

# ANALYTICAL TABLE OF CONTENTS

## VOLUME I

<i>Preface</i>	xxxiii
i. The Book	xxxiii
ii. The Background	xxxvii
<b>1. SETTING THE SCENE</b>	<b>1</b>
1.i. Mind and its Place in Nature	1
a. Questions, questions . . .	2
b. How to find some answers	3
c. Never mind minds?	8
1.ii. The Scope of Cognitive Science	9
a. Of labels and cans	10
b. Two footpaths, many meadows	12
c. Why computers?	14
d. What's in, what's out	16
1.iii. <i>Caveat Narrator</i>	18
a. Beware of Whig history	19
b. Losing the Legend	21
c. The counter-cultural background	26
d. The counter-cultural somersault	31
e. Hardly hero worship	37
f. Discovering discoveries	39
g. So what's new?	42
h. Rhetoric and publication	46
i. An explanatory can of worms	49
1.iv. Envoi	50
<b>2. MAN AS MACHINE: ORIGINS OF THE IDEA</b>	<b>51</b>
2.i. Machine as Man: Early Days	52
a. Ancient automata and Dark Age decline	53
b. In fashion again	54

2.ii. Descartes's Mechanism	58
a. From physics to physiology	59
b. Science as cooperation	61
c. Cartesian cooperation develops	64
d. Descartes on animals—	68
e. —but just what did he mean?	69
f. Vivisection revived	71
g. Human bodies as machines	73
2.iii. Cartesian Complications	74
a. The mind is different	74
b. Birth of a bugbear	76
c. The prospects for AI	80
2.iv. Vaucanson's Scientific Automata	81
a. Fairs and flute-players	82
b. Theories in robotic form	84
c. Robotics, not AI	86
2.v. Mechanism and Vitalism	87
a. Animal experiments: Are they needed?	87
b. Holist chemistry	89
2.vi. The Neo-Kantian Alternative	90
a. Kant on mind and world	91
b. Biology, mechanism, teleology	93
c. Philosophies of self-realization	95
d. Goethe, psychology, and neurophysiology	96
e. The birth of morphology	99
f. Goethe's eclipse	101
2.vii. The Self-Regulation of the Body	102
a. Automatic equilibria	102
b. The embarrassing embryo	104
c. Creative evolution	105
2.viii. The Neurophysiological Machine	107
a. Getting on one's nerves	107
b. Reflections on the reflex	108
c. From nerves to neurones	110
d. Integration in the nervous system	114
e. How do neurones work?	115
f. Brains and machines	117

2.ix. Strictly Logical Automata	119
a. Early gizmos	119
b. Logic, not psychology	122
2.x. Psychology as Mechanism—But Not as Machine	123
a. Visions of a scientific psychology	123
b. Non-empiricist psychologies	128
<b>3. ANTICIPATORY ENGINES</b>	131
3.i. Miracles and Mechanism	132
a. Babbage in the round	132
b. Religion and science	135
3.ii. Differences that Made a Difference	138
a. Division of labour, again	138
b. Design and disappointment	140
3.iii. Analytical Engines	142
a. From arithmetic to algebra	142
b. Programs . . . and bugs	144
3.iv. Had Wheelwork Been Taught to Think?	146
a. For Lovelace read Babbage throughout	146
b. What Lovelace said	149
c. Babbage and AI	151
3.v. Electronic Babbage	152
a. A soulmate in Berlin	152
b. Call me MADM	155
c. Intimations of AI	157
d. Turing's invisibility	158
e. Von Neumann's contribution	160
3.vi. In Grandfather's Footsteps?	162
a. Conflicting evidence	163
b. So what's the verdict?	167
<b>4. MAYBE MINDS ARE MACHINES TOO</b>	168
4.i. The Turing Machine	169
a. Turing the man	169

b.	Playing the game	171
c.	What computation is	173
d.	Only programs, not computers	176
4.ii.	From Maths Towards Mind	177
a.	Computers and computers	177
b.	Commitment to the claim	179
c.	But what about the details?	181
4.iii.	The Logical Neurone	182
a.	McCulloch the Polymath	182
b.	Experimental epistemology	184
c.	Enthused by logic	186
d.	The young collaborator	189
e.	Mind as logic machine	190
f.	Initial reception	193
4.iv.	The Functionalist Neurone	195
a.	From calculus to computer	195
b.	Function, not implementation	197
4.v.	Cybernetic Circularity: From Steam-Engines to Societies	198
a.	Feedback, way back	198
b.	Infant interdisciplinarity	200
c.	Biological roots	202
d.	Information theory	204
e.	Bateson, Pask, and a sip of Beer	205
4.vi.	Brains as Modelling Machines	210
a.	A Cambridge cyclist	211
b.	Similarity isn't enough	214
c.	Craik and cognitive science	215
d.	Might-have-beens	217
4.vii.	Feedback Machines	218
a.	Purposes of war	218
b.	Post-war projects	220
4.viii.	Of Tortoises and Homeostats	222
a.	Robots at the festival	223
b.	Of wheels and whiskers	225
c.	Less sexy, more surprising	228
d.	How the Homeostat worked	230

4.ix. Schism	232
a. All too human	233
b. Adaptation or meaning?	235
<b>5. MOVEMENTS BENEATH THE MANTLE</b>	<b>237</b>
5.i. Newtonianism	238
a. The six assumptions	238
b. What sort of revolution was it?	240
5.ii. Psychology's House	241
a. Sitting tenants with personality	242
b. Sitting tenants with knowledge	247
c. Sitting tenants with biology	252
5.iii. Soft Centres	257
a. Mentalism goes underground	257
b. Behaviourism softens	260
c. Behaviourist machines	262
5.iv. Neurology Creeps In	264
a. Hierarchies in the brain	265
b. Connectionism named	268
c. The cell assembly	271
d. Beyond perceptual learning	274
e. Hebb's originality?	276
f. Loosening the mantle	278
<b>6. COGNITIVE SCIENCE COMES TOGETHER</b>	<b>282</b>
6.i. Pointers to the Promised Land	283
a. Informed by information	283
b. Miller and magic	286
c. Going with the flow	289
d. Information and computation	293
e. Chomsky comes on the scene	296
6.ii. The New Look	298
a. Coins and cards	299
b. A study of thinking	304
c. Computational couture	307
d. Costume change	311
e. Will seeing machines have illusions?	313

6.iii. From Heuristics to Computers	317
a. The economics of thought	318
b. A meeting of minds	320
c. A new dawn	323
6.iv. The Early Church	328
a. Consciousness raising	328
b. A trio of meetings	330
c. The manifesto	336
d. The first mission station	343
e. Missionary outposts	348
f. The sine qua non	351
6.v. Spreading the Word	354
a. Training sessions	354
b. Library tickets	356
c. Journal-ism	363
<b>7. THE RISE OF COMPUTATIONAL PSYCHOLOGY</b>	<b>366</b>
7.i. The Personal Touch	368
a. The return of the repressed	369
b. Argus with 100 eyes	373
c. From scripts to scriptis	376
d. Emotional intelligence	381
e. Architect-in-waiting	385
f. Of nursemaids and grief	388
g. Free to be free	394
h. Some hypnotic suggestions	397
i. An alien appendage	402
7.ii. The Spoken Word	404
a. Psychosyntax	405
b. Up the garden path	406
c. You know, uh, well . . .	409
d. Meaning matters	412
7.iii. Explanation as the Holy Grail	416
a. Competence and performance	417
b. Three levels, two types	419
c. The sweet smell of success	421
d. Chasing a will-o'-the-wisp?	422

7.iv. Reasoning and Rationality	427
a. Simon's ant	429
b. Productions and SOAR	430
c. The ACTs of Anderson	435
d. Models in the mind	439
e. The marriage of Craik and Montague	442
f. Irrationality rules—or does it?	444
g. Evolved for success	446
h. Give thanks for boundedness	449
7.v. Visions of Vision	451
a. Icons of the eyes	451
b. Vision from the bottom up	456
c. Maths and multimodels	459
d. The fashion for Mexican hats	462
e. Direct opposition	465
f. Let battle commence!	469
7.vi. Nativism and its Vicissitudes	472
a. The words of Adam and Eve	473
b. Some surprises from ethology	475
c. From Noam to Nim	477
d. Modish modules	481
e. But how many, exactly?	484
f. Theory of Mind	486
g. The third way	492
h. What makes higher thinking possible?	496
i. The modularization of modules	499
7.vii. Satellite Images	503
a. A telescopic vision	504
b. Forking footpaths	507
c. The Newell Test	509
d. Low focus	512
e. The bustling circus	513
<b>8. THE MYSTERY OF THE MISSING DISCIPLINE</b>	<b>515</b>
8.i. Anthropology and Cognitive Science	516
a. The beginnings of cognitive anthropology	517
b. Peoples and prototypes	519
c. Hopes and a hexagon	522
d. More taxonomies (and more Darkness than light)	523
e. And modelling, too	526

8.ii. Why Invisibility?	530
a. Psychology sidelined	531
b. Skirmishes in the science wars	534
c. Top dogs and underdogs	537
d. What's in a name?	539
e. Barkow's baby	540
8.iii. Minds and Group Minds	543
a. Models of seamanship	544
b. Networks of navigation	547
8.iv. Mechanisms of Aesthetics	549
a. From Savanna to Sotheby's	549
b. The seductiveness of symmetry	552
c. Universality in variety	553
8.v. Cultural Evolution	556
a. Evolution in the third world	556
b. A new mantra: BVSr	558
c. The meme of memes	562
d. Cloak uncloaked	565
8.vi. The Believable and the Bizarre	568
a. An epidemiology of belief	569
b. Religion as a cultural universal	573
c. Symbolism	577
d. The extraordinary out of the ordinary	579
e. Anything goes?	583
f. The impurity of induction	587
<b>9. TRANSFORMING LINGUISTICS</b>	<b>590</b>
9.i. Chomsky as Guru	592
a. The tenfold Chomsky myth	592
b. A non-pacific ocean	593
9.ii. Predecessors and Precursors	594
a. Why Chomsky's 'history' matters	594
b. The rationalist background	596
c. The puzzle of innate ideas	597
9.iii. Not-Really-Cartesian Linguists	600
a. Descartes's disciple	600
b. Arnauld and the abbey	602

c. The Port-Royal <i>Grammar</i>	602
d. Deaf-mutes and Diderot	605
9.iv. Humboldt's Humanism	606
a. Language as humanity	607
b. Languages and cultures	609
c. Humboldt lives!	610
d. A fivefold list	611
e. Origins	612
f. Creativity of language	614
g. The inner form	616
9.v. The <i>Status Quo Ante</i>	618
a. Two anti-rationalist 'isms'	618
b. The shock of structuralism	620
c. The formalist Dane	622
d. Tutor to Chomsky	624
e. Not quite there yet . . .	626
9.vi. Major Transformations	627
a. Chomsky's first words	627
b. The need for a generative grammar	628
c. Beyond information theory	631
d. Transformational grammars	634
e. So what?	637
9.vii. A Battle with Behaviourism	638
a. Political agenda	639
b. That review!	641
c. Nativist notions	643
d. Universal grammar?	645
9.viii. Aftermath	647
a. Polarized passions	648
b. Revisions, revisions . . .	650
c. Semantics enters the equation	652
9.ix. Challenging the Master	654
a. Linguistic wars	655
b. Who needs transformations?	656
c. Montagovian meanings	657
d. Transformations trounced	660
e. Why GPSG matters	662
f. Computational tractability	665
g. Linguistics eclipsed	666

9.x.	The Genesis of Natural Language Processing	669
	a. Ploughman crooked ground plough plough	669
	b. Shannon's shadow	671
	c. Love letters and haikus	674
	d. Wittgenstein and CLRU	674
	e. Is perfect translation possible?	677
	f. Is adequate translation achievable?	678
9.xi.	NLP Comes of Age	680
	a. MT resurrected	681
	b. Automatic parsing	683
	c. 'What I did on my holiday'	688
	d. Semantic coherence	689
	e. The seductiveness of semantic networks	692
	f. Whatever will they say next?	695
	g. A snippet on speech	698

## VOLUME II

<b>10.</b>	<b>WHEN GOFAI WAS NEWFAI</b>	701
10.i.	Harbingers	702
	a. When is a program not a program?	702
	b. The first AI program—not!	705
	c. How a program became a program	708
	d. First-footings	710
	e. The book of Samuel	713
	f. Programmatics	715
	g. First 'Steps'	719
	h. The harbinger in the Bush	725
	i. Spacewar	729
	j. The empty chair at the banquet	730
10.ii.	Establishment	731
	a. First labs	731
	b. The ripples spread	736
	c. New waves	739
10.iii.	The Search for Generality	739
	a. SIP spawns KR	741
	b. A resolution to do better	749
	c. Planning progresses	752

d. Early learning	759
e. 'Some Philosophical Problems'	769
10.iv. The Need for Knowledge	775
a. A triumph, and a threefold challenge	776
b. Clearer vision	781
c. Expert Systems	794
10.v. Talking to the Computer	799
a. Psychology outlaws binary	799
b. Entering the lists	801
c. LISPing in 'English'	805
d. Virtual cascades	808
e. NewFAI in parallel	811
f. It's only logical!	814
10.vi. Child's Play	817
a. The power of bugs	817
b. Complication and distribution	820
c. Pointers to the future	821
<b>11. OF BOMBS AND BOMBSHELLS</b>	<b>822</b>
11.i. Military Matters	823
a. Nurtured in war	825
b. Licklider as a military man	828
c. Star Wars and AI qualms	832
d. <i>Les mains sales?</i>	835
11.ii. Critics and Calumnies	838
a. The outsider	838
b. Scandal	841
c. After Alchemy	846
d. Dreyfus and connectionism	848
e. The neighbour	850
f. A sign of the times	852
g. The unkindest cut of all	855
11.iii. A Plea for Intellectual Hygiene	857
a. The insider	858
b. Natural Stupidity survives	861
11.iv. Lighthill's Report	864
a. A badly guided missile	865
b. Clearing up the rubble	869

