assume epiphenomenalism, antecedently or otherwise. Instead, he *argues* for it by constructing a neat thought experiment in which a logically omniscient achromatopsic person is given all the physical facts and asked to figure out how red things appear to those with normal colour vision. Since it is overwhelmingly plausible that achromatopsic Mary will never discover how red things look, the colour facts could not be determined by the physical facts, *pace* materialism. Whence, as those with normal colour vision apparently know something Mary does not know, there are non-physical properties of experience, qualia, and materialism is false. Epiphenomenalism is then presented as the best account of qualia.

There is nothing unfair or question-begging from the materialist's point of view in the thought experiment. So if the materialist is to resist the conclusion that materialism is false, s/he must either question the validity of the inference from premises to conclusion or explain why one of the premises is false. The argument is clearly valid, as its formal rendition shows and we have already narrowed down the choices for false premises to (2) or (5) – either Mary succeeds in figuring out what normal people are adverting to when they call ripe tomatoes red, in which case (5), which asserts that Mary does not know this, is false or else, compatibly with materialism, the colour facts are *not a priori* deducible from the physical facts, in which case (2) which asserts they are, turns out false.

The proponent of the Old Fact/New Knowledge view singles out (5) as the false assumption. That means Mary *does* know what others are talking about when they call ripe tomatoes red – *pace* Jackson, she can and has figured out how red things look. This is a very bold claim. To back it up one needs at least two things: (i) a physical theory of phenomenal character and (ii) some strong empirical evidence supporting such a theory.

Although some proponents of the Old Fact/New Knowledge view such as Michael Tye have indeed proposed physical theories of phenomenal character, it would be more than a little optimistic to claim they have any empirical evidence favouring their theories over nomological dualist or epiphenomenalist alternatives. In fact, there is a real question as to whether there could even be such empirical evidence in the form of some "critical experiment(s)" differentially confirming the materialist's physical theory.

My conclusion is that we should reject the Old Fact/New Knowledge response. It is almost impossible to read Jackson's own formulation of the Knowledge Argument and still insist that Mary merely gains new ways of seeing something she knew all along. At the very least, she becomes directly acquainted with a new property that appears to be associated with the surfaces of ripe tomatoes, a property of which she had no inkling in her achromatopsic state.⁴

What of the Ability Hypothesis, as it is known? According to this hypothesis, all that distinguishes achromatopsic Mary from visually normal Tom when it comes to colour vision is a cluster of cognitive abilities which Tom has and she lacks: the ability to form red after-images; to construct a mental image of something red; to recall in memory how red things look, etc. (Lewis 1990, 1994).

There is a real question whether this response is so much as to the point. According to Jackson's own formulation of the thought experiment, nothing in the information Mary has about the brain states and neural states caused by visual presentations of red alerts her to the fact that Tom's visual experiences of red have a certain phenomenal character. It is this fact that Mary is alleged not to know. Call this fact F. To prove Jackson wrong on this score, the materialist has to show that achromatopsic Mary does indeed know F.

Yet the Ability Hypothesis does not even address this question. Just like the Old Fact/New Knowledge response, it helps itself to the assumption that Mary already knows F by virtue of knowing all the physical facts. Since that is precisely what's at issue, it is similarly question-begging. Once we see this, we can effectively dismiss the discussion about abilities and the acquisition of propositional knowledge as a red herring.

Perhaps, though, this is not what the Ability Theorist has in mind? Perhaps the elusive fact F concerning the phenomenal character of Tom's visual experiences of red things just *consists of* the congeries of Tom's imaginative and recognitional abilities with respect to red things, cognitive abilities that are essentially non-propositional?⁵ What then? Well the question turns to why factually and logically

4. It may be, as representationalists aver, that the property that appears to inhere on the surface of the ripe tomato is in reality a representation of its surface in the mind of the agent viewing the surface. Neither colour objectivism that holds the property of redness is a property of the tomato nor colour realism which holds that the property is a mind-independent one need be assumed in order for Mary to learn a new fact when she sees her first tomato. Whatever their source, whatever their nature, colours are properties of which achromatopsic Mary is blissfully unaware according to Jackson. If so, Mary learns a new fact when she encounters redness for the first time.

5. In fact, there are empirical reasons for thinking that such abilities may crucially depend on a propositional representation in the brain. Chomsky (2000: 51) points to the existence of speakers who lose the practical ability to speak a language through injury or disease and then regain

omniscient Mary cannot acquire these abilities. That the abilities in question involve no propositional representations in normal subjects is beside the point. For all the anti-materialist cares, the target abilities might be stochastically generated in Mary using Monte Carlo simulations.

So even though the debate has centred on whether Mary learns a new true proposition when she sees her first ripe tomato, the issue of propositional as opposed to non-propositional representation is simply irrelevant. Someone who learns where the gold is hidden on the island by being shown its location on a map and someone who learns the same thing by being given an explicit description of its whereabouts both enjoy access to the *same fact* even if their *modes* of access differ – non-propositional in the first case, propositional in the second.

It might be helpful to see how a cognate issue was addressed in semantics some time ago. Donald Davidson long ago confronted the complaint that radical interpreters (field linguists grappling with a hitherto uninterpreted and culturally alien language) who modelled the native semantic abilities of competent speakers by means of arcane truth theories were distorting the ability they purported to model by producing a representation of it that those very speakers could not recognise as correct.

He responded that the task was not to model those propositions, if any, that speakers actually grasped when they learned a language but rather to produce a representation that a theorist could use under certain favourable conditions to interpret the language. This was the only theoretically relevant sense in which the theorist was modelling the same ability as the native speakers possessed.

Davidson's claim was that a radical interpreter who was able to derive so-called T-sentences of the form "‘la neige est blanche’ is true if and only if snow is white" as theorems of a formally and materially adequate Tarskian Truth Theory for the jungle language would *thereby* come to possess the very same semantic ability a native speaker possessed by virtue of learning that language (French in the sample T-sentence) as his mother tongue.

Just as for the radical interpreter, so for Mary. In fact achromatopsic Mary's problem is precisely one of radical interpretation. Her task is to find out what it is that those with normal colour vision mean to advert to when they call ripe tomatoes red. Jackson's view is that Mary would know what this property of redness was

it. The most plausible empirical explanation is that although they have temporarily lost the practical ability, they have retained the internal representation in the brain of the relevant procedure, Chomsky argues: "In all such cases, something remains constant, some property K, while ability to speak, understand etc. varies. In ordinary usage, we say that K is knowledge of language; thus Jones's knowledge remained constant while his ability to put his knowledge to use improved, declined, recovered etc. The account in terms of internal representation of a generative procedure accords with informal usage in this case."

if she were able, as he thinks she has to be in order for materialism to be true, to deduce the relevant propositions from her vast store of all physical knowledge. He is not committed to the claim that what distinguishes achromatopsic Mary from chromatically normal Tom is that Tom knows propositions that Mary does not know. For as Davidson famously put it:

Kurt utters the words 'Es regnet' and under the right conditions we know that he has said that it is raining. ... What could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same as the question what we *do* know that enables us to interpret the words of others. For there may be easily something we could know and don't know of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation. (Davidson 1984: 125)

Thus, even if there are no propositions Tom knows which enable him to visualise, recall, or imagine red things, this does not even begin to show that knowledge of the relevant propositions describing them would not suffice for Mary to acquire these abilities. If Mary has identified the neural tokens activated by the abilities in question and *ex hypothesi* knows everything there is to know about the functional and physical behaviour of these tokens, it is a mystery why she cannot acquire the relevant abilities. The Ability Hypothesis is therefore a failure.

Surely it is clear that Mary *does* learn something new when she has her colour vision restored and she sees her first ripe tomato? How could cramming an achromatopsic's head full of facts and sharpening their logical abilities bring them any closer to knowing what the colours look like? If materialism cannot afford to admit that Mary does indeed learn something new, it is unworthy of our credence.

Can the materialist do so? Yes. For, I submit, it is not (5), the assumption that Mary learns something new that is refuted by the Knowledge Argument, but rather (2) the assumption that the colour facts are *a priori* deducible from the physical facts. We have already seen that this assumption rests upon the anti-realist's K=T Conjecture and have also seen why that conjecture is dubious in its full generality. But for all that has been shown so far, materialists might be committed to maintaining that the colour facts at least are deducible from the neurophysiological (or computational) facts even if the neurophysiological facts are not in turn deducible from more fundamental physical facts.

So are materialists committed to the *a priori* deducibility of colour facts from neurophysiological facts? It has to be acknowledged that the intuition that says they are is a powerful one. If you knew all the colour-determining neurophysiological facts and were logically omniscient what could possibly *prevent* you figuring

out how the colours look? Isn't it sheer obfuscation to maintain otherwise? Jackson expresses the point with characteristic force and verve:

It is implausible that there are facts about very simple organisms that cannot be deduced *a priori* from enough information about their physical nature and how they interact with their environments, physically described. The physical story about amoeba and their interactions with their environment is the whole story about amoeba. Mary would not lack knowledge about them. But according to materialism, we differ from amoeba essentially only in complexity of ingredients and their arrangement. It is hard to see how that kind of difference could generate important facts about us that in principle defy our powers of deduction, and the fact that we have a phenomenal psychology is certainly an important fact about us. Think of the charts in biology classrooms showing the evolutionary progression from single-celled creatures on the far left to the higher apes and humans on the far right: where in that progression can the materialist plausibly claim that failure of *a priori* deducibility of important facts about us emerges? Or, if it comes to that, where in the development of each and every one of us from a zygote could the materialist plausibly locate the place where there emerge important facts about us that cannot be deduced from the physical story about us? (Jackson 1995: 189)

As compelling as Jackson's challenge to materialism is, I believe it can be met. The point where "there emerge important facts about us that cannot be deduced from the physical story" is the point where we come to master the folk mental concepts and use them to describe our conscious experiences and other mental episodes. For it is what the folk concepts apply to that "cannot be deduced from the physical story about us".

If all that were significant about folk mental concepts was their functional role, as analytic materialists such as Lewis and (post-conversion) Jackson aver, it would indeed be hard to resist the demand for *a priori* deducibility of the colour facts from the neural facts. But colour concepts and sensation concepts have both a complexity and a simplicity that ill-suits them to the type of functionalist reduction demanded by analytic materialism. On the one hand, there is no saying *which* property 'red' names without adverting to physical samples. This suggests the folk concept of red is of some primitive unanalysable property, leaving it hostage to further scientific and philosophical analysis that might reveal the true nature of that property. On the other hand, no philosophical analysis of red can replace the folk concept. This is because a folk concept is needed not only to pick out the salient patterns in our experience but also to make sense of disputes about the nature of the property to which the folk concept applies.

This view of folk mental concepts has been defended independently by Robert Stalnaker and John Hawthorne (cf. Stalnaker 2002; Hawthorne 2002).⁶ Stalnaker arrives at this position by inquiring into the theoretical loading the notion of sensory awareness bears. Suppose that our best theory of visual awareness is a neurocomputational theory which holds that to be visually aware is to experience a certain type of neurocomputational activity in the pyramidal cells of the V4 region of the cortex. Now suppose that although the theory is highly confirmed by a wealth of experimental evidence, it turns out to be mistaken. Would we be led to deny as a result that people ever see anything?

The very question strikes us as ludicrous – we would have no inclination whatsoever to deny the facts of visual experience simply because our best theoretical account of them had turned out to have been wrong. But if so, the folk concept of visual awareness could not be theoretical to the same degree that the concept of oxygen or of phlogiston is theoretical. Further, the very fact that we can comprehend and evaluate theoretically incompatible accounts of the nature of consciousness such as materialist, epiphenomenalist, nomological dualist or dual attribute theories attests to the non-theoretical status of the folk notion. We feel no inclination whatsoever to say of one or the other such theory “this is not consciousness that this theory is discussing, it is something else”, neither do we feel that the plurality of theories of consciousness explicate a plurality of different folk notions. The reason is that it is not the function of folk concepts of experience to speculate on the inner nature of various features of experience – their function is to identify and draw attention to those commonalities in experience that mark us out as distinctively human.

Yet if these reflections are on the right track, the analytic functionalist account of folk mental notions that underlies analytic materialism must be incorrect. For the analytic functionalist contends that folk mental concepts are implicitly defined by the non-mental terms in the postulate formed by conjoining all the platitudes of the folk theory of mind. Folk mental concepts are genuinely theoretical concepts on this view – indeed, they have no meaning outside of the theory in which they occur. Then, if we find a physical state that fills the functional role etched out by the folk theory of visual experiences of red things, brain state R say, we will have reduced the mental state to a physical state in the form of brain state R.

The mistake is to think of the folk mental concepts as being of a high enough grade of theoreticity to permit such a reduction. For a general and pervasive feature of folk concepts is that they appear to be metaphysically neutral or tolerant. We can reflect on the facts of colour experience whilst considering various hypotheses about the nature of that experience – perhaps it is, as epiphenomenalists

6. A version of this view is also defended independently in Khlentzos (2004).

aver, a matter of sensing emergent non-physical qualia; alternatively, perhaps, as materialists contend, it is all just a matter of neural firings? Such musings would be dubiously coherent if the folk concept of a visual experience of red referred to an essentially non-physical property as dualists such as David Chalmers contend or to an essentially physical property as analytic functionalists such as Lewis, Shoemaker and the born-again materialist Jackson maintain.

Following John Hawthorne's suggestion, we can represent the metaphysical neutrality of folk mental concepts partly in terms of indicative conditionals (cf. Hawthorne 2002):⁷

- (Q) *If the world is not purely physical and colour experience directly acquaints us with emergent non-physical properties or qualia then if Q_R is the quale associated with experiences of red the folk concept of red applies to Q_R .*
- (P) *If the world is purely physical and P_R is the physical property that plays the red role in folk colour theory then the folk concept of red applies to P_R .*

Hawthorne points to a powerful consideration in favour of this view: Suppose you believed that 'red' named a non-physical property and an omniscient oracle were to inform you that the world was purely physical. Would you conclude that you had never seen anything red? Clearly not. *Pari passu* if you believed 'red' named a physical property and the oracle were to inform you that the world contained emergent non-physical properties in the form of phenomenal properties.

Now if this view is right, and it seems to make much better sense of folk practice than the functionalist alternative endorsed by analytic materialists, it will not be possible for achromatopsic Mary to deduce from Φ alone whether 'red' picks out a physical property or an emergent non-physical property. At least she will not be able to do so until she learns which of materialism or dualism is true. But it was the whole purpose of the thought experiment to decide this very question!

So achromatopsic Mary *does* indeed learn something new when, colour vision restored, she sees a ripe tomato for the first time. Either she learns that 'red' names a physical property, one of the properties with which she may already be familiar, or else she learns that 'red' names none of the properties with which she may be familiar. Since which she learns depends upon which one of materialism or dualism is true, the Knowledge Argument cannot be used to decide that matter.

The materialist thus has a plausible reply to the Knowledge Argument:

- (i) *Mary did indeed know all the physical facts determining the colour facts.*
- (ii) *Yet Mary did not know the colour facts, physical facts though these are.*

7. Hawthorne's example involves pain since he is discussing the Zombie Argument but his point carries over straightforwardly to colours.

- (iii) *For Mary was not able to deduce all the physical facts from Φ using the background folk theory.*
- (iv) *This was because the folk concept of red left it an open question whether 'red' refers to a physical or a non-physical property.*
- (v) *Hence achromatopsic Mary could not rule out that 'red' refers to a non-physical property.*
- (vi) *She thus could not know that red things looked the physical way ripe tomatoes look.*
- (vii) *Therefore, once her colour vision is restored, Mary will **learn** how red things look when she sees her first ripe tomato.*

Why was Mary not able to deduce all the physical facts from Φ using the background folk theory? Because the folk theory was compatible with the two radically different metaphysical proposals presented above in (P) and (Q): Namely, that if it is a *physical* property P_R that played the role that the concept of redness plays in folk colour theory, then 'red' referred to that physical property P_R ; but if, on the other hand, it is a *non-physical* property Q_R that plays this role then 'red' refers to that non-physical property Q_R . Thus even though, if materialism is true, it is really P_R and not Q_R that 'red' refers to, Mary who is omniscient about the physical but not the metaphysical facts cannot deduce this from the information she has since she does not know which of materialism or dualism is true.

So the upshot of our discussion of the Knowledge Argument is very different from the one usually drawn. The Knowledge Argument does indeed refute one version of materialism – the analytic materialism of David Lewis and (the latter day) Jackson which holds that all physical facts are *a priori* deducible from the fundamental physical facts. For, given the twin assumptions of materialism and *a priori* deducibility it follows deductively that achromatopsic Mary learns nothing new when, colour vision restored, she sees her first ripe tomato, a conclusion that simply beggars belief.

Moreover we can now see what is right about the Knowledge Argument – it works as a powerful *reductio* of the assumption that all facts are *a priori* deducible from the physical facts. In so doing, it provides us with another reason, independent of the one based on Fitch's Argument, for thinking that the K=T Conjecture that is definitive of anti-realism and on which the assumption of *a priori* deducibility depends is false.

Materialists who are naturalistic realists, rather than unwitting anti-realists (as are analytic materialists), can thus plausibly argue that the Knowledge Argument is unsound *without* taking for granted that any new fact Mary learns must be physical. This is because they reject on realist rather than materialist grounds the K=T

Conjecture and its specification in the assumption of *a priori* deducibility of all facts from the fundamental physical facts.

However they should not get too excited by this result. For the *challenge* that underlies the Knowledge Argument has yet to be met by materialists of any stripe – that challenge was to discover the physical properties that are to be identified with the phenomenal properties.

References

- Bentham, Johan van. 2004. What one may come to know. *Analysis* 64(282). 95–105.
- Chalmers, David. 1996. *The conscious mind: In search of a fundamental theory*. New York, NY: OUP.
- Chomsky, Noam. 2000. *New horizons in the study of language and mind*. Cambridge: CUP.
- Davidson, Donald. 1984. *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Dummett, Michael. 2001. Victor's error. *Analysis* 61(269). 1–2.
- Edgington, Dorothy. 1985. The paradox of knowability. *Mind* 94. 557–568.
- Fitch, Frederick. 1963. A logical analysis of some value concepts. *Journal of Symbolic Logic* 28. 135–142.
- Hawthorne, John. 2002. Advice for physicalists. *Philosophical Studies* 109.17–52.
- Jackson, Frank. 1995. What Mary didn't know. In Paul K. Moser & James D. Trout (eds.), *Con-temporary materialism: A reader*, 180–189. London: Routledge.
- Jackson, Frank. 1998. *From metaphysics to ethics: A defense of conceptual analysis*. Oxford: Clarendon Press.
- Khlentzos, Drew. 2004. *Naturalistic realism and the antirealist challenge* (Representation and Mind). Cambridge, MA: Bradford Books.
- Lewis, David. 1990. What experience teaches. In William G. Lycan (ed.), *Mind and cognition: A reader*, 499–519. Oxford: Blackwell.
- Lewis, David. 1994. Reduction of mind. In Samuel D. Guttenplan (ed.), *A companion to the philosophy of mind*, 412–431. Oxford: Blackwell.
- Perry, John. 2001. *Knowledge, possibility and consciousness*. Cambridge, MA: The MIT Press.
- Stalnaker, Robert. 2002. What is it like to be a zombie? In Tamas Szabo Gendler & John Hawthorne (eds.), *Conceivability and possibility*, 385–400. Oxford: Clarendon Press.
- Tennant, Neil. 2002. Victor vanquished. *Analysis* 62 (274). 135–142.
- Williamson, Timothy. 2002. *Knowledge and its limits*. Oxford: OUP.

Vagueness, supertranslatability, and conceptual schemes

Arcady Blinov

A case is developed against Donald Davidson's thesis that any attempt to give meaning to the idea of conceptual relativism is doomed. Even under Davidson's assumption that untranslatability is a necessary condition of a difference between two schemes, it appears to be possible to conceive a situation in which one language is untranslatable for another, but still accessible for truth testing, which undermines Davidson's main argument. One further argument by Davidson is countered by the observation that even if we play the game of radical interpretation strictly on Davidson's terms, we may well encounter a situation in which the only rational choice is to allow for partial failures of translation, and thus for a difference between the two conceptual schemes.

1 Introduction

What linguists call "categorisations" or "categorical systems" philosophers often call "conceptual schemes". Thus, what Donald Davidson undertook to do in his famous paper Davidson (1984) is to undermine the very idea of two languages having different categorical systems.

To support his subversive project, Davidson developed more than one argument. Correspondingly, numerous critics of Davidson's project may, and often really do, differ from one another, firstly, in their respective interpretations of what they perceive as anti-conceptual-schemes arguments in Davidson's writings, and secondly, in their rankings of Davidson's arguments by their perceived centrality for the anti-scheme project.

Being a friend of (moderately different) conceptual schemes, I will have to begin my chapter with fulfilling the same job of selection and interpretation: I will select two arguments from Davidson (1984) which I think are absolutely central for Davidson's project. Then, after having provided an interpretation of them, I

will argue that: (1) both arguments are flawed in that they both – each in its own manner – oversimplify and, by the same token, distort the conceptual situation at issue; (2) after due improvement, they can (at the very most!) only demonstrate that *radical* differences between conceptual schemes are impossible, leaving unscathed a possibility of *non-radical* ones; (3) because I can't see any further serious threats for a possibility of non-radical differences in the offing, this leaves us with the upshot that non-radically different conceptual schemes are possible, and in actual fact they are featured by different languages.

In Sections 2 and 3, I will reconstruct and discuss the two Davidson's arguments I mentioned above. I will show that they are both flawed. In Section 4, I will adapt to my purposes a thought experiment by Georg Simmel. In Sections 5–8, I will develop, step by step, an argument to the effect that the two languages from my Simmelian thought experiment have non-radically different conceptual schemes, thus presenting a case against Davidson's attack on the very idea of a conceptual scheme.

2 Davidson's first argument against conceptual schemes

One central feature of Davidson's project is establishment of an intimate link between the idea of otherness of a conceptual scheme and the idea of untranslatability between two languages. Because of this link, any argument against untranslatability becomes an argument against the very possibility of different conceptual schemes, and Davidson presents us with (at least) two such arguments: One is to the effect that we cannot make sense of *total failure* of translatability because of our inability to understand the notion of truth independent of the notion of translation (Davidson 1984: 185–195); the other (cf. Davidson 1984: 195–197), that even *partial failures* of translation fail “to give a solid meaning to the idea of conceptual relativism, and hence to the idea of a conceptual scheme” (Davidson 1984: 197) – this time because of the Charity Principle that is inbuilt into methodology of Radical Interpretation.

That total failure of translatability is unintelligible seems intuitively plausible. On the other hand, to deny either that partial failures are frequently met in the real-life practice of translators or that sometimes such failures may mark a (not necessarily radical) difference between the two conceptual schemes seems to fly in the face of another set of deeply-entrenched intuitions of ours.¹ So what I think would be worth having is a theory of conceptual schemes that could reconcile these two sets of intuitions. Actually, the point of this chapter is to make a couple of steps toward such a reconciliatory theory.

1. For an interesting discussion of this issue, see Bar-On (1994).

Davidson's argument against total failure of translatability, and thus against (radically different) conceptual schemes, is – as far as I was able to reconstruct it from Davidson (1984: 185–195) – as follows:

1. That a conceptual scheme (construed as a theory) is different from ours means that it is largely true, but not translatable.
2. But we can't even understand the notion of truth, as applied to language, independent of the notion of translation.
3. Any test whether a conceptual scheme is different from ours should involve a test whether the scheme is largely true. [from (1)]
4. But any test for truth does involve, in one form or another, T-sentences.²
5. Any form of involvement of T-sentences presupposes translatability of the object-language sentences under test.
6. Thus, a test whether the scheme is largely true presupposes translatability. [from (4)-(5)]
7. Any test whether a conceptual scheme is different from ours presupposes translatability. [from (3) and (6)]
8. Thus, any test whether a conceptual scheme is different from ours presupposes that the scheme under test is not different from ours. [from (1) and (7)]
9. Thus, we can't even make sense of the idea of a different conceptual scheme. [from (8)]

2. T-sentences are metalinguistic devices for characterising truth conditions of sentences from an object language. E.g. with the English sentence "Snow is white" there is associated a T-sentence "'Snow is white' is true if and only if snow is white'. The following excerpt from Davidson helps to understand Davidson's view on the significance of T-sentences as a link between the notion of truth and the notion of translation:

We recognize sentences like "'Snow is white' is true if and only if snow is white' to be trivially true. Yet the totality of such English sentences uniquely determines the extension of the concept of truth for English. Tarski generalized this observation and made it a test of theories of truth: according to Tarski's Convention T, a satisfactory theory of truth for a language L must entail, for every sentence *s* of L, a theorem of the form '*s* is true if and only if *p*' where '*s*' is replaced by a description of *s* and '*p*' by *s* itself if L is English, and by a translation of *s* into English if L is not English. This isn't, of course, a definition of truth, and it doesn't hint that there is a single definition or theory that applies to languages generally. Nevertheless, Convention T suggests, though it cannot state, an important feature common to all the specialized concepts of truth. It succeeds in doing this by making essential use of the notion of translation into a language we know. Since Convention T embodies our best intuition as to how the concept of truth is used, there does not seem to be much hope for a test that a conceptual scheme is radically different from ours if that test depends on the assumption that we can divorce the notion of truth from that of translation. (Davidson 1984: 194–195)

It should be understood from the outset that no considerations related to radical interpretation play any (direct) role in this argument. The argument is about the process of truth testing, not the process of interpretation. Correspondingly, the main figure here is that of the Truth Tester, not of the Radical Interpreter. The assumption is that somehow or other the translations of “alien” sentences are available to the Truth Tester, and she is going to use them in order to test whether the “alien” conceptual scheme is largely true.

Now, I claim that Davidson’s argument is flawed in that statement (6) is false. I will argue that some varieties of partial failures of translatability are such that they allow to establish a relation between the two languages that is not translatability proper, but is sufficiently similar to translatability to enable the Truth Tester to test whether the conceptual scheme is largely true. Because in such cases the failures of translatability are partial, not total, the “alien conceptual scheme”, even if different from ours, cannot be *radically* so. But a non-radical difference is still a difference; it is sufficient to demonstrate that different conceptual schemes are a viable possibility, and thus that Davidson’s attack on the very idea of a conceptual scheme must be flawed.

The fact that Davidson ignored, in his first argument, a possibility that two languages that are not strictly intertranslatable still may stand to each other in a relation similar enough to translatability to enable the Truth Tester to test whether the “alien” conceptual scheme is largely true, amounts in my view, to a distorting oversimplification of the conceptual situation at hand. In Sections 4–8 below, I will endeavour to rectify the oversimplification.

3 Davidson’s second argument against conceptual schemes

Davidson’s second argument against conceptual schemes, the one to the effect that even *partial failures* of translation fail “to give a solid meaning to the idea of conceptual relativism, and hence to the idea of a conceptual scheme” can be found in Davidson (1984: 195–197). Below is my reconstruction of it:

1. Given the underlying methodology of radical interpretation, the interpretation can only be successful if the interpreter endeavours to optimise agreement in belief up to explicable error.
2. When others think differently from us, no appeal to evidence can force us to decide that the difference lies in our beliefs rather than in our concepts, or vice versa.
3. Therefore, we ought also to optimise agreement in conceptual schemes (= optimise translatability) up to explicable untranslatabilities. [from (1) and (2)]

4. Therefore, there are no chances that the process of radical interpretation can result in detecting radical untranslatabilities. [from (3)]
5. Therefore, given the underlying methodology of interpretation, we could not be in a position to judge that others had a conceptual scheme radically different from our own. [from (4)]

In my view, the most unsatisfactory thing about the second argument is its conclusion. It looks unsatisfactory because it is too similar to – indeed, almost indistinguishable from – the conclusion of the first argument. Both conclude that *radical* differences in conceptual schemes are ruled out. But to convince us of that, one argument – the first one – is quite enough, and, hence, the second one seems superfluous.

But, interestingly, parts (1)–(3), together with some additional considerations, enable one to infer a conclusion that usefully differs from (5). Let me add those considerations.

If (i) we should optimise both agreement in belief and translatability and (ii) whenever a problem is met, we are free to decide, without appeal to evidence, whether the problem is due to error in belief or to untranslatability, then in our decision we should be only guided by considerations of explicability. If the error in belief is explicable, whereas the untranslatability is not, then we should decide for the former rather than the latter, and vice versa. Given that, what is crucial for our purposes here is that there seems to be no guarantee that, guided by considerations of explicability, we will always be able to decide for error in belief rather than for untranslatability. It is quite conceivable that we can meet a situation in which the error in belief would be inexplicable, whereas for the untranslatability in question we will be able to find a satisfactory explanation. Then we will *have to* opt for the untranslatability.³ Thus, we may well find ourselves in a situation in which an optimal decision would consist of accepting a non-radical untranslatability, which amounts to a non-radical difference in conceptual schemes.⁴ But that means that *non-radical* differences in conceptual schemes are *not* ruled out by the second argument, and this is exactly the different conclusion that I meant above.

That Davidson fails to infer this conclusion is, I think, a further distorting simplification of the conceptual situation at hand that has to be rectified.

To sum up our discussion so far, two claims have been made regarding Davidson's arguments against conceptual schemes: (1) that some varieties of partial

3. By the way, there seems to be no guarantee, either, that we won't meet a situation in which either of the two decisions will result in an *inexplicable* disagreement. Then, we will have to get content with an inexplicable non-radical disagreement in belief or with an inexplicable non-radical untranslatability.

4. Mind you, *radical* untranslatabilities have already been ruled out by the first argument.

failures of translatability are benign enough to allow the interpreter to test whether the “alien scheme” (construed as a theory) is largely true; (2) that non-radical differences of conceptual schemes are not ruled out by Davidson’s second argument. Now I will try to substantiate the claims.

4 An adaptation of Simmel’s thought experiment

Let us begin with considering our ability to clarify possible sources of explanation for non-radical untranslatabilities. Generally speaking, there may well be more than one source, or one pattern, of such explanations. Rescher (1980) mentions two major sources: (1) different cultures and different intellectual traditions; (2) different arrangement of sensory systems of cognitive agents. For my purposes here, the latter is of immediate interest.

This is the relevant extract from Rescher’s paper:

[C]onsider a *Gedankenexperiment* suggested by Georg Simmel in the last century – that of approaching the problem of cognitive adequacy from the standpoint of an entirely different sort of cognitive being.⁵ Imagine intelligent and actively inquiring creatures (animals, say, or beings from outer space) whose experiential modes are quite different from our own. Their senses are highly responsive to quite different physical parameters – relatively insensitive, say, to heat and light, but substantially sensitized to various electromagnetic phenomena. Such intelligent creatures, Simmel held, could plausibly be supposed to operate within a largely different framework of empirical concepts and categories – the events and objects of the world of their experience might be very different from those of our own – their predicates, for example, might have altogether variant domains. (Rescher 1980: 323)

Using Simmel’s thought experiment as a springboard, let me develop my own. It will retain some similarities with Simmel’s, but on the whole will be tailored to suit the purposes at hand.

Imagine, then, intelligent creatures (say, beings from outer space; let us dub them “Simmelians” for the sake of a label) with sensory systems that have the following properties. Firstly, their sensory systems support roughly the same division of the experienced world into middle- and large-sized things as ours do. Correspondingly, one common feature of our languages is that they both are thing languages that distinguish roughly the same sorts of things. But, secondly, their sensory systems are such that they immediately experience electromagnetic radiation of the part of the spectrum that is visually perceived by humans as colours. The

5. See Georg Simmel (1895: 40–41).

crucial thing is that unlike humans, the Simmelians immediately experience electromagnetic waves of this part of the spectrum *as such* – that is, *qua* electromagnetic waves. Thirdly, they experience them in such a manner that the numerical expression (say, in nanometres) of the length of the experienced wave pops up in their minds with roughly the same immediacy with which the number of, say, visually perceived apples on a table, if that number is equal to or less than three, pops up in our human minds.

Because of that, if in the Simmelian philosophical culture there is an analogue of our logical positivism, the Simmelian logical positivists would count, among their Simmelian *Protokollsätze*, such Simmelian sentences that their English translations are, e.g. ‘This surface is (now) reflecting electromagnetic wave of the length of 680 nm’ and the like.

Fourthly, their sensory systems are such that they do not allow human colour phenomenology, so that not only Simmelian translations of English sentences like ‘This surface is red’ are not Simmelian *Protokollsätze*, but even it is not immediately clear that such English sentences as ‘This surface is red’ do have adequate Simmelian translations at all. (We will discuss this issue in some detail below.)

Keeping further details of Simmelians’ sensory system indefinite, I will proceed on the assumption that there is a consistent way of completing the description of their sensory system such that the complete description would be compatible with the list of features I am going to impute to the Simmelian language. The list is this: In the Simmelian language

1. there are such predicates that can be translated with English predicates of the form ‘_ is (now) reflecting electromagnetic wave of the length of x nm’;
2. there are syntactic and lexical means sufficient for constructing such complex predicates that can be translated with English predicates of the form ‘_ is (now) reflecting electromagnetic radiation with the wavelength within the range between x nm and y nm’;
3. if in the Simmelian language there are means of translating English sentences of the form ‘ x is red’, ‘ x is green’ and the like, then the means are those characterised in clauses (1) and (2) above.

(Note that clause (3) only implies that there are no *one-word* Simmelian translations of English colour terms like ‘red’, ‘green’, etc. It does *not*, by itself, predetermine an answer to the crucial question of whether English colour sentences have *any* Simmelian translations at all.)

Again I would like to keep further details of the Simmelian language indefinite in order to avoid any unnecessary loss of generality.

5 Are Simmelian and English intertranslatable?

Suppose that the Simmelian sentence S_s is a translation of the English sentence S_e . Then S_s and S_e must have the same truth conditions.

Consider now an English colour sentence, say, 'This surface is red'. Suppose that in Simmelian there are words that can serve as adequate (at least adequate for the context under consideration) Simmelian translations of English 'this', 'surface' and 'is'. Denote them 'Simm(this)', 'Simm(surface)' and 'Simm(is)', respectively. Suppose further that the Simmelian grammar is such that if there is an adequate translation of 'This surface is red', then it begins with the concatenation of words 'Simm(this) Simm(surface) Simm(is) ...'. But how could this beginning be completed?

From (1)–(3) it follows that the completion should have either the form 'is (now) reflecting electromagnetic wave of the length of x nm', or 'is (now) reflecting electromagnetic radiation with the wavelength within the range between x nm and y nm'.

But if it has the former form, then truth conditions of the two sentences will not coincide, whatever number we may substitute for x . Thus, suppose that we have substituted '680 nm' for x . Then if the length of the reflected wave is, say, 678 nm, then the English sentence would be true, whereas its Simmelian counterpart, false.

The latter form won't do either, as the completion. For whatever numbers we substitute for x and y , the truth conditions of the resulting Simmelian sentence would differ from the truth conditions of the English sentence. Suppose we have substituted '630 nm' and '700 nm' for x and y , respectively. Then, if the length of the reflected wave is, for example, 630 nm, then the English sentence will be neither true nor false, whereas its Simmelian counterpart, true. The reason why 'This surface is red' will be neither true nor false is that the English word 'red' is vague, that is, some points on the colour spectrum are borderline cases, as far as its application is concerned: They are neither red nor non-red, and the point of 630 nm is exactly one of borderline cases.

Now, because – due to point (3) – no other form of completion would result in an adequate translation, 'This surface is red' is untranslatable to Simmelian. Of course this failure of translation is immediately explained by the fact that such English colour words as 'red' are vague, whereas their closest Simmelian semantic counterparts are numerically precise.

6 Super-translatability introduced

But of course English and Simmelian are not *completely* semantically sealed for one another: There is the possibility of arranging a mapping between the two languages which, even if it is not a translation in the literal sense of the word, still is

such as to enable the interpreter to test whether the conceptual scheme under interpretation (construed as a theory) is largely true.

Suppose that the conditions of application of 'red' are as follows:

620/640 nm	borderline cases
640/695 nm	applicable
695/705 nm	borderline cases ⁶

Then, define set ST('This surface is red') as the set of all Simmelian sentences of the form 'This surface is (now) reflecting electromagnetic radiation with the wavelength within the (closed) range between x nm and y nm', with x running over (the Simmelian names of) all real numbers from 620 to 640, and y over (the Simmelian names of) all real numbers from 695 to 705. Define, finally, that (1) ST('This surface is red') is *true* if all of its members are true in Simmelian; (2) ST('This surface is red') is *false* if all of its members are false in Simmelian; (3) ST('This surface is red') is *neither true nor false*, if some of its members are true, while some are false in Simmelian.

Thus defined, ST('This surface is red') does specify truth conditions of the English sentence 'This surface is red'; actually, the former has the same truth conditions as the latter. ST('This surface is red') would, thus, deserve to be called a Simmelian translation of that English sentence, if not for the embarrassing fact that ST('This surface is red') is *not a sentence*! It is a *set* of sentences, instead. To mark this fact and the obvious links with the method of supervvaluations, we will call ST('This surface is red') a Simmelian *supertranslation* of the English sentence 'This surface is red'.⁷ This seems to be the most we can do by way of restoring the

6. This is still an oversimplification, because the boundaries between the borderline areas and the areas of applicability and non-applicability are themselves vague. To take that into account, we would have to move to fuzzy sets, which would complicate our discussion without making any principal difference to the final result.

7. In the formal semantics of vagueness, supervvaluationism is a method of treating vague predicates. A vague predicate fails to divide things into two sets: the set of true predications, and the set of false predications, because some predications are neither true nor false. This failure is considered by some semanticists to reveal a deficiency in the meaning of a vague predicate. The deficiency can be removed by assigning definite truth values to those predications that are neither true nor false. Any such assignment would be one possible way of making a vague predicate precise – one possible precisification of it. But because possible precisifications are many, to choose any one of them would be an arbitrary decision. One should, rather, come up with such a characterisation of the vague predicate at issue that would take account of all possible precisifications. Supervvaluationism proposes to do this by holding a predication sentence true iff it is true on all possible precisifications, false iff false on all precisifications, and neither true nor false iff true on some possible precisifications and false on others. For further details on supervvaluationism, see, e.g. Williamson (1994: 142–164), and Keefe and Smith (1996: 23–35), but note that no further details are needed for the purposes of understanding the remaining part of this chapter.

semantic traffic between English and Simmelian.⁸ In particular, we can't hope to convert an ST-set to a single compound sentence with members of the ST-set as its ultimate constituents.

7 Super-translatability enables a Simmelian truth-tester to test whether the English conceptual scheme is largely true

To remind, the crucial feature of the situation under discussion is that such English sentences as 'This surface is red' have no Simmelian translations because no Simmelian sentence has the same truth conditions as the English sentence 'This surface is red'. Still, I claim that it does not prevent a Simmelian truth-tester, equipped with the set ST('This surface is red'), to successfully test whether 'This surface is red' is true in English as relativised to some individual surface x . To do that, she has to check the truth values (in Simmelian) of all the members of ST('This surface is red'), as relativised to the selfsame surface x . If all the Simmelian sentences from ST('This surface is red') are true in Simmelian (relative to x), then 'This surface is red' is true in English (relative to x). If all of them are false in Simmelian, then 'This surface is red' is false in English. If some are true and some false, then 'This surface is red' is neither true nor false in English (relative to x) because x presents a borderline case between being red and not being red.

The gist of the matter is that the set ST(S), where S is an English sentence with occurrences of vague colour terms, is no worse a tool for testing the truth value of S than a Simmelian translation of an English sentence without vague terms. And because all English sentences are either translatable or super-translatable to Simmelian, the Simmelian truth-tester, if she knows translations and/or super-translations of all English sentences, should meet no insurmountable obstacles in determining their truth-values in English.

That completes a demonstration of my claim that statement (6) in Davidson's first argument is false, and the whole argument is, thus, flawed.

Some comments might be in order at this junction. Firstly, the failure of translation due to the mismatch between English vagueness and Simmelian (relative) precision is one-way. While English sentence 'This surface is red' goes without a proper Simmelian translation, Simmelian sentences about wavelengths were *introduced* with the help of their English translations (see Section 4 above), and actually I would be hard put if I were asked to invent some other method of their introduction. In actual fact, English is strictly richer in this respect than Simmelian, because in English we have both *vague* colour words and their *precisifications* in

8. Even more can be done, if we are prepared to use fuzzy sets and fuzzy logic, but see Note 7.

terms of wavelengths. That being so, a natural question arises: 'Why can't Simmelians enrich their language and, thus, remove the untranslatability at issue by, say, just borrowing English colour words like 'red', 'blue' etc. together with their vague English meanings?' The answer is, 'They certainly can', but this answer comes with a couple of caveats.

Caveat One They can borrow English vague colour words from English, and they can borrow Hopi vague colour words from Hopi, and if you name any other language with vague colour words, they can borrow them from that language. They can borrow vague colour words from *any* language that has such words. But to conclude from that that they can borrow vague colour words from *all* actual and possible languages is to commit the modal version of the fallacy of composition, because all actual and possible languages are indefinitely many so that it is not a foregone conclusion that the number of words to be borrowed is finite or even that all their (vague!) meanings could be recursively generated on the basis of some finite stock.⁹ And if they can't borrow vague colour words from all actual and possible languages, then the existence of a pair of languages with partial failures of translatability between them is perpetuated for an indefinitely long time.

Caveat Two Because of the differences between our and Simmelians' sensory systems, the linguistic situations of the two respective communities are *not* symmetrical. We English-speaking humans have *both* our eyes that are naturally attuned to vague colour predicates *and* our equipment for measurement of wavelengths, so that both sub-languages of English – that of vague colour words and that of wavelengths – are grounded either in our biological or in our technological equipment. But Simmelians only have their biological devices for measurement of wavelengths, so that borrowing vague colour terms from other languages would leave the use of such terms by Simmelians referentially ungrounded.

Caveat Three What is perhaps most important is that the lack of vague colour terms and the ensuing partial untranslatability does not prevent Simmelians from successful radical interpretation of us English-speaking humans with our vague colour terms. The procedure has basically the same stages, the only complication being that at the stage of seeking for appropriate assignments of truth-conditions to individual English sentences, the Simmelian radical interpreter will have to keep in mind a possibility that some English sentences are such that their Simmelian counterparts will be sets of Simmelian sentences rather than single sentences.¹⁰

9. But cf. interesting findings and discussion thereof in Berlin and Kay (1969) that might perhaps be used to construct an argument against this point of mine.

10. Actually, there are more than one method of representing linguistic vagueness with non-vague means. Supertranslations are just one of them. For more advanced methods, see, first of all, Black (1937). Black's paper is an early precursor of the methods of fuzzy logic.

8 English and Simmelian, two languages with non-radically different conceptual schemes

As against Davidson's first argument against conceptual schemes, our thought experiment produced an example of two languages with partial failures of translatability between them. But even in the presence of partial untranslatabilities, a Simmelian truth tester finds herself in a position to successfully test whether the English conceptual scheme is largely true. Because of untranslatability, the two conceptual schemes cannot be one, but difference between them is not radical in that successful interpretation is possible in both directions.

Still there is some unfinished business. Given our modification of Davidson's second argument against conceptual schemes, we have yet to check that the Simmelian radical interpreter's decision to interpret the English sentence 'This surface is red' as strictly untranslatable to Simmelian is the only rational option to her, given the principle of explicability. The problematic case for the radical interpretation process is the expression of indecision, doubt, suspension of judgment and the like, on the side of English speakers, when they are presented with what they perceive as a borderline case of applicability of the vague predicate 'red'. The persistence and regularity of the cases of such expression has to duly perplex the Simmelian interpreter, because, as we know by now, the closest Simmelian analogues of English colour predicates have no borderline cases at all (at least not on the scale that would be detectable by the human eye). Gradually she gets aware that she faces a problem. In principle, there are two possible solutions: The first one being to attribute the frequent expressions of indecision to an error of opinion on a mass scale, the second one, to admit that 'This surface is red' is, strictly speaking, untranslatable to Simmelian. The preference has to be given to that one which is supported by a better explanation.

Now, an error of opinion on a mass scale, given no changes whatsoever in light conditions, alertness, willingness to cooperate and other relevant factors, is puzzling indeed. On the other hand, if the interpreter is aware of the drastic difference in the design of sensory systems of humans and Simmelians, that awareness constitutes a solid ground for an excellent explanation of imputed untranslatability: The design of the human eye and vision resulted in vagueness of ordinary English colour concepts and colour words; their vagueness amounts to availability of borderline cases; the availability of borderline cases explains expression of indecision on a mass scale. Ultimately, the mismatch between English vagueness and Simmelian (relative) precision explains the imputed untranslatability.

Thus, in actual fact, the Simmelian interpreter is *forced*, by Davidson's principle of explicability, to decide in favour of untranslatability between the two languages.

Now at last we have satisfied ourselves that, in the light of those very principles that Davidson established in his theory of radical interpretation, English and Simmelian have been shown to have non-radically different conceptual schemes.

9 Concluding remarks

It stands to reason that one's answer to the question whether there can be intermediate cases between the two extremities: The one of sameness of conceptual scheme in different languages and the one of radical difference of two schemes resulting in mutual unintelligibility – one's answer to this question depends on many things: In particular, on how one understands the notions of radical and non-radical differences between schemes, what is one's stance on the issue of relations between translatability, interpretability, and intelligibility, and so on. But I think it is methodologically advantageous to have a theory of conceptual schemes that would allow for a whole spectrum of possible relations between languages, not just for extreme cases.

There may be at least three reasons why the Simmelian thought experiment that was developed and discussed in this chapter may prove useful for the purposes of developing such a theory. First of all, even if the thought experiment itself might seem to be very remote from real-life languages, involving, as it does, such exotic, science-fiction creatures as Simmelians, in fact part of the significance of the Simmelian thought experiment lies in the fact that a language like Simmelian might well serve as a mediating language, a fine-grained calibration device, for a pair of mundane languages that are mutually untranslatable due to mismatches in vagueness profiles of their respective colour words, even if the order of vagueness could be roughly the same in both languages.

Second, vagueness is a pervasive phenomenon in natural languages going far beyond colour lexicons. Hence a wide range of potential applications of the device of supertranslations.

Last but not least, *raison d'être* of the device of supertranslatability is to serve as a pretty regular, systematic *Ersatz* for unavailable (due to the vagaries of vagueness) translations between two conceptual schemes that are different but not radically so. Now, vagueness is just one of several major sources of partial untranslatability between natural languages. Can it be the case that *Ersatzes* of comparable degrees of systematicity could be found for other major sources of partial untranslatability? It is hard to say at this stage, but the very idea of seeking such systematic *Ersatzes* for each of several major sources of partial untranslatability may prove to be a promising strategy.

References

- Bar-On, Dorit. 1994. Conceptual relativism and translation. In Gerhard Preyer, Frank Siebelt & Alexander Ulfig (eds.), *Language, mind and epistemology*, 145–170. Dordrecht: Kluwer.
- Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Black, Max. 1937. Vagueness: An exercise in logical analysis. *Philosophy of Science* 4. 427–455.
- Davidson, Donald. 1984. On the very idea of conceptual scheme. In *Inquiries into truth and interpretation*, 183–198. Oxford: OUP.
- Keefe, Rosanna & Peter Smith. 1996. Introduction: Theories of Vagueness. In Rosanna Keefe & Peter Smith (eds.), *Vagueness: A reader*, 1–57. Cambridge, MA: The MIT Press.
- Rescher, Nicholas. 1980. Conceptual schemes. In Peter A. French, Theodore E. Uehling, Jr. & Howard K. Wettstein (eds.), *Studies in epistemology: Midwest studies in philosophy* V, 323–345. Minneapolis, MN: University of Minnesota Press.
- Simmel, Georg. 1895. Über eine Beziehung der Selektionslehre zur Erkenntnistheorie. *Archive für systematische Philosophie und Soziologie* 1. 34–45.
- Williamson, Timothy. 1994. *Vagueness*. London: Routledge.

Visual representation in a natural communication system

What can signed languages reveal about categorisation across different modes of representation?

Dorothea Cogill-Koez

Both signed and spoken communication systems exploit multiple modes of representation, including abstract symbolisation and manual depiction. This study outlines the highly structured system of manual depiction that is argued to exist, alongside the regular linguistic system of abstract symbols, in signed languages from America to Thailand, and explores the significance of this system for categorisation in natural communication systems. It is shown that mode of representation (manual depiction, rather than manually-presented abstract symbols) has real, but only limited, effects on the predominance and the content of mental categories. Comparisons with hearers' iconic gesture, spoken language classifier systems, and systems of drawing and modelling make these unique signed systems of manual depiction a valuable resource for research into categorisation.

1 Introduction

Does the human mind work on the basis of categories? Does categorisation work in the same way across the domains of human language and behaviour? The aim of this study is to outline the relatively elaborated system of manual depiction that is found, alongside abstract symbolic systems, in signed communication from America to Thailand, and to consider some early implications and potential of this pictorial strategy of representation for helping to answer the above questions.

Both signed and spoken communication systems exploit multiple modes of representation. Obviously, speech-centred communication systems draw on abstract symbols (such as the word 'circle'). However, among further possibilities, they also routinely incorporate depictions created with the hands. Across different

cultures there is common gestural habit, while speaking, of tracing or placing in the air contextually-relevant shapes (such as a circle). After a period of strong focus on the presence and nature of their abstract symbol system, signed communication systems are now also increasingly modelled as including a second system of manual depictions. However, I have argued that in signed systems the realisation of manual depiction shows vastly more formal structure than is found in the manual depictions of speakers. This indicates that in signing this depictive system of representation exhibits distinctive cognitive features relative to its less-sophisticated speech-accompanying cousin.

From this starting point, it will be shown that when signed communication is based on the principle of depiction, there may at times be little evidence of categorisation as a controlling factor in the signed form produced and hence, by implication, in the underlying mental representation. (This implication can be drawn in part because, unlike some aspects of spoken language phonology, categorical perception does not appear to be a part of signed language processing; Newport 1982.) Nevertheless, when signed systems of manual depiction are considered as a whole, the evidence suggests that mental categories are indeed active: That while signers' manual depiction can dispense with them, it usually does not. In this, signers' manual depiction is in fact similar to other and better-known forms of pictorial representation, such as children's drawings, cartoons, and indeed the sketching of most adults.

Both in itself and in concert with partially similar, partially different "sister" systems of representation, manual depiction in signed communication systems may thus be an extremely fertile field for testing and developing models of categorisation in human thought.

2 Representation in speech-centred communication: Abstract symbols and depiction

The first step is to describe the double representational system present, in different forms, in both speech-based and signed communication systems. This will require the clarification of some terminology, an area fraught with pitfalls in cross-disciplinary discussion. The generally more familiar forms of speech-based communication systems will help to introduce not only the relevant terms and concepts used here, but also some important initial assumptions about representation which speech-based systems have tended to impose on researchers in signed languages. Those familiar with gesture research, in particular, will be aware that the following outline glosses over great and in some ways relevant complexity, but the purpose here is to quickly lead readers unfamiliar with signed languages into the

problems and the possibilities these languages pose regarding categorical form and thought.

That said – English speakers, at least, are accustomed to making a distinction between gesture and speech. The primary distinction is one of what will be called channel, or the physical means of production and perception: Speech uses the oral/aural channel, while gesture is manual/visual.