11.v. The Fifth Generation	873
a. A warning shot from Japan	873
b. Self-defence in the USA	875
c. Lighthill laid to rest	879
11.vi. The Kraken Wakes	881
a. Small fry and sleeping draughts	881
b. Competition	881
<b>12. CONNECTIONISM: ITS BIRTH AND RENAISSANCE</b>	<b>883</b>
12.i. Lighting the Fuse	885
a. A long gestation	885
b. Turing and connectionism	886
c. 'How We Know Universals'	887
d. From logic to thermodynamics	890
12.ii. Infant Implementations	892
a. B24 bricolage	893
b. Self-organizing networks	894
c. Connections with the Ratio Club	897
d. Pandemonium	898
e. The perceptron	903
f. Excitement, and overexcitement	907
g. Enter the Adaline	909
12.iii. Attack Without Apology	911
a. The devilish duo	911
b. The opening salvo	912
c. Intransigence	916
d. The hybrid society of mind	917
e. Were they to blame?	921
12.iv. Lamps Invisible	923
a. Relegation to the background	924
b. Run and twiddle	926
c. Reinforcement and purpose	926
12.v. Behind the Scenes	928
a. Left alone to get on with it	928
b. A problem shared . . . ?	930
c. How large is your memory?	931
d. Disillusion on distribution	934
e. Linear associative memories	935

f.	The physicists have their say	936
g.	The power of respectability	940
h.	Hinton relaxes	942
i.	Passing frustrations	943
12.vi.	Centre-Stage	945
a.	The bible in two volumes	945
b.	Bowled over by Boltzmann	948
c.	Backprop hits the headlines	952
d.	Backprop anticipated	953
e.	Wonders of the past tense	955
f.	Escaping from the black box	957
12.vii.	The Worm Turns	959
a.	Joyful jamborees	959
b.	DARPA thinks again	962
12.viii.	<i>A la recherche . . .</i>	963
a.	Emulating the ancestors	965
b.	Recurrent nets	966
c.	Start simple, develop complex	968
d.	Pathways for representation	969
e.	The importance of input history	972
12.ix.	Still Searching	972
a.	Assemblies of cell assemblies	973
b.	Hands across the divide	975
c.	Constructive networks	979
d.	What had been achieved?	980
12.x.	Philosophers Connect	982
a.	A Pulitzer prelude	982
b.	Connectionist concepts	984
c.	The proper treatment of connectionism?	986
d.	The old ways defended	989
e.	Microcognition and representational change	991
f.	Non-conceptual content	993
g.	An eye to the future?	996
12.xi.	Pointing to the Neighbours	1000
<b>13.</b>	<b>SWIMMING ALONGSIDE THE KRAKEN</b>	1002
13.i.	Later Logicism	1003
a.	Less monotony	1003

b. More naivety	1006
c. The AI en-CYC-lopedia	1007
13.ii. Choppy Waters	1013
a. Apostasy	1013
b. Can the fox catch the rabbit?	1015
c. Matters-in-law	1020
d. Judgements about judges	1024
13.iii. Advance and Attack	1027
a. Gelernter revived	1027
b. Planning attacked—	1029
c. —and defended	1035
d. Agents and distributed cognition	1038
e. Social interaction and agents	1043
f. Technology swamps psychology	1046
13.iv. Explaining the Ineffable	1052
a. Creativity ignored	1053
b. Help from outside	1054
c. In focus at last	1059
13.v. Outreach to Everyman	1069
a. Papert and the media lab	1069
b. The H in HCI	1072
c. Good ideas in hibernation	1074
d. The human face of the interface	1076
13.vi. Virtual Reality	1081
a. Intimations of VR	1082
b. VR as a practical aid	1084
c. VR in art and play	1087
d. Computerized companions	1092
e. Psychology and avatars	1096
13.vii. Coda	1100
a. Is AI a discipline?	1100
b. Has GOFAI failed?	1105

<b>14. FROM NEUROPHYSIOLOGY TO COMPUTATIONAL NEUROSCIENCE</b>	1110
14.i. Notes on Nomenclature	1111
a. The naming of neuroscience	1112
b. The computational species	1113

14.ii.	Very Non-Neural Nets	1114
	a. Too neat	1114
	b. Too simple	1115
	c. Too few	1116
	d. Too dry	1117
14.iii.	In the Beginning	1121
	a. Computational questions	1121
	b. Computations in the brain	1125
	c. Formal synapses	1128
14.iv.	A Fistful of Feature-Detectors	1130
	a. Bug-detectors	1130
	b. And more, and more . . .	1134
	c. But how?	1136
	d. Monkey business	1138
14.v.	Modelling the Brain	1140
	a. The Mars robot	1140
	b. The musician in the spare room	1143
	c. Secrets of the cerebellum	1145
	d. Audience reaction	1149
	e. Beyond the cerebellum	1151
	f. A change of tack	1154
14.vi.	Realism Rampant	1157
	a. A voice in the wilderness	1158
	b. Adaptation—and feature-detectors	1161
	c. ARTful simulations	1164
	d. Avoiding the black box	1167
14.vii.	Whole Animals	1169
	a. CNE—what is it?	1169
	b. A wizard from Oz	1170
	c. <i>Rana computatrix</i> and its scheming cousins	1172
14.viii.	Representations Galore	1177
	a. What's the problem?	1178
	b. From probabilities to geometries	1179
	c. Emulation and subjectivity	1184
	d. The philosophers worry	1187
14.ix.	Computation Challenged	1189
	a. Structure without description	1189
	b. Dynamics in the brain	1193

c. Epigenesis	1196
d. Neural selection	1199
e. Grandmother cells	1205
f. Modelling modulation	1210
g. Time blindness—and glimmers of light	1213
14.x. Cartesian Correlations	1216
a. Consciousness comes in from the cold	1216
b. Cognitive neuroscience	1220
c. The \$64,000 question	1224
d. Philosophical contortions	1230
14.xi. Descartes to the Tumbrils?	1236
a. Describing the mind, or inventing it?	1237
b. A computational analysis	1237
c. The other side of the river	1240
d. Lions and lines	1242
e. Hung jury	1244
<b>15. A-LIFE IN EMBRYO</b>	1247
15.i. Life, Mind, Self-Organization	1249
a. Life and mind versus life-and-mind	1249
b. Self-organization, in and out of focus	1249
15.ii. Biomimetics and Artificial Life	1251
a. Artificial fish	1251
b. What is A-Life?	1253
15.iii. Mathematical Biology Begins	1254
a. Of growth and form	1254
b. More admiration than influence	1258
c. Difficulties of description	1259
15.iv. Turing's Biological Turn	1261
a. A mathematical theory of embryology	1261
b. History's verdict	1264
15.v. Self-Replicating Automata	1268
a. Self-organization as computation	1268
b. Why the delay?	1271
15.vi. Evolution Enters the Field	1274
a. Holland, and mini-trips elsewhere	1274
b. Awaiting the computers	1278

c. The saga of SAGA	1280
d. Open-ended evolution	1284
15.vii. From Vehicles to Lampreys	1286
a. Valentino's vehicles	1287
b. Of hoverflies	1289
c. Playing cricket	1292
d. Evolving lampreys	1298
15.viii. Parallel Developments	1299
a. Artificial ants	1300
b. New philosophies of biology	1304
c. Dynamical systems	1307
15.ix. Order and Complexity	1309
a. The four classes of CA	1309
b. K for Kauffman	1310
c. Morphology revived	1313
d. Discussions in the desert	1316
15.x. Naming and Synthesis	1317
a. The party	1317
b. Simulation or realization?	1322
15.xi. After the Party	1325
a. Resurrection of the Homeostat	1325
b. Analysing dynamics	1327
<b>16. PHILOSOPHIES OF MIND AS MACHINE</b>	<b>1334</b>
16.i. Mid-Century Blues	1337
a. Interactionist squibs	1337
b. Puffs of smoke and nomological dangles	1338
c. Dispositions and category mistakes	1339
d. Questions of identity	1343
16.ii. Turing Throws Down the Gauntlet	1346
a. Sketch of a future AI	1346
b. The gauntlet spurned	1349
c. The Turing Test: Then and now	1351
16.iii. Functionalist Freedoms	1356
a. Just below the surface	1357
b. The shackles loosened	1359

16.iv.	Three Variations on a Theme	1362
	a. Content and consciousness	1363
	b. From heresy to scandal	1367
	c. Must angels learn Latin?	1369
	d. Fodorian frills	1374
	e. Eliminative materialism	1376
16.v.	Counter-moves	1379
	a. Gödel to the rescue?	1379
	b. Consciousness and zombies	1381
	c. That room in China	1382
	d. Neuroprotein and intentionality	1385
	e. How multiple is multiple?	1387
	f. Subconsciousness attacked	1388
16.vi.	Betrayal	1389
	a. Friendly fire	1390
	b. Crossing the river	1392
16.vii.	Neo-Phenomenology—From Critique to Construction	1394
	a. Where Dreyfus was coming from	1395
	b. Hands-on Heideggerians	1398
	c. Flights from the computer	1399
	d. Computation and embodiment	1404
16.viii.	Mind and “Nature”	1407
	a. No representations in the brain	1407
	b. Mind as second nature	1410
	c. Mind and VR-as-nature	1412
16.ix.	Computation as a Moving Target	1414
	a. Three senses of computation	1414
	b. Physical symbol systems	1419
	c. From computation to architecture	1420
	d. The bit in “three and a bit”	1422
	e. A philosophy of presence	1423
	f. The moral of the story	1428
16.x.	What’s Life Got To Do With It?	1429
	a. Life in the background	1430
	b. Functionalist approaches to life	1434
	c. The philosophy of autopoiesis	1438
	d. Evolution, life, and mind	1440

<b>17. WHAT NEXT?</b>	1444
17.i. What's Unpredictable?	1444
17.ii. What's Predictable?	1447
17.iii. What's Promising?	1448
17.iv. What About Those Manifesto Promises?	1451
<i>References</i>	1453
<i>List of Abbreviations</i>	1587
<i>Subject Index</i>	1593
<i>Name Index</i>	1613

# SETTING THE SCENE

*Once upon a time there was a teddy bear called Twink*—and with those few words, the scene is set. We know what we’re talking about. Twink’s story can begin . . .

This story can’t begin so quickly, however. For we *don’t* yet know what we’re talking about. Some readers may know very little about cognitive science at this stage. Even more to the point, those who are already familiar with it think of it in varying ways. That was true right from the start, and it’s even more true now. (So it’s no accident that the summary chapter of a recent book is subtitled: ‘It’s Cognitive Science—But Not As We Know It’—M. W. Wheeler 2005: 283.)

One of the founders of the field, when asked to define it, confessed that “Trying to speak for cognitive science, as if cognitive scientists had but one mind and one voice, is a bum’s game” (G. A. Miller 1978: 6). And twenty years afterwards, two long-time leaders edited a book called *What Is Cognitive Science?* (Lepore and Pylyshyn 1999). You’d think they’d know by now! But no: even in the textbooks, never mind coffee conversations and idle chat, definitions differ.

I shan’t list them: the boredom barometer would shoot through the roof. However, the differences do make a difference. This will become clearer throughout the following pages, as we see how theory and practice have changed over the years (in some cases, coming full circle). Meanwhile, before starting the story, some scene setting may be helpful.

One way of saying what we’re talking about is to give some examples of the wide-ranging questions studied by cognitive science. I’ll do that in Section i. And I’ll do it in everyday language: the technicalities can wait until later.

Another is to give a definition of the field, even if this can’t be presented as *the* universally agreed definition. I’ll do that in Section ii. This, I hope, will help to show why I’ve decided to tell the story in the way I do.

Finally, in Section iii, I’ll identify a number of traps that lie in wait for anyone discussing the field’s intellectual history.

## 1.i. Mind and its Place in Nature

A host of intriguing questions about mind and its place in nature occur to most thinking people. (The FAQs of the mind, Web-users might say.) As explained in the Preface, some have puzzled me for almost as long as I can remember—and I usually found that

my friends were puzzled by them too. They centred on the nature of mind and the mind–body problem; the evolution of mind; freedom and purpose; and how various psychopathologies are possible.

Most of the topics studied in cognitive science fall under one of these broad categories. And those which don't, such as the nature of *computation*, are closely related to them.

### a. Questions, questions . . .

We're intrigued by consciousness, for example. We know there are close correlations between brain events and conscious states—but why is that so? The answer seems to be that our brains generate our consciousness. But how do they do this, in practice? Even more puzzling, how *can* they do this, in principle?

Or maybe we only *think* we know this? Some people argue that it *doesn't even make sense* to suggest that there are correlations between conscious states and brain states. How could anyone with any common sense be led to make such a deeply counter-intuitive claim? Perhaps “common sense” itself is radically misguided here (and was radically different in other historical periods)?

What about dogs and horses: are they conscious? And snails, flies, newts . . . ? For that matter, what about newborn babies: are they conscious in *anything like* the sense in which adult humans are? What of machines? Could a machine be conscious—and if not, why not?

People often wonder whether a creature has to have a brain, or something very like one, to be intelligent. If so, why? Is a brain (as well as eyes) needed to *see*, for example? What do the visual brain cells do that the retinal cells don't? What about intelligent *action*? How, for instance, does the brain convert an Olympic diver's intention to dive into the finely modulated bodily movements that ensue? If we knew this, could we drop talk of intentions and refer only to brains instead?

Consider chimps, or cats: what can *their* brains do, and what can't they do? And what can they do without the mammalian (and avian) glory, the cerebral hemispheres? Given that *Homo sapiens* evolved from lower animals, what does this tell us about our mental powers? Can anything interesting be learnt about the human mind by studying distantly related species such as frogs, or insects?

As for machines, just how—if at all—must an artifice resemble a real brain if it's even to *seem* to support a mind? And *even if* studying insects can teach us something about ourselves, what about studying inanimate tin cans—like a Mars robot, or an automatic controller in a chemical factory? How could these things (*sic*) possibly be relevant?

What mental powers does a human brain provide, and how does it manage to do so? How is free will possible? And creativity? Are creative ideas unpredictable, and if so why? What are emotions—and do they conflict with rationality, or support it?

Are our abilities inborn, or determined by experience? And how does the brain get its detailed anatomical structure: from genetics or from the environment—or perhaps even from spontaneous self-organization? (Is that last suggestion mere hand-waving, more magic than science?)

Do we all share psychological properties that mould every human culture? Perhaps the same underlying sense of beauty: maybe in symmetry, or expanses of water? Or

the same tendency towards religious belief? If so, is that because we've evolved that way? Or are evolutionary explanations of human psychology mere Just So stories, no more plausible than the delightful tale about The Cat Who Walked By Himself (Kipling 1902)? Superficially, at least, cultures are hugely diverse . . . but can they harbour *just any* conceivable idea?

In mental illnesses of various kinds, what's gone wrong: something in the brain, or something in the mind? What's the difference?

Sometimes, people say that only living things can have a mind. Is that true? If so, why? What is life, anyway? And how did it arise in the first place? Could a living thing be created by us?

Last, but by no means least, coffee-table chat abounds with puzzles about language. For instance, people wonder what *counts* as a language: why not birdsong? Can any non-human animals learn a language? If not, is that merely because we're better at learning, or because language is a human instinct? And what, exactly, does that mean? Is language needed for thought, or can some dumb animals think?

Can two different languages ever express exactly the same thought? Or is perfect translation impossible? Could a machine converse with us in English, or French—and would it understand us, even if it did? Imagine a machine that appeared to be solving problems and using language just like us: would that prove that it was truly intelligent?

None of these questions is new. (That's largely why listing them is a scene-setting equivalent of saying "Once upon a time, there was a teddy bear . . .".)

Some date back to Aristotle. Many, including those about language-using machines, were discussed in the 1630s by René Descartes. Others were considered by Immanuel Kant, Johann von Goethe, or Wilhelm von Humboldt in the late eighteenth century. The rest surfaced in the nineteenth, or very early twentieth, century (see Chapter 2).

Originally, then, most were discussed by philosophers. Some still are (the difference between *mind* and *brain*, for example). But even those need to be considered in light of the scientific data available.

Most of our Twink-questions were later developed—and some answered—by traditional scientific research in psychology, anthropology, neurophysiology, or biology. Since the 1940s, however, *every one* has been further sharpened by work in cognitive science.

## b. How to find some answers

Cognitive science tries to answer these questions in two closely related ways. Both of them draw on machines. But the machines in question are very unlike what used to be thought of as a machine.

Forget steam-engines and telephones: these new machines can be hugely more complex even than an E-type Jag, or a jet plane. Indeed, the capacities of modern jets, from the much-lamented Concorde to stealth bombers, are largely due to their having these new machines inside them. It follows that to think of minds *as* machines, as cognitive scientists in general do, isn't so limiting—nor so absurd—as it may seem to someone who has only pre-1950 machines in mind.

Specifically, cognitive science uses abstract (logical/mathematical) concepts drawn from artificial intelligence (AI) and control theory, alias cybernetics (see Section ii.a, below).

- \* AI tries to make computers do the sorts of things that minds can do. These things range from interpreting language or camera input, through making medical diagnoses and constructing imaginary (virtual) worlds, to controlling the movements of a robot.
- \* Control theory studies the functioning of self-regulating systems. These systems include both automated chemical factories and living cells and organisms.

These concepts (of computation and control) sharpen psychological questions because they can express ideas about mental processes more clearly than verbal concepts can. Moreover, when implemented in computer models they can test the coherence and implications of those ideas more rigorously. Often, they show that a previously favoured theory has unsuspected gaps in it. Sometimes, they suggest how those gaps might be filled. They can show that a theory *might be, could be*, true—although to know whether it *is* true, we need psychological and/or neuroscientific evidence as well. Some important questions have been answered in this way which couldn't have been answered otherwise.

Consider language and machines, for instance. It's now clear that computers can (seem to) use natural language, up to a point. What's not yet clear is just where, in practice or principle, that point lies. How good can we expect future computer prose, or computer conversation, to be? And what problems will have to be overcome to get there? For that matter, what are the problems which have already been overcome, to get to where we are now? And are these problems linguistic, psychological, or philosophical—or perhaps a mixture of all three?

Thirty years ago, a medical friend told me he'd spent the afternoon visiting an immigrant family from India, whose 8-year-old son had been translating across three languages for his elders. "You'd never get a computer to do that!" he said.—Maybe, maybe not. But if not, why not? And if so, how?

Only five years after this pessimistic comment, the European Union's translation system achieved 78 per cent intelligibility for its 'raw' text, and 98 per cent for the tidied-up version. Unlike the little boy, this program could handle only two languages at a time. Ten years later, however, another one could switch between forty-two different language pairs. But the boy—by then, in his early twenties—still had the edge. He could translate remarks about anything, within reason, whereas these programs could deal only with relatively specialized topics.

What my friend didn't seem to realize was that if AI research could enable a computer to use even *one* language properly, translating it into another would be easy by comparison. Or rather, translating it helpfully, usefully, acceptably... would be relatively easy. Translating it perfectly is another matter. But then, it's not clear that a human being, whether 8 years old or 80, could produce a *perfect* translation of anything interesting. Even *Please give me six cans of baked beans* will cause problems, if one of the languages codes the participants' social status by the particular word chosen for *Please*.

Nor did he stop to ask how the 8-year-old did it—still less, how his own children had learnt their mother tongue. He simply took it for granted that language learning happens. But how? After all, vocabulary isn't the only problem: there's grammar, too.

Different languages have different grammars. Or at least, they appear to. (The order of adjective and noun varies, for instance: think of *the red house* and *la maison rouge*.) But perhaps all languages share some underlying ‘universal’ grammar? If so, what is it? And how is it related to the syntactic rules that bedevil us when we encounter a new tongue? How did the family’s young translator manage to cope with three distinct grammars, from different language groups? As for the rules of one’s mother tongue, how are these learnt, and how are they represented in the mind/brain?

All these questions, and many more, have had to be faced by cognitive scientists working in psycholinguistics and/or natural language processing (NLP). And a great deal has been learnt in the process, even if many mysteries—and some bitter controversies—remain (see Chapters 9 and 12.vi.e and x.d–e).

“Not too fast!” you may say. “Computers can handle language to some extent, and even translate it usefully too. But do they *understand* the language they use?” (Don’t let’s stop, here, to ask what it is for *human beings* to understand the language they use—but see Chapters 7.ii.d, 12.x.g, and 16.)

You may even mention the Chinese Room, an intriguing idea that’s hit the mass media worldwide (16.v.c). This example is intended to show that the answer to your question is “No”. A monoglot English speaker could spend weeks following formal rules for shuffling slips of paper bearing *squiggles* and *squoggles*, without ever realizing that they are Chinese characters which, for readers of Chinese, can be used to deliver true answers to meaningful questions. The moral is supposed to be that AI programs are intrinsically meaningless, and that for understanding you need a brain.—And, it’s often added, you need a brain for consciousness too: a robot, no matter how human-like, would be a non-conscious zombie.

An equally well-known argument claims that even if the language produced by a computer program, or a robot, were indistinguishable from that produced by a human being, that wouldn’t prove that the thing was really intelligent. “Passing the Turing Test”, as this is called, wouldn’t guarantee intelligence, understanding, or consciousness. The Test-passer might simply be a zombie.

Both these arguments have been hotly debated within cognitive science—although each is much less important *for the practice of AI* than most people imagine (see 16.ii.c). There’s still no unanimity on whether they’re well founded. Indeed, some cognitive scientists hold that *there can be no such thing* as a zombie—not because the technology is too difficult (although in fact it may be), but because the very notion is incoherent. On this view, science-fiction novels and Hollywood scenarios about zombies are, literally, non-sense (14.xi and 16.iii–v).

Whether the technology really is too difficult is disputed also. The vast majority of cognitive scientists would say that it is, at least for the foreseeable future. But one leading research team, initially with a prominent philosopher on board, is betting that it isn’t. They’re hoping to build a (literally) conscious robot, with a mind like that of a young child (see Chapter 15.vii.a).

(*An aside:* That last sentence was true when I wrote it, in the mid-1990s. Now, in 2005, the project has ground to a halt. The roboticist team leader always had other fish to fry in his research time, and is now buckling under a heavy administrative load as well; as for the research students who were working on it, in snatched moments of their spare time, they’ve left to take up jobs elsewhere. However, the leader still believes

that the project is feasible, and he might even revive it some day. Given that fact, the following paragraphs can stand, as though the plan were still being actively pursued.)

Even ignoring the issue of consciousness, they face hugely challenging problems. To build a robot that even *seems* to have the intelligence of a 5-year-old, they must provide all the relevant perceptual discriminations, motor skills, learning power, problem-solving ability, and language mastery.

For each of these, they must depend on work done by other cognitive scientists. For example, they need a computer vision system modelling the child's visual powers: so they need to know *what these are* and *how they work* (see Chapters 7.v, 14.iv and vi.b–d). They need a powerful theory of perceptuo-motor control, for generating appropriate movements of the eyes, head, and fingers (14.vii and x). They should also enable the robot to switch smoothly between stable gaits, such as crawling, standing, walking, and running (14.v and ix.b). (In fact, they've avoided those problems by giving their robot a pedestal in place of legs.) The system's capacity for learning must be built on the work that's been done in this area (see Chapters 10.iii.d, 12, 13.iii.d, and 14). As for enabling the robot to develop language, they must rely on research into some of the psycholinguistic questions outlined above (Chapters 7.ii and vi, 9, and 12.vi.e).

Strictly, they should also simulate the temper tantrums of 'the terrible twos', a stage of infancy that all parents will remember with a shudder. And, if only to preserve their own sanity, they should enable the robot to develop the greater self control—which is to say, the greater freedom (7.i.g)—of the 5-year-old. But the control of temper tantrums is even more difficult to model than stable walking or running is.

Indeed, you may think that the appropriate word here isn't "difficult", but "impossible". Certainly, many people believe that a computational psychology can't have anything to say about emotions, still less freedom.

Well, it can, and it does (see Chapter 7.i). This challenge was mounted over forty years ago, and was soon taken up by Herbert Simon, one of the high priests of computational psychology. At that time, too, a computer simulation of neurosis was developed in which different levels of 'anxiety' selected different defence mechanisms to repress the 'troubling' thought. In 1983 the authors of the Gifford Lectures on Natural Religion gave a computational analysis of personality and freedom (and religious belief) in which emotion figured prominently. Several philosophers have analysed human freedom in terms of a certain type of computational (cognitive and emotional) complexity. And a very recent program models the emotion-guided activity of a nursemaid caring for a dozen babies, each of whom has to be fed, watered, changed, cuddled, entertained, and prevented from falling into the river or crawling towards a busy road.

A nursemaid is free to choose what to do at every moment. But her choices are far from random. To the contrary, they're constrained by the goals she wants to achieve (which may conflict: she only has two hands); by the priorities she holds (feeding is necessary, lullabies aren't); by her deliberations about consequences (no-cuddles will produce an unhappy baby); by her judgements of urgency (even the hungriest baby can be temporarily ignored, if another is nearing the road); and by her emotional reactions (sometimes, she must rescue the baby *immediately*, without stopping to think). On some occasions, she 'has no choice': the danger *must* be averted, and it must be done *now*; and the baby *must* be fed, soon. But the sense in which she (sometimes) has no choice is fundamentally different from the sense in which a non-human animal, such as a cricket

for example, (always) has no choice about what to do next (see Chapter 15.vii). She's free, it isn't. Moreover, her freedom doesn't depend on randomness, or on mysterious spiritual influences: to the contrary, it's an aspect of *how her mind works*.

The nursemaid research group has even analysed the computational structure of grief. The emotion of grief is more than mere feeling. It involves irrational behaviour driven by obsessional thoughts, continual distraction, depression, anger, and guilt—all of which gradually pass, over many months, as mourning does its work. (Just what “work” is that? These cognitive scientists suggest an answer: Chapter 7.i.f.)

Grief is possible only for humans, although dogs sometimes seem to *sorrow*. A cricket simply cannot grieve. It lacks the necessary mental architecture: the concepts, knowledge, motives, values, and social commitments required to generate—or to overcome—the deeply disturbing emotion of grief. And unlike a human baby, who can't grieve either (even though it can ‘miss’ an absent carer), it has no way of developing them. Its mind, if one wants to use that term at all here, is very simple. It can't even learn to recognize objects or patterns as one of a *general class*—such as another cricket.

To be sure, crickets manage. They've survived. They even do some apparently clever things, such as locating a potential mate at a distance. However, they do this unthinkingly. They rely on a hardwired biological trick, an anatomical detail evolved for this function alone. Similarly, a frog locates its food by relying on perceptuo-motor reflexes linking cells in its retina and brain to muscles that make it jump to just the right spot (Chapter 14.iv and vii).

People, too, sometimes use such biological tricks—for instance, in locating the source of a sound (14.viii.c). But their perception and learning involves much, much, more. Even without language, mammals (and birds) can do what crickets cannot: they can learn to recognize new stimulus patterns, and can generalize those patterns over different class members (see Chapters 12 and 14).

Moreover, some of the detailed brain structures that enable mammals to *see*, or to *hear*, arise by spontaneous self-organization in the womb. (So the fact that a newborn baby, or kitten, already has a certain perceptual ability *doesn't* prove that it was specifically coded in the genes.) This may seem surprising, even magical. But computer modelling has shown how such anatomical self-organization is possible (Chapter 14.vi.b and ix.c).

You may be sceptical. You may feel, for instance, that this general approach is merely an example of what Donna Haraway (1944– ) calls “cyborg science”, more a mark of the times than of the truth (Haraway: 1986/1991).

As she puts it, the many sciences currently informed by the concepts of information and computation involve a “reinvention of nature”. They express a pervasive world-view, or “lived social reality”, in which human minds and human beings “are constructed as [jointly] natural–technical objects”. In her opinion, this view couldn't have arisen without post-Second World War military technology and its aggressive political background.

That last charge is true (see Chapters 4.vi.a, 11.i, and 12.vii.b). To a large extent, the “reinvention” charge is true also. Whether it follows that cognitive science is, as Haraway claims, deeply suspect and epistemologically compromised is quite another matter (Section iii.b–d, below).

Even if you're not an admirer of Haraway's writings, which include many provocative claims about the late twentieth-century Zeitgeist, you may nevertheless be sceptical about cognitive science. You may simply suspect that cognitive scientists have been seduced by the technology. Perhaps they're like the proverbial 'hacker' (11.ii.e), or those people from all walks of life who sit hunched over their bedroom computer for hours on end? Computers, after all, are only too capable of enticing people to waste their time. (Although, sadly, "waste" isn't always the right word: taking control over computers, which are usually much more predictable than other human beings, provides some devotees with their main source of ego strength and contentment—Shotton 1989: chs. 8 and 10.)

However, this type of research, vulgarly trendy though it may appear, is driven by a philosophical view of understanding and explanation that has deep and ancient roots. That is, it's an expression of the "maker's knowledge" (*verum factum*) tradition. This holds that in order to understand something properly one has to be able to make it. In other words, observation and abstract argumentation aren't enough.