For users of a speech-centred communication system, though, this distinction in channel then maps fairly well onto many others. First among these is principle of representation. Throughout this chapter, “principle of representation” will refer to an inferred cognitive process; it is the cognitive strategy by which a sign, word, gesture or other communicative act evokes or is linked to its meaning. The principle of representation is therefore distinct from the observable sign, word or gesture itself.

In spite of growing interest in iconicity in spoken language, it remains undisputable that the vast majority of spoken words are related to their referents in an arbitrary fashion. I will call the cognitive strategy of arbitrary association the “abstract principle of representation”, and call the forms that are so processed “symbols”. The reader is urged to note carefully both the narrowly Peircian and the fundamentally processing-oriented meaning attached to “symbol” in this study.

Some classes of gestures, on the other hand, clearly rely on a different strategy to represent meaning. They depict; they evoke their referent by means of an online, detected structural match between salient aspects of the form of the gesture, on one hand, and selected, significant aspects of the form of the referent on the other. Such forms will be called depictions or pictures, and the strategy of structural matching utilised, the principle of depiction.¹

Over and above this contrasting principle of representation, there are further differences between speech and gesture. For the purposes of this study it is significant that words give abundant evidence of being distinguished from each other formally by discrete, categorical contrasts, and in the process, of pointing to

1. In previous work I have called these forms “visual representations”, following the terminology used in the study of drawing. However, as an anonymous reviewer has pointed out, in yet other fields the term “visual representation” simply means a representation which is detectable visually – a most unhelpful clash, in the context of signed languages. Given my Peircian use of the term “symbol”, it might seem logical to use his corresponding term “icon”. However, in linguistics – including signed language linguistics – as regards the crucial contrast that will emerge between superficial form and underlying principle of representation, the word ‘icon’ is riddled with multiple beguilingly similar yet deeply conflicting meanings.

Ultimately, the best solution may be to abandon the symbol/picture distinction altogether; it will already be obvious that the contrast between “symbol” and “picture” here is not inherent in the form used, but in how the mind processes the form. Perhaps when the underlying processes of cognitive decoding are appropriately recognised and fully distinguished, no distinctive terminology need nor (logically) should be applied to the realised forms themselves.

underlying mental categories. Speech-accompanying gestures, on the other hand, lack comparable formal structure; in particular, depictive gestures are usually said to utilise analogue form, in which the shape of the gesture is improvised on the spot in an attempt to produce a one-off imitation of the form of the referent (McNeill, Levy & Pedelty 1990; McNeill 1992). With this lack of discrete, particulate structure comes a lack of evidence that the gestures point to distinct mental categories. For example, when a speaker supports his or her verbal description of their memory of a particular church window's shape with depictive gestures that outline the shape, the most parsimonious interpretation is that the gestures don't refer to any intervening mental category of shape, but are modulated to depict the unique visual memory of this particular church window's outline.

The binary contrasts above have influenced thinking about human communication not only in themselves – i.e. as paired contrasts – but also as sets. The dominance of speech-based systems is perhaps the reason why it seems not uncommon to assume some necessary connection between speech, abstract principles of representation, discrete and discrete-combinatorial structure, and categorical thought itself as one set, and on the other hand, between the elements that make up the bundle speakers call “gesture”: The use of the hands; pictorial principles of representation; analogue structure; the lack of mental categories. I would suggest that much of the history of signed language research has been controlled by the holding, and the breaking, of these two assumed sets.

3 Representation in signed communication: Abstract symbols

Thus for much of the twentieth century, the manual/visual signs of signed languages were assumed to consist of nothing more than a collection of “gestures”; that is, they could be assumed to convey meaning via improvised, analogue forms that could only be interpreted as pictures. To a user of speech-based systems this seems natural enough at first glance, even more so since in citation form many signs do seem to depict their referent, such as the sign from Auslan (Australian Sign Language) illustrated in Figure 1.

However, in the process of guessing at the meaning of this sign (which is deliberately left unlabelled at this point), the reader will notice that even when the form is approached as being a depiction, a very wide range of referents is still possible. From a pictorial point of view the actual gloss is certainly appropriate (and it is undisputed that the sign was originally *invented* as a depiction), but the form of the sign still underdefines the referent; that is, the meaning of the sign in Auslan is not confidently deducible solely from the picture created by the sign. (In the capitals that conventionally gloss elements of signed language lexis, it is TABLE.) Given

this little thought experiment, the reader can probably also accept that on the limited resources of the body and hands it would be difficult at least – impossible more likely – to create an adequate total vocabulary that relied entirely on depicting each referent well enough to create or map onto a picture of it in the interlocutor's mind. If pictorial representation on the body were that reliably informative, the game of charades would never have been invented.



Figure 1. A regular or “lexical” sign from Auslan: Following the usual convention, arrows outline the movement of the hands

Are the signs of signed languages, then, not simply pictures produced *de novo*, but *conventional* pictures; pictures the interpretation of which is supported by a single conventionally-agreed meaning? This is a possibility which will be revisited later in connection with the particular form of signing that is the focus of this study – but for now and for the bulk of regular or “lexical” signs, the answer is still no. For from about 1960 onwards, evidence accumulated to show that in spite of their often superficially pictorial form, at a deeper cognitive level most signs of signed languages were processed, like most spoken words, in an essentially form-blind fashion.

The lines of evidence are many. First was the finding that the signs of a given signed language were not in fact analogue in form; a substructure existed which could be analysed exhaustively in terms of a language-specific set of discrete, digital subelements within each of the core formal parameters of handshape, movement, location and orientation. Further, these subelements usually did not carry detectable independent meaning, and thus the overall pattern amounted to evidence of a phonemic level of form in sign. Next, of course, many signs show absolutely no deducible pictorial resemblance to their meanings anyway; fourth, even where a “picture” can be seen in the citation form of a sign (as for TABLE above), in real signed conversation, regular assimilation processes very frequently destroy the pictorial qualities of the physical form produced. This leaves no picture to be seen, conventional or otherwise. Fifth, regular grammatical patterns of inflexion and derivation also apply regardless of the potential picture in a sign, and similarly

can obscure or destroy it. Sixth, historical drift of sign forms shows little regard for retaining the fidelity of the “picture” but plenty of evidence for the influence of formal structure (i.e. phonological tendencies and constraints). Finally, damage to the brain that destroys pictorial skills leaves the bulk of signing unaffected – while damage that produces sign aphasia (stemming from damage to the classic language areas of the left hemisphere) conversely leaves pictorial skills intact.

These and further findings have all pointed towards the same conclusion. In spite of any pictorial look they may possess when one reflects upon the citation form, during normal conversation (or “online”) most signs of signed languages are not mentally processed by matching between their physical structure and a mental visual image of the referent. Rather, at a cognitive level, most signs are simply and directly linked to their meanings in the same way that a spoken word like *table* logically must be (Stokoe 1960; Klima & Bellugi 1979; Poizner, Klima & Bellugi 1987). Cognitively, they are symbols as defined herein, drawing on the principle of abstract representation.

By means of this discovery, signed language linguistics long ago drew a clear distinction between superficial form and deeper cognitive processing. This is arbitrariness of the sign with a new twist, Peircian “icons” and “symbols” with a cross-cutting layer of meaning which stresses, not what one can see in the overt representation when it is studied at leisure, or even its likely historical origin, but rather how the mind associates meaning with the representation during actual use. (It is a distinction with many implications, and one that may be submerged or even lacking in discussions of iconicity in spoken language.)

Two assumptions, easily made against the familiar background of speech-based communication, have been overturned here. The first is the assumption that manual channel equals depictive representation; the second is the assumption that any form which bears a structural match to its referent is necessarily processed, in online use, by means of those structural matches.

4 Representation in signed communication: Depiction

Work in the last decade, however, has begun to reveal greater complexity in modes of representation in signed languages. In particular, a class of signing historically called classifier predicates are now being isolated and reinterpreted as being in some respects true pictorial representations as defined above. The reader is warned that the nature of these “classifier predicates” is possibly the most controversial topic in signed language research at the moment, and the pictorial interpretation outlined below is rejected by some sign language linguists (who still hold to a fully symbolic model of virtually all signed representation), held in significantly different

forms by others (including, importantly, those who would restrict the pictorial interpretation to only the analogue components of classifier predicates), while the model below might be considered overly modest by still others, who call on a somewhat different concept of “iconicity” to cover, not only signed classifier predicates, but also many other aspects of signing. In keeping with the dispute over their true nature, the number of terms now used to label this mode of signing almost equals the number of researchers who address them, so it is for the sake of ease of the reader, rather than accuracy of terminology, that the old label “classifier predicate” and the acronym “CP” will be used here.

That said, CPs are a class of signing used exclusively in the representation of visual-spatial situations and events. They are used typically only after lexical signs or clear context have identified the real-world objects or participants in the events (tables, perhaps, as opposed to square stools, or sheds); thus, if CPs are manual depictions, the problem we have noted of adequately identifying their referents from the limited depictive resources of the body is circumvented. CPs are used to then add further information about the shape, location, movement or manipulation of these objects.

Formally, like lexical signs, CPs draw very heavily on limited sets of discrete, repeatedly-used handshapes, movements and, in some respects, orientations and even locations (Supalla 1982). This finding was key in a shift towards analysing CPs as being cognitively symbolic rather than depictive. As a result, for about two decades, CPs too have been analysed using exactly the same grammatical models as are applied to the bulk of signed language. However, there are a number of significant differences between CPs and the regular lexical signs described above.

First, it is now increasingly recognised that some elements of CPs can take on meaningful analogue variation. Prominent here are the realisations of location and of orientation. For this reason and others, just those elements of CP form which *cannot* be interpreted in terms of a finite set of discrete, recognisable, repeatedly-used formal elements are now modelled by some as pictorial (Liddell & Metzger 1998; Liddell 2003; Schembri 2001; Schembri 2003; see also Taub & Galvan 2001).

There is a logical difficulty apparent in modelling the physical location and orientation of a handshape as depicting that of a physical referent, without some prior pictorial cognitive process in mapping the handshape itself on to the referent (Cogill 2003).² Apart from this however, and most interestingly, there is abundant

2. There are also further difficulties with the partially-pictorial model in its current realisation (Liddell 2003), chief among them being that it necessitates a geometric increase in the number of morphemic elements that must be nominated in CPs (Cogill 2003). Uncontrollable elaboration of necessary morphemic elements has long been troublesome within fully morphemic models of CPs (Liddell 1995), but partially-pictorial models to date suffer from the same problem to an even higher degree.

further evidence that the discrete, repeatedly-used elements of CPs are also created and interpreted using the pictorial cognitive strategy.

The evidence for their genuinely pictorial online processing comes from the same diverse range of sources as has long proven the genuinely abstract cognitive processing of the bulk of signed languages (Cogill-Koez 2000a, 2000b). First, with some caveats that will appear below, CPs do not sometimes but always retain a physical structural correspondence with their referents. Additionally, as the formal elements relate on a one-to-one basis to aspects of form of the referent, evidence for a phonemic level is lost. Third, although CPs show conventions of use, with conventional choices varying slightly from one signed language to another, these conventional choices always retain visual-spatial features that allow them to remain functional as depictions or models of their referents – much as styles and conventions of art may vary from culture to culture, while still remaining comprehensible depictions. Fourth, damage to the brain that affects pictorial skills strikes also at these “classifier predicates”, while left hemisphere damage that produces sign aphasia leaves them untouched. This is true even within a single individual, a hearing interpreter of American Sign Language (ASL) who sustained such damage to her right hemisphere (Emmorey 1996). Her subsequent loss of ASL CPs was not accompanied by loss of control of English sentences of identical content (such as ‘put the pen on the paper’), suggesting that it is not simply the visual-spatial subject matter of CPs which makes them dependent on intact pictorial skills; rather, it is indeed the pictorial mode of representation itself. Finally, (as will be discussed further below), children’s and indeed most adults’ drawing relies largely on discrete, standard, repeatedly-used graphic forms to construct a picture, just as is seen in CPs.



Figure 2. An Auslan CP, used here to show size and shape of a carton: Despite formal similarity with TABLE in Figure 1, in online use the former draws on an abstract form-meaning link, while the CP is argued to use a pictorial cognitive strategy

Again, note that the superficial iconicity of the forms, while clearly essential now to *any* model of CPs as depictions, is in fact a minor part of the evidence on which the comprehensive pictorial model rests. Rather, just as was the case for lexical sign, multiple diverse lines of evidence all point towards a particular basic strategy of cognitive processing for both analogue and discrete forms in CPs. However, where for lexical sign the underlying cognitive process appears to be that of direct linking with meaning, for CPs the same lines of evidence point to a pictorial strategy for their creation and decoding. Comparison of Figures 1 and 2 will highlight the limited relevance of iconicity to arguments about deeper cognitive processing.

The interpretation of CPs as being the product of a pictorial strategy, both in their analogue and in their discrete elements, has implications for categorisation in natural communication systems. It also highlights alternatives regarding the possible causes or sources of categorical form in pictorial representations, and shows how the trend towards modelling CPs as depictions may allow the investigation of diverse further specific questions regarding representation and mental categories. These last will be covered below in an illustrative range only, as it is too early in the development for CP data and models to bear much weight of further theoretical claims. However it is hoped that the remainder of this chapter and the illustrative cases raised at the end will whet the interest of a cross-disciplinary audience, showing how CPs, as the only known form of highly structured pictorial representation that can be used in natural conversational flow, can be used to help triangulate on otherwise elusive questions of how humans think about events and objects.

4.1 Analogue representation strategies: No evidence for visual-spatial categories

The first thing CPs demonstrate is that within a natural communication system, representation of visual-spatial form, and hence by implication human thinking about such form, can be category-free. Whether we wish to depict that an object had this shape rather than another, was located here rather than anywhere else, moved in just this way or rotated thus far, no more and no less, pictorial representation usually offers formal possibilities for creating a unique attempt at mapping between the visual-spatial form of the representation and the visual-spatial aspects of the referent. When the strategy of construction in a depiction is to improvise a form, with the aim to control the form of the representation at every point to make it correspond with the form of the referent, pictorial representation is said to use an analogue strategy. Importantly for the argument here, when one constructs depictions – or even just some aspect of it – by using the analogue strategy to convey detailed meaning, it is possible that one is also drawing on an underlying conception

of the relevant aspects of the referent's visual-spatial form that is also free of categorical contrasts.

As mentioned above, it is increasingly acknowledged that across different signed languages, CPs do indeed exploit analogue variation in form in a meaningful way. In fact, they can do so not just on some but on all of the formal parameters by which signs have been traditionally described; i.e. movement, location, orientation and handshape. For example, it appears that all signed language CP systems described so far can show detailed aspects of the location and movement of one person relative to another using variations of location and movement such as those in Figure 3.

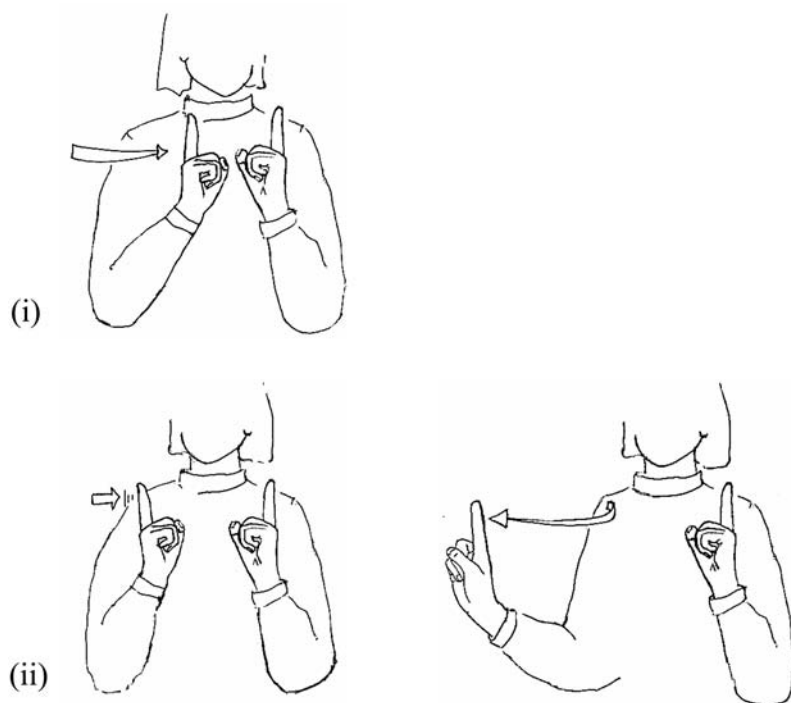


Figure 3. (i) One person approaches another: A CP signing, that uses only standard CP elements in all CP systems described to date. It could be used in most (though not all) signed languages to represent one person approaching another. (ii) One person starts to approach another, hesitates... then turns and goes away: An equally common possibility that appears to use meaningful analogue form

In all CP systems described to date, the signing in Figure 3 (i) can be formed largely or entirely from nomifiable, discrete CP options regarding the realisation of

handshape and movement – even perhaps orientation and relative location. However, equally widely, the meaningful deployment of variations in location and movement in Figure 3 (ii) challenge traditional CP inventories, appearing to use non-standard, analogue improvisation. This is true even while the handshape choice remains standard.

Within the physical limitations of the hands, form itself can also be modified. In Figure 4, a child has been discussing the actions of a little toy aeroplane seen in a video animation. The video showed the peculiar event of one of the model aeroplane's wings "unfolding" from a tucked-up position against the fuselage to the normal extended position for an aeroplane wing. The child shows this by beginning with a handshape in which the thumb was tucked up against the hand. The child then straightened the thumb to create a pictorially-meaningful transition to the final handshape. The final form is the standard ASL choice for CPs representing aeroplanes, but the particular meaningful change in handshape that preceded it has no standard precedent in ASL CP handshapes.

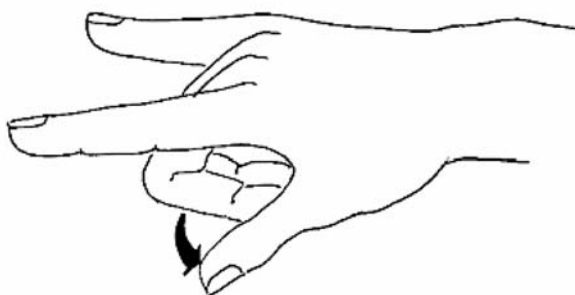


Figure 4. An ASL CP used to represent the unfolding of a toy aeroplane's wing. Based on a description of an ASL signing by Supalla (1982)

Similarly, Emmorey and Herzig (2003) found that when signers wish to show the precise size of a medallion hanging from a chain around someone's neck, they can do so with surprising communicative accuracy. By producing a CP depicting a rounded shape located on their own chests and adjusting the amount that their fingers are extended or curled, signers were shown to convey size contrasts well enough for a signing interlocutor to pick out and match up the corresponding size of medallion from a range of possibilities. Thus, even where the depiction of shape drew on two standard options for depicting circles, the depiction of the *size* of the medallion appears to have been achieved without reference to standard categorical contrasts. Furthermore, signers were more successful than nonsigners at the same

task. This suggests that, with the routine reliance on CPs as a major mode of representation, signers are more attuned to these uses of noncategorical pictorial contrasts than are nonsigners.

The first important point about CPs, then, is that they are a natural communication form that can and does convey meaning about the visual-spatial aspects of objects and events using noncategorical form. In these cases, signers must certainly be drawing on a principle of similarity in order to construct (and decode) a depiction of a referent, but it does not appear necessary that this principle of similarity creates or reflects any higher or more significant, independent mental category or concept to which the shape of the referent and the shape of the depiction can both be referred. The implication is that in CPs, communicators are not actually constrained to think about visual-spatial reality in categorical ways, either. Even in the midst of a flow of conversation, they can both represent and think about visual-spatial aspects of events and objects in ways that are category-free.

I call this point “important” rather than interesting in itself. This is because, after all, the depictive gestures that accompany speech often use the same analogue strategy, and also within the time frame of conversation – leading to a model of “thinking for speaking” that may also be of interest to those pursuing issues of categorisation (McNeill & Duncan 2000). The fact that signers can and do use an analogue strategy of manual depiction becomes important, rather, when one contrasts it with the presence of other depictive strategies present in CPs.

4.2 Discrete combinatorial representation strategies in CPs:

Depicting categories of visual-spatial form

Across different signed languages, CPs do additionally show abundant evidence of using standard, discrete forms. Standardised, conventional and to this degree categorical representation appears at two levels in CPs; at the level of general depiction of visual-spatial form as such, and at the level long called Semantic (cf. Section 4.2.2) in signed language linguistics (Supalla 1978); that is, categories such as “human”, “animal” or “vehicle”, which are not normally held to be defined primarily by visual-spatial qualities. This in turn suggests that signers have a basic preference for approaching visual-spatial phenomena in category-based ways, even in the presence of a realised alternative pattern of thought and representation.

4.2.1 *Depicting visual-spatial information using pictorial exemplars or prototypes*

For every CP system studied to date it has been possible to nominate a set of discrete, standard, repeatedly-used handshapes, movements and perhaps even (in a more restricted sense) locations (Cogill 2003). Each of these standard contrasts obviously has distinctive visual-spatial qualities; consequently, it can work as a

pictorial exemplar of these qualities. The resulting standard set of elements can be drawn upon in the construction of a CP “picture” or manual model, much as one assembles a model object from Lego parts by matching up aspects of the referent’s with the form of the available standard parts.

The set of nominated handshapes varies with the signed language, and although the subset of most-frequently-used CP handshapes may not be very large, the entire set of CP handshapes is now also typically rather long – substantially longer than that used for the phonological contrasts of nonCP signing in the same signed language. A much smaller set of repeatedly-used movements is usually nominated as being drawn on to depict movement (as in Figure 3) or shape (as in Figure 2). Movement of the hand in a straight line, in an arc and in a circle are cross-linguistically common as standard options in a signed language’s CPs, as is stasis, sometimes preceded and apparently emphasised by a short, sharp downward motion as if “setting down” the object referred to by the handshape, at that particular location (‘this object is/was *here*’). Turning to location, it may even be possible to identify simple contrasts of location which are used repeatedly to depict contrasts such as ‘ground level’ versus ‘high’ (Cogill 2003), while work for American Sign Language (ASL) has also nominated routine contrasts in the spatial relationship between two handshapes where these represent the spatial relationship between two objects (Supalla 1982).

The use of a standard option results in a choice of form that, clearly, cannot have been *exactly* like that of the referent in real life, or in the signer’s visual memory of a unique event or object. The final depiction is also imperfect to this degree (that is, as a depiction of a unique event). Nevertheless, as depictions, the strategy and its products are quite adequate to convey a great deal of everyday information about objects and events. The same strategy and the same effect is seen when the limited range of plastic forms in Lego are assembled to produce a variety of inexact but quite recognisable models, or when figures in cartoons such as South Park are displayed with largely standard orientations and in repetitive movement patterns to produce highly stylised but effective depictions of people walking, running, jumping and falling over.

As depiction, then, the standard CP element is working as an approximation of a selected, contextually-relevant visual-spatial feature of the referent. Cognitively, though, it seems to be ‘depicting a category’ of visual-spatial form; an idealised prototype or perhaps a maximally typical exemplar of some aspect of visual-spatial shape or movement (see Figure 5).

4.2.2 *Depicting semantic categories using pictorial exemplars or prototypes*

Additionally, certain assemblages of elements are routinely drawn upon in the depiction of certain classes of object, such as people, or cars, that are not usually

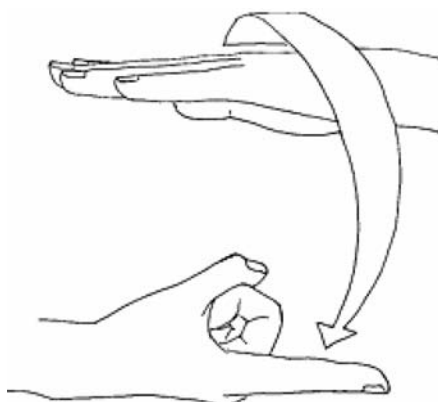


Figure 5. The point hand, flat open hand and arc movement are standard elements in all CP systems to date. Mapped onto visual-spatial features of objects and events – such as a pen falling from a shelf – they imply categorisation of visual-spatial form

considered to be defined entirely by their visual-spatial qualities. These assemblages can then be combined with other CP elements (either digital or analogue) to create a depiction of the total visual-spatial event the signer wants to represent. In diverse models of CPs, this further level of conventionalisation is typically taken as evidence of categorisation at work at a level beyond the visual-spatial, and such CPs have long been called “semantic”. I will capitalise this as Semantic (although the capitalisation is not the custom in signed language linguistics), in order to distinguish this label of a class of CP elements from the more ordinary uses of the term. I will however retain the label itself, because it has been in use for as long as CP analysis itself and has been very influential in almost all CP models.

The Semantic class of CP handshapes has its origin in Supalla’s foundation model of CPs (Supalla 1978, 1982). This model strove to interpret all aspects of CP form in terms of abstract cognitive representational strategies, and thus in the case of Semantic CPs the imperfect match between form and referent was extremely important; together with certain regularities in patterns of use of the Semantic handshapes, it was taken as evidence that the form referred to a category quite independently of the visual-spatial qualities of exemplars of that category.

However, in a fully pictorial model of CPs, the imperfect formal match is seen rather as the effect of use of a prototype. Thus it is qualitatively no different from what has been seen in other discrete CP elements so far. For the vast majority of these standard assemblages or potential Semantic-containing CPs, the evidence of influence of an underlying category that relies crucially on non-visual-spatial features of the referent now lies only in a somewhat subjective assessment of the degree of conventionalisation of choice and assembly of elements.

For example, the use of a straight index finger in an upright orientation can be used to depict many objects that share a shape that is long in one dimension, and is positioned vertically. It is also a standard option to show the movement or location of a person in Auslan, American, and British Sign Language CPs (among many other signed languages), as seen for example in the CPs in Figure 3. The choice of upright index finger here may at first glance seem extremely natural and obvious on purely visual-spatial grounds. After all, an upright person is also long in one dimension and oriented vertically: If the upright index is a normal choice for objects with such visual-spatial qualities, its choice in depicting the motion or location of a person seems to require no further categorisation of the referent on non-depictable grounds. However, the upright little finger and the upright thumb are used as the standard options for the human figure in Taiwan Sign Language, with the little finger representing the typically slither female and the more robust thumb the man (Smith 1990). In Auslan, either choice would be most odd.

This gives a taste of how CPs might work as conventional depictions of categories of object, not just categories of visual-spatial form. The conventionalised representation chosen in each case is certainly adequate to work as a depiction of the salient visual-spatial features of an upright person, but the level of conventionalisation in the choice suggests that the choice is influenced by considerations beyond the purely visual-spatial. One might say, then, that the category of human being is *depicted* in signed language CPs via the visual-spatial qualities of a default exemplar or a prototype of the category. In this way, categories such as “human”, or even “man” and “woman”, may help to shape signed pictorial representations.

4.3 Symbolic tags added to depictions:

Categories both depicted and symbolised

Across signed languages described to date, two known CP standard assemblages effectively contain distinctions or elements of form that cannot also be mapped onto a pictorial prototype. In this way, they contain elements of form the purpose of which is only to tag the depiction via abstract means, marking it as belonging to narrower category within the range of possibilities depicted by the highly-imperfect medium of the hands. The forms on which these “symbolic tags” appear also possess all the visual-spatial features necessary to work as a depiction of the objects which they model, so the assemblage can then be oriented, located or moved about in the normal CP fashion to show the orientation, movement or location of the specific referent.

It is a minor pattern; we will see later that in CPs symbolic tagging is less common than its presence in other genres of pictorial representation might predict. However, the interpretation of the standard assemblages above, as depicting

Semantic as well as visual-spatial categories, is supported by the apparent “symbolic tagging” of just a few Semantic classes.

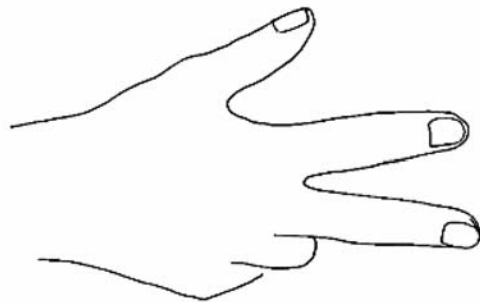


Figure 6. The distinctive ASL handshape-plus-orientation assemblage used to depict movement and location of vehicles: This assemblage is now appearing as a borrowing in numerous unrelated signed languages

One example is the ASL vehicle classifier (Figure 6). In folk etymology, this tri-branching handshape has been said to originally depict the mast and spars of a sailing ship. However, in its common contemporary use for the depiction of location, orientation and movement of a land vehicle such as a car, these elements of its shape have no visual-spatial correspondent at all.³ Yet in other visual-spatial respects, the handshape has the same qualities of the open flat hand that is a far more common cross-linguistic choice for the representation of vehicles. That is, the ASL ‘vehicle’, like the cross-linguistically common open flat hand, has visual-spatial qualities that can map unequivocally onto a top, bottom, front and back of a vehicle, both handshapes have one prominently extended dimension that can be mapped onto the similarly-extended length of a vehicle (as opposed to its height or width), and both handshapes are easily and naturally moved in the direction of the fingertips thus marked ‘forward’. For the ASL signers’ handshape choice, then, its as if, given a rather limited range of basic Lego shapes, one decides that sticking a small white block on top of a long red and otherwise all-purpose block means that the pair is from now on a ‘vehicle’.

The second example is given by the Danish Sign Language use of the fist with protruding upright thumb – a handshape and orientation complex familiar to many from the ‘thumbs up’ gesture of approval. In Danish Sign Language this form

3. Though Schembri (pers. com), notes the occasional humorous use that identifies the structure of the protruding thumb with the upper body of the seated driver, such that when the ‘driver’ gets out of the ‘car’, the thumb of the usual ASL vehicle handshape in Figure 6 is quickly dropped down parallel to the index finger to depict the driver’s absence, now, from the car.

is used to show the orientation, location and movement of people when thought of in a particular way (Engberg-Pedersen 1993), one which might be crudely summarised as moving in company or in competition. Here, the structural qualities of the handshape do make it a reasonable choice for the representation of a person on the limited resources of the hands, but in Danish Sign Language the choice of this handshape, rather than one of the language's two other conventional handshapes for people, symbolises something about the relationships of the people depicted – something which has no correspondence in the structure of the referent at all. Sticking with the Lego analogy, one could compare it to choosing green-coloured upright blocks, rather than red or white ones, to flag the people thus depicted as being engaged in sporting activities.

The Semantic categories, symbolically tagged or not, are not numerous in CPs, but across different signed languages one or more has so far always proven nomina-ble for people and for at least some classes of vehicle. Aeroplanes, trees, and animals have also been suggested to be depicted in some languages by conventional assemblages, indicative of underlying categorisation that is not purely visual-spatial (Supalla 1982).

CPs, then, depict rather than symbolise, and they depict largely via the use of standard visual-spatial forms. At the most basic level these appear to categorise only visual-spatial qualities. However, for some classes of object these forms are then regularly combined into standard conventional depictions, indicating the presence and influence of categories of thought beyond the purely visual-spatial, and the interpretation of these models as marking Semantic categories in the minds of signers is supported by two known cases of additional genuinely symbolic tagging of the manual depiction. Even though signed pictorial representation can abandon such discrete forms and apparent categorical thinking, for many aspects of representation in CPs it is the discrete standard forms which appear to be the unmarked option; signers are commonly described as falling back on analogue strategies when under the pressure of communicative demands for pictorial precision that is greater than the standard, discrete forms will allow.⁴ The observed preference for categorical strategies of representation and thought is all the more striking for its evident optionality.

Within instance theory one could see each standard CP element and standard assemblage as working as a pictorial exemplar (Brooks 1978; Medin & Schaffer 1978), perhaps the default instance of a mental category. Within prototype

4. Thus, for example, in the depiction of the medallion in the study of Emmorey and Herzig (2003), the signer apparently had two standard 'circle' representations to draw on, a smaller and a larger. These both used standard CP handshape options. The handshapes were only modified more finely, using analogue deformations of the standard shape, when the communicative task could only be fulfilled by the depiction of finer distinctions in size.

approaches one could see the standard CP form as a depiction of a prototype (Rosch 1975; Rosch et al. 1976; cf. also Lakoff 1987). Although the phenomenon might seem intuitively less compatible with classic feature-based approaches, within these, the elements of CPs might potentially be modelled depicting visually- and spatially-minimal bundles of one or more of such features, while the depiction of Semantic categories would be interpreted as producing pictorial correspondences to selected significant visual-spatial features of the Semantic category. In any approach, though, the use of standard elements to construct a depiction suggests that even when they depict, signers are thinking about both visual-spatial form and movement, and objects that possess visual-spatial form, in terms of categories.

5 Tracing the stimulus to visual-spatial categories: CPs and sister systems

What is the stimulus to categorical representation in CPs? And what controls the individual forms of category prototypes? Do these prototypes take their nature from patterns of thinking about the visual-spatial world, or from patterns of thinking for representation, from thinking for *pictorial* representation, or perhaps merely from representing with the hands? Comparison of CPs with other, “sister” systems of representation that are partially similar, partially different, has clear potential for answering these and many other questions regarding categorical patterns – both in CPs, and more generally.

Given that the very nature and role of pictorial representation in CPs (and signed languages generally) is under dispute in sign language linguistics, and given also the limited amount of research that has been done in this area, it would be wiser to avoid piling uncertainty on uncertainty here by elaborating many specific arguments. Consequently only relatively straightforward deductions, potential methodological strategies and main heads of existing patterns of argument are highlighted here, and the aim is primarily to illustrate the potential role of the unusual phenomenon of CPs in investigating categorisation.

5.1 The role of the hands in the presence and form of categorical representation

A depressing possibility that must be confronted (depressing in being the most circumscribed in relevance) is that the patterns seen in signed pictorial representation arise simply from the fact that such pictorial representation is occurring on the hands. Perhaps given the physical limitations of the hands, assembly of the depiction from a limited range of standard forms somehow routinely allows greater ease in the manual depiction, regardless of how the signer is thinking about the

visual-spatial world at a deeper level, and this is what provides the stimulus to categorical strategies as such. Perhaps, further, the starting point for each visible categorical contrast in CPs is a set of manual constraints, and the signer is simply assigning a point on a mental visualised continuum of shape or movement to the “nearest” of a limited practicable set of manual options for depiction. In this case the existence of categories at some level would still have to be invoked, but their precise form could be secondary, shaped by the nature of production and perception on the hands alone.

Fortunately both the above scenarios seem unlikely to be the case, at least as a total explanation of what is seen in CPs. Regarding the stimulus to categorical strategies of representation as such, the presence of optional analogue strategies in all parameters of CPs as mentioned above shows that manual representation need not inevitably take categorical form. The same is demonstrated by speakers’ gesture. I am not aware of any published studies of speakers’ gesture that separates the issue of categorical or discrete-combinatorial structure in depictive gesture (often called *iconic gesture*) as opposed to gesture that operates from other principles of representation. Most speech-accompanying gesture, though, is described as being “holistic” and lacking in substructure or conventionalisation. Where consistent structure of various sorts does emerge in the gesture of those who rely on speech-centred communication in everyday life (Acredolo & Goodwyn 1990; Goldin-Meadow, McNeill & Singleton 1996; Wilkins 1999; Boyes Braem & Bräm 2000; Kendon & Versante 2003), it does so under the pressure of what might be summarised as increasing communicative load (cf. “Kendon’s continuum” [Kendon 1982] as conceptualised by McNeill 1992).⁵ It seems clear, then, that categorical strategies in CPs do not emerge solely by virtue of the representation being on the hands.

Regarding the role of the manual channel in controlling the actual form of CP category prototypes, it seems reassuringly unlikely that this input could account for 100%, at any rate, of the patterns seen. If the manual channel were paramount in controlling CP realisations, then one would expect that the list of CP elements would correspond in some detail with the set of phonemic contrasts used in

5. Comparing spoken and sign-based communication systems, I have also suggested that greater conscious awareness of production, greater weight of independent propositional load (i.e. how often information is conveyed uniquely in only one mode of representation) and temporal independence from other channels and other principles of representation (i.e. alternating with, rather than accompanying, a symbol-laden speech stream or its signed equivalent) may all contribute to the emergence of structure in the representational mode, and by implication here, categorical cognition (Cogill-Koez 2000b). In this light, it is interesting that current research that looks for regular patterns of structure in speakers’ gesture of all sorts often does so using data from Italian speakers, a culture in which gesture is seen as relatively independent, frequent and prominent.

regular, lexical signing (these being free of any online demands for structural correspondence with visual-spatial contrasts, and hence far more likely to follow and reveal any tendencies in manual convenience that might exist). I emphasise that as physically convenient tendencies, “phonological” patterns of manual co-ordination do seem to have some influence on common CP usage; they emerge as default formational strategies when they can be accommodated within the demands of contextually-significant aspects of the depiction. However, individual CP realisations also show patterns that, taken together, are not obviously those expected if some natural optimum points in ease of manual construction were the *only* factor that determined their form. Starting with the minor oddities and moving to the more significant ones: First, lists of CP elements can greatly exceed those used for phonological contrasts in the same signed language. This is strikingly so for the parameter of location, in which the set for CPs has usually actually been modelled as infinite; it is increasingly acknowledged to be true also in the parameter of handshape. Second, location and movement (of the hand as a whole) are both parameters which would seem on the face of it to be equally free of the mass of influences that the muscular, skeletal and neural construction of the hand could potentially impose, yet across signed languages, movement is the CP parameter which typically displays much the shortest list of standard elements, while location displays by far the largest. Third, CPs notoriously break the phonological rules that govern the combination of two hands in lexical signs. Indeed, when pictorial requirements demand it, they can transgress even some that are physically quite a challenge to break.

Overall, then, the medium of the hands may certainly produce some tendencies within categorical representation in CPs, but at this stage it seems impossible that categorical representation as such can be entirely accounted for as merely the inevitable by-products of representation on the hands, and most unlikely that *all* of the particular categories that emerge reflect nothing but manual constraints. Furthermore, if one wishes to isolate just what does trigger the shift to categorical strategies of representation, the contrast between speakers’ depictive gesture and signers’ CPs clearly affords a good starting point.

5.2 Categorical representation in depictions: CPs, drawing and the schematic strategy of pictorial representation

Both categorical representation strategies in CPs, and some of the categories that emerge, may be common features of cognition at the level of “thinking for pictorial representation”. The following section will illustrate that CPs’ reliance on categorical strategies is unexceptional, functional and relatively natural for humans

when we represent via depiction, while at least a few of the specific categories seen in CPs are also prominent in another channel of depiction.

Turning, then, to Western drawing (the best-studied mode of pictorial representation to date); here, formal reliance on standard elements is a pervasive strategy of depiction that has been called “schematic”. Far from being an unusual or even a sophisticated achievement, this is the first strategy to predominate in children’s drawings (Goodnow 1977; Fenson 1985; Thomas & Silk 1990; Cox 1992). These are composed largely from circles, arcs, lines and dots; it is notable that these “graphic primitives” (van Sommers 1984) are identical with the core set of CP movement elements nominated in most signed languages.

In drawing just as in CPs, the standard elements are then assembled to construct a depiction. Further, in depicting some classes of object there is an additional strong tendency to fit certain elements together into a standard assemblage. In drawing the result has long been called a “schema” (Gombrich 1962; obviously here with no reference to the use of the term in action categories). Not surprisingly, schemas appear to predominate in drawings of the human face and figure.

In spite of their conventionalisation, cognitively drawings are not simple symbols to the children who draw them. Just like signers with their CPs, even very young children will try to break their schemas when the communicative needs demand it, i.e. when some other element of the object or event not routinely depicted in a conventional schema becomes salient to them. Thus for example most children, presented with a real cup as a model, will draw a cup with a handle even if the model’s handle is hidden from sight. Conversely they will very likely exclude from their drawing of the cup a flower painted on the glaze of the real cup. However, both these aspects of their standard representation of a cup can change if they are simply primed to want to distinguish *this* cup from another, differently-positioned or flower-free one (Cox 1992). Astonishingly, as many as 50% of children as young as four can be induced to depict only part of a robber, the rest of whom is hidden behind a wall (Thomas & Silk 1990). Looking at their routine schema for the human figure at this age, such complexity of depiction of the human form would seem well beyond them – but the key lies in presenting the story to the children in such a way that the fact of part of the robber being hidden is highly important to them. On one hand, then, drawings of children are highly structured and conventionalised, yet on the other they are demonstrably still “drawings” to them, still pictorial representations.

The schematic strategy is not restricted to children. Unless one trains in drawing, the reliance on the schematic strategy usually persists on into adulthood. Here it has been shown to influence everything from preference for standard realisations of geometrical elements of shape, through to the generation of photographically accurate perspective (Arnheim 1954; Gombrich 1982; Hagen 1985; Feldman

2000). Most of us still find schematic depiction fairly easy, and analogue depiction, unsupported by sub-elements or conventions of assembly, extremely hard: Compare the ease with which most of us can represent human action using stick figures with the embarrassingly difficult task of drawing the human form without recourse to such conventions.

Furthermore, even the use of the occasional outright symbol as a dedicated part of a culturally-transmitted depictive genre has its parallel in drawing. For example, cartooning (Goodnow 1977) relies heavily on cross-culturally incomprehensible symbols to “depict” – or more accurately here, to symbolise – things like pain (e.g. via a series of radiating lines around the “painful” area of a figure), speed (via “speed lines”), or the contrast between what is thought and what is said (the “thought bubble” versus the smooth-outlined “speech bubble”). These system-specific symbols are placed adjacent to or on figures with which the action or sensation is associated. So common are such symbolic tags, and so great is the ease with which we decode them within the overall depiction, that it can be hard to realise that one is really interpreting an integrated mix of pictures and symbols, products of depictive and symbolic strategies, each realised in culture- and genre-specific ways.

Finally, the comparison with drawing also helps to clarify the possibly opaque issue of how a pictorial representation can stand for a category of thought, be it a purely visual-spatial category or not. Each standard graphic primitive in drawing, like each discrete element in CPs, amounts to a pictorial exemplar of a category of visual-spatial thought and by this means evokes the corresponding visual-spatial category, while similarly each conventional assemblage amounts to a depiction of the visual-spatial qualities associated with a Semantic category and in this way evokes the corresponding Semantic category. Each graphic element or assemblage of elements would “refer” to its corresponding mental category (whether spatial or Semantic) by depicting its prototype, much as, from Rosch’s prototype model, a British speaker of English might be expected to depict the category verbally labeled *bird* by drawing something that has the overall proportions of a robin, rather than something that has the proportions of a penguin or an emu.

What of the representation of ‘male’ and ‘female’ in Taiwan Sign Language CPs? Even the representation of this seemingly nonvisual or (in signed language linguistic terms), semantic categorical distinction also has its parallel in drawing. In the course of developing their depiction of the human figure, children not only gradually master a depiction that is increasingly elaborate and full (but still conventionalised); following this, they will next typically work through phases of distinguishing female from male by means of height distinctions, triangular “dress-shaped” bodies, hair ribbons and hair length. In Australia, street signs that signal ‘elderly people crossing’ draw on some of the same pictorial prototypes. It is not necessarily that children (or street sign designers, or the viewers of such signs)

believe that all men are taller than all women, that all women have long hair, or that all female humans wear dresses – or again, that the “pictures” are cognitively really processed as nothing more than abstract symbols. The suggestion here is rather that the prototypical visual image associated with the category of gender can include a height, a hair length and/or a dress distinction, and that these can then be exploited as conventional visual indicators of gender in children’s pictorial representations and in street signs. Thus the pictorial representation of the posited mental category is routinely achieved by means of some prototypical, nonarbitrary visual-spatial feature of the real-world objects embraced by the category.

It seems significant to categorical patterns of “thinking for pictorial representation” that drawing and CPs are in many respects very different embodiments of the principle of pictorial representation, yet at all levels their categorical strategies correspond closely. The depiction of visual-spatial prototypes to represent categories of visual-spatial reality does not seem exceptional. As regards details of the strategy, there is common use of categorisation both at the level of purely visual-spatial form and at higher levels. Within the signed language linguistics’ traditional visual-spatial versus Semantic contrast, even the list of standard graphic elements in children’s schematic drawing maps well onto the core movement elements of many CP inventories, while within the depictions of Semantic categories, schemas for the human figure, the most commonly researched and recognised subject in drawing, have their parallel in the pervasive use of standard assemblages in CPs to depict the human figure.

Which of these aspects of categorical form and thought are to be sheeted home to pictorial representation strategies, and which to “thinking for representation” or simply “thinking”? To answer these questions it will be necessary to make comparisons with further sister systems of communication that use partly similar, partly different representational strategies.

5.3 Categorical representation independent of principle of representation: CPs and spoken languages

Such systems are close at hand. Cross-comparisons of the signed pictorial representations of CPs with categorisation in spoken abstract representational systems is only the starting point. Other sister systems include signed regular lexis, the gesture of speakers that uses abstract symbols (such as the Western ‘thumbs up’ gesture that signals approval), and the abundant graphic symbols that tag categories that lack visual-spatial form, both in drawing and elsewhere – be they the radiating “impact lines” around a cartoon figure that indicate a collision, the red cross on an ambulance, or the blue and white checked stripe along the side of a British or Australian police car. The following merely sketches work that has been

done so far in relation to CPs. Although there is limited agreement even between those working within the same model of signed languages, the few existing enquiries amply illustrate the potential of CPs and sister systems for triangulating on sources and patterns of categorisation.

Supalla (1978, 1982, 1986), drawing on Allan's classic study (Allan 1977), has suggested that categories which Allan identifies as salient cross-linguistically in spoken classifier systems are also prominent in signed CPs. Within purely visual-spatial aspects, parameters nominated as being in common include contrasts such as 'long', 'flat', 'rounded'. Within his Semantic categories, Supalla points out that conventional ASL CP forms distinguish the human figure from animals, tagging categories which Allan states are also commonly marked by distinctive classifiers in spoken languages. In Supalla's model of CPs, in which CPs are held to refer abstractly rather than depictively, this is evidence of the uniform action of a linguistic level of categorisation: In the model of CPs as structured pictorial representation, these same patterns would be modelled as having their causal origin deeper, in more general levels of "thinking for representation", or in thinking as such.

Supalla (1990) also suggested that signed CPs show evidence of constraints on the combination of Talmy's proposed elements of spatial categories (Figure, Ground, Path etc.; Talmy 1985). The CP constraints differ in detail from those nominated from spoken languages, but are similar in limiting overall semantic complexity of the verb (in Supalla's model) or depiction (in emerging models such as that espoused here) in a way that is language-specific. Although Talmy appears to reject Supalla's specific application of his model (Talmy 2003a), subsequently he has used Supalla's conceptualisation of CPs (Talmy 2003b), circumventing the issue of abstract versus pictorial representational principles via a model in which commonalities of CP and spoken language semantic structure stem from a shared mental module, while aspects which differ are assigned to language-specific (or channel-specific) modules.

Taub and Galvan (2001) have also offered arguments and evidence against Supalla's proposed parallel between signed CPs and Talmy's nominated patterns for spoken languages, arguing instead that the CPs of adult signers can show pretty much any combination here – they give the example of a signer representing someone edging sideways along a ledge to rescue a cat, using a single CP that combines within it Figure, Path, Manner and Motion. Nevertheless, Taub and Galvan accept Supalla's combinations as valid at the level of tendencies, suggesting that they therefore have their origin in "general cognitive and perceptual functioning rather than arbitrary [here, grammatical] restrictions" (Taub & Galvan 2001: 180). Furthermore, they add to this an observed trend towards separating path and manner of locomotion, both in signed systems' CPs and in the gestures of speaking-hearing children up to 12. Taub and Galvan suggest that the avoidance stems

from “a cognitive universal, not a language-specific one, and that the patterns of motion verbs in ASL derive from the same cognitive structures as spoken-language gestures” (2001: 195). Taub’s model of CPs in signed languages commits her only to a historical role for depiction in generating these forms, the issue of online processing being left open (Taub 2001), but like the current author, Taub and Galvan conclude that “the discretization of the representation of space and movement in ASL would not then derive from separate ‘morphemes’ in the traditional sense but separate cognitive representations” (2001: 195).

Whatever the facts may prove to be in here, the above illustrates how contrasts in possibilities within different sister systems may allow one ultimately to trace the input that a specific mode of cognitive processing or representation can have on categorisation.

6 Conclusion

Does the human mind work on the basis of categories? The evidence presented here suggests that the answer is “not inevitably – but often by preference, even when other strategies are available”. Does categorisation work in the same way across the domains of human language and behaviour? CPs show that, within natural human communication systems, the principle of representation used can indeed affect the degree to which we draw on categories. Nevertheless, it is significant that where categories do emerge, they may sometimes wear much the same face on the hands as in speech, and in depictions as they do in classic symbolic representation. CPs, drawing, speakers’ gesture, and the symbolic systems of regular signed and spoken lexis seem to amount to a set of mutually-controlling natural experiments with interesting potential for teasing apart the factors that drive categorisation.

References

- Acredolo, Linda P. & Susan W. Goodwyn. 1990. Sign language among hearing infants: The spontaneous development of symbolic gestures. In Virginia Volterra & Carol J. Erting (eds.), *From gesture to language in hearing and deaf children*, 68–78. Berlin: Springer.
- Allan, Keith. 1977. Classifiers. *Language* 53. 285–311.
- Arnheim, Rudolf. 1954. *Art and visual perception*. Berkeley, CA: University of California Press.
- Boyes Braem, Penny & Thuring Bräm. 2000. A pilot study of the expressive gestures used by classical orchestra conductors. In Karen Emmorey & Harlan Lane (eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, 143–167. Mahwah, NJ: Lawrence Erlbaum.