One leading proponent of *verum factum* was Giambattista Vico (1668–1744), who famously argued that only the humanities can provide us with genuine knowledge, because they study the creations (not of God but) of human beings (Perez-Ramos 1988: 189–96; Miner 1998). Specifically, history—for Vico, the key to the understanding of human minds and cultures—involves an active re-creation of the thoughts of the people being studied. The eighteenth- and nineteenth-century Romantic philosophers used essentially similar arguments to prioritize art over science (see 2.vi.c and 9.iv). Others were more inclusive, applying *verum factum* to the natural sciences too. So for many of the early modern scientists, a scientific experiment was seen as a *construction*, and theory-based technology was an intellectual justification of 'pure' science.

In short: if you can build it, you can understand it. Cognitive scientists would agree—and their constructions include not only theories but computer models too.

### c. Never mind minds?

There's an even more difficult question, one which threatens to undermine the rationale of cognitive science as a whole. Namely: Maybe we'd be better off if we avoided talk of 'mind' altogether?

One way of doing that would be to avoid psychological language entirely (cf. Chapter 5.i.a). But at what cost? Gossip would be impossible—an advance in morality, perhaps, but not in the gaiety of nations. And scientific studies of the topics that fascinate gossips would be impossible too. We could describe the bodily movements, but couldn't say what *action* was being performed, or what *purpose* was being followed. Similarly, we could say how the brain cells are responding, but not what they're *doing*.

Less radically, one could retain psychological language but gloss it in purely behavioural terms, or perhaps in the abstract, functionalist, terms of information processing. Then, scientific studies (of these types) would be justified. The second of these is the position taken by the vast majority of cognitive scientists.

Or—an option that's recently grown increasingly popular *within* cognitive science, as we'll see—one could say that 'mind' was *invented* by Descartes, not just described

by him (Rorty 1979: 17–69). On that view, this Cartesian fiction (*sic*) separates the individual both from their own body and from other human beings—and the physical environment, too (see 2.iii.a–b). The implication is that cognitive science should stress embodiment rather than intellectualist reasoning, and social engagement rather than individualistic action and thought.

A much more radical approach would be to argue that both ‘mind’ *and* ‘body’ are concepts constructed on some deeper philosophical base, and are highly misleading when taken—by scientists, for example—as fundamental realities (see 14.xi and 16.vi–viii). That would eliminate many puzzles, but only by dismissing hope of *any* scientific explanation of psychology (and any naturalistic account of meaning). Cognitive science, on this view, wouldn’t just be difficult: it would be non-sense.

You don’t have to be a rocket scientist to guess that I don’t share that last view. Perhaps you don’t share it, either. But I’m not going to counter it yet. Indeed, we shan’t consider it at length until Chapter 16.vi–viii (although it will push its nose above the surface in Section iii.b below, and also in Chapter 14.xi).

Even there, I shan’t be able to give a knockdown argument against it: it’s perhaps the deepest division in philosophy. To make things worse, it’s often closely allied with a form of relativism that would undermine *all* scientific knowledge (see Section iii.b, below). However, many people—including many scientists—aren’t even aware of this division, and don’t take it seriously if they are. Throughout most of these pages, then, I’ll continue to speak of ‘mind’ and ‘mental’ phenomena as though such talk were relatively unproblematic. That is, I’ll assume that minds and mental phenomena do exist, even while admitting that there are many disagreements about just how they should be described.

Similarly, I’ll assume (until Chapter 16) that *some* scientific psychology is in principle possible. Even if it isn’t, the first fourteen chapters won’t be irrelevant. For a science-denier should offer an alternative interpretation of the facts discovered by science—for which task, they need to know something about what these facts are (see 16.viii.a).

All the examples I’ve mentioned in this section fall under the centuries-old questions about the mind that were listed at the outset. As remarked there, most of us have mused on these at some time or other. For cognitive scientists, they’re a prime concern. And as we’ll now see, they ask them in a particular way, which was inconceivable before the late 1930s.

## 1.ii. The Scope of Cognitive Science

Cognitive science is a catholic field, in three ways:

- \* First, it covers all aspects of mind and behaviour. (That was illustrated by the wide range of questions listed above.)
- \* Second, it draws on many different disciplines in studying them.
- \* And third, it relies on more than one kind of theory. Broadly speaking, it’s the study of *mind as machine*—a definition that covers various types of explanation, as we’ll see.

### a. Of labels and cans

In a neat and tidy world, where every label fitted what's inside the can, cognitive science would be the science of cognition (knowledge). Indeed, it's often defined that way. However, things aren't so simple.

In fact, cognitive science deals with all mental processes. Cognition (language, memory, perception, problem solving . . .) is included of course. But so are motivation, emotion, and social interaction—and the control of motor action, which is largely what cognition has evolved *for*.

You may feel that these types of psychological process aren't clearly distinguishable. If so, you're in good company. The 'holistic' belief that they're intimately intertwined is both very old-fashioned and very new. Its heyday was 200 years ago (see 2.vi). It never died out entirely: the cybernetic psychoanalyst Lawrence Kubie, for instance, said that "the various areas of psychic life are so interdependent that no one of them can be [experimentally] assayed alone and apart from the others" (1953: 48). However, it did go out of favour with the scientific community, resurfacing only very recently (Chapters 14.x.c, 15.vii.c, and 16.vii.c). Even now, it's an unorthodox view. For the moment, then, let's go along with the common assumption that cognition, motivation, emotion, social interaction, and bodily action can be considered separately.

Given that cognitive science isn't focused only on cognition, the label is highly misleading. Why, then, were these words chosen in the first place?

Today, they don't trip off Everyman's tongue. At the time, however, they were less arcane than one might think. Both had recently been popularized by social psychologists discussing "cognitive dissonance" (Festinger 1957). That terminology had even entered the media. Many journalists had summarized their explanations of the power of advertising, and of high pricing, on consumer behaviour. And the newspapers had had a field day in rehashing the social psychologists' reports about a recent cult in the mid-West of the USA (Festinger *et al.* 1956). These people had expected to be rescued from The End Of The World on a certain day by aliens in spaceships—only to see no EOTW, no spaceships, and *no* diminution of their faith in the cult leader (see 7.i.c).

In any event, professional psychologists were perfectly familiar with "cognition" as a technical term. But it had originally been coined, some two centuries earlier, specifically to *exclude* motivation and emotion. So, again, why choose it?

One of the two men mainly responsible—George Miller and Jerome Bruner—has explained it like this:

In reaching back for the word "cognition", I don't think anyone was intentionally excluding "volition" or "conation" [aka motivation] or "emotion" (Hilgard 1980). I think they were just reaching back for common sense. In using the word "cognition" we were setting ourselves off from behaviorism. We wanted something that was *mental*—but "mental psychology" seemed terribly redundant. (Miller 1986: 210)

In short, they intended "cognitive" science to address cognition *and more*.

A glance through *Plans and the Structure of Behavior* (G. A. Miller *et al.* 1960) confirms this. That inspirational book (see Preface, ii, and Chapter 6.iv.c) discussed animal behaviour, instinct, and learning, as well as human memory, language, problem solving, personality, mental illness, and hypnosis. Social and cultural matters were touched on also. No aspect of mental life was excluded.

(It was left to another early volume, however, to highlight political, bureaucratic, and economic behaviour: Guetzkow 1962. The editor, Harold Guetzkow, had been a close friend of Simon when they were both graduate students at Chicago. At that time, Simon's interests—like Guetzkow's—had been in economics and management science, not psychology: see 6.ii.a.)

The breadth of coverage in *Plans and the Structure of Behavior* was seen—by readers who were sympathetic at all—as being just as it should be. Virtually all the founding fathers of cognitive science (Noam Chomsky excepted) had asked how motives and emotions interact with cognition, and several had also mentioned psychopathology. In short, these more sexy matters (literally!) were often discussed in the early days.

That didn't last. Because motivation, emotion, and social interaction (whether in small groups or in societies) are even more difficult to study—and to simulate—than cognition is, they were soon put onto the back burner. They were left there for thirty years, while cognition got almost all the attention. Vast amounts of research were done on perception, language, problem solving, concepts, belief, memory, and learning. The name of the field reflects this.

In fact, the label on this particular can was changed several times. In the early 1960s, the field was known by the more neutral “computer simulation”. A Harvard graduate course was run under this rubric, and books appeared with titles such as *Computer Simulation of Personality* (Tomkins and Messick 1963), *Simulation in Social Science* (Guetzkow 1962), and *Computer Simulation of Behaviour* (M. J. Apter 1970). As research on cognition became more dominant, however, three new names emerged: cognitive studies, cognitive sciences, and cognitive science.

“Cognitive *Studies*” was chosen in 1961 by Bruner and Miller (the lead author of *Plans and the Structure of Behavior*) to name their new research centre at Harvard. This comprised a wide variety of psychologists, leavened by a few linguists and computer specialists and the occasional interdisciplinary philosopher. Nelson Goodman (1906–98), who co-founded Harvard's “Project Zero” studying representation and education in art, was in house when I was there, for example. These psychologists didn't do simulation as such, although Miller had co-published with Chomsky on mathematical models of language (9.vi.a). But they typically used ideas drawn from early AI, and from information theory, in their seminars and experiments (6.iv.d).

The “Cognitive” in the centre's title reflected the two co-founders' main interests: perception, language, memory, and problem solving. Even when Bruner studied values, he focused on their effect on *perception* (6.ii.a). Nevertheless, Miller later admitted that “conative” and “affective” phenomena (i.e. motives and emotions) should also be mentioned in the definition (see the quotation above, and also G. A. Miller 1978: 9).

“Studies” had become “sciences” by 1973. Already used in everyday chat by an Edinburgh research group for a couple of years, “cognitive sciences” first appeared in print in a defence of AI-based psychology, then under attack by a world-famous mathematician (Longuet-Higgins 1973: 37; cf. 11.iv). And the singular version—cognitive *science*—appeared soon afterwards, in two widely read collections of papers (Bobrow and Collins 1975, p. ix; Norman and Rumelhart 1975: 409).

Now, that's the label which is used most often. Even so, the editors of the recent *MIT Encyclopedia of the Cognitive Sciences* (R. A. Wilson and Keil 1999) chose the plural

version—which highlights the fact that several very different disciplines are involved in the field.

The singular form, by contrast, highlights the intellectual links between them. That’s why I’ve chosen to use it here. For as we’ll see, there have been countless instances of work in one discipline being radically influenced by work in another. That’s not surprising. To understand the mind (mind/brain) properly, one doesn’t only need to look at it from all directions: one must also *integrate* the various views.

## b. Two footpaths, many meadows

The field would be better defined as the study of “mind as machine”. For the core assumption is that the same type of scientific theory applies to minds and mindlike artefacts. More precisely, cognitive science is *the interdisciplinary study of mind, informed by theoretical concepts drawn from computer science and control theory*.

These concepts change, as time passes. (Many examples of such change are described in later chapters.) So cognitive scientists don’t believe that today’s computer-related concepts suffice to explain the mind. Rather, they believe that they’re a good beginning, and that later explanations will use concepts drawn from what then happens to be the best theory of what computers do (see Chapter 16.ix.f).

My “two-footpaths” definition, above, carries a health warning. As we’ll see later, one highly influential alternative definition of the field specifically *excludes* control theory. It allows only explanations in terms of formal symbol manipulation (see Chapters 12.x.d and 16.iv.c–d). Cybernetics, and even connectionism, is therefore said to lie outside cognitive science.

For reasons which I hope will become clear throughout the narrative, I regard that definition as much too narrow. It’s true, however, that cognitive science has seen—and is still seeing—competition, as well as cooperation, between computer science and cybernetics as ways of thinking about the mind. Indeed, the pendulum-swings between these two intellectual sources are a central, and fascinating, aspect of the story (see especially Chapters 4, 10, and 12–15).

As remarked in the Preface, the main disciplines involved are psychology, linguistics, AI, A-Life, neuroscience, and philosophy—and, though it’s relatively rarely mentioned, anthropology. Certain areas of biology, such as ethology and evolutionary theory, are also included (and, at the fringes, some aspects of biochemistry are relevant too: 15.x.b). Moreover, the many examples in Section i.a (above) imply that the relevant research ranges all the way from mate finding in crickets to grammar, and even grief, in human beings.

The history of cognitive science is marked by a deep, and continuing, interdisciplinarity. This is a more intellectually intimate relationship than mere multidisciplinary. Again and again, researchers in one area have borrowed *theoretical ideas*, not just *data*, from another.

Certainly, many specialist sub-areas (and sub-sub-areas . . .) have emerged over the past half-century. Each has its own conferences, journals, and textbooks. Moreover, their personnel rarely communicate. “Fair enough!” you may say. “If someone’s interested in depth vision or learning, why should they bother with English grammar?” Well, they don’t need to, in order to tackle their current problems. It remains true,

nevertheless, that stereopsis and learning are studied in the way they are today partly because of mid-century work on syntax and mid-1980s research on past-tense verbs (see Chapters 7.vi.a and 12.vi.e). In short, even the most ‘separate’ specialisms share life-giving historical roots.

They share some central assumptions, too. All areas of cognitive science are informed by computational concepts, and driven by computational questions. In other words, all cognitive scientists use such concepts as core theoretical terms. This isn’t the same thing as using computers. Biochemists or geologists, and non-computational psychologists too, often use computers as research tools (to do statistics, for example)—but their *theories* aren’t computational. (Nor is it the same thing as building computer models: many cognitive scientists do this—but many don’t.)

Broadly speaking, computational concepts are of two main types. On the one hand, they’re drawn from computer science, AI, and software engineering. On the other hand, they hail from information theory and control engineering—in a word, cybernetics.

This dual definition, like my catholic definition of the field (above), carries a health warning. ‘Computation’ is often understood as Alan Turing defined it (see Chapter 4.i.b–c). Indeed, his definition remains the only rigorous one. And it *doesn’t* cover cybernetics, nor even connectionist AI. Nevertheless, many people today—including computer scientists—use the term more widely. In other words, ideas about what computation *is* have become more extensive. (We’ll see in Chapter 16.ix that, despite the undeniable loss of rigour, there are good reasons for this.) I’m one of many who use the term more widely than purist symbolists do. In general, the context should show when I’m using the term to refer to ideas from only one of these two sources.

The two sides of the computational coin were first clearly distinguished in the mid-twentieth century (see Chapter 4), although each had been prefigured much earlier (2.vii–x). Over the years, the theoretical concepts involved have developed into a varied group, defining many different types of information processing, virtual architecture, and computer model. This development, which hasn’t been without hiccups, has involved both competition and cooperation between the two sides—first seen as competing in the late 1950s (Chapter 4.viii).

For example, research based on ‘dynamical systems’ falls on the cybernetic side of the fence, and is typically peppered with disparaging remarks about symbolic AI (14.ix.b, and 15.viii.c–d and xi). In particular, dynamicists claim that they can explain the temporal aspects of cognition, which earlier approaches ignored. But many people who work in this area were trained in AI, and depend heavily on it—for instance, when using genetic algorithms as ways of evolving dynamical systems (15.v). Similarly, most connectionist AI is closer to cybernetics, and has often been fiercely opposed to GOFAI—that is, to Good Old-Fashioned AI (Haugeland 1985: 112). Nevertheless, some researchers have tried to combine these two approaches (7.i.e–f, 12.viii–ix, and 13.iii.c).

As for the hiccups, Chapter 12 describes the birth *and renaissance* of connectionism—and the Sleeping Beauty phase in between. A-Life, too, had its Sleeping Beauty phase, from which it awoke one year later than connectionism. Psychology and philosophy have reflected these changes, offering very different theories of mind at different times.

The two computational pathways wound through many disciplinary meadows. The meadows were close neighbours at the beginning. Indeed, in the 1940s and 1950s—when distinct disciplines were being deliberately, outrageously, juxtaposed—highly *inclusive* consciousness-raising meetings were important (see Chapters 4.v.b and 6.iv.a–b). From the mid-1960s, however, the specialisms reasserted themselves. A second phase of “outrageous” interdisciplinarity was launched in 1987, at a party in the New Mexico desert (see 15.x). But most of the party-goers, though newly enthused, went back home to work in their own specialist houses.

This affects how our tale can be told. Chapters 2 to 6, by and large, move along a single time line, taking us from antiquity up to the mid-twentieth century. In Chapters 7 to 16, the time line branches. Each discipline has its own chapter (although AI has three, and we’ll backtrack about 500 years for linguistics). Even so, most of the important topics feature in *several* ‘disciplinary’ chapters. For the same two computational pathways are there throughout, connecting the different disciplinary meadows with each other.

### c. Why computers?

Computers as such are *in principle* less crucial for cognitive science than computational concepts are.

To be sure, computer technology (both digital and analogue) is an important player in the narrative. Computer modelling has a prominent role because it’s often needed *in practice* to confirm—or even to discover—the full implications of a computational theory. Indeed, advances in software design (especially high-level programming languages: 10.v) and computer engineering may be needed before such theoretical modelling can be attempted. Turing himself was unable to develop many of his ideas because of the primitive state of computers in his day (15.iv). As he put it:

At my present rate of working I produce about a thousand digits of programme a day, so that about sixty workers, working steadily through about fifty years might accomplish the job, if nothing went into the wastepaper basket. Some more expeditious method seems desirable. (A. M. Turing 1950: 61)

But cognitive scientists don’t always build computer models. Chomsky’s linguistics, John von Neumann’s cellular automata, and David Marr’s early brain theories, for example, were formal models—not functioning simulations (see Chapters 9.vi, 15.v, and 14.iv, respectively).

Some highly influential discussions weren’t even *formal*. Marvin Minsky’s “society of mind” theory (12.iii.d) and his earlier account of “frames” (10.iii.a), Michael Arbib’s schema diagram for control of the hand (14.vi.c), and Robert Abelson’s work on the structure of belief systems (7.i.a) are all cases in point. Indeed, the two seminal papers co-authored by Warren McCulloch and Walter Pitts (4.iii.e, 12.i.c, and 14.iii.a) were published in the early to mid-1940s, before the first modern digital computer had been built. It was another ten years before computer simulations of psychology were feasible. Some people would argue that *serious* simulation wasn’t possible before the late 1980s—if then (see Chapter 14.vi.d).

Because computational concepts are essential, AI—or AI/A-Life—is a central discipline. Not all of AI is germane, however. Workers in AI—and A-Life—can have either

of two motives. (Some have both.) The first is to build computer systems that are useful in some way. These range from automatic translators and financial networks to robot toys and remote-controlled surgeons. The second is to use software and/or robotics to help us understand human and animal minds (or life), or even *all possible* minds (or life). Let's call these 'technological' and 'psychological' (or 'biological') AI/A-Life, respectively. Only the latter project falls squarely within cognitive science.

Occasionally, this project has a further motive: to build, or anyway to start on the road towards building, a *real* intelligence, or a *real* living thing. These aims have driven some of the most well-known AI/A-Life research. And they've been very widely discussed for over fifty years—in terms (for example) of the Turing Test, strong AI, and strong A-Life (Chapter 16.ii, v.b–c, and ix.b). Nevertheless, they're minority tastes.

Despite the sensational quarter-truths peddled by the media ever since the 1950s, most researchers in AI/A-Life haven't argued—and probably haven't believed—that an AI program could actually be an intelligent mind, or that a merely virtual 'creature' could really be alive. Some have even denied it, claiming that *embodiment* is needed for life and intelligence (see Chapters 15 and 16.vii and x.) In short, this third motive has sometimes played a role in psychological AI/A-Life, but it isn't essential to it.

Technological AI is usually irrelevant to cognitive science—so is only rarely mentioned in my narrative—because it seeks to do something *irrespective* of how the mind/brain does it. The developers of IBM's Deep Blue, which beat the world chess champion Gary Kasparov in New York on 11 May 1997 (winning a prize of \$100,000 dollars in the process: see 16.ii.c), were happy to use dedicated computer chips. These enabled the program, processing 200 million positions per second, to rely on exhaustive look-ahead over eight moves. Anyone studying how human beings play chess would avoid this biologically unrealistic hardware.

There's one type of situation, however, where even purely technological AI is relevant: namely, if someone believes that certain tasks simply *cannot* be done by computers.

For instance, Hubert Dreyfus's judgement (in 1965) that no program could play even "amateur" chess was falsified only a year later, when he himself was defeated by a program (H. L. Dreyfus 1965: 10; Papert 1968, para. 1.5.1). And his claim that no computer would ever play chess at a human level unless it could distinguish perceptually between "promising" and "threatening" areas of the board (H. L. Dreyfus 1972, pp. xxix–xxxiii, 208) was decisively refuted by the performance of Deep Blue. The fact that its exhaustive "counting out" strategy, to use Dreyfus's term, isn't one that humans can use is irrelevant.

Admittedly, the distinction between the two types of AI isn't clear-cut. For instance, I said in the Preface that Margaret Masterman's pioneering work on machine translation and classification was technological, but also guided by strong intuitions about how people process language. It's not that she wasn't interested in how the mind works—although, as a post-Fregean philosopher, she was wary of 'psychologism' (Chapter 2.ix.b). But detailed psychological questions would have been premature. There was no experimental evidence enabling her to decide, for example, that one of two thesauri was the more realistic. She had to rely on other, more intuitive, criteria.

Even today, fifty years later, most technological AI is grounded in intuitions about human thinking. What the writers of expert systems call "knowledge engineering", for example, includes a method of questioning human experts, to help them make

their expertise explicit (10.iv.c). Sometimes, data from experimental psychology and/or neuroscience influence the program too. Some industrial applications even use special-purpose hardware chips modelled on the mammalian visual system (12.v.f). But the interpretation of the AI systems as models of actual mental processes isn't the object of the exercise.

#### d. What's in, what's out

Not all of *psychology* is germane to cognitive science, either. All psychological *data* are relevant, in the sense that cognitive science, if it is to succeed, must one day explain them. But many psychological theories aren't computational. This history narrates how some theoretical psychology became computational—and the dramatis personae are selected accordingly. (The same applies, *mutatis mutandis*, to anthropology, linguistics, and neuroscience.)

I'll say relatively little, for example, about the behaviourists, or Sigmund Freud—despite their importance for the history of psychology in general. With respect to computational psychology, behaviourism was significant as something to be reacted against, not developed (see Chapters 5 and 6.i–iii).

As for Freud, his psychodynamics (Chapter 5.ii.a) inspired some early AI simulations of neurosis, and a model of the effects of anxiety on speech (7.i.a and ii.c). In a broad sense, it informed Minsky's work on the society of mind (12.iii.d), and spread from there to Daniel Dennett's philosophy of consciousness (14.xi.b and 16.iv.a–b). It contributed to Arbib's schema theory, and especially to the application of schema theory to religion (7.i.g). And it provided examples and ideas that fed into Aaron Sloman's work on the architecture of grief (see Chapter 7.i.f). So I mention Freud in all those contexts—but I don't focus closely on him.

Even “cognitive psychology” doesn't always fall within cognitive science (Chapter 6.v.b). For instance, David Clark's (1996) explanation of—and therapy for—anxiety disorders (such as phobias, panic attacks, and post-traumatic stress disorder) analyses them in terms of the person's underlying beliefs about danger: it's cognitive, but doesn't involve explicit reference to computational concepts or theories. Admittedly, the term “cognitive psychology” was first defined in a computational context (Neisser 1967). But some cognitive psychologists, including that author himself in later years, specifically reject computational theories (7.v.e–f).

You may be surprised to see neuroscience included in this account. For a recent dictionary of psychology states that “cognitive scientists rarely pay much attention to the nervous system”, and that cognitive science and neuroscience are “almost mutually exclusive” (N. S. Sutherland 1995: 83). The explanation given there is that “cognitive science deals with the brain's software, neuroscience with its hardware”.

As a quick summary, that's correct. The cognitive scientists of the 1960s and 1970s adopted, or defined, an *abstract* (functionalist) philosophy of mind, and most still do (see 16.iii–iv). But facts about the brain have inspired various forms of AI, and of its cousin, cybernetics (Chapters 4.iii–vii and 12). Moreover, the intellectual traffic is increasingly two-way. Computational ideas inspired one of the most famous papers in neurophysiology, ‘What the Frog's Eye Tells the Frog's Brain’ (Lettvin *et al.* 1959), and they've been used for thirty years to model the brain (see Chapter 14). Neuroscientists

today regularly use computational ideas, asking not only which cells and neurochemicals are involved but also what functions they're computing and/or modulating.

You may not expect A-Life to be counted as a member discipline, either. For A-Life researchers usually make a point of distancing themselves from traditional cognitive science, rejecting its representational (“Cartesian”) view of mind. Even when they do so, however, they may admit that they “count as cognitive scientists” in so far as they try to create “working artifacts that demonstrate basic cognitive abilities” (I. Harvey 2005: first page).

Moreover, A-Life bears a new name on an old bottle. It dates back to mid-century cybernetics, and its three founding fathers—von Neumann, Turing, and W. Ross Ashby—were crucial also for the rise of other areas in cognitive science. In addition, many people believe that life and/or embodiment is *necessary* for mind (Chapter 16.vii and x). If they're right, then A-Life would be essential to our discussion even without the strong historical links.

The inclusion of anthropology may surprise you also. For it's rarely mentioned as being part of cognitive science. But we'll see in Chapter 8 that this is truer today than it was forty years ago, and that part of the reason is that much of the relevant work now appears under a different label—namely, evolutionary psychology.

As for philosophy, it plays a key role in the story. Indeed, it's rarely absent. Besides having its own chapter (Chapter 16), it has dedicated sections in several others (see Chapters 2.ii–iii, vi, and ix.b; 7.iii; 9.iv–v and viii; 10.iii.f and vi.a–d; 12.x; and 14.viii and xi). And it provides many passing remarks throughout the story.

This isn't merely because, before the scientists got their act together, only philosophers were saying anything about the mind. It's also because the scientists themselves have repeatedly raised philosophical questions—and they still do.

That's why there are so many different definitions of cognitive science, and so many competing camps (and changing fashions) within it. My pages don't tell a happy tale of consensual colleagues harmoniously seeking the same sort of truth. Personal rivalries and self-regard have played their part, of course: we're talking about human beings here, after all. But the deepest divisions have been at philosophical fault lines, so that *what could possibly count* as the truth, or as an explanation, isn't always agreed.

For example, cognitive scientists differ profoundly on how the mind–body distinction should be understood. A few even reject the distinction entirely. They also differ on what can count as a *representation*, a term that's often used within the field—and in some influential definitions of it.

Moreover, some philosophers outside the field argue that cognitive science can't illuminate *any* aspect of mind—not even cognition (or representation) itself. Such fundamental disputes have their own history, recounted in Chapter 16.iii–viii. Meanwhile (as explained in Section i.c, above), I'll continue to use “mind” as a useful shorthand, setting these basic philosophical questions aside until then.

The existence of philosophical disagreements between cognitive scientists may partly explain N. Stuart Sutherland's acid comment that cognitive science is “[an] expression [that] has come into being mainly in order to allow workers who are not scientists to claim that they are” (N. S. Sutherland 1995: 83). “Surely”, someone might say in Sutherland's defence, “a *science* must have some fundamental philosophy on which all its practitioners agree?” This objection is too confident: even quantum physics

would be excluded, given the notorious disagreement among physicists about its interpretation. With respect to cognitive science, it's partly because the theoretical foundations are still in dispute that it's so fascinating.

In sum, my view of cognitive science is relatively broad. It covers seven disciplines, and various forms of explanation. In particular, it includes both symbolic and cybernetic theories, despite the differences between them. Many people define the field in a much narrower way (see 12.x.d, 14.ix.b, and 16.iv.c–d). To do that, however, is both philosophically controversial and historically misleading. Cognitive science is a rich intellectual tapestry, woven over the years from many different threads.

### 1.iii. *Caveat Narrator*

Before embarking on my story of how cognitive scientists came to think as they do today, eight caveats are needed—for me as author, and for you as reader. (So it's *caveat lector* as well.)

The first three concern the temptations of Whig history (Butterfield 1931) and of over-idealized views of science. The fourth warns against the seductiveness of heroic accounts of creativity. The next two point out the difficulty in identifying an idea as a discovery—or even as new (which isn't the same thing). The seventh focuses on the role of rhetoric and publication in gaining a place in history. And the last reminds us that whether a new idea is favoured can depend on what sort of thing people with influence are prepared to count as an 'explanation'.

These are very *general* dangers. They make history-telling inherently problematic, a matter of more than just good plain fact. Indeed, a pessimist might argue that *Joe Bloggs was born in year x* is as near to plain fact as one can possibly get. I don't think the situation is quite as hopeless as that (see subsection b, below). But the historical claims in the following chapters, whether made by me or quoted from someone else—in their personal recollections, perhaps—are all prey in principle to the dangers listed here.

I'll mention some examples in this section as illustrations—but only very briefly: each one is discussed at greater length later. And if you return to this chapter after having read all the others, you'll be able to add many further instances.