- Brooks, Lee. 1978. Nonanalytic concept formation and memory for instances. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and categorization*, 169–221. Hillsdale, NJ: Lawrence Erlbaum.
- Cogill-Koez, Dorothea. 2000a. Signed language classifier predicates: Linguistic structures or schematic visual representation? *Sign Language and Linguistics* 3(2). 153–208.
- Cogill-Koez, Dorothea. 2000b. A model of signed language 'classifier predicates' as templated visual representation. *Sign Language and Linguistics* 3(2). 209–236.
- Cogill, Dorothea. 2003. Signed language classifier predicates as templated visual representation. Armidale, NSW: University of New England PhD dissertation.
- Cox, Maureen. 1992. *Children's drawings*. London: Penguin.
- Emmorey, Karen. 1996. The confluence of space and language in signed languages. In Paul Bloom, Mary A. Peterson, Lynn Nadel & Merrill F. Garrett (eds.), *Language and space*, 171–209. Cambridge, MA: The MIT Press.
- Emmorey, Karen (ed.). 2003. *Perspectives on classifier constructions in sign languages*. Mahwah, NJ: Lawrence Erlbaum.
- Emmorey, Karen & Melissa Herzig. 2003. Categorical versus gradient properties of classifier constructions in ASL. In Emmorey (ed.), 221–246.
- Engberg-Pedersen, Elizabeth. 1993. *Space in Danish sign language*. Hamburg: Signum.
- Feldman, Jacob. 2000. Bias toward regular form in mental shape spaces. *Journal of Experimental Psychology: Human Perception and Performance* 26(1). 152–165.
- Fenson, L. 1985. The transition from construction to sketching in children's drawings. In Freeman & Cox (eds.), 374–386.
- Fischer, Susan D. & Patricia Siple (eds.). 1990. *Theoretical issues in sign language research, Volume 1: Linguistics*. Chicago, IL: The University of Chicago Press.
- Freeman, Norman H. & Maureen V. Cox (eds.). 1985. *Visual order*. Cambridge: CUP.
- Goldin-Meadow, Susan, David McNeill & Jenny Singleton. 1996. Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review* 103(1). 34–55.
- Gombrich, Ernst H. 1962. *Art and illusion: A study in the psychology of pictorial representation*. London: Phaidon.
- Gombrich, Ernst H. 1982. *The image and the eye: Further studies in the psychology of pictorial representation*. Oxford: Phaidon.
- Goodnow, Jacqueline. 1977. *Children's drawing*. Cambridge, MA: Harvard University Press.
- Hagen, Margaret A. 1985. There is no development in art. In Freeman & Cox (eds.), 59–77.
- Kendon, Adam. 1982. The study of gesture: Some observations on its history. *Recherches Semiotiques/Semiotic Inquiry* 2. 45–62.
- Kendon, Adam & Laura Versante. 2003. Pointing by hand in Neapolitan. In Sotaro Kita (ed.), *Pointing: Where language, culture and cognition meet*, 109–137. Hillsdale, NJ: Lawrence Erlbaum.
- Klima, Edward & Ursula Bellugi. 1979. *The signs of language*. Cambridge, MA: Harvard University Press.
- Lakoff, George. 1987. Cognitive models and prototype theory. In Ulric Neisser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*, 63–100. Cambridge: CUP.
- Liddell, Scott K. 1995. Real, surrogate and token space: Grammatical consequences in ASL. In Karen Emmorey & Judy Reilly (eds.), *Language, gesture and space*, 19–41. Hillsdale, NJ: Lawrence Erlbaum.

- Liddell, Scott K. 2003. *Grammar, gesture and meaning in American Sign Language*. Cambridge: CUP.
- Liddell, Scott K. & Melanie Metzger. 1998. Gesture in sign language discourse. *Journal of Pragmatics* 30(6). 657–697.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL: The University of Chicago Press.
- McNeill, David & Susan D. Duncan. 2000. Growth points in thinking-for-speaking. In David McNeill (ed.), *Language and gesture*, 141–161. Cambridge: CUP.
- McNeill, David, Elena T. Levy & Laura L. Pedelty. 1990. Speech and gesture. In Geoffrey Hammond (ed.), *Cerebral control of speech and limb movements*, 203–256. Amsterdam: North Holland.
- Medin, Douglas L. & Marguerite M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85. 207–238.
- Newport, Elissa. 1982. Task specificity in language learning? Evidence from speech perception and ASL. In Eric Wanner & Lila R. Gleitman (eds.), *Language acquisition: The state of the art*, 450–486. Cambridge: CUP.
- Poizner, Howard, Edward Klima & Ursula Bellugi. 1987. *What the hands reveal about the brain*. Cambridge, MA: The MIT Press.
- Rosch, Eleanor. 1975. Cognitive reference points. *Cognitive Psychology* 7. 532–547.
- Rosch, Eleanor, Carolyn Mervis, Wayne Gray, David Johnson & Penny Boyes Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8. 382–439.
- Schembri, Adam. 2001. Issues in the analysis of polycomponential verbs in Australian Sign Language (Auslan). Sydney NSW: University of Sydney PhD dissertation.
- Schembri, Adam. 2003. Rethinking ‘classifiers’ in signed languages. In Emmorey (ed.), 3–34.
- Smith, Wayne H. 1990. Evidence for auxiliaries in Taiwan Sign Language. In Fischer & Siple (eds.), 211–228.
- Sommers, Peter van. 1984. *Drawing and cognition: Descriptive and experimental studies of graphic production processes*. Cambridge: CUP.
- Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication system of the American deaf* [Studies in Linguistics: Occasional Papers 8]. New York, NY: Dept. Anthropology and Linguistics, University of Buffalo.
- Supalla, Ted. 1978. Morphology of verbs of motion and location in American Sign Language. In Frank Caccamise & Doin Hicks (eds.), *National Symposium on Sign Language Research and Teaching* 2, 27–45. Silver Springs, MD: National Association of the Deaf.
- Supalla, Ted. 1982. Structure and acquisition of verbs of motion and location in American Sign Language. University of California PhD dissertation.
- Supalla, Ted. 1986. The classifier system in American Sign Language. In Colette Craig (ed.), *Noun classes and categorization*, 181–214. Amsterdam: John Benjamins.
- Supalla, Ted. 1990. Serial verbs of motion in ASL. In Fischer & Siple (eds.), 127–152.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description: Grammatical categories and the lexicon*, Vol. 3, 57–149. Cambridge: CUP.
- Talmy, Leonard. 2003a. The representation of spatial structure in spoken and signed language. In Emmorey (ed.), 169–195.
- Talmy, Leonard. 2003b. The representation of spatial structure in spoken and signed language: A neural model. *Language and Linguistics* 4(2). 207–250.

- Taub, Sarah. 2001. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: CUP.
- Taub, Sarah & Dennis Galvan. 2001. Patterns of conceptual encoding in ASL motion descriptions. *Sign Language Studies* 1(2). 175–200.
- Thomas, Glyn V. & Angele M. J. Silk. 1990. *An introduction to the psychology of children's drawings*. New York, NY: Harvester Wheatsheaf.
- Wilkins, David. 1999. Spatial deixis in Arrernte speech and gesture: On the analysis of a species of composite signal as used by a Central Australian Aboriginal group. *Workshop on Deixis, Demonstration and Deictic Belief, at European Summer School in Logic, Language and Information* (ESSLI XI), <http://www.lili.uni-bielefeld.de/~deixis/reader/node24.html>.

Hidden units in child language

Stephen Crain, Takuya Goro and Utako Minai

By three or four, children can judge the truth or falsity of sentences, in accord with the judgments of adults. Both learning-theoretic and nativist accounts of such phenomena have been offered. This chapter introduces several aspects of children's linguistic competence that have not been previously documented in the "nature versus nurture" debate. We describe cross-linguistic studies of children's understanding of logical expressions, including negation, disjunction, conjunction, and focus expressions (e.g. English *only*). The findings suggest that children understand the "hidden" content of these logical expressions in the absence of any realistic database by which children's understanding could be obtained by general methods of learning. The findings offer insights into the contributions of human nature to language acquisition.

1 Introduction

This chapter describes two hidden properties of natural language. They have something in common, namely the interpretation of the natural language counterparts to the (Boolean) connectives of classical logic. It will be useful, therefore, to review the relevant aspects of classical logic. The review consumes the remainder of the introduction.

The Boolean connectives of classical logic are negation ' $\neg$ ', disjunction ' $\vee$ ', and conjunction ' $\wedge$ '. The counterparts to these connectives in English are *not*, *or*, and *and*. We start with disjunction. Natural language disjunction figures prominently in the studies in the sections that follow. In classical logic, disjunction ' $\vee$ ' is assigned truth conditions associated with inclusive-*or*. This means that formulas of the form $A \vee B$ are true if A is true (but not B), or B is true (but not A), or if both A and B are true. This is illustrated in Table 1.

Because disjunction is assigned the truth conditions of inclusive-*or*, the truth conditions for a formula in which disjunction appears in the scope of negation,

Table 1. Truth table (disjunction)

A	B	$A \vee B$
0	0	0
0	1	1
1	0	1
1	1	1

$\neg(A \vee B)$, will be just the reverse of those for $A \vee B$. The formula $\neg(A \vee B)$ will be true if and only if both A and B are false. In other words, $\neg(A \vee B)$ and $\neg A \wedge \neg B$ are logically equivalent; the formulas have exactly the same truth conditions (see Table 2). This is one of De Morgan’s laws: $\neg(X \vee Y) \Rightarrow \neg X \wedge \neg Y$. We will call this law the ‘conjunctive’ entailment of disjunction in the scope of negation.

Table 2. Truth table (De Morgan’s Law)

A	B	$A \vee B$	$\neg(A \vee B)$	$\neg A$	$\neg B$	$\neg A \wedge \neg B$
0	0	0	1	1	1	1
0	1	1	0	1	0	0
1	0	1	0	0	1	0
1	1	1	0	0	0	0

Sometimes, sentences from natural languages exhibit truth conditions that resemble this one of De Morgan’s laws. For example, when the English word *or* is interpreted within the scope of the word *not*, the sentence can be recast into a one in which conjunction (*and*) presides over both of the terms that formed a disjunction in the original sentence. For example, disjunction under negation in Example (1) licenses the conjunctive entailment stated in (2).

- (1) John does not speak Spanish *or* French.
- (2) John does not speak Spanish *and* John does not speak French.

The conjunctive entailments of sentences with disjunction are not only licensed in sentences with negation, but also with a variety of other natural language expressions (Zwart 1996). These expressions are referred to as *downward entailing* (or *monotone decreasing*). In (3), we provide examples of other expressions that license conjunctive entailments when they have scope over disjunction (see Crain et al. 2002). According to this diagnostic, therefore, these expressions are downward entailing.

- (3) a. *Every*
 Every student who speaks French or Spanish likes to travel.
 $\Rightarrow$ *every student who speaks French likes to travel*
and every student who speaks Spanish likes to travel
- b. *None of the NPs*
 None of the students in this class speaks French or Spanish.
 $\Rightarrow$ *none of the students in this class speaks French*
and none of the students in this class speaks Spanish
- c. *Before*
 John went to Europe before learning French or Spanish.
 $\Rightarrow$ *John went to Europe before learning French*
and before learning Spanish
- d. *Without*
 John went to Europe without learning French or Spanish.
 $\Rightarrow$ *John went to Europe without learning French*
and without learning Spanish

The defining property of downward entailing expressions is that they make valid inferences from claims about things, to claims about subsets of those things.¹ We call this type of inference: *downward entailment*. The examples in (4) demonstrate that all of the linguistic expressions in (3) have the defining property of downward entailment, since it is valid to substitute claims about sets of things (being a Romance language) with claims about subsets of those things (French, Spanish, Italian, ...).

- (4) a. Every student who speaks a Romance language likes to travel.
 $\Rightarrow$ *every student who speaks French likes to travel*
- b. None of the students in class speaks a Romance language.
 $\Rightarrow$ *none of the students in class speaks French*
- c. John went to Europe before learning a Romance language.
 $\Rightarrow$ *John went to Europe before learning French*
- d. John went to Europe without learning a Romance language.
 $\Rightarrow$ *John went to Europe without learning French*

1. The pattern of inference is in a “downward” direction when a set-denoting expression (e.g. noun phrase) appears in the scope of any downward entailing expression. Replacing the set-denoting expression by one that denotes a subset of that set does not affect the truth-value of the whole sentence. In the absence of any downward entailing expression, by contrast, the pattern of inference is “upward” more often than not – inferences are sanctioned from claims about sets of things to claims about supersets of those things. So, for example, *Mary bought a Ferrari* entails *Mary bought a car*, but not vice versa.

So far we have seen that downward entailing expressions have two properties. They license downward entailments (i.e. inferences from sets to their subsets), and they create conjunctive entailments when they are combined with disjunction. This provides us with two tests for any candidate for a downward entailing operator. In the following paragraphs, we look at the focus operator *only* (in English), and its interaction with disjunction, to see if it is downward entailing.

As proposed by Horn (1969), the meaning of a sentence with *only*, such as (5), can be decomposed into two conjoined propositions. The first proposition pertains to the focus element, *Bruce*. The content of this proposition is the truth-conditional meaning of the original sentence, absent the focus operator, *only*. So, (6) represents the first meaning component of (5). We will call this meaning component its *presupposition*.

(5) Only Bruce speaks a Romance language.

(6) Bruce speaks a Romance language.

The second meaning component is a proposition that is entailed by the original sentence. The content of this proposition comes from the focus operator *only*. In virtue of the focus operator, it is entailed that the property (speaking a Romance language) being attributed to the individual in focus is not also a property of anyone else. So, the second meaning component of (5) can be represented as (7). We will call this the *assertion*.²

(7) For all x [$x \neq \text{Bruce}$], it is not the case that x speaks a Romance language.

Now let us ask if *only* is downward entailing. As we saw, downward entailing expressions endorse inferences from claims about sets of things to claims about subsets of those things. The entailment from a set to its subsets does not hold for sentences with the focus operator *only*, however. English speakers typically deny that (8) entails (9), on the grounds that Bruce might speak Spanish or Italian, and not French. Based on observations such as this, von Stechow (1999) argues that the first meaning component of *only*, the presupposition, does not contain a downward entailing operator. This explains why the standard diagnostic for downward entailment fails. For example, *Bruce speaks a Romance language* does not entail

2. Everyone agrees that the second proposition is directly asserted/entailed by the original sentence, but the status of the first meaning component is more controversial: Horn (1969, 1996) and von Stechow (1999) argue that it is a presupposition, whereas Atlas (1993, 1996) and Herburger (2000) take it to be the assertion of the sentence. We will adopt Horn's terminology. We will refer to the first proposition, about the element in focus, as the "presupposition"; and we will refer to the second proposition, about the individuals being contrasted with the element in focus, as the "assertion". It should be understood that we intend no theoretical commitments by adopting this terminology.

Bruce speaks French. What about the second meaning component, the assertion? It turns out that, in this meaning component, sentences with the focus operator *only* do license inferences from a set to its subsets. For example, (8) entails (10). So, the second meaning component of the focus operator *only* apparently does contain a downward entailing expression. We assume that the downward entailing expression is negation, or the semantic equivalent of negation (as below the technical rendering of the assertion). Thus, although sentences with *only* (e.g. (10)) lack an overt downward entailing expression, they contain a *covert* downward entailing expression. The covert downward entailing expression appears in the assertion.

- (8) Only Bruce speaks a Romance language.
- (9) # Only Bruce speaks French.
- (10) For all x [$x \neq \text{Bruce}$], it is not the case that x speaks French.
(All the others being contrasted with Bruce do not speak French.)
(None of the others being contrasted with Bruce speak French.)

Another test of downward entailment, as we saw, is the licensing of conjunctive entailments when disjunction appears in the scope of a downward entailing expression. Let's apply this test to sentences with the focus operator *only*. Consider (11). Notice first, that (11) does not license the conjunctive entailment in (12). There are circumstances in which (11) is true, but where (12) is false, such as the circumstances in which Bruce speaks just French, or just Spanish, but not both. So, disjunction in the presupposition of sentences with *only* does not create conjunctive entailments, hence *only* is not downward entailing in this meaning component. By contrast, (11) does entail the conjunctive statement in (13). So, disjunction in the assertion of sentences with *only* does license conjunctive entailments of disjunction, hence *only* proves, again, to pass another test for being downward entailing, but just in one of its meaning components. We might say that the focus operator *only* is 'partially' downward entailing.

- (11) Only Bruce speaks French or Spanish.
- (12) # Only Bruce speaks French *and* only Bruce speaks Spanish.
- (13) Everyone else does not speak French *and* everyone else does not speak Spanish.

We have seen that, at least in part, natural languages mirrors logic in English sentences with negation and disjunction. The conjunctive entailment created by negation is also found in sentences with other natural language expressions, such as *every*, *before*, *none of the*, and *without*. Lumped together (with negation), these expressions make up the class of downward entailing expressions. The focus operator *only* also belongs to this class, but care must be taken here, because *only* has

two meaning components, and only one of them contains a downward entailing expression. The focus operator *only* creates a “partially” downward entailing context. Or, as von Stechow (1999) puts it, *only* is “Stalnaker decreasing”.

The interpretation of logical words is also influenced by norms of conversation. We saw that *or* is interpreted as inclusive-*or* if it is in the scope of a downward entailing expression, as in classical logic. In non-downward entailing environments, by contrast, the use of *or* does not conform to classical logic. In these cases, the use of *or* implies “not both”, but does not entail it. In other words, a statement of the form $A \text{ or } B$ is pragmatically odd (infelicitous) as a description of a situation in which both A and B are true. The implicature of “exclusivity” stems from the availability of another statement, namely $A \text{ and } B$, which is more informative. $A \text{ and } B$ is more informative than $A \text{ or } B$, because $A \text{ and } B$ is true in only one set of circumstances, whereas $A \text{ or } B$ is true in other circumstances as well. Due to this overlap of truth conditions, the expressions *or* and *and* form a scale based on information strength, with *and* being more informative than *or* (in contexts that verify sentences with either expression). A pragmatic principle *Be Cooperative* (cf. Grice 1975) entreats speakers to be as informative as possible. Upon hearing someone use the weaker term on the scale *or*, listeners infer that the speaker was being cooperative and, hence, was not in a position to use the stronger term *and*. So the speaker is taken to imply the negation of the stronger term: *not both A and B*. Adult use of *or* is clearly governed by this *scalar implicature*, so adults avoid using $A \text{ or } B$ in situations in which both A and B are true. Consequently, the vast majority of children’s experience is consistent with the conclusion that natural language disjunction is exclusive-*or*, not inclusive-*or* (see Crain, Goro & Thornton 2006).

This completes our review of the logical expressions that we discuss. We will now proceed to a consideration of children’s acquisition of the meanings of these logical expressions. Our emphasis is on components of meaning that are well hidden, but which children find nonetheless.

2 The semantic subset principle

The first hidden property of natural language is manifested in children’s adherence to classical logic in languages where the input to children runs counter to classical logic. The example compares the acquisition of English with that of Japanese and Chinese. As promised, the phenomenon under investigation is the interpretation of disjunction under negation. We have seen that, when disjunction *or* appears in the scope of negation in sentences of English, these sentences license a conjunctive entailment, as in De Morgan’s law: $\neg(X \vee Y) \Rightarrow \neg X \wedge \neg Y$. For example, *Max’s*

computer did not come with Ichat or Isync entails that Max's computer did not come with Ichat *and* did not come with Isync.

There have been a number of experimental studies on English-speaking children's interpretation of disjunction in the scope of negation and in the scope of quantificational expressions such as *none of the* and *every* (Chierchia et al. 2001; Gualmini et al. 2001; Gualmini & Crain 2002; Crain et al. 2002; Gualmini & Crain 2004, among others). These studies have revealed that 4- to 5-year-old English-speaking children are aware of the "conjunctive" entailment of disjunction when it appears in the scope of a downward-entailing operator. A representative example is an experiment by Crain et al. (2002) using the Truth Value Judgment Task in the prediction mode (see Crain & Thornton 1998 for extensive discussion of this experimental methodology). On a typical trial of the experiment, sentence (14) was uttered by a (wizard) puppet as a prediction about how events would unfold in a story. It subsequently turned out that the girl who stayed up late received a jewel, but not a dime. English-speaking children (ages 3;11–5;9, mean 5;0) correctly rejected sentences like (14) 92% of the time in experimental contexts such as this. Children's stated reason for rejecting (14) was that the girl who stayed up late had received a jewel. It is evidence that, in children's grammars, (14) entails that the girl would receive neither a dime nor a jewel, i.e. she would not receive a dime *and* she would not receive a jewel. This conjunctive entailment follows from (14) only if the disjunction operator *or* is assigned a Boolean inclusive-*or* interpretation, as in classical logic.

(14) The girl who stayed up late will *not* get a dime *or* a jewel.

As Crain, Goro and Thornton (2006) point out, children have little direct evidence for the Boolean interpretation of disjunction, because the vast majority of the input to children consists of *positive* statements, where disjunction is governed by an implicature of exclusivity. It comes as no surprise, therefore, that the same children accepted sentences like (15) 87% of the time in the same contexts.

(15) The girl who didn't go to bed will get a dime *or* a jewel.

There is even more compelling evidence that young children's interpretation of negated disjunction is not influenced by properties of the input. The evidence comes from a study of Japanese-speaking children, by Goro and Akiba (2004), replicated in Chinese by Jing, Crain and Hsu (2005). These studies are based on the observation by Goro (2004) that, in Japanese, simple negative sentences with disjunction do not license conjunctive entailments. In Japanese, the translation of the English sentence *Max's computer did not come with Ichat or Isync* makes the assertion that Max's computer didn't come with Ichat or it didn't come with Isync. For adult speakers of Japanese, therefore, simple negative sentences lack the conjunctive entailments associated with De Morgan's laws. Thus, adult Japanese-speakers interpret

(16) to mean that the pig didn't eat the carrot *or* didn't eat the pepper. Despite the appearance of disjunction *ka* within the scope of sentential negation in surface syntax, *ka* is interpreted by adults as having scope over negation.

- (16) *Butasan-wa ninjin ka pi'iman-wo tabe-nakat-ta*
 pig-TOP pepper or carrot-ACC eat-NEG-PAST

Literally: 'The pig didn't eat the pepper or the carrot'

Meaning: 'The pig didn't eat the pepper *or* the pig didn't eat the carrot'

Based on considerations of language learnability, Goro (2004) hypothesized that Japanese-speaking children would, nevertheless, interpret the disjunction operator *ka* as licensing conjunctive entailments in simple negative sentences like (16). The prediction was that Japanese-speaking children would interpret such sentences in the same way as English-speaking children and adults, despite the absence of this interpretation for adult speakers and, hence, the absence of evidence for this interpretation in the input to children. In brief, Goro's proposal is that the semantics of natural language disjunction is innately specified as Boolean inclusive disjunction. However, the interaction of disjunction with negation is subject to cross-linguistic variation, as proposed by Szabolcsi (2002). In one class of languages, including English, disjunction may be interpreted within the scope of local negation, whereas it must be interpreted outside the scope of local negation in another class of languages, including Japanese, regardless of its surface position in such languages. In short, Goro proposed that disjunction is a *positive polarity item* in Japanese (like *some* in English), but not in English. In Japanese, then, the disjunction operator *ka* appears to have the truth conditions associated with exclusive-*or* in simple negative sentences (due to the conversational implicature of exclusivity), whereas *or* creates conjunctive entailments in the corresponding sentences of English.

Goro's next observation was that the alternative values of the "positive polarity parameter" for disjunction stand in a subset/superset relation, with English exemplifying the subset value of the parameter, and Japanese the superset value. He reasoned that this situation would lead to a "subset problem" unless children acquiring Japanese initially select the parameter value corresponding to English. Since adult speakers do not make such entailments, it is unlikely that children learn to make them based on the adult input. The reason children should appear more logical than adults in Japanese, Goro suggested, is that children adhere to a principle of language acquisition: the semantic subset principle (Crain, Ni & Conway 1994).

The semantic subset principle enforces an ordering on the values of certain parameters, where one value makes a sentence true in a subset of the circumstances that make it true on the other value. The semantic subset principle compels children to adopt the subset value of the parameter as their initial interpretation; this value is abandoned only on the basis of positive evidence in the local language.

If children adopted the superset value instead, they would generate sentences that are not in the local language, in addition to sentences in the local language. This raises a familiar learnability problem: In the absence of negative evidence, it is difficult to see how children would purge their grammars of the means for generating sentences that are not acceptable in the local language. To avoid this problem, the semantic subset principle orders the value of parameters.

In simple sentences with negation and disjunction, according to Goro (2004), a parameter governs the scope relations between these two expressions. The semantic subset principle establishes the default setting, according to which negation has scope over disjunction. English retains the default setting, whereas speakers of Japanese and Chinese adopt the alternative setting of the parameter, with disjunction having scope over negation; this makes it look as if these languages violate De Morgan's law in simple negative sentences (see Goro 2004, for a complete analysis).

To investigate this solution to the "logical problem of language acquisition", Goro and Akiba (2004) examined Japanese children's interpretation of negated disjunctions in sentences like (16) using the Truth Value Judgment Task. They interviewed thirty Japanese-speaking children (age 3;7–6;3, mean 5;3) as well as a control group of Japanese-speaking adults. On a typical trial, subjects were asked to judge whether or not (16) was an accurate description of a situation in which the pig had eaten the carrot but not the green pepper. The findings were precisely as anticipated. Japanese-speaking adults uniformly accepted the target sentences (such as 2), whereas children rejected them 75% of the time. The findings are even more compelling once the data from four children, who responded like adults, were set aside. The remaining 26 children rejected the target sentences 87% of the time. The pattern of responses by Japanese-speaking children are difficult to explain on a "data driven" account of language development, since Japanese-speaking children interpreted negated disjunctions as licensing conjunctive entailments, whereas Japanese-speaking adults did not. On the other hand, the findings are consistent with the continuity hypothesis, according to which child language is expected to diverge from the local adult language, but only in ways that adult languages can differ from each other (see e.g. Crain 1991; Crain & Pietroski 2001, 2002; Crain, Goro & Thornton 2006; Thornton 1990). It is worth pointing out, in addition, that the disjunction-as-polarity-item parameter is an example of a "specific contingent fact" about natural language; it is part of the distinctive way in which human systems for associating sounds and meanings have evolved, and does not follow from innate knowledge of logic.

The experimental study of the development of the interpretation of disjunction in Japanese has recently been extended to Chinese (Jing, Crain & Hsu 2005). In adult Chinese, the expression corresponding to logical disjunction is *huozhe*. In

ordinary contexts, where there is no overt or covert downward entailing operator, *huozhe* is interpreted like English *or* and Japanese *ka*. We assume that its basic meaning is inclusive-*or*, and that it is subject to an implicature of exclusivity.

When the Chinese expression *huozhe* appears under negation, it patterns like Japanese *ka*, but not like English *or*. So, for example, (17) does not have the conjunctive entailment that Kermit does not speak English *and* does not speak Chinese. Instead, (17) can be paraphrased as saying that Kermit does not speak English or he does not speak Chinese, but the speaker is not sure which of these languages Kermit does not speak.