There's a ninth, higher-order, warning too: the accepted history of the field is itself *part of the field*, helping to shape researchers' judgements and aims. However, because of the difficulties noted in the other caveats, there's no such thing as 'the' accepted history. For cognitive scientists with different theoretical agendas often differ about who did what, why they did it, and what the consequences were.

One illustration concerns the twenty-year connectionist 'winter': was it due to the undeveloped state of technology, to theoretical weakness, to the self-serving activities of two highly combative men, or to the fact that one of them had a very old friend in very high quarters? I'll ascribe it in some measure to all four (Chapter 12.iii.e). But that's not what it looked like at the coal face—and even there, it looked different depending upon which seam one was working on.

In brief: myths matter. History gets told over the coffee cups, as well as (sometimes) being made there—and the telling can affect the making.

## a. Beware of Whig history

I don't assume nor, I hope, imply the Whiggist view that everything gets progressively better and better until we reach today's date.

Still less do I assume that it gets better smoothly. I've already mentioned the Sleeping Beauty phase of connectionism, for instance. Hotly disputed theoretical alternatives have surfaced, disappeared, and resurfaced over the years, and one aim of this book is to help readers understand why.

The Whiggist assumption, within the history of science, leads to two common types of anachronism.

On the one hand, ideas that don't fit into the progressivist narrative are ignored. This is usually because, from the historian's own viewpoint, they're incorrect or misguided. They may even be overlooked because they've become near-unintelligible. It can be difficult to understand what questions scholars and scientists of the past were asking, and why, if their concerns didn't align with what counts today as a scientific enquiry (N. Jardine 1991).

On the other hand, past ideas may be tendentiously misrepresented. This may be done by stating/implying that earlier thinkers were trying to solve *our* problems, and/or by using today's terminology in describing their work. It's all too easy to project our own concerns onto ancient writings that bear some superficial resemblance to ours, in order to make a progressivist story appear more plausible.

It's also possible, of course, to be correct in attributing something very like our interests to a past thinker, but incorrect in assuming—because of Whiggist pre-suppositions—that their work was actually influential. Conversely, if someone was unrepresentative of their age, we may fail to look closely enough to find interesting parallels between their thought and today's ideas.

Within cognitive science, there are specific examples that should warn us against all these varieties of Whiggism. They include arguments relating to fundamental issues, such as disputes about whether Descartes's account of mind was an advance on Aristotle's (Chapters 2.iii and 14.x–xi), or whether Goethe's biological views were well understood, and rightly scorned, by most of his scientific successors (2.vi.d–f, and 15.iii.a and vii.c). They also include narrower issues, such as whether Humboldt's views on language were as similar to Chomsky's as Chomsky himself has claimed (9.v.d–g). In addition, some work should have been more influential than it has been (7.i.e–f). An uncritically Whiggist approach would either exaggerate its early influence or ignore it entirely.

As for backwards projection of our own concepts onto the past, Descartes—contrary to what is now widely believed—did not deny (what we call) conscious experience to animals. Nor did he ascribe it to them, either. This wasn't because he was agnostic about the facts of the matter, but because he lacked the particular concept of consciousness used in formulating the question (see Chapter 2.ii.d–e). I argue also that we shouldn't attribute any version of the modern idea of 'man as machine' to those early Greek philosophers who posited materialism, nor to any ancient engineers inspired by them (Chapter 2.i). It's similarly anachronistic to assume that Charles Babbage had any intention of likening minds to machines, for he didn't: he was *not* an early cognitive scientist (see 3.iv).

Some historical figures with (some) ideas remarkably close to our own had a negligible—perhaps even regressive—influence on the later development of those ideas. Usually, this was because their ideas were so far ahead of their time that their contemporaries couldn't appreciate them.

Babbage, again, is (arguably) an example. He's often described as a crucial player in a Whiggist story about the development of computer technology, but some experts believe that knowledge of his work actually delayed the invention of modern computers (Chapter 3.vi.a). And Jacques de Vaucanson is often wrongly assumed to have had no scientific intent in building his automata, largely because most of his fellow artificers didn't—and because he committed the sin of exhibiting them for money (2.iv). Most of his contemporaries were under the same illusion, so even they didn't go to him for scientific inspiration.

As for more recent examples, Petr Smirnov-Troyanskii's pioneering ideas about machine translation, part-patented in 1933, were ignored even in his native Russia (9.x.a). John Clippinger's intriguing mid-1970s model of the effects of anxiety on speech fell into a black hole from which it hasn't yet emerged (7.ii.c). And even Karl Lashley's now famous discussion of 'The Problem of Serial Order in Behavior' (K. S. Lashley 1951*a*) caused barely a ripple when it was first delivered in 1948, or first published in 1951 (5.iv.a). It came into its own in 1960, when—thanks to late 1950s work in AI and linguistics—cognitive scientists were at last in a position to appreciate it.

Nevertheless, there are many areas where one can point to definite progress. And the value I have in mind here—given that “progress” can't be defined in a value-neutral way—is *closeness to the truth in describing and explaining aspects of the world*. This includes both how real people *actually* manage to think about real and imaginary worlds, and how it's *possible* for any creature—man, mouse, or Martian—to do so.

The growth of our understanding of cellular automata is one clear example of progress (15.v–viii). The advance of neuroscience is another (2.viii, 7.v.b–d, and 14), and of interactive human–computer interfaces yet another (9.xi.f and 13.v–vi). A fourth (related) example concerns the idea of mind-expanding 'cognitive technologies'. These were prefigured by a physicist in the 1940s, named by an experimental psychologist in the 1960s, and clarified/complexified over the next forty years by psychologists, computer scientists, and philosophers (10.i.h, 6.ii.c, 13.v, and 16.vii.d respectively).

Even when significant controversy remains, much may have been found out. That applies, for instance, to the degree of 'modularity' in human minds (7.vi.d–i); to the relation between autism and 'Theory of Mind' (7.vi.f); to how we understand religious concepts (8.vi); and to how we compensate for the fact that our rationality is limited (6.iii, 7.iv, and 8.i.b). Indeed, *every* chapter instances—quite apart from the amassing of new empirical data—the gradual clarification and development of theoretical ideas.

This constitutes scientific advance, even if the idea eventually turns out to be an explanatory cul-de-sac. (As Karl Popper put it, science grows by conjectures *and* refutations: K. R. Popper 1963; likewise Francis Bacon, in *The New Organon*, “Truth emerges more readily from error than from confusion”: Bacon 1620.)

I make no apology, therefore, for the strong whiff of Whiggism that attends many parts of my text.

## b. Losing the Legend

That claim—that we’ve already seen some progress in understanding the real world—may raise a few readers’ eyebrows. Indeed, it may raise their blood pressure too. For, as evidenced by the ‘science wars’ that have raged for the last thirty years, social constructivists—supporters of ‘the strong programme’ in the sociology of knowledge—reject the possibility of scientific progress (as just defined) *in principle*.

That’s because they deny that science is the study of some objective reality which exists, and which has certain properties rather than others, independently of human minds. In other words, they reject *realist* accounts of science, and *objectivist* accounts of truth.

One root of this position, which was officially inaugurated by Edinburgh’s Science Studies group in the early 1970s (Bloor 1973), and backed by a quasi-anthropological study of the Salk Institute soon afterwards (Latour and Woolgar 1979), is the work of Thomas Kuhn. And cognitive science, arguably, offers examples illustrating Kuhn’s remark that science “progresses” because old scientists die (1962: 150). I say “arguably” partly because most of the field’s leaders are still alive. But partly, too, because the new ideas resisted to the end by some of those who’ve died—connectionism, in the case of the late Simon and Allen Newell for instance—are still *competing theoretical positions*, rather than new *paradigms* unquestioningly accepted by everyone else (see Chapters 12.viii–ix and x.d, and 15.viii.b–c).

Another root is neo-Kantian philosophy (introduced in Chapter 2.vi), or what’s often called Continental philosophy—including the version of it known as postmodernism. For the science wars are a special case of what Simon Blackburn (2005) dubs “the truth wars”, which oppose relativism to realism in respect of *all* areas of knowledge/belief. He sees this opposition as “arguably the most exciting and engaging issue in the whole of philosophy” (2005, p. xx). Certainly, it’s fundamental to the philosophy of mind, as well as of science: that’s why I flagged it very early on, in Section i.c above. Indeed, some disputes within and about cognitive science today are grounded in this philosophical debate (7.vii.c, 14.xi.a, and 15.viii–xi).

However, neo-Kantianism is so radically different from the Cartesian–empiricist Anglo-Saxon approach that the two are, to borrow Kuhn’s term, virtually incommensurable. The one prioritizes sophisticated interpretation, or hermeneutics; the other relies on scientific objectivity, as understood in the experimental tradition. Consequently, they have very different ideas on what counts as a proper scientific attitude, especially with respect to the biological sciences—and a fortiori the human sciences of psychology and anthropology.

Both philosophical positions can be defended, to be sure (see 16.vi–viii). And historical evidence can be marshalled for each of them: the case for constructivism in seventeenth-century science, for example, has been made with some spirit, and much fascinating detail, by Steven Shapin and Simon Schaffer (Shapin and Schaffer 1985; Shapin 1994). But there’s no clincher argument on either side. At some point, one must opt (*sic*) for one or the other.

For my part, I regard the constructivists’ position as fundamentally irrational, even though they did have some important insights—which need to be remembered if one

is to understand the history of this, or any other, scientific field (see below). The case against them was put in a nutshell nearly 400 years ago by Bacon:

After the human mind has once despaired of finding truth, everything becomes very much feebler; and the result is that they turn men aside to agreeable discussions and discourses, and a kind of ambling around things, rather than sustain them in the severe path of inquiry. (Bacon 1620: 56)

Bacon himself, of course, was largely responsible for defining what is now familiar as well-honed scientific method. So today's constructivists are even more shocking than Bacon's contemporaries, in their rejection of "the severe path of inquiry".

The irrationality of their approach has been clearly summarized by philosophers Susan Haack (1996) and Noretta Koertge (2000), and put more tendentiously—but with panache—by the sociologist Ernest Gellner (1992). A serious philosophical rebuttal can't merely ask, "If science is mere cultural convention, why would anyone ever board an aeroplane?", and leave it at that. The arguments for constructivism, and also those against it, are more subtle. But this isn't the place to recount them (for an excellent discussion, see Blackburn 2005: esp. chs. 6–7). So I'll merely say that I see relativist and anti-realist philosophies of science as both self-defeating and fundamentally implausible. The relativism undermines every philosophical claim; and the rejection of realism, despite science's practical successes, ignores what philosophers of science call IBE, or inference to the best explanation (Harman 1965; Lipton 1991).

If realism is correct, as the vast majority of practising scientists (and most readers of this book?) assume, then scientific progress—in the sense defined above—is in principle possible. And in cognitive science, it has actually happened. In one discipline, it's happened *despite* opposition from highly influential constructivists in the area concerned (Chapter 8.ii.b–c).

So where's the caveat? Well, even where there has been progress, it hasn't always happened in a 'purely scientific' way. Personal and political interests sometimes influence the generation and/or acceptance of ideas in science, as they do in the arts. This is part of what the constructivists have been pointing out.

In other words, science in practice *doesn't* fit what the physicist, and science-policy expert, John Ziman (1925–2004) has called "the Legend". The Legend is "the stereotype of science that idealizes its every aspect", representing it as wholly objective and rational (Ziman 2000a: 2). If that mismatch isn't recognized, one's historical view of the field will be misleading.

In Chapter 2, for example, I describe the rise of Cartesianism within science in general and neurophysiology in particular. This took place very quickly: "From 1700 on, [such ideas] were taken to 'go without saying'; and, in practice, they often went unsaid" (Toulmin 1999: 108). The rapid spread of Cartesian ideas happened partly because they supported the sorts of experiment and theory already pioneered by Galileo Galilei and William Harvey, and taken up by the nascent Royal Society in London and Académie royale des sciences in Paris. (Three cheers for the Legend!) But there were sociopolitical reasons, too.

People weary of the religious disputes and bigotry that had been brewing in Europe since the late sixteenth century, and which had recently led to the ruinous—and destabilizing—Thirty Years War, welcomed Descartes's stress on clarity and agreement

(even certainty) independent of religion. This was especially true in England and France, and in ‘establishment’ circles (Toulmin 1999: 119–25). And, of course, the gentlemen members of the two “Royal” societies were establishment figures par excellence. So they had more than purely scientific reasons for favouring Descartes.

That’s not to say that these nebulous political considerations were consciously recognized by the individual thinkers concerned—still less, that the novel ideas were *generated* because of them. Even far less nebulous political influences may be invisible to the people whose creative work is being accepted largely because of them. For example, we’ll see below (in subsection c) that Cold War politics encouraged the explosion of Abstract Expressionism in post-war New York, and fostered its acceptance by art critics around the world. However, this fact has come to light only relatively recently. At the time, it was hidden *even from the artists themselves*.

Sometimes, political influences are more easily visible. So, for instance, progress in cybernetics and AI has been (and still is) largely fuelled by military aims and funding, as opposed to the disinterested pursuit of truth (Chapters 4.vi.a, 9.x.a, 11.i, and 12.vii.b). That’s a large part of what Haraway was saying (Section i.a, above), although she supplemented it with much more questionable claims—about the “need” for a feminist philosophy of science, for example (for a sensible rebuttal, see Koertge 2000).

Personal interests have entered the picture too. A hugely influential Report on AI prepared for the UK’s Science Research Council in 1972 would probably never have been commissioned, and for sure would have been very different, if the personality of the UK’s leading AI scientist had been other than it was (see 13.iv.a–b). A disinterested document this was not.

Much the same applies to an equally influential, and apparently highly abstract, attack on connectionist AI (12.iii). The published version pulled no punches, and the draft had contained many vitriolic remarks which friends persuaded the authors to remove.

Perhaps you’re surprised, even shocked? The more one buys into the Legend, the more surprised one will be. The Legend has had powerful defenders—above all, among followers of Popper (1935, 1963). Popper himself was a sophisticate: he saw his theory as an idealization (a “rational reconstruction”) of science, an account of how it should be done rather than a description of how it is in fact done. (Predictably, he dismissed Kuhn as a *philosopher* of science, seeing him as a sociologist/historian instead.) But cruder versions of the Legend still abound, often written by practising scientists (e.g. L. Wolpert 1992).

In such accounts, personal prejudices and sociopolitical factors are glossed—where they’re even recognized—as unwanted subjective intrusions into objective study. While they may, and perhaps should, often determine which problems will be considered by science, they’re held to be irrelevant to the content of the solutions offered. James Watson (1968), for instance, admits to the personal ambitions that drove him and Francis Crick in their search for the structure of DNA, but regards these as irrelevant to what they said in their scientific papers. The claim is that theory is value-free, even if choice of research topic isn’t.

But that’s too quick. For one thing, choice of research topic is hugely relevant in understanding the history of an entire scientific field. For another, even the driest theories

are sometimes infected by personal/political values. (So a claim about the superiority of one programming language, or programming style, over another—*Boring!*—was typically stated, during the years of the Cold War, in terms pitting liberal democracy and egalitarianism against authoritarian regimes: Chapter 10.iv.a.) As for theories carrying implications regarding the nature of human beings, the scope for personal/political issues to prejudice clear judgement is significant (see Chapter 8.ii). Third, the core point of the constructivists' case, science is influenced also by *conceptual* assumptions, which can even affect researchers' *perception* of the 'data' (see 16.iv.e).

Ziman gives countless examples (mostly drawn from physics and chemistry)—and careful arguments—showing that the Legend not only isn't true, but couldn't be true. It underplays the sociology and psychology of science, and distorts its epistemology.

However, the Legend's polar opposite (constructivism) isn't true either—as I've said, above. Indeed, Ziman wrote his book to *defend* science against the most extreme of these attacks. Besides his many anti-Legend examples, he also cites a host of cases showing how science, as a body of theory and as a social institution, works as a relatively objective, rational, enterprise—despite the frailties of its human practitioners.

Why did Ziman feel the need to do this? After all, he wasn't a professional philosopher. Can't science be left to look after itself? Well, apparently not:

Science is under attack. People are losing confidence in its powers. Pseudo-scientific beliefs thrive. Anti-science speakers win public debates. Industrial firms misuse technology. Legislators curb experiments. Governments slash research funding. Even fellow scholars are becoming sceptical of its aims. (Ziman 2000a: 1)

The “fellow scholar” mentioned here was the historian Gerald Holton, a long-time friend of science whose recent work had, perhaps surprisingly, given some credit to the “anti-science” position (Holton 1992, 1993).

Ziman, too, gave constructivism some credit. He even went so far as to declare himself a postmodernist:

Our investigation thus arrives at a paradoxical conclusion. Academic science, the spearhead of modernism, is *pre-modern* in its cultural practices: and yet it turns out to be *post-modern* in its epistemology.

Contrary to the Legend, science is not a uniquely privileged way of understanding things, superior to all others. It is not based on firmer or deeper foundations than any other mode of human cognition. Scientific knowledge is not a universal “metanarrative” from which one might eventually expect to be able to deduce a reliable answer to every meaningful question about the world. It is not objective, but reflexive: the interaction between the knower and what is to be known is an essential element of the knowledge. And like any other human product, it is not value-free, but permeated with social interests. (2000a: 327)

He hastened to add, however, that “terms such as ‘modernism’ and ‘post-modernism’ are very ill-defined . . . Most scientists only know of them as slogans, uttered wholesale by the partisans of the most diverse fashions and fads.”

That's true. Even the “partisans” use the term postmodernism in differing ways. When it was used by the French philosopher Jean-François Lyotard (1924–98) in the late 1970s, in his study of “knowledge in computerized societies”, he'd attacked all overarching philosophies (“meta-narratives” or *grands récits*) of human progress. As though his little (110-page) book weren't succinct enough, he put it in a nutshell:

“Simplifying to the extreme,” he said, “I define *postmodern* as incredulity towards metanarratives” (1984, p. xxiv). So he attacked Marxism and Christianity, modernist aesthetics (which trumpeted the restorative ‘transcendence’ of art), and scientism too. His main target, indeed, was the “legitimacy” of scientific knowledge, and its “mercantilization” owing to “the computerization of society” (1984: 3–9). In other words, postmodernism was explicitly opposed to the growing role of science and computers in the late twentieth century.

Indeed, science had already been targeted by the literary scholar Roland Barthes (1915–80). In his essay on ‘The Death of the Author’ (written in May 1968, during the revolutionary events on the streets of Paris), he’d rejected its ‘quasi-theological’ claim to objectivity:

Literature . . . by refusing to assign a “secret”, an ultimate meaning, to the text (*and to the world as text*), liberates what may be called an anti-theological activity, an activity that is truly revolutionary since to refuse to fix meaning is in the end to refuse God and his hypostases—*reason, science, law*. (Barthes 1968/1977: 147; italics added)

Renaissance scholars, too, had seen the world as a text, to be interpreted by us. But they’d supposed it to be God’s text, expressing His hidden meaning. The scientific revolution had consisted, in large part, of rejecting these hermeneutic (intentional) accounts of the world in favour of empirical ones—hence the subtitle of Bacon’s 1620 book: *or, A True Guide to the Interpretation of Nature* (and cf. Glanville 1661–76). Now, four centuries later, Barthes was describing the world as a text with *no* author, whose multifarious “meanings” aren’t actual messages (or scientific laws) for us to discover but possible conceptualizations for us to construct.

Lyotard himself admitted that his thesis “makes no claims of being original”—“or”, he added (as a relativist must), “even true” (p. 7). It revived the neo-Kantian contrast between scientific and humanist knowledge emphasized by Wilhelm Dilthey (1833–1911) and Max Weber (1864–1920), and by many philosophers today (see Chapter 16.viii.b). Where Dilthey (1883) had spoken of empirical/scientific versus historical/hermeneutic knowledge, and Weber of *Naturwissenschaft* versus *Geisteswissenschaft* (i.e. natural science versus human, or interpretative, science: Shils and Finch 1949), Lyotard spoke of science versus narrative. (And often of discourse rather than knowledge, for the legitimization of a given type of discourse *as* knowledge was being questioned.) Moreover, he was offering a political critique (of global capitalism) as much as an exercise in ‘pure’ epistemology.

Lyotard’s influence on political and aesthetic discourse (and on the practice of art) was immense. But science, and its computerized offshoots, didn’t escape. And Barthes, even though he’d focused primarily on literary texts, engendered hostility to science—more specifically, to science’s claims to objectivity—in his readers.

As postmodernism flourished in the 1980s and 1990s, becoming (to traditional minds) ever more outrageous and bizarre, some followers of Lyotard and Barthes—often encouraged by the Science Studies work mentioned above—passed from deep scepticism to genuine incredulity, *even with respect to scientists’ answers to ‘properly’ scientific questions*. They didn’t confine their doubts to the social/behavioural sciences, which many philosophers believe lie outside science’s reach (see 16. viii.a–b), for even the objectivity of physics was suspect. Calls were made, for example, for a specifically

feminist philosophy—and practice—of science (e.g. Harding 1986; Longino 1990). As one illustration, the feminist anthropologist Emily Martin (1991) made a bitter sociopolitical attack on theories of reproductive biology (for a sanity-restoring reply, see P. R. Gross 1998). It was as though, in these postmodernists' eyes, the metanarrative was not just scientism but science itself.

That was many steps too far for Ziman. He was an opponent of scientism—but not of science. His version of postmodernism, then, is very different from that of the virulently anti-science faction in the science wars.

In short, the Legend is false. The disinterested pursuit of truth is rarer than many choose to believe, and the conceptually innocent ditto is impossible. On those points at least, the constructivists have had important things to say—things which must be borne in mind when thinking about the history of cognitive science.

### c. The counter-cultural background

The science wars were part of a more general intellectual/political movement, which the sculptor Theodore Roszak (1969) dubbed the counter-culture. This was politically prominent in the 1960s and early 1970s, and still intellectually prominent in the 1980s and 1990s—by which time it had largely metamorphosed into the postmodernist movement. On the whole, counter-culturalism favoured non-scientific, and even explicitly anti-scientific, ways of thinking. (For examples of counter-cultural anti-science diatribes, see P. R. Gross and Levitt 1994/1998; Parsons 2003.) It wasn't only the arts that were celebrated. Religions in general, and 'New Age' spirituality, flourished. Crystals were more respected as amulets than as chemicals on the laboratory bench.

Computers were even less favoured than chemicals. Then available only to very large organizations, they were seen by most 1960s–1970s counter-culturalists not as a useful tool but as a threat. And the image of mind as machine was the deepest threat of all: "Technocratic assumptions about the nature of man, society, and nature [i.e. cybernetics]", said Roszak, had "warped" the experience of scientists, scholars, and policy-makers at source. "In order to root out those distortive assumptions, *nothing less is required than the subversion of the scientific world view*" (1969: 50; italics added).

This movement had philosophical roots in the nineteenth and early twentieth centuries (Romanticism, Marxism, and Continental phenomenology). In that guise, the culture it was countering was Cartesian modernism, and especially Enlightenment optimism about the reach of science.

The philosophical arguments were reinforced/overlain by mid-century disillusion about some of the intended (e.g. Hiroshima) and unintended (e.g. pollution) effects of science. Rachel Carson's environmentalist book *The Silent Spring* (1960), and her *New Yorker* articles on the ill effects of industrial chemicals, led to heated debates in the US Congress. It wasn't only decisions about technology that were questioned: explicit attacks were mounted against taxpayers' money being given for science education and research in general (R. C. Atkinson 1999). (Science education in the USA had been hugely boosted around 1960, because of the shock presented by the Soviets' Sputnik, a football-sized satellite orbiting the earth every 90 minutes: see 11.i, preamble.)

But the anti-science campaign also drew major strength and inspiration from political events—above all, the Cold War. This dominated Western, and especially American, politics and military planning from 1947 on. (There was some lessening of tension during the 1970s, but it was ratcheted up again by Ronald Reagan in the 1980s: see Chapter 11.i.c.) So the counter-culture was driven by a growing unease about the effects of technology (from nuclear weapons to agrochemicals), and about the arms race and MAD (Mutually Assured Destruction) calculations of the Cold War.

It was further inflamed by the passions involved in the mid-1960s civil rights movement in the USA. Not least, it was fostered by discontent/disillusion about the Vietnam war—in which more US bombs were dropped on that undeveloped country “than had been dropped by all combatants in *all previous wars combined*” (P. N. Edwards 1996: 137). Huge sums of money for scientific (including AI) research aiding the Vietnam adventure were made available from the mid-1960s to early 1970s, although lack of success in both the research and the war led to a tightening in the early 1970s.

Encouraged by intellectual/political leaders such as Chomsky (9.vii.a), Herbert Marcuse (1964), and Angela Davis, young people were enthused to protest against these political forces. Student activism in the USA began in 1964 with the Free Speech Movement at Berkeley, which led to overreaction by the authorities and eventually to riots and tear gas on campus (McGill 1982). Those confrontations, and many others across the world, were fuelled by the student-led *événements* in Paris in May 1968, which are still frequently mentioned today. Even those young people who didn’t join in the violence were often broadly sympathetic.

Science and technology were seen as enemy forces. Both were crucial to Cold War rhetoric, and to Cold War military preparations, whether offensive or defensive—and generous funding was provided by the US (and UK) governments accordingly. (Some of it reached AI and other areas of cognitive science: 6.iv.f and 11.i.) Moreover, that funding was clearly visible to the public, and thus provocative to those whose politics might lead them to be provoked. Roszak himself was one of those, and his *Counter-culture* book provided a host of references to the angry and/or despairing writings of many others (1969; cf. 1986).

He wasn’t concerned only—or even primarily—with bombs. Scorning the “commonplace contemporary idiocies which small minds are now busily elaborating into a *Weltanschauung*”, he rued “the degradation of human personality” that he believed was resulting from the use of Wiener’s cybernetic metaphors for mind.

This effect, he said, was no mere philosophical bagatelle. For it was influencing military policy too: “Not even Jonathan Swift could have invented such pernicious lunacy as the balance of terror or thermonuclear civil defense” (1969: 295). The “lunacy” was rooted in scientific and technological research in general, and especially in the work of RAND, the Stanford Research Institute, “and ever so many other military–industrial–university think-tanks”. (Both RAND and SRI were crucial to the rise of AI, as we’ll see in Chapters 10.ii.a and 11.i.)

Sciences having no direct military relevance were despised too, especially if they were seen to imply a non-humane image of mankind. Even at the outset of the Cold War, social science in general, and behaviourism in particular, was already being lambasted in the popular press. For example, an editorial in *Life* magazine commented viciously on Burrhus Skinner’s utopian (or dystopian?) novel *Walden Two* (1948). Its author,

the cultural commentator John K. Jessup, also reviewed the book for *Fortune*, where he said:

If social scientists share Professor Skinner's values—and many of them do—they can change the nature of Western civilization more disastrously than the nuclear physicists and biochemists combined. (quoted in Skinner 1979: 369)

In the counter-culture's opinion, then, the sciences were deeply compromised (and technology, even more so). The arts, and religion, were seen as the saviours of civilization. For in this neo-Romantic version of the *verum factum* tradition (i.b above), the creative arts were both truly free and essentially intelligible. In particular, they were untouched by the contaminating fingers of the military–industrial complex.

That's a historical irony, since Cold War influences on the arts in the USA were in fact even deeper, if less costly, than those on science. But they were much less provocative, because apart from Senator Joseph McCarthy's early 1950s banning of “subversive” authors from government-funded libraries (and his committee's protests at a US drama group's staging “two productions of some Russian guy called Anton Chekhov” —Caute 2003: 62), they were deliberately—and for many years, successfully—hidden.

Or rather, they were hidden in the USA. In the Soviet Union, comparable pressures on artistic style and content were clearly explicit. The dramatist Konstantin Simonov declared that “in defense of Communism, most sacred of all things, all powers must be employed, including art” (Caute 2003: 108). This was official policy. Lenin himself had banned Russian modernism (e.g. Marc Chagall and Vasily Kandinsky) in the early 1920s, insisting on “revolutionary realism” instead; by mid-century, painters had to join the Union of Soviet Artists, whose rules prescribed adherence to socialist realism (Caute 2003: 510, 519). And in 1947 the Director of the Academy of Arts had announced: “In educating young artists we must make it absolutely clear that penetration of the walls of Soviet art schools by this or that decadent influence from the capitalist West is absolutely out of the question” (Caute 2003: 514). He meant modernist painting, from Paul Cézanne (and even the Impressionists) onwards—and especially Abstract Expressionism, which was making its mark in New York at that time.