- (17) *Kermit bu hui yingwen huozhe zhongwen*
 Kermit not know English or Chinese
 ‘Kermit does not know English or Chinese’

Another parallel between Chinese and Japanese is the way one achieves the *neither* reading associated with the conjunctive entailment that is missing in (17). As in Japanese, Chinese speakers insert the lexical item corresponding to conjunction, *he*, in simple negated sentences in order to convey the “neither” interpretation. This is illustrated in (18).³

- (18) *Kermit bu hui yingwen he zhongwen*
 Kermit not know English and Chinese
 ‘Kermit does not know English and Chinese’

Based on the reasoning in Goro (2004), and the findings of Goro and Akiba (2004), it was conjectured that Chinese adults adopt a parameter setting for disjunction that takes *huozhe* to be a Positive Polarity Item (PPI), as *ka* is in Japanese. By definition, a PPI cannot be interpreted in the scope of local negation, which would explain the absence of the conjunctive entailment in (17). A further conjecture was that Chinese-speaking children would initially interpret *huozhe* in conformity with De Morgan’s laws in simple negated sentences like (17), despite the absence of this interpretation for adults.⁴

Following the logic of the Goro and Akiba study of Japanese speaking children, Chinese-speaking children would appear to be more logical than adults, because children would obey De Morgan’s laws (as in English, and as in child

3. Notice that this interpretation of conjunction under negation is not required by De Morgan’s laws, since “not (A and B)” is logically equivalent to “not A or not B”, where (18) is interpreted as requiring both A and B to be false. We return to this in Sections 3 and 4.

4. One can only speculate on why languages have evolved in this way, turning disjunction into a PPI. One possibility is that this adds to the expressive power of natural languages by carving up the truth conditions that are associated with disjunction (in classical logic) into distinct, complementary sets of circumstances.

Japanese). The reason for this is that children adhere to the subset condition for setting the “disjunction-as-positive-polarity-item” parameter, whereas adults have reset the parameter to the superset value. This prediction was tested in an experiment with 21 monolingual Chinese-speaking children (ages 4;4–5;3, mean 4;10). On one of the test trials, Donald Duck was awarded a silver medal because he was able to lift up at least one heavy object, a table or a TV. He would have been awarded a gold medal if he had been able to lift both objects, but he would have received no medal if he had failed to lift either object. In this circumstance, both children and adults were asked to judge whether (19) accurately described why Donald had been awarded a silver medal. Adult Chinese-speakers accepted (19) as an accurate assessment, whereas 20 of the 21 Chinese-speaking children rejected it on the grounds that Donald Duck had succeeded in lifting the table.

- (19) *Tangtaoya meiyou ju-qi zhuozi huozhe dianshiji*
 Donald Duck not-PAST lift-up table or TV
 ‘Donal Duck did not lift up the table or the TV’

This finding is interpreted as further evidence of children’s obedience to the semantic subset principle, which governs the acquisition of parameters whose values are in a subset/superset relation. And the finding provides further evidence for the continuity assumption, which supposes that child language can differ from that of adults, but only in ways in which adult languages can differ from each other.

3 Entailments of focus operators

A second hidden property of natural language is revealed in the entailments of sentences that contain a focus operator (e.g. English *only*; Japanese *dake* and *shika-nai*; Chinese *zhi-you* and *dou-shi*). These entailments straighten out several misalignments that occur between logic and language. We already observed two such misalignments, involving natural language expressions for disjunction. One occurs in adult Japanese, where the expression for disjunction (*ka*) does not conform to De Morgan’s laws. Another misalignment is due to a scalar implicature, which yields the “not both” interpretation of disjunction, except in contexts that include a downward entailing expression. Because of the implicature of “exclusivity” for disjunction in the vast majority of linguistic environments, children encounter little evidence of the Boolean nature of the natural language expressions corresponding to disjunction in classical logic. In our introduction to logic, however, we noted that sentences with the focus operator *only* license the conjunctive entailments associated with De Morgan’s laws, in the covert meaning component. Our next research question is whether or not children know that sentences with

the focus operator *only* contain a hidden downward entailing operator (negation, or its semantic equivalent).

As noted earlier, there is mounting evidence that 4- to 5-years-old children appear to know that *or* licenses conjunctive entailments in downward entailing contexts. So, children’s interpretation of *or* was used to assess their knowledge of the semantics of *only* (Goro, Minai & Crain 2004). One of the test sentences was (20).

(20) Only Bunny Rabbit will eat a carrot or a green pepper.

Under the decomposition analysis, the meaning of (20) can be partitioned into the two conjoined propositions in (21).

- (21) a. *Presupposition*: Bunny Rabbit will eat a carrot or a green pepper.
- b. *Assertion*: Everyone other than Bunny Rabbit will not eat a carrot or a green pepper.

Within the *presupposition* component, the disjunction operator *or* yields *disjunctive* truth conditions: *Bunny Rabbit will eat a carrot or will eat a green pepper*. Suppose, first, that children assign the correct interpretation to *or* within the *presupposition* component. If so, children should avoid the conjunctive entailment of disjunction in the presupposition, so they should accept (20) in the situation where Bunny Rabbit ate a carrot but not a green pepper. The entire truth conditions are schematically represented in Table 3.

Table 3. Condition I

	Carrot	Green Pepper
Winnie the Pooh	*	*
Bunny Rabbit	√	*
Cookie Monster	*	*

Table 4. Condition II

	Carrot	Green pepper
Winnie the Pooh	*	*
Bunny Rabbit	√	*
Cookie Monster	*	√

By contrast, within the assertion meaning component of (21), *or* appears in a downward entailing environment. Therefore, it licenses the conjunctive entailment – that everyone else will not eat a carrot and they will not eat a green pepper.

Consequently, if children assign the correct interpretation to *or* within the assertion, they should reject (21) in the situation represented in Table 4 on the grounds that Cookie Monster ate a green pepper (while, again, Bunny Rabbit ate a carrot but not a green pepper).

The findings were as follows. In Condition I, 21 English-speaking children (age 3;6–5;8) accepted test sentences like (20) 93% of the time. (An adult control group accepted them 100% of the time.) The high acceptance rate in Condition I suggests that children assigned disjunctive (*not both*) truth conditions to *or* within the presupposition meaning component. If children had incorrectly assigned the conjunctive entailment of disjunction, then they should have rejected the test sentences. In contrast, the same children rejected the test sentences in Condition II 90% of the time. (An adult control group rejected them 100% of the time.) The high rejection rate by children in Condition II suggests that they know that the disjunction operator *or* creates conjunctive entailments within the assertion component of the test sentences. Overall, the contrasting pattern of children's behavior in the two conditions suggests that they treated *or* as if it had distinct truth conditions in the same test sentences. This in turn suggests that children have adult-like knowledge about the semantics of *only*, and are able to compute its complex semantic interaction with the interpretation of *or*. These findings are difficult to explain on experience-based accounts of language development. In particular, it is hard to see how such accounts could explain the fact that the same lexical item is interpreted in two different ways in the same sentence. No straightforward learning algorithm would do the trick.

4 Covert negation and conjunction in adult language

Our next goal was to see whether or not English-speaking children assign the Boolean interpretation to natural language conjunction, e.g. *and* in English, ... *mo* ... *mo* in Japanese. Adopting the same research strategy as before, the obvious place to start was to see if children generate the disjunctive (*not both*) entailment in response to simple negated sentences like (22). If so, such sentences should invoke *disjunctive* entailments, as in another of De Morgan's laws of propositional logic: $\neg(X \wedge Y) \Rightarrow \neg X \vee \neg Y$. The simplest experiment that came to mind was to ask children to judge the truth or falsity of (22) in a situation where the Smurf jumped over the tree, but not the pond.

(22) The Smurf didn't jump over the tree and the pond.

If children assign the Boolean interpretation to *and*, then they should accept (22) as a description of the situation, since the disjunctive entailment of (22) is that the

Smurf didn't jump over the tree *or* the Smurf didn't jump over the pond. However, there is a possible non-Boolean interpretation of *and*, according to which conjunction fails to have this disjunctive entailment. On this non-Boolean interpretation, conjunction engenders an entailment according to which both conjuncts are negated: $\neg(X \wedge Y) \Rightarrow \neg X \wedge \neg Y$. If children assign the non-Boolean (*neither*) interpretation to conjunction, then they should reject (22), on the grounds that the Smurf jumped over the tree.

We began by testing a control group of English-speaking adults. We used sentences like (22) in a context in which the character mentioned in the sentences performed only one of the two actions that were mentioned. We tested 38 undergraduates at University of Maryland at College Park. We found that the majority of them disfavored the Boolean (*not both*) interpretation of *and* in the experimental contexts. They rejected the test sentences over two-thirds of the time. We determined, in another experiment, that adults accepted the Boolean (*not both*) interpretation of conjunction if focal stress was placed on *and*. However, in light of adults' acceptance of the non-Boolean interpretation of conjunction without stress, it would hardly be surprising if children, too, assigned the non-Boolean (*neither*) interpretation to negated *and*.

Although the extent of the preference for the non-Boolean interpretation of conjunction was unexpected, the finding was consistent with the claim by Szabolcsi and Haddican's (2004) that English definite conjunctions can be given the same denotation as definite plurals and, therefore, can yield a non-Boolean interpretation under negation. Without stressing *and*, then, simple negated sentences like (22) would not provide an appropriate testing ground for assessing children's semantic knowledge about conjunction in English. If children are like adults, then we would not expect their responses to be consistent with the non-Boolean interpretation of conjunction.

We considered using sentences in which *and* receives focal stress as a way of assessing children's awareness of the Boolean character of conjunction. However, in previous research (e.g. Gualmini & Crain 2002), children were found to perform poorly in detecting the semantic effects of intonation, especially stress. Thus, we conjectured that children might reject the Boolean (*not both*) interpretation for sentences like (22) even with stress on *and*, due to their insensitivity to the interpretive consequences of stress. If so, children's failure to access the Boolean interpretation of conjunction in such an experiment would not provide much insight into their semantic knowledge – it could be that children have the Boolean interpretation of conjunction, but simply fail to access it, favoring the non-Boolean interpretation instead even when conjunction receives focal stress. To resolve the dilemma, we adopted different experimental materials, namely sentences with the focus operator *only*.

Let us illustrate how the covert negation in sentences with *only* interacts with a definite conjunction. Consider sentence (23).

- (23) Only the Smurf jumped over the tree and the pond.

Under the decomposition analysis, the “assertion” of (23) is as in (24).

- (24) For all x such that $x \neq$ the Smurf, it is not the case that x jumped over the tree and the pond.

If the definite conjunction in (23) receives the non-Boolean (*neither*) interpretation within the covert proposition, then (24) effectively means that everyone other than the Smurf failed to jump over either obstacle. By contrast, if the definite conjunction in (23) receives the Boolean (*not both*) interpretation in the covert meaning component of (24), then it means that everyone other than the Smurf failed to jump over both of the obstacles. Following Szabolcsi and Haddican (2004), we assume that both interpretations are available to adult English speakers. However, in sentences containing *only*, such as (23), we had the intuition that the Boolean (*not both*) interpretation within the covert negative proposition would be much easier for people to access than it proved to be in simple negative sentences such as (22), at least when conjunction does not receive focal stress.

To confirm this expectation, we ran an experiment using the Truth Value Judgment Task with 26 adult speakers of English (undergraduates at UMCP). The crucial test condition was a situation in which the Smurf jumped over the tree and the pond and Winnie the Pooh jumped over the pond. In this condition, adults accepted test sentences such as (23) 96% of the time, in the absence of focal stress on *and*. So, adult subjects assigned the Boolean (*not both*) interpretation to the conjunction in the covert negative proposition associated within sentences with *only*. If they had preferred the non-Boolean (*neither*) interpretation of *and*, as they had with simple negative sentences like (22), then they should have rejected the test sentences. The experimental findings revealed that, for adults, the preference that was observed earlier for the non-Boolean (*neither*) interpretation of non-stressed *and* was neutralized in sentences containing *only*, where negation is contained in a covert proposition (see Goro 2004 for analysis).

5 Covert negation and conjunction in child language

This brings us back to our goal of assessing children’s interpretation of *and* under negation. It is evident that we can use the linguistic context in which adult speakers consistently accepted the Boolean interpretation of negated conjunction to

examine children's knowledge about the interaction of *only* and the logical connectives, including *and*.

An experiment was designed to test children's access to the Boolean (*not both*) interpretation of the conjunction operator *and* within the covert negative proposition associated with sentences containing the focus operator *only*. The experiment employed the Truth Value Judgment Task in the description mode. The theme of the experimental story-line was a magic competition, in which three would-be magicians (Aladdin, Neptune, and Genie) attempted to perform feats of magic using their vocabulary of magic words. Toy figures of the three would-be magicians were manipulated by one experimenter, while another experimenter coordinated a computer-generated animation presentation of feats of magic, which was implemented as the toy figures pronounced their magic words. The subject witnessed the figures uttering magic words as they gestured towards objects displayed on a computer screen. Subsequently, objects were animated by the experimenter working the computer (e.g. boxes shook and opened, cars flipped over, a frog turned into a princess). Each of the magicians was awarded prizes if he succeeded in performing one of the feats of magic. These prizes served as reminders for children whose task it was to judge the truth or falsity of test sentences. The test sentences were produced by a puppet, Kermit the Frog (manipulated by a third experimenter) who watched the magic competition along with the child subject.

There were nine trials, presented to children in a fixed order. Five were test trials and four were filler trials. At the end of each trial, Kermit the Frog said what he thought happened on the trial, using a filler sentence or one of the test sentences containing *only* and *and*, as exemplified in (25):

- (25) Only Aladdin opened the blue box and the black box.

Under the analysis we are following, the meaning of the sentence can be decomposed into the two conjoined propositions in (26):

- (26) a. Aladdin opened the blue box and the black box.
 b. For all x such that $x \neq \text{Aladdin}$, it is not the case that x opened the blue box and the black box.

What is crucial is children's interpretation of the second *assertion* component of (26). On the non-Boolean (*neither*) interpretation of *and*, (26) is false if someone other than Aladdin (say Genie) succeeded in opening one of the boxes (but not the other). However, the sentence is true in the same situation on the Boolean (*not both*) interpretation of *and*, since Genie didn't open both of the boxes. The crucial test condition is schematically represented as follows:

Table 5. Condition I

	Blue box	Black box
Aladdin	√	√
Neptune	*	*
Genie	√	*

Recall that adults consistently accepted the test sentences in this condition. If children have the Boolean interpretation of *and*, then they should also accept the test sentences. However, if children assign a non-Boolean semantics to *and*, then they should reject the test sentences. In addition to Condition I, there were two control conditions, diagrammed below.⁵

Table 6. Condition II

	Blue box	Black box
Aladdin	*	√
Neptune	*	*
Genie	√	*

Table 7. Condition III

	Blue box	Black box
Aladdin	√	√
Neptune	*	*
Genie	√	√

The test sentences were false (on the adult interpretation) in these control conditions.⁶ Twenty English-speaking children (age: 3;10–6;3, mean 4;11) participated in the experiment. The subjects were interviewed individually at the Center for Young Children at University of Maryland at College Park. The overall results are summarized in Table 8.

5. Different objects were used in each of the conditions in the actual experiment. In Table 6 and 7, the labels “blue box” and “black box” are used only for ease of exposition.

6. Depending on one’s theory of the nature of the decomposed propositions for sentences with *only*, the test sentence under Condition II may not technically be false, but rather a case of presupposition failure. We set this issue aside.

Table 8. Results

	% Acceptance
Condition I (True)	95 (38/40)
Condition II (False)	15 (3/20)
Condition III (False)	10 (4/40)

The main finding is that children consistently accepted the test sentences in Condition I. The pattern of children's responses was like that of adults. Contrast this with the pattern of children's responses in the experiment using *only* and *or*. Recall that in the Goro, Minai and Crain (2005) study, children consistently rejected test sentences with *only* and *or* when someone other than the focused individual performed an action on one of the objects. This suggests that disjunction licensed a conjunctive entailment (*neither*) for children within the assertion component of the test sentences. The experiment with conjunction used the same type of circumstances, yet children consistently accepted the test sentences. If children were not aware of the Boolean character of *and*, and assigned it the non-Boolean (*neither*) interpretation, then they should have rejected the test sentences, as in the earlier study. We would add that the overall accuracy of children's responses to the control conditions demonstrated that they clearly understood the task.

The same type of poverty-of-stimulus situation arises in the acquisition of conjunction, as in the acquisition of disjunction. As we saw, adult English speakers prefer a non-Boolean interpretation of (unstressed) conjunction in simple negative sentences, which means that the input to children lacks much evidence of the Boolean character of conjunction. On the assumption that children's linguistic knowledge is tied to experience (cf. Tomasello 2003), the fact that children adhered to the Boolean interpretation of *and* is difficult to explain, because children apparently ignore the non-Boolean interpretation that is prevalent in the input.

Children's adherence to the Boolean interpretation of disjunction and conjunction has also been demonstrated in Japanese. Goro, Minai and Crain (2004) carried out the Japanese version of the current experiment, and found that Japanese children also consistently accepted the Boolean (*not both*) interpretation of Japanese conjunction ... *mo* ... *mo* within the covert negative proposition associated with sentences containing the focus operator *dake* ('only'). Like disjunction *ka*, Japanese conjunction ... *mo* ... *mo* is never interpreted within the scope of local negation in simple negative sentences (see Goro 2004 for details). Therefore, the vast majority of the input to Japanese-speaking children fails to exemplify the Boolean character of ... *mo* ... *mo*; nonetheless, these children too were aware of the Boolean character of conjunction.

6 Conclusion

The present studies of the acquisition of disjunction and conjunction reveal that children are not misled by input. Although the input concerning the semantics of conjunction in English and Japanese is misleading to differing degrees (cf. Goro 2004), children acquiring these languages *uniformly* adhere to the “logical” Boolean interpretation of conjunction. We see no straightforward way for an experience-based account to explain the pattern of children’s behavior. We are led to endorse an alternative account, according to which the semantics of conjunction is innately specified in Universal Grammar.

§Atlas, Jay D. 1993. The importance of being “only”: Testing the neo-Gricean versus neo-entailment paradigms. *Journal of Semantics* 10. 301–318.

Atlas, Jay D. 1996. ‘Only’ noun phrases, pseudo-negative generalized quantifiers, negative polarity items, and monotonicity. *Journal of Semantics* 13. 265–328.

Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini & Luisa Meroni. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In Ann H.-J. Do, Laura Domínguez & Aimee Johansen (eds.), *Boston University Conference on Language Development* (BUCLD) 25, 157–168. Somerville, MA: Cascadilla.

Crain, Stephen. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 4. 597–650.

Crain, Stephen & Paul Pietroski. 2001. Nature, nurture and Universal Grammar. *Linguistics and Philosophy* 24. 139–185.

Crain, Stephen & Paul Pietroski. 2002. Why language acquisition is a snap. *The Linguistic Review* 19(1/2). 163–183.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: The MIT Press.

Crain, Stephen, Takuya Goro & Rosalind Thornton. 2006. Language acquisition is language change. *Journal of Psycholinguistic Research* 35(1). 31–49.

Crain, Stephen, Weijia Ni & Laura Conway. 1994. Learning, parsing, and modularity. In Charles Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives on sentence processing*, 443–467. Hillsdale, NJ: Lawrence Erlbaum.

Crain, Stephen, Amanda Gardner, Andrea Gualmini & Beth Rabbin. 2002. Children’s command of negation. *Tokyo Conference on Psycholinguistics* (TCP) 3, 71–95. Tokyo: Hituzi.

Fintel, Kai von. 1999. NPI-licensing, Strawson entailment, and context dependency. *Journal of Semantics* 16. 97–148.

Goro, Takuya. 2004. The emergence of Universal Grammar in the emergence of language: The acquisition of Japanese logical connectives and positive polarity. Ms., University of Maryland at College Park.

Goro, Takuya & Sachie Akiba. 2004. The acquisition of disjunction and positive polarity in Japanese. *West Coast Conference on Formal Linguistics* (WCCFL) 23, 251–264. Somerville, MA: Cascadilla.

- Goro, Takuya, Utako Minai & Stephen Crain. 2004. Boolean conjunction in child Japanese. Poster presented at the 1st GALANA conference, University of Hawai'i at Manoa.
- Goro, Takuya, Utako Minai & Stephen Crain. 2005. Two disjunctions for the price of only one. In Alejna Brugos, Manuella R. Clark-Cotton & Seungwan Ha (eds.), *Boston University Conference on Language Development* (BUCLD) 29, 228–239. Somerville, MA: Cascadilla.
- Grice, Paul. 1975. Logic and conversation. In Peter Cole & Jerry Morgan (eds.), *Syntax and semantics 3: Speech acts*, 41–58. New York, NY: Academic Press.
- Gualmini, Andrea & Stephen Crain. 2002. Why no child or adult must learn De Morgan's laws. In Barbara Skarabela, Sarah Fish & Anna H-J. Do (eds.), *Boston University Conference on Child Language Development* (BUCLD) 26, 367–378. Somerville, MA: Cascadilla.
- Gualmini, Andrea & Stephen Crain. 2004. Operator conditioning. In Alejna Brugos, Linnea Micciulla & Christine E. Smith (eds.), *Boston University Conference on Child Language Development* (BUCLD) 28, 232–243. Somerville, MA: Cascadilla.
- Gualmini, Andrea, Stephen Crain, Luisa Meroni, Gennaro Chierchia & Maria Teresa Guasti. 2001. At the semantics/pragmatics interface in child language. In Rachel Hastings, Brendan Jackson & Zsófia Zvolenszky (eds.), *Proceedings of Semantics and Linguistic Theory* 11, 231–247. Cornell, NY: CLC Publications.
- Herburger, Elena. 2000. *What counts: Focus and quantification*. Cambridge, MA: The MIT Press.
- Horn, Laurence. 1969. A presuppositional approach to *Only* and *Even*. *Chicago Linguistic Society* (CLS) 5, 98–107.
- Horn, Laurence. 1996. Presupposition and implicature. In Shalom Lappin (ed.), *Handbook of contemporary semantic theory*, 299–319. Oxford: Blackwell.
- Jing, Chunyuan, Stephen Crain & Ching-Fen Hsu. 2005. The interpretation of focus in Chinese: Child vs. adult language. *Tokyo Conference on Psycholinguistics* (TCP) 6, 165–190.
- Szabolcsi, Anna. 2002. Hungarian disjunctions and positive polarity. In István Kenesei & Péter Siptár (eds.), *Approaches to Hungarian* 8, 217–241. Budapest: Akadémiai Kiadó.
- Szabolcsi, Anna & Bill Haddican. 2004. Conjunction meets negation: A study in cross-linguistic variation. *Journal of Semantics* 21(3), 219–249.
- Thornton, Rosalind. 1990. Adventures in long-distance moving: The acquisition of complex Wh-questions. University of Connecticut PhD dissertation.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Zwart, Jan-Wouter. 1996. N-Feature checking in Germanic verb second configurations. In Hoskuldur Thráinsson, Samuel David Epstein & Steven Peter (eds.), *Studies in comparative Germanic syntax II*, 257–275. Dordrecht: Kluwer.

Name index

A

Acredolo, Linda P. 265
 Aiello, Leslie C. 35
 Aitchison, Jean. 125
 Akiba, Sachie 281, 283, 284
 Akins, Kathleen 83
 Allan, Keith 270
 Allen, James 54
 Allman, John 65
 Altman, Nolan R. 173, 178
 Ann, Jean 151
 Arbib, Michael 155
 Argue, Debbie 29
 Arnheim, Rudolf 267
 Atlas, Jay D. 278

B

Badler, Norman I. 126
 Bar-On, Dorit 234
 Barsalou, Lawrence W. 156, 167
 Bar-Yosef, Ofer 22, 27
 Bear, D. 58
 Bear, Mark 66
 Bearzotti, Federica 175
 Bechtel, William 83
 Beer, Randall 76
 Bekkering, Harold 165
 Bellugi, Ursula 252
 Bentham, Johan van 218
 Beran, Michael 45, 63
 Berlin, Brent 105, 243
 Berthoz, Alain 166
 Bickerton, Derek 69
 Biederman, Irving 124
 Bierwisch, Manfred 207
 Birdsall, Joseph 39
 Black, Max 243
 Blinov, Arcady 7, 8
 Boesch, Christophe 52, 61
 Boesch, Hedwige 52, 61
 Bonfiglioli, Claudia 5
 Bonnicksen, Robson 15
 Bordaz, Jacques 22
 Borges, François 22
 Borgatti, Renato 6, 175, 176,
 184, 185

Borghi, Anna M. 5, 153, 154, 156
 Boronat, Consuelo B. 167
 Boyes Braem, Penny 265
 Brace, C. Loring 38
 Braffman, Bruce H. 173, 178
 Bräm, Thuering 265
 Brett, Matthew 154
 Broadfield, Douglas C. 46, 47,
 64, 65
 Brooks, Lee 263
 Brooks, Rodney 76
 Brown, Francis H. 48
 Brown, Peter 11, 12, 37, 55, 57
 Brumm, Adam 11, 12, 22, 29
 Bub, Daniel N. 154, 166
 Bucy, Richard S. 99
 Bukach, Cindy M. 154, 166
 Burgess, Paul W. 66
 Buxbaum, Laurel J. 154, 166

C

Callahan, Errett 21
 Caplan, David 174
 Cappa, Stefano 172
 Casterline, Dorothy C. 151
 Castiello, Umberto 155
 Chalmers, David 216, 217, 230
 Chang, Mario 113
 Chao, Linda L. 154
 Chella, Antonio 106
 Chilosi, Anna Maria 172
 Chomsky, Noam 63, 225, 226
 Christiansen, Morten H. 19, 20
 Churchland, Paul 84, 89, 90
 Clark, Andy 80
 Clark, Graham 48
 Cogill-Koez, Dorothea 3, 8, 253,
 254, 258, 259, 265
 Cohen, Michael 175
 Collins, Mike 3
 Colosimo, Pamela F. 66
 Colpo, Giovanni 178
 Connors, Barry W. 66
 Conway, Christopher M. 19, 20
 Conway, Laura 282

Corballis, Michael C. 109
 Cornoldi, Cesare 178
 Courchesne, Eric 185
 Coventry, Kenny R. 153
 Cox, Maureen 267
 Crabtree, Don E. 14
 Crain, Stephen 8–10, 276,
 280–283, 286, 288, 292
 Craver, Carl 83
 Crawford, John R. 178, 183
 Creem, Sarah H. 166
 Croft, William 148
 Croneberg, Carl G. 151
 Cronin-Golomb, Alice 174
 Crosson, Bruce 172
 Cummins, Robert 89, 90

D

Darden, Lindley 83
 Davidson, Donald 7, 8, 130, 226,
 227, 233–238, 242, 244, 245
 Davidson, Iain 2, 3, 21, 36, 38,
 39, 48–50, 53
 De Fabritiis, Paolo 178
 De La Coste-Lareymondie,
 Marie C. 47
 Deacon, Terence W. 64, 65, 69
 Del Giudice, Ennio 178
 Delagnes, Anne 21, 51
 DeLong, Mahlon 172
 Denier van der Gon, J. J. 99
 Dennett, Daniel C. 17
 DeVore, Irven 52
 Dibble, Harold L. 22
 Dixon, Peter 156
 Dow, Robert S. 173–175
 Doya, Kenji 99
 Duffner, Patricia K. 175
 Duhamel, Jean René, 154
 Dumais, Susan T. 166
 Dummett, Michael 218
 Dunbar, Robin I. M. 35
 Duncan, Susan D. 258
 Dunn, Nathan A. 81, 83, 84

- E
 Ebbinghaus, Heinz-Dieter 203
 Edgington, Dorothy 218
 Ellis, Rob 155, 156
 Emmorey, Karen 254, 257, 263
 Engberg-Pedersen, Elizabeth 263
- F
 Fabbro, Franco 6, 172–178, 185
 Fagg, Andrew H. 155
 Falk, Dean 12, 29, 38, 55, 57, 58, 64, 68
 Fauconnier, Gilles 195
 Feldesman, Marc R. 37
 Feldman, Jacob 267
 Fenson, L. 267
 Ferrée, Thomas C. 81–83
 Fiez, Julie A. 173–175
 Fintel, Kai von 278, 280
 Fitch, Frederick 218, 219, 231
 Fleage, John G. 48
 Flum, Jörg 203
 Fodor, Jerry 166
 Foley, Robert A. 11, 27, 35, 38, 46, 48
 Forbus, Kenneth D. 197
 Forde, Emer M. E. 161
 Frigaszy, Dorothy M. 19
 Frixione, Marcello 106
- G
 Gaglio, Salvatore 106
 Gallese, Vittorio 167
 Gallistel, Charles R. 79, 88, 89
 Galvan, Dennis 253, 270, 271
 Gamut, L. T. F. 134
 Gangitano, Massimo 156
 Ganter, Bernhard 203
 Gärdenfors, Peter 4, 5, 96, 98, 100, 102, 103, 105, 106, 109
 Gardner, Robert J. 178, 186
 Garner, Wendell R. 103
 Garrod, Simon C. 153
 Garthwaite, Paul H. 178, 183
 Gentilucci, Maurizio 156
 Gentner, Dedre 191, 197
 Gerlach, Christian 154, 161
 Ghez, Claude 173
 Gibson, Kathleen R. 20, 45, 63
 Gilboa, Itzhak 117
 Giorgi, Cesare 175
 Givón, Talmy 20
 Glenberg, Arthur M. 156, 167
 Glover, Scott 156
 Goldin-Meadow, Susan 265
 Gombrich, Ernst H. 267
 Goodall, Jane 61
 Goodnow, Jacqueline 267, 268
 Goodwyn, Susan W. 265
 Goro, Takuya 8–10, 280–284, 286, 289, 292, 293
 Grafton, Scott T. 154, 165
 Grice, Paul 280
 Griffiths, Ruth 177, 180
 Groves, Colin 29
 Gruber, Geoffrey 130, 147
 Grupp, M. T. 178
 Grush, Rick 99, 100, 115
 Gualmini, Andrea 281, 288
 Gulz, Agnetha 109
 Gust, Helmar 6, 192, 193, 197, 200, 201
- H
 Haddican, Bill 288, 289
 Hagen, Margaret A. 267
 Hahn, Ulrike 102
 Hakeem, Atiya 65
 Hampton, James 102
 Handy, Todd C. 157
 Hawthorne, John 229, 230
 Hayden, Brian 22
 Hayes, Patrick J. 109
 Heaton, Robert K. 177
 Henneberg, Maciej 35, 36
 Herburger, Elena 278
 Herholz, Karl 175
 Herzig, Melissa 257, 263
 Hessel, Germund 100
 Holliday, Trenton W. 35
 Holloway, Ralph L. 35, 46, 47, 64, 65
 Holyoak, Keith J. 191
 Horn, Laurence 278, 294
 Hsu, Ching-Fen 281, 283
 Hughes, George E. 134
 Humphrey, Nicholas K. 99
 Humphreys, Glyn W. 156, 161
 Hutchings, W. Karl 22
- I
 Indurkha, Bipin 194
 Itihara, M. 54
 Ito, Masao 99
- J
 Jackendoff, Ray 5, 123, 124, 130, 131, 139, 140
 Jackson, Frank 216, 220, 221, 224–226, 228, 230, 231
 Jacob, Teuku 12
 Jäger, Gerhard 105
 Jeannerod, Marc 155
 Jerison, Harry J. 45, 46, 65
 Jing, Chunyuan 281, 283
 Job, Remo 158
 Johnson, Mark H. 172
 Johnson-Pynn, Julie 19, 20
 Justus, Timothy 173, 176, 186
- K
 Kaas, Jon H. 64
 Kalman, Rudolf E. 99, 100, 114, 115, 117, 118
 Karol, Alankar 117
 Kaschak, Michael P. 156
 Kawato, Mitsuo 99, 115
 Kay, Paul 105, 243
 Keefe, Rosanna 241
 Keener, Matthew T. 155
 Kellenbach, Marion L. 154
 Kendon, Adam 265
 Khlentzos, Drew 7, 216, 229
 Klatzky, Roberta L. 156
 Kleckner, Geoff 37
 Klima, Edward 252
 Kohn, Marek 12
 Kokinov, Boicho N. 191
 Komosinski, Maciej 84
 Köpf-Maier, Petra 151
 Kossler, H. Branch 155
 Kuhn, Steven L. 22, 27
 Kühnberger, Kai-Uwe 6, 192, 193, 197, 200, 201
 Kulvicki, John 90
- L
 Lahr, Marta Mirazón 11, 27
 Lakoff, George 129, 148, 167, 264
 Landauer, Thomas 166
 Lang, Ewald 207
 Law, Ian 154, 161
 Law, Keith 197
 Lee, Phyllis C. 38
 Lee, Richard B. 52
 Lee, Sang-Hee 35
 Leggio, Maria G. 174–176, 185
 Leiner, Alan L. 173–175
 Leiner, Henrietta 173–175
 Levisohn, Lisi 174
 Levy, Elena T. 250
 Lewin, Roger 21, 35, 46, 48

Lewis, David 216, 222, 225, 228,
230, 231
Liddell, Scott K. 253
Lloyd, Dan 84
Lockery, Shawn R. 81–83
Lotto, Lorella 158
Lovell, Mary S. 67
Lundy, John K. 37

M

Machamer, Peter 83
MacNeillage, Peter F. 20
Maddox, W. Todd 103
Maedche, Alexander 203
Mandik, Pete 3, 80, 84, 87
Mariën, Peter 175
Marr, David 5, 124
Martin, Alex 154, 167
Martin, Robert D. 12, 27
Masson, Michael E. J. 154, 166
Matthews, Robert 90
McCarthy, John 109
McCawley, James D. 129
McDougall, Ian 48
McHenry, Henry M. 37, 46
McNeill, David 250, 258, 265
Medin, Douglas L. 263
Melara, Robert D. 103
Mellars, Paul 21
Metzger, Melanie 253
Miguel, Carmen de 35, 36
Minai, Utako 8–10, 286, 292
Mitchell, Catherine 40
Molinari, Marco 174–176, 185,
186
Moore, Mark W. 2, 11, 12, 18, 19,
22, 28, 29, 40, 50, 51, 59, 60
Moruzzi, Giuseppe 174
Morwood, Michael J. 2, 3, 11, 22,
29, 30, 39, 40, 53–55, 57, 59, 62

N

Nadeau, Stephen E. 172
Naidich, Thomas P. 173, 178
Newport, Elissa 248
Ni, Weijia 282
Nicoletti, Roberto 5
Nishida, Toshisada 61
Nishina, Hiroyuki 5, 126, 134,
151
Noble, William 38, 39, 50, 53

O

O'Connor, Sue 39, 54
Oosterzee, Penny van 39

P

Paradis, Michel 177, 178, 181
Paradiso, Michael 66
Parsons, Terence 130
Patterson, Karalyn 154
Paulson, Olaf B. 154, 161
Pecher, Diane 156, 167
Pedelty, Laura L. 250
Pelcin, Andrew W. 22
Pelegriin, Jacques 12
Perry, John 223, 224
Peters, Stephen 203
Petersen, Steven E. 173–175
Peylo, Christoph 203
Phillips, Cary B. 126
Phillips, Jacqueline 156
Pickett, Emily R. 173
Pietroski, Paul 283
Pinker, Stephen 69
Piñon, Denise 19
Plotkin, Gordon D. 197
Poizner, Howard 252
Poncet, Michel 154
Postal, Paul M. 129, 130
Preuss, Todd 64
Proffitt, Dennis R. 166
Pulvermüller, Friedemann 156
Pylyshyn, Zenon 166

Q

Qi, Huixin 64

R

Ramscar, Michael 102
Rescher, Nicholas 238
Rey, André 177, 179, 184
Reynolds, Tim 22
Ricciardelli, Paola 5
Richardson, Robert 83
Riddoch, M. Jane 161
Rightmire, Philip 12
Riva, Daria 175
Roberts, Richard G. 54
Roche, Hélène 21, 51
Roebrooks, Wil 52
Romano, Alfonso 178
Rooij, Robert van 105
Rosch, Eleanor 264
Rubichi, Sandro 5
Ruff, Christopher B. 35
Rumbaugh, Duane 45, 63
Rumiati, Raffaella 155

S

Saffran, Eleanor M. 155

Sandgathe, Dennis M. 22
Sapthomo, E. Wayhu 30
Savage-Rumbaugh, Sue 21, 61
Schaffer, Marguerite M. 263
Schembri, Adam 253, 262
Schick, Kathy D. 12, 13, 51, 65
Schmahmann, Jeremy D. 173,
174, 184
Schmeidler, David 117
Schmid, Ute 6, 192–194, 197,
200, 201
Schwartz, Myrna F. 154
Scott, Richard B. 175
Shallice, Tim 153, 177
Shanker, Stuart 61
Sherman, Janet C. 173, 174, 184
Silk, Angele M. J. 267
Silveri, Maria Caterina 174–176,
185
Simmel, Georg 234, 238
Singleton, Jenny 265
Sirigu, Angela 154
Sivik, Lars 105
Skoyles, John R. 40
Smith, Peter 241
Smith, Wayne H. 261
Soejono, Radien P. 30
Sommers, Peter van 267
Spinnler, Hans 177
Spoor, Fred 37
Spriggs, Matthew 39
Stalnaker, Robert 229
Stanford, Craig B. 52, 61
Sternberg, Saul 166
Stickel, Mark E. 204
Stiner, Mary C. 40
Stokoe, William C. 151, 252
Stout, Deitrich 13, 65
Stumme, Gerd 203
Suarez, Mauricio 90
Suddendorf, Thomas 109
Supalla, Ted 253, 257–260, 263,
270
Suppes, Patrick 90
Swann, Philip 20
Swoyer, Chris 92
Szabolcsi, Anna 282, 289

T

Taft, Charles 105
Talmy, Leonard 129, 270
Taub, Sarah 253, 270, 271
Tavano, Alessandro 6, 175, 176,
178

Taylor, Talbot J. 61
Tebbi, Cameron K. 175
Tennant, Neil 218
Tettamanti, Marco 173
Thach, William T. 173
Thieme, Hartmut 53
Thomas, Glyn V. 267
Thornton, Rosalind 280, 281, 283
Titomanlio, Luigi 178
Tobias, Phillip V. 46, 47
Tognoni, Gianni 177
Tomasello, Michael 20, 292
Toribio, Josefa 80
Torre, Ignacio de la 51
Toth, Nicholas 12–14, 51, 65
Trauner, Doris 172
Trinkaus, Erik 35
Trouillas, Paul 177, 178
Tucker, Mike 155, 156
Turner, Mlachebekjeark 195

U
Ullman, Michael T. 174, 176, 186

V
Vallar, Giuseppe 172
Veramonti, Tracy 154
Vereschagin, Alex 3
Verhoeven, Theodore 29
Versante, Laura 265
Veth, Peter 39, 54
Vicari, Stefano 172
Vos, John de 54

W
Wakefield, Jennie 64, 65
Walter, Christoph 204, 206
Warrington, Elizabeth 153
Watson, Karli 65
Webber, Bonnie L. 126
Wechsler, David 177, 179
Weiner, Steve 40
Wenban-Smith, Francis F. 22
Whittaker, John C. 21
Wilkins, David 265
Wilkins, Wendy K. 64, 65
Wille, Rudolf 203

Williams, Mary-Anne 4, 5, 96, 97, 102, 106
Williamson, Timothy 218, 241
Wohlschläger, Andreas 165
Wolpert, Daniel M. 99
Wolpoff, Milford H. 35
Wong, Kate 12
Wurz, Sarah 22
Wynn, Thomas 21, 52, 59

Y
Young, David E. 15
Yuan, Michael S. 46, 47, 64, 65

Z
Zatsiorsky, Vladimir 126
Zee, Emile van der 148–150, 153
Zeelenberg, Renž, 167
Zettin, Marina 175, 176, 185
Zilles, Karl 47
Zinnes, Joseph L. 90
Zwaan, Rolf A. 156
Zwart, Han-Wouter 276

Subject index

A

a posteriori materialism 217
a priori deducibility 217, 218,
 220–222, 224, 227, 228, 231,
 232
 abstraction 192, 193, 197, 208,
 210
 accessibility 123, 125, 134–138,
 144–146
 acquired cerebellar lesions 171,
 174, 176, 185
 action 1, 5, 6, 11–15, 20, 21, 38,
 48–50, 52, 53, 62, 96, 98–101,
 107–110, 114, 116–118, 123–152,
 153, 154–157, 160, 165–167, 184,
 211, 267, 268
 adaptive behaviour 65, 77, 80,
 210
 affect 173–175
 agent 4, 5, 95–121, 126, 138–140,
 147, 150, 210, 215, 225, 238
 agrammatism 171, 173–176, 185
 agreement 176, 181, 185
 AIBO 111, 112, 116
 analogue 221, 239, 244, 250,
 251, 253, 255–258, 260, 263,
 265, 268
 analogy 191, 194, 195, 197, 198
 animal 1, 3, 38, 45, 49, 52, 53, 59,
 61–63, 69, 70, 75, 77–80, 83,
 97–99, 101, 107–109, 125, 126,
 154, 161, 204, 205, 208, 210,
 238, 258, 263, 270
 anomia 174, 183
 anticipatory planning 109, 110
 anti-instances 199
 anti-unification 197, 199
 ape 47, 51, 56–59, 61, 63, 69, 97,
 109, 228
 aphasia 172, 176, 177, 181, 252,
 254
 apraxia 177–179, 184
 assertion 278, 279, 281, 286, 287,
 289, 290, 292

ataxia 174, 177, 178
 attention 20, 21, 174, 180, 184
 auditory sequential 175
 australopithecines 39, 57, 63–65
Australopithecus 37, 46, 57
 autistic-like behaviour 174, 175
 automatic activation 155, 165,
 167, 168
 automaticity 20, 21

B

balance 173, 174
 basal ganglia 6, 171–173
Be Cooperative 280
 bilateral torque 47, 57
 biological agent 95
 bird 55, 58, 98, 204, 205, 268
 blending 195
 body size 2, 29, 35, 37–39, 45,
 56, 57, 110
 brain 1–5, 13, 20, 21, 38, 45–47,
 50, 51, 57, 58, 62–69, 75,
 99–101, 154, 165, 172, 175, 223,
 225, 226, 229, 252, 254
 brain reorganisation 43, 57, 61,
 63–65, 67
 brain size 2, 3, 12, 29, 35–42, 43,
 45, 46, 57, 62–64
 brain stem 174
 Broca's area 20, 21, 46, 47, 57, 65,
 68, 172–174
 Brodmann's area 10 58, 62,
 65, 66

C

Caenorhabditis Elegans (*C. Elegans*), 81, 82, 84–86, 89
 categorisation 1, 7, 8, 101, 102,
 104, 106, 126, 139, 154, 157,
 161, 192, 196, 233, 247, 248,
 255, 258, 260, 261, 263, 264,
 269–271, 296
 category similarity 161, 165
 causative analysis 129–131, 133,
 150

central pattern generator 84
 Cerebellar Cognitive Affective
 Syndrom 174
 cerebellar lesions 171, 173–176,
 185
 cerebellar structures 172
 cerebellar tumor 174
 cerebellum 6, 99, 171–189
 cerebral hemispheres 172, 175
 cerebrum 99
 chemotaxis 3, 4, 77, 80–88
 childbirth 38
 chimpanzees 2, 4, 20, 43, 46, 51,
 52, 57, 58, 61–65, 69
 Chinese 9, 280, 281, 283–285
 classical logic 8, 275, 280, 281,
 284, 285
 classifier predicate 8, 252–254
 closed class word 176, 185
 cognition 1–3, 5, 6, 11, 12, 19, 21,
 27, 28, 35, 40, 43, 44, 49, 51, 52,
 54, 57, 63, 65, 66, 68, 69, 92, 95,
 98, 99, 123, 125, 126, 151, 172,
 174, 175, 178, 192, 196, 265, 266
 cognitive capabilities 95–97, 101,
 111, 116, 119
 cognitive evolution 20
 cognitive operations 50, 62
 cognitivist theories 166, 167
 collaboration 4, 95–121
 colour 7, 8, 90, 101–106, 110, 113,
 114, 153, 155, 213, 215, 220–231,
 238–240, 242–245
 combinatorial explosion 167
 communication 1–4, 8, 43, 54,
 70, 95–121, 171, 175, 247–250,
 252, 255, 258, 265, 269, 271
 compatibility paradigms 155
 competition 38, 39, 263
 comprehension 6, 66, 171, 172,
 176, 181, 183, 185, 186
 computation 7, 79, 80, 86, 118,
 174, 191–193, 206
 computational capacity 6, 172, 173

concepts 5, 49, 62, 96–98,
 100–103, 105, 106, 113, 114,
 116–119, 153, 155, 157–159,
 165–167, 191–197, 201, 203–211,
 214, 216, 223, 228–231, 235,
 236, 238, 244, 253, 258
 concepts as arbitrary sym-
 bols 167
 conceptual schemes 7, 233–246
 conceptual spaces 4, 5, 95, 97,
 101–107, 109, 110, 114, 116–118
 conceptual structure 5, 102,
 119, 123–125, 130, 131, 138–140,
 147–150
 conjunction 8, 9, 218, 275, 276,
 284, 287–290, 292, 293
 continuity hypothesis 283
 conventional 192, 195, 211, 251,
 254, 258, 261, 263, 267–270
 cooperation 61, 62
 coordination 6, 66, 101, 173,
 174, 184
 cortical 6, 171, 172
 cranial capacity 35–37, 45, 46
 creativity 192
 cued representations 5, 95–97,
 101, 106, 109, 110, 114, 119

D
 De Morgan's laws 8, 9, 276, 280,
 281, 283–285, 287
 depiction 247–250, 252–255,
 257–271
 derived features 11, 58, 65, 66
 detached representations 5,
 95–97, 101, 106–110, 114, 116,
 119
 development 19–21, 38, 47, 69,
 107, 172, 176, 178–180, 283, 287
 disjunction 8, 9, 275, 276,
 278–287, 292, 293
 downward entailment 277–279
 drawing 5, 153, 157–159, 162,
 165, 166, 247–249, 252, 254,
 266–271
 dysarthria 174, 180, 185
 dysprosodia 174

E
 electric circuit analogy 197, 198
 embodied theories of cogni-
 tion 157, 166
 endocast 47, 55, 57, 58

English 8, 9, 148, 149, 176,
 235, 239–245, 249, 254, 268,
 275, 276, 278–285, 287–289,
 291–293
 epiphenomenalism 224
 evolution 1–4, 11, 12, 17, 20, 35,
 37, 38, 40, 43–45, 49–52, 55, 57,
 63, 66–70, 75–77, 86, 89, 95,
 98, 106, 115, 214
 exclusive-or 280, 282
 exemplar 5, 102, 196, 258–261,
 263, 268
 existential movement formu-
 la 124, 125, 134, 136, 137, 144
 expensive organs 38

F
 feedback 82, 84, 85, 98, 99, 107,
 115, 116, 159
 fire use 3, 40, 51, 52, 67
 Fitch's Argument 218, 219, 231
 flake unit 2, 11, 12, 14–22, 27–29, 50
 Flores 2, 3, 11, 12, 29, 39, 40,
 43–73
 focal stress 288, 289
 frame problem 109
 frontal lobe 58, 65, 67, 69, 174,
 185
 functional features 153, 154
 functional neuroimaging 154

G
 gaze 173
 generalisation 191–193, 197,
 199–201, 204–209
 generation 95, 154, 171, 173,
 175, 181
 gesture 24, 29, 110, 154–156, 166,
 247–250, 258, 262, 265, 266,
 269–271
 groundedness 97, 114
 grounding 97

H
 heat flow analogy 197–202
 hemisphere 47, 64, 172, 173, 175,
 252, 254
 Heuristic-Driven Theory Pro-
 jection (HDTTP) 7, 191–193,
 197, 200, 201, 206, 211
 heuristics 90, 193, 200, 201
homo erectus 11, 12, 29, 43, 44,
 46–48, 52, 54–59, 64–66, 68

homo floresiensis 2, 3, 11–33,
 35–42, 43–73
 homomorphism 90, 91
 human cognition 6, 151, 192, 196
 humans 2, 58, 64, 77, 97, 98, 105,
 107, 111, 126, 200, 210, 214,
 221, 228, 238, 239, 243, 244,
 255, 266
 hunting 3, 40, 51–53, 59–61,
 66, 67

I
 iconicity 249, 252, 253, 255
 idiom 194, 195
 implicature of exclusiv-
 ity 280–282, 284, 285
 inclusive-or 275, 280, 281, 284
 information 3–6, 8, 52, 75, 79,
 80, 83–85, 87–92, 95–98,
 100–103, 106, 109, 110, 113–118,
 123–125, 140, 148, 153–157, 160,
 161, 165–168, 172, 173, 185, 186,
 225, 228, 231, 258, 259
 inner world 95–99, 107–109, 111,
 114, 119
 instrumental 139, 140
 intentions 110, 111, 117, 119, 210
 interaction 5, 39, 61, 97, 98, 110,
 184, 228
 interference 157, 161, 165, 166,
 175
 irony 194
 isomorphism 75, 87–92
 Italian 158, 176, 178, 180–182, 185,
 265, 277, 278

J
 Japanese 9, 139, 148, 150,
 280–285, 287, 292, 293

K
 Kalman filter 99, 100, 114, 115,
 117, 118
 kinematics 118, 156
 Knowledge Argument 7,
 213–216, 220–225, 227, 230–232

L
 language acquisition 6, 171, 172,
 174, 176, 186, 275, 282, 283
 language comprehension 6, 171,
 172, 176, 181, 183, 185, 186
 language processing 6, 171–173,
 176, 186, 248

- language production 6, 171, 172, 176, 178
- learning 6, 9, 20, 21, 38, 78, 89, 105, 171, 173, 174, 176, 177, 185, 186, 192, 196, 210, 211, 226, 275, 287
- left ventral premotor cortex 154
- lexicon 6, 171, 186
- Liang Bua 2, 11, 22–30, 54, 60, 62
- lithic design space 11–33
- lithic technology 11, 22
- lobes 13, 46, 66, 173
- location 4, 79, 81, 105, 113–115, 117, 118, 131, 139, 147, 156, 251, 253, 256–259, 261–263, 266
- locomotion 124, 173, 270
- lunate sulcus 46, 57, 64, 65
- M**
- mammal 45, 46, 54, 62, 98, 207, 208
- manipulability 5, 153–170
- materialism 213–218, 220–224, 227–231
- meat acquisition 38
- memory 4, 65, 66, 77–81, 84, 86–88, 98, 174–177, 179, 180, 184–186, 211, 225, 250, 259
- mental representation 3–5, 75, 77, 90, 215, 216, 248
- metaphor 6, 107, 191–212
- microliths 48, 49, 60
- minimalist approach 49–51, 60
- modal logic 5, 123–152, 218
- modality specific domains 167
- model 2, 3, 5, 11, 12, 14, 16, 20, 23–28, 35, 39, 40, 43–45, 55, 59–70, 77, 81, 82, 86, 87, 90, 98, 102, 103, 114, 115, 117, 123–126, 134–136, 196, 207, 216–218, 248, 252, 253–255, 257–260, 263, 267, 268, 270, 271
- modern humans 2, 14, 20–22, 27, 29, 39, 47–50, 53–55, 58, 61, 65
- modulation 172, 174, 177
- mood 172
- morphological features 57, 65
- morphosyntax 6, 171, 176
- motor preparation 156, 165
- motor program 157, 161, 165, 166, 168
- motor skills 19, 20, 174, 184
- movement 3–5, 8, 80, 82, 86, 99, 124, 126–128, 132–136, 138, 140, 142, 146–151, 155, 157, 165, 172–174, 184, 251, 253, 256–267, 269, 271
- muscles 80, 83, 84, 87, 99, 100, 184
- mutism 174, 175
- N**
- narrative 178, 183, 185, 186
- natural objects 153, 154, 157–159, 161–165, 167
- naturalism 7, 79, 184, 213–216, 231
- negation 8, 9, 275, 276, 279–284, 286–289, 292
- negative evidence 283
- nematode 3, 4, 81
- neural network 4, 75, 81, 84, 85
- nutrition 38, 61
- O**
- object permanence 100, 101
- on-line dorsal system process 155
- only (focus operator), 278–280, 285, 286, 288, 290
- ontology 7, 103, 191–212
- oracle 193, 201, 230
- orientation 14, 85–87, 158, 177, 179, 251, 253, 256, 257, 259, 261–263
- outer world 97, 98, 102, 110, 112–115, 117, 118
- P**
- parameter 118, 174, 183, 222, 238, 251, 256, 265, 266, 270, 282–285
- perceptions 4, 79, 95–121
- phenomenal 8, 213, 214, 216, 217, 221–225, 228, 230, 232
- phonology 6, 171, 173, 176, 181, 248
- phototaxis 80
- picture 5, 107, 154–163, 165, 166, 178, 184, 249–252, 259, 268, 269
- planning 1, 2, 4, 16, 52, 53, 58, 65, 66, 69, 78, 95–98, 101, 107–110, 114, 116, 173, 174, 184
- plans 95–99, 108–110, 114, 116, 117, 119
- plasticity 69, 175, 186
- pons 178
- positive polarity item 9, 282, 284, 285
- possible world 92, 123, 134, 216
- posterior fossa 174, 175
- posture 61, 98, 126, 127, 132–135, 141, 156, 173, 174
- presupposition 77, 278, 279, 286, 287, 291
- primary visual cortex (PVC) 46, 64
- primitive 11, 58, 99, 126, 173, 267, 268
- procedural 6, 171, 173, 176, 177, 180, 184, 186
- processing 3, 6, 64, 79, 101, 113, 116, 117, 155, 161, 165, 166, 171–173, 176, 177, 186, 248, 249, 252, 254, 255, 271
- production 6, 23, 26, 47, 48, 50, 51, 59, 60, 67, 83, 161, 171, 172, 174, 176, 179, 249, 265
- property 6–8, 51, 79, 88, 90–92, 96, 100, 103, 105, 106, 131, 191, 193–195, 197, 200, 202, 203, 206–210, 213, 223–226, 228, 230–232, 238, 277, 278, 280, 281, 285
- prototype 8, 102, 105, 258–261, 263–265, 268, 269
- prototype theory 102
- Q**
- qualia 7, 223, 224, 230
- qualitative physics 197
- quality dimensions 103–105, 117
- R**
- radical interpretation 226, 233, 234, 236, 237, 243–245
- reading 168, 178, 180, 181, 183, 184, 284
- realism 214–216, 218, 225, 231
- recurrent network 82, 85
- reorganisation 43, 57, 64, 65, 67
- representational content 75, 78, 87–89
- response times 155, 157, 161, 165
- right hand 158–160, 162–165, 177
- robot soccer 95, 96, 102, 104, 106, 111, 113, 114, 116–118

robots 4, 5, 95–97, 100–102, 105,
106, 110, 111–119, 121
rotation function 128, 131,
133–137, 142, 145

S

schema 267, 269
selection 17, 35, 36, 38–40, 43, 51,
55, 66, 84, 118, 125, 201, 233
selective pressure 12, 29, 38,
65, 66
semantic primitive 129, 131, 146
semantic subset principle 280,
282, 283, 285
semantics 123, 124, 129, 130, 131,
155, 181, 192, 207, 217, 226, 241,
282, 286, 287, 291, 293
sensation 7, 96, 97–101, 113, 114,
213, 228, 268
sequencing 6, 173, 184
shape 22, 28, 48, 49, 51, 54, 57,
60, 90, 101, 106, 110, 153, 166,
192, 195, 206–210, 248, 250,
253–255, 257–259, 261–263,
265, 267
signed languages 8, 247–274
similarity 5, 8, 66, 75, 89,
102–104, 117–119, 161, 165, 166,
179, 184, 192, 206, 211, 258
similes 194, 195
simulations 5, 75–77, 80, 81,
84–87, 89, 95–121, 126, 226
simulator 99, 100, 107, 109, 167
skeleton 22, 37, 123, 126–128,
131–135, 140, 147, 150, 151
Soa Basin 53, 54
social provisioning 62
sound 7, 20, 103, 110, 174, 184,
283

source 6, 7, 23, 26, 56, 80, 102,
147, 155, 191–195, 197, 199–202,
207, 209–211, 220, 225, 238,
255, 270
specific activation of the motor
system 156, 165
speech 47, 172, 174–176, 178,
183–185, 247–250, 252, 258,
265, 271
stature 36, 37
steamroller problem 204, 206
stem 174, 184
stone flaking 2, 11, 29
stone tools 2, 12, 21, 38–40,
47–51, 53, 55, 56, 59–61
structural description 192, 193,
197, 200
subset/superset relation 282, 285
substitution 129, 133, 142, 143,
176, 183, 185, 186, 197, 199,
201, 202
subsumption relation 203, 204,
205, 209
Sulawesi 39, 56
supertranslatability 233–246
supertranslation 8, 241, 243, 245
supratentorial structures 174
synonym 175, 185
syntax 149, 173, 181, 184, 282

T

target 6, 7, 49, 51, 89, 191–195,
197, 199–202, 207, 209–211
technological evolution 11
temporal lobe 46, 62, 65, 66,
69, 174
thalamus 172
time 103, 104
timing 6, 173, 176, 184, 186

Timor 39
Truth Value Judgement
Task 281, 283, 289, 290
tsunami 39, 56

V

vagueness 8, 233–246
valuation 128, 134–136, 138,
144–146
ventral system 157, 166
verbal fluency 175
verbal memory 177, 180
verbs of action 124, 125, 138, 139
vermis 173–175
vision-to-action route 155, 157
vision-to-semantics-to-action
ventral route 155
visual awareness 217, 229
visual experience 222, 223, 225,
229, 230
visual processing 64, 101
visual-spatial category 255, 262,
264, 268
visual-spatial features 254,
259–261, 264, 269
visual-spatial form 255, 256,
258–261, 263, 264, 269
voice 174, 180, 185
Voronoi tessellations 105

W

watercraft 39
word 5, 8, 20, 21, 105, 124, 125,
153, 155–163, 165, 166, 171, 173,
176, 179–181, 183, 185, 186, 239,
240, 249, 251, 252, 276, 280
word generation 171, 175
word order 176
working memory 84, 174, 185, 186

Table of contents of volume 2

Preface	VII
List of contributors	IX
1. Mental categories in natural languages <i>Drew Khlentzos and Andrea C. Schalley</i>	1
2. A culture-neutral metalanguage for mental state concepts <i>Cliff Goddard</i>	11
3. Shape and colour in language and thought <i>Anna Wierzbicka</i>	37
4. Universal and language-specific aspects of “propositional attitudes”: Russian vs. English <i>Anna Gladkova</i>	61
5. Mental states reflected in cognitive lexemes related to memory: A case in Korean <i>Kyung-Joo Yoon</i>	85
6. Taste as a gateway to Chinese cognition <i>Zhengdao Ye</i>	109
7. “Then I’ll huff and I’ll puff or I’ll go on the roff!” thinks the wolf: Spontaneous written narratives by a child with autism <i>Lesley Stirling and Graham Barrington</i>	133
8. Interaction between language and cognition in language development <i>Heather Winskel</i>	173
9. What figurative language development reveals about the mind <i>Herbert L. Colston</i>	191

10. Would you rather ‘embert a cudsert’ or ‘cudsert an embert’? How spelling patterns at the beginning of English disyllables can cue grammatical category <i>Joanne Arciuli and Linda Cupples</i>	213
11. Ethnobiological classification and the environment in Northern Australia <i>Brett Baker</i>	239
12. Events masquerading as entities: Pseudorelative perception verb complements in Mawng (Australian) and Romance languages <i>Ruth Singer</i>	267
13. Word and construction as units of categorization: The case of change predicates in Estonian <i>Renate Pajusalu and Ilona Trägel</i>	289
14. Categories and concepts in phonology: Theory and practice <i>Helen Fraser</i>	311
15. You can run, but: Another look at linguistic relativity <i>Roger Wales</i>	331
Name index	351
Language index	355
Subject index	357
Table of contents of Volume 1	361

Studies in Language Companion Series

A complete list of titles in this series can be found on the publishers' website, www.benjamins.com

- 97 **DOLLINGER, Stefan:** New-Dialect Formation in Canada. Evidence from the English modal auxiliaries. xxii, 355 pp. *Expected January 2008*
- 96 **ROMEO, Nicoletta:** Aspect in Burmese. Meaning and function. xv, 289 pp. *Expected January 2008*
- 95 **O'CONNOR, Loretta:** Motion, Transfer and Transformation. The grammar of change in Lowland Chontal. 2007. xiv, 251 pp.
- 94 **MIESTAMO, Matti, Kaius SINNEMÄKI and Fred KARLSSON (eds.):** Language Complexity. Typology, contact, change. xiv, 340 pp. + index. *Expected January 2008*
- 93 **SCHALLEY, Andrea C. and Drew KHELENTZOS (eds.):** Mental States. Volume 2: Language and cognitive structure. x, 362 pp.
- 92 **SCHALLEY, Andrea C. and Drew KHELENTZOS (eds.):** Mental States. Volume 1: Evolution, function, nature. 2007. xi, 304 pp.
- 91 **FILIPOVIĆ, Luna:** Talking about Motion. A crosslinguistic investigation of lexicalization patterns. 2007. x, 182 pp.
- 90 **MUYSKEN, Pieter (ed.):** From Linguistic Areas to Areal Linguistics. vii, 274 pp. + index. *Expected February 2008*
- 89 **STARK, Elisabeth, Elisabeth LEISS and Werner ABRAHAM (eds.):** Nominal Determination. Typology, context constraints, and historical emergence. 2007. viii, 370 pp.
- 88 **RAMAT, Paolo and Elisa ROMA (eds.):** Europe and the Mediterranean as Linguistic Areas. Convergencies from a historical and typological perspective. 2007. xxvi, 364 pp.
- 87 **VERHOEVEN, Elisabeth:** Experiential Constructions in Yucatec Maya. A typologically based analysis of a functional domain in a Mayan language. 2007. xiv, 380 pp.
- 86 **SCHWARZ-FRIESEL, Monika, Manfred CONSTEN and Mareile KNEES (eds.):** Anaphors in Text. Cognitive, formal and applied approaches to anaphoric reference. 2007. xvi, 282 pp.
- 85 **BUTLER, Christopher S., Raquel HIDALGO DOWNING and Julia LAVID (eds.):** Functional Perspectives on Grammar and Discourse. In honour of Angela Downing. 2007. xxx, 481 pp.
- 84 **WANNER, Leo (ed.):** Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In honour of Igor Mel'čuk. 2007. xviii, 380 pp.
- 83 **HANNAY, Mike and Gerard J. STEEN (eds.):** Structural-Functional Studies in English Grammar. In honour of Lachlan Mackenzie. 2007. vi, 393 pp.
- 82 **ZIEGELER, Debra:** Interfaces with English Aspect. Diachronic and empirical studies. 2006. xvi, 325 pp.
- 81 **PEETERS, Bert (ed.):** Semantic Primes and Universal Grammar. Empirical evidence from the Romance languages. 2006. xvi, 374 pp.
- 80 **BIRNER, Betty J. and Gregory WARD (eds.):** Drawing the Boundaries of Meaning. Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn. 2006. xii, 350 pp.
- 79 **LAFFUT, An:** Three-Participant Constructions in English. A functional-cognitive approach to caused relations. 2006. ix, 268 pp.
- 78 **YAMAMOTO, Mutsumi:** Agency and Impersonality. Their Linguistic and Cultural Manifestations. 2006. x, 152 pp.
- 77 **KULIKOV, Leonid, Andrej MALCHUKOV and Peter de SWART (eds.):** Case, Valency and Transitivity. 2006. xx, 503 pp.
- 76 **NEVALAINEN, Terttu, Juhani KLEMOLA and Mikko LAITINEN (eds.):** Types of Variation. Diachronic, dialectal and typological interfaces. 2006. viii, 378 pp.
- 75 **HOLE, Daniel, André MEINUNGER and Werner ABRAHAM (eds.):** Datives and Other Cases. Between argument structure and event structure. 2006. viii, 385 pp.
- 74 **PIETRANDREA, Paola:** Epistemic Modality. Functional properties and the Italian system. 2005. xii, 232 pp.
- 73 **XIAO, Richard and Tony McENERY:** Aspect in Mandarin Chinese. A corpus-based study. 2004. x, 305 pp.
- 72 **FRAJZYNGIER, Zygmunt, Adam HODGES and David S. ROOD (eds.):** Linguistic Diversity and Language Theories. 2005. xii, 432 pp.
- 71 **DAHL, Östen:** The Growth and Maintenance of Linguistic Complexity. 2004. x, 336 pp.
- 70 **LEFEBVRE, Claire:** Issues in the Study of Pidgin and Creole Languages. 2004. xvi, 358 pp.

- 69 **TANAKA, Lidia:** *Gender, Language and Culture. A study of Japanese television interview discourse.* 2004. xvii, 233 pp.
- 68 **MODER, Carol Lynn and Aida MARTINOVIC-ZIC (eds.):** *Discourse Across Languages and Cultures.* 2004. vi, 366 pp.
- 67 **LURAGHI, Silvia:** *On the Meaning of Prepositions and Cases. The expression of semantic roles in Ancient Greek.* 2003. xii, 366 pp.
- 66 **NARIYAMA, Shigeko:** *Ellipsis and Reference Tracking in Japanese.* 2003. xvi, 400 pp.
- 65 **MATSUMOTO, Kazuko:** *Intonation Units in Japanese Conversation. Syntactic, informational and functional structures.* 2003. xviii, 215 pp.
- 64 **BUTLER, Christopher S.:** *Structure and Function – A Guide to Three Major Structural-Functional Theories. Part 2: From clause to discourse and beyond.* 2003. xiv, 579 pp.
- 63 **BUTLER, Christopher S.:** *Structure and Function – A Guide to Three Major Structural-Functional Theories. Part 1: Approaches to the simplex clause.* 2003. xx, 573 pp.
- 62 **FIELD, Fredric:** *Linguistic Borrowing in Bilingual Contexts. With a foreword by Bernard Comrie.* 2002. xviii, 255 pp.
- 61 **GODDARD, Cliff and Anna WIERZBICKA (eds.):** *Meaning and Universal Grammar. Theory and empirical findings. Volume 2.* 2002. xvi, 337 pp.
- 60 **GODDARD, Cliff and Anna WIERZBICKA (eds.):** *Meaning and Universal Grammar. Theory and empirical findings. Volume 1.* 2002. xvi, 337 pp.
- 59 **SHI, Yuzhi:** *The Establishment of Modern Chinese Grammar. The formation of the resultative construction and its effects.* 2002. xiv, 262 pp.
- 58 **MAYLOR, B. Roger:** *Lexical Template Morphology. Change of state and the verbal prefixes in German.* 2002. x, 273 pp.
- 57 **MEL'ČUK, Igor A.:** *Communicative Organization in Natural Language. The semantic-communicative structure of sentences.* 2001. xii, 393 pp.
- 56 **FAARLUND, Jan Terje (ed.):** *Grammatical Relations in Change.* 2001. viii, 326 pp.
- 55 **DAHL, Östen and Maria KOPTJEVSKAJA-TAMM (eds.):** *Circum-Baltic Languages. Volume 2: Grammar and Typology.* 2001. xx, 423 pp.
- 54 **DAHL, Östen and Maria KOPTJEVSKAJA-TAMM (eds.):** *Circum-Baltic Languages. Volume 1: Past and Present.* 2001. xx, 382 pp.
- 53 **FISCHER, Olga, Anette ROSENBAACH and Dieter STEIN (eds.):** *Pathways of Change. Grammaticalization in English.* 2000. x, 391 pp.
- 52 **TORRES CACOULOS, Rena:** *Grammaticization, Synchronic Variation, and Language Contact. A study of Spanish progressive -ndo constructions.* 2000. xvi, 255 pp.
- 51 **ZIEGELER, Debra:** *Hypothetical Modality. Grammaticalisation in an L2 dialect.* 2000. xx, 290 pp.
- 50 **ABRAHAM, Werner and Leonid KULIKOV (eds.):** *Tense-Aspect, Transitivity and Causativity. Essays in honour of Vladimir Nedjalkov.* 1999. xxxiv, 359 pp.
- 49 **BHAT, D.N.S.:** *The Prominence of Tense, Aspect and Mood.* 1999. xii, 198 pp.
- 48 **MANNEY, Linda Joyce:** *Middle Voice in Modern Greek. Meaning and function of an inflectional category.* 2000. xiii, 262 pp.
- 47 **BRINTON, Laurel J. and Minoji AKIMOTO (eds.):** *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* 1999. xiv, 283 pp.
- 46 **YAMAMOTO, Mutsumi:** *Animacy and Reference. A cognitive approach to corpus linguistics.* 1999. xviii, 278 pp.
- 45 **COLLINS, Peter C. and David LEE (eds.):** *The Clause in English. In honour of Rodney Huddleston.* 1999. xv, 342 pp.
- 44 **HANNAY, Mike and A. Machtelt BOLKESTEIN (eds.):** *Functional Grammar and Verbal Interaction.* 1998. xii, 304 pp.
- 43 **OLBERTZ, Hella, Kees HENGVELD and Jesús SÁNCHEZ GARCÍA (eds.):** *The Structure of the Lexicon in Functional Grammar.* 1998. xii, 312 pp.
- 42 **DARNELL, Michael, Edith A. MORAVCSIK, Michael NOONAN, Frederick J. NEWMAYER and Kathleen M. WHEATLEY (eds.):** *Functionalism and Formalism in Linguistics. Volume II: Case studies.* 1999. vi, 407 pp.

- 41 **DARNELL, Michael, Edith A. MORAVCSIK, Michael NOONAN, Frederick J. NEWMAYER and Kathleen M. WHEATLEY (eds.):** Functionalism and Formalism in Linguistics. Volume I: General papers. 1999. vi, 486 pp.
- 40 **BIRNER, Betty J. and Gregory WARD:** Information Status and Noncanonical Word Order in English. 1998. xiv, 314 pp.
- 39 **WANNER, Leo (ed.):** Recent Trends in Meaning-Text Theory. 1997. xx, 202 pp.
- 38 **HACKING, Jane F:** Coding the Hypothetical. A comparative typology of Russian and Macedonian conditionals. 1998. vi, 156 pp.
- 37 **HARVEY, Mark and Nicholas REID (eds.):** Nominal Classification in Aboriginal Australia. 1997. x, 296 pp.
- 36 **KAMIO, Akio (ed.):** Directions in Functional Linguistics. 1997. xiii, 259 pp.
- 35 **MATSUMOTO, Yoshiko:** Noun-Modifying Constructions in Japanese. A frame semantic approach. 1997. viii, 204 pp.
- 34 **HATAV, Galia:** The Semantics of Aspect and Modality. Evidence from English and Biblical Hebrew. 1997. x, 224 pp.
- 33 **VELÁZQUEZ-CASTILLO, Maura:** The Grammar of Possession. Inalienability, incorporation and possessor ascension in Guaraní. 1996. xvi, 274 pp.
- 32 **FRAJZYNGIER, Zygmunt:** Grammaticalization of the Complex Sentence. A case study in Chadic. 1996. xviii, 501 pp.
- 31 **WANNER, Leo (ed.):** Lexical Functions in Lexicography and Natural Language Processing. 1996. xx, 355 pp.
- 30 **HUFFMAN, Alan:** The Categories of Grammar. French *lui* and *le*. 1997. xiv, 379 pp.
- 29 **ENGBERG-PEDERSEN, Elisabeth, Michael FORTESCUE, Peter HARDER, Lars HELTOFT and Lisbeth Falster JAKOBSEN (eds.):** Content, Expression and Structure. Studies in Danish functional grammar. 1996. xvi, 510 pp.
- 28 **HERMAN, József (ed.):** Linguistic Studies on Latin. Selected papers from the 6th International Colloquium on Latin Linguistics (Budapest, 23–27 March 1991). 1994. ix, 421 pp.
- 27 **ABRAHAM, Werner, T. GIVÓN and Sandra A. THOMPSON (eds.):** Discourse, Grammar and Typology. Papers in honor of John W.M. Verhaar. 1995. xx, 352 pp.
- 26 **LIMA, Susan D., Roberta L. CORRIGAN and Gregory K. IVERSON:** The Reality of Linguistic Rules. 1994. xxiii, 480 pp.
- 25 **GODDARD, Cliff and Anna WIERZBICKA (eds.):** Semantic and Lexical Universals. Theory and empirical findings. 1994. viii, 510 pp.
- 24 **BHAT, D.N.S.:** The Adjectival Category. Criteria for differentiation and identification. 1994. xii, 295 pp.
- 23 **COMRIE, Bernard and Maria POLINSKY (eds.):** Causatives and Transitivity. 1993. x, 399 pp.
- 22 **McGREGOR, William B.:** A Functional Grammar of Gooniyandi. 1990. xx, 618 pp.
- 21 **COLEMAN, Robert (ed.):** New Studies in Latin Linguistics. Proceedings of the 4th International Colloquium on Latin Linguistics, Cambridge, April 1987. 1990. x, 480 pp.
- 20 **VERHAAR, John W.M. S.J. (ed.):** Melanesian Pidgin and Tok Pisin. Proceedings of the First International Conference on Pidgins and Creoles in Melanesia. 1990. xiv, 409 pp.
- 19 **BLUST, Robert A.:** Austronesian Root Theory. An essay on the limits of morphology. 1988. xi, 190 pp.
- 18 **WIERZBICKA, Anna:** The Semantics of Grammar. 1988. vii, 581 pp.
- 17 **CALBOLI, Gualtiero (ed.):** Subordination and Other Topics in Latin. Proceedings of the Third Colloquium on Latin Linguistics, Bologna, 1–5 April 1985. 1989. xxix, 691 pp.
- 16 **CONTE, Maria-Elisabeth, János Sándor PETŐFI and Emel SÖZER (eds.):** Text and Discourse Connectedness. Proceedings of the Conference on Connexity and Coherence, Urbino, July 16–21, 1984. 1989. xxiv, 584 pp.
- 15 **JUSTICE, David:** The Semantics of Form in Arabic. In the mirror of European languages. 1987. iv, 417 pp.
- 14 **BENSON, Morton, Evelyn BENSON and Robert F. ILSON:** Lexicographic Description of English. 1986. xiii, 275 pp.
- 13 **REESINK, Ger P.:** Structures and their Functions in Usan. 1987. xviii, 369 pp.
- 12 **PINKSTER, Harm (ed.):** Latin Linguistics and Linguistic Theory. Proceedings of the 1st International Colloquium on Latin Linguistics, Amsterdam, April 1981. 1983. xviii, 307 pp.
- 11 **PANHUIS, Dirk G.J.:** The Communicative Perspective in the Sentence. A study of Latin word order. 1982. viii, 172 pp.

- 10 **DRESSLER, Wolfgang U., Willi MAYERTHALER, Oswald PANAGL and Wolfgang Ullrich WURZEL:** Leitmotifs in Natural Morphology. 1988. ix, 168 pp.
- 9 **LANG, Ewald and John PHEBY:** The Semantics of Coordination. (English transl. by John Pheby from the German orig. ed. 'Semantik der koordinativen Verknüpfung', Berlin, 1977). 1984. 300 pp.
- 8 **BARTH, E.M. and J.L. MARTENS (eds.):** Argumentation: Approaches to Theory Formation. Containing the Contributions to the Groningen Conference on the Theory of Argumentation, October 1978. 1982. xviii, 333 pp.
- 7 **PARRET, Herman, Marina SBISÀ and Jef VERSCHUEREN (eds.):** Possibilities and Limitations of Pragmatics. Proceedings of the Conference on Pragmatics, Urbino, July 8–14, 1979. 1981. x, 854 pp.
- 6 **VAGO, Robert M. (ed.):** Issues in Vowel Harmony. Proceedings of the CUNY Linguistics Conference on Vowel Harmony, May 14, 1977. 1980. xx, 340 pp.
- 5 **HAIMAN, John:** Hua: A Papuan Language of the Eastern Highlands of New Guinea. 1980. iv, 550 pp.
- 4 **LLOYD, Albert L.:** Anatomy of the Verb. The Gothic Verb as a Model for a Unified Theory of Aspect, Actional Types, and Verbal Velocity. (Part I: Theory; Part II: Application). 1979. x, 351 pp.
- 3 **MALKIEL, Yakov:** From Particular to General Linguistics. Selected Essays 1965–1978. With an introduction by the author, an index rerum and an index nominum. 1983. xxii, 659 pp.
- 2 **ANWAR, Mohamed Sami:** BE and Equational Sentences in Egyptian Colloquial Arabic. 1979. vi, 128 pp.
- 1 **ABRAHAM, Werner (ed.):** Valence, Semantic Case, and Grammatical Relations. Workshop studies prepared for the 12th International Congress of Linguists, Vienna, August 29th to September 3rd, 1977. xiv, 729 pp. *Expected Out of print*