The rapid rise of Abstract Expressionism in the 1950s, and the displacement of Paris by New York as the ‘hub’ of modern art, weren't due to an excess of creative genius in the studio lofts of SoHo. To be sure, Mark Rothko and Jackson Pollock painted as they did (and Roszak sculpted as he did), even before the Cold War started, for their own reasons: aesthetic, not political. (Similarly, scientists developed their theories for their own reasons: scientific, not political—although some of the *applications* were specifically military.) These artists were drawn to abstraction, and to individual expression, by pressures internal to art and their own psychology. But their huge public success wasn't due only to aesthetic values, nor even to enthusiastic art critics/collectors like the notorious Clement Greenberg. It was hugely encouraged by US government investment. For the politicians, it wasn't art for art's sake but art for America's sake.

The government's interest wasn't mere cultural imperialism, a wish to be top dog in the ateliers of the world. Admittedly, Americans would take pride in the fact that painters were increasingly looking to New York, not Paris, for inspiration. The art critic Robert Hughes says, “It would be foolish to claim that 1945–70 in New York rivaled 1870–1914 in Paris. America has never produced an artist to rival Picasso or Matisse,

or an art movement with the immense resonance of Cubism” (Hughes 1991: 3–4). Nevertheless, he remembers:

In the early 1960s, when I was a baby critic in Australia, it seemed that faraway New York had become a truly imperial culture, heir to Rome and Paris, setting the norms of discourse for the rest of the world’s art. . . . One saw this triumph from afar. . . . In Australia one’s response to it came out as a sigh—resignation to one’s own cultural irrelevance. (Hughes 1991: 3, 4)

Thirty years ago, Abstract Expressionism was pretty well a mandatory world style. We in Australia looked at it with awe. . . .

This act of unwonted humility was made by thousands of people concerned with the making, distribution, teaching, and judgment of art, not only in places like Australia but throughout Europe and—not incidentally—in America in the mid-1960s. They resigned themselves to an imperial situation. (pp. 5–6)

When Americans in the fifties and sixties eagerly claimed that their art had superseded that of Europe, their eagerness itself was a period phenomenon. . . . The idea that Europe was culturally exhausted was an important ingredient of American self-esteem. (p. 7)

However, there was more to it than that. Specifically, there was also a Cold War dimension. If Rothko and Pollock had painted tractors, or even meticulous still lifes, the government money wouldn’t have been forthcoming. For there were two unspoken political messages. On the one hand, there was a clear distinction and a choice: Abstract Expressionism was the aesthetic opposite of the Realist representational art favoured in—or allowed by—the Soviet Union. On the other hand, the American artists’ freedom to shock, to counter accepted artistic canons and to express their individuality in doing so, was a visible sign of the freedom generally available in the West.

But such messages, to be truly effective, had to be unspoken. If it was only the Soviet Union who directed their artists, individual thought in the West being no affair of government, then US government support had to be invisible. (A second reason was that modernist art wasn’t popular with the US taxpayer. An ex-CIA man later admitted that “It had to be covert because it would have been turned down if it had been put to a vote in a democracy”—Caute 2000: 550.)

Accordingly, much of the investment came via supposedly independent bodies, such as the Museum of Modern Art (MOMA) and the National Endowment for the Arts (NEA), founded in 1965 (Guilbaut 1983). These institutions had access to Rockefeller largesse. (Governor Nelson Rockefeller, who had been supporting modernist painting since the mid-1940s, had set the ball rolling in 1960 by initiating the New York State Council on the Arts, having failed to persuade Congress to sponsor a national body.) They not only bought the New York School’s canvases, but also sponsored Expressionist exhibitions across the USA and abroad.

Europe was the prime target, because of its proximity to (and sympathies for) communism. But non-aligned India and Japan were targeted too. For example, a huge exhibition of American art was sent to New Delhi in 1967, and a two-day seminar, led by Greenberg, was held in the hope of encouraging young Indian artists to study in New York rather than Paris—or, worse, Moscow (D. Guthrie, personal communication). The intention was to impress a political moral on an international audience, by contrasting the rigid social realism of Soviet painting with the liberated artistic expression of ‘the Free World’. (After the Cold War was declared ‘won’ in

the early 1990s, NEA funding for the visual arts plummeted accordingly: D. Guthrie, personal communication.)

Analogous activities went on in the literary and musical worlds and in film, theatre, and dance too (F. S. Saunders 1999; Cauter 2003, pts. II–IV). Large sums of money were silently channelled through supposedly neutral bodies such as the Rockefeller and Ford Foundations, and the Congress for Cultural Freedom—which was set up for these specific purposes. The National Endowment for the Humanities (NEH) supported work by writers whose left-leaning politics would have led to their being spurned in the McCarthyite period a decade earlier.

(Over the years, the NEA and NEH were hugely beneficial to the arts in America. A recent NEA chairman reported that, since 1965, the number of state-supported arts agencies had grown from 5 to 56, and local ones from 400 to 4,000; non-profit theatres had burgeoned from 56 to 340; symphony orchestras had grown from 980 to 1,800, opera companies from 27 to 113, and dance companies had multiplied eighteen times—Ivey 2000: 3.)

Where literature was concerned, spreading the message internationally was a problem. The NEH couldn't send exhibitions of writing around a multilingual world. Nor, as a "National" body, could they support foreign writers directly. So the CIA stepped in: the transatlantic literary journal *Encounter*, and several 'progressive' European magazines including Germany's *Der Monat* and France's *Preuves*, were funded in large part by the CIA. CIA money (\$750,000) was used also for the New York Metropolitan Opera's tour of Europe in 1956, and the Agency sponsored the Boston Symphony Orchestra's 1956 European visit.

What's relevant for our purposes here is that the covert political agenda in the arts wasn't fully uncovered until the mid-1980s. Its discovery caused outrage, among the people who had long been suspicious only of science.

To be sure, *Encounter's* political stance had already been questioned in the 1960s, largely because it "dealt gently, if at all, with topics such as race and Vietnam" (Collini 2004: 10). Conor Cruise O'Brien, who then had no inkling of the financial arrangements, commented in 1963 that the editorial policy seemed consistently designed to support the US government. As he said later, when the truth had come out: "the beauty of the operation... was that the writers of the first rank, who had no interest in serving the power structure, were induced to do so—unwittingly" (quoted in Collini 2004: 10). The magazine's funding-cover was conclusively blown in 1967, leading the English co-editor Stephen Spender, who'd been innocent of the CIA involvement, to resign (J. Sutherland 2004). But the *visual* arts remained relatively untouched by such rumours.

Only "relatively": in the early 1970s, Max Kozloff (1973) highlighted the political implications of the sudden success of the New York school, and the American muralist Eva Cockcroft (1974) had some even more pungent things to say—linking MOMA, the CIA, and the Rockefellers. But the cat was well and truly let out of the bag in the 1980s by the French art historian Serge Guilbaut, in his fascinating book *How New York Stole the Idea of Modern Art* (1983). The major scandal arose from his July 1984 article of the same title in the art magazine *Commentary*, and from back-up articles by others, including Cockcroft, in the August 1986 number (several are reprinted in Frascina 2000). Since

then, evidence of comparable—indeed, cooperative—cultural/intellectual meddling by Britain’s intelligence services has come to light (Dorril 2000).

Given that the counter-culturalists of the 1960s and 1970s didn’t realize that art was (unknowingly) acting in the service of capitalism, they couldn’t be enraged by this. Their political animus was reserved, rather, for *science*. And after all, it was the scientists and technologists, not the artists, who were more directly involved in the governmental war machine.

The history of cognitive science was coloured as a result. It wasn’t to be all hostility, however. As we’ll now see, late-century theoretical shifts within the field led the attitude of the counter-culture to change: from being firmly against cognitive science, to being (to some extent) in favour of it.

#### d. The counter-cultural somersault

Even an anti-science movement can have preferences *within* science. For example, the mentalistic aspect of the cognitive revolution was more attractive to the early counter-culturalists than behaviourism was. However, computers—given their place in military–industrial technology, and in the growth of global communication—were a special focus of suspicion. (We’ve already seen, for instance, that Lyotard’s key text was explicitly aimed at “knowledge in computerized societies”). In particular, mind-as-machine was anathema (Roszak 1986).

That’s an over-simplification. For instance, it doesn’t apply to one of the most influential members of the counter-culture, the biologist–artist Stewart Brand (1938– ). Brand’s first *Whole Earth Catalog* (1968), a compendium of “tools” for environmentally friendly living, inspired a host of back-to-nature projects worldwide. (The 1972 edition sold over one and a half million copies, and won a US National Book Award.) He had criticisms of technology aplenty, but he wasn’t an enemy of technology as such. Far from it. He’d helped Douglas Engelbart to demonstrate the first computer mouse at a now-famous meeting in San Francisco (see Chapter 10.i.h). And he shared Engelbart’s faith that personal computers, and IT in general, could help towards more ecologically viable lifestyles. Computer software was featured in the first *Catalog*, and increasingly in the following ones. So one of the early heroes of the counter-culture was also a computer enthusiast and computer visionary.

Moreover, a few brave artists in the mid-1960s were beginning to experiment with computer art (see Chapter 13.vi.c). Indeed, a major international exhibition (the first) was held in London in 1968 (Reichardt 1968). Many of the visitors were excited—though not the art critic of *The Guardian*, who described it as a “frivolous activity” (A. Sutcliffe, personal communication). But it was the relative *weakness* of the counter-culture in late 1960s Britain that had enabled the show to go ahead. One of the artists involved recalls that it “could at this time not have taken place in Paris. The revolutionary students would have swept it away” (Nake 2005: 59). The organizer herself has said: “The same venture in Paris would have needed police protection” (quoted in Klutsch 2005: 109). And a historian has commented:

Could it be that the ICA’s “happy accidents” flourished so well because they were staged in an atmosphere of breathtaking *naïveté*? Only a few lone voices seem to acknowledge the more serious and inevitably unhappy accidents that litter the history of cybernetics. . . . [In Great

Britain] the subversive momentum of 1968 never unfurled in the same way, with the same force, as it did in continental Europe or the United States. . . . Against this backdrop, [the London exhibition] offered a light-hearted view of the modern world without raising too many (if any) objections or stirring fears. (Usselman 2003: 391 ff.)

AI/cybernetics, by contrast, did stir fears—and was a prime target for criticism accordingly. From the late 1970s on, it suffered high-profile counter-cultural attacks by Dreyfus and Joseph Weizenbaum (11.ii). And cognitive science as a whole, which has AI at its intellectual core, was targeted too (Roszak 1986).

Chomsky's specifically political attacks on orthodox social scientists, and on their policy advice to the US government, were partly responsible: after all, he was highly respected *as a cognitive scientist*. But since the field had unfashionable—and apparently threatening—things to say about human beings and human nature, it would have been attacked by the counter-culture even without him. And, significantly, it wasn't only the intellectuals who disapproved of AI and cognitive science. These powerful cultural forces led US politicians to approve a temporary drop in military funding for AI, and a skewing towards civilian applications (see 11.i.b).

The component disciplines were each affected. So in the philosophy of mind and of psychology, neo-Kantianism (including the newly published Kuhn) soared in popularity in the UK and USA. Cognitive anthropology was nipped in the bud in the early 1970s, and all but destroyed by the 'literary' turn in the discipline (8.ii.b–c). At much the same time, many social psychologists reacted strongly against experimental, computational, and information-theoretic approaches (6.i.d). A feisty best-seller written by a well-known professional psychologist in Great Britain was called *The Cult of the Fact*, a title that says it all (Hudson 1972). One leading cognitive scientist was so worried by these professional developments that he wrote a book, and organized a high-visibility conference, to protest against them (see 6.i.d).

In short, cognitive science in its early years faced hostility from counter-cultural critics. In its later years, however, they *welcomed* certain intellectual changes that took place within the field.

The change of heart was initiated by the public availability of the personal computers that had been dreamt of by Brand and Engelbart, and by the development of the Internet, which allowed for what Haraway called "network identities" (cf. Chapter 13.v.d and vi.e). But this intellectual somersault was soon reinforced by specific theoretical aspects of late-century cognitive science.

For instance:

- \* Computational psychology highlighted situatedness, embodiment, and epigenesis (Chapters 7.iv.g, v.b–f, and 15.vii–viii).
- \* In the late 1970s, the AI researcher Terry Winograd left MIT spiritually as well as geographically to join Fernando Flores and Hubert Dreyfus in California (9.xi.b), and
- \* Minsky and Seymour Papert started drafting their theory of the decentralized "society" of mind (12.iii.d).
- \* Dennett, following Minsky, described the "self" not as a unitary thing but as a laboriously constructed—and not fully coherent—"narrative" schema that helps to guide our choices (7.i.e and 14.xi.b).

- \* In the 1980s Rodney Brooks and Randall Beer rejected GOFAI in favour of situated robotics (13.iii.b, 15.vii and viii.a).
- \* GOFAI researchers started studying “distributed cognition”, wherein coherent behaviour emerged from the action of many separate “agents”—with no central controller (13.iii.d).
- \* Similarly, PDP connectionism flourished, and seduced the public at large—and even some postmodernist philosophers (Globus 1992; Canfield 1993; E. A. Wilson 1998)—largely because of its decentralized approach (12.vi and x).
- \* Much the same happened when A-Life hit the scene, offering what some saw as a near-magical alternative to traditional cognitive science (15.x.a, and S. R. L. Clark 1995).
- \* The magic was seemingly underlined by the emergence of unexpected properties through computerized evolution (15.vi).
- \* Some anthropologists used the ideas of situated action and distributed cognition to analyse the behaviour of human groups (8.iii).
- \* Brian Cantwell Smith started work on a radically new, and admittedly highly eccentric, “participatory” account of the nature of computation as such (16.ix.e).
- \* The availability of personal computers encouraged a wide range of experiments in computer art, including interactive and/or evolutionary art (13.vi.c).
- \* And, by the end of the century, the technology of ‘virtual reality’ enabled people to experiment with the presentation, and perhaps even the construction, of *self* in very ‘non-Cartesian’ ways (13.vi. d–e).

These ideas fitted well with certain aspects of the late-century counter-culture. For example, postmodernists in literary and aesthetic circles—who’d already proclaimed “the death of the author” (Barthes 1977)—welcomed the implications of non-hierarchical control, and of user-directed computer technologies such as hypertext and interactive art (Chapters 10.i.h and 13.v.d and vi.c).

Alongside their undermining of the authorial signature had been their undermining of the unitary self. So, likewise, they had some sympathy with theoretical work that presented the mind and/or self as a virtual machine, consisting of many interacting agents as opposed to a unitary Cartesian centre (Chapters 7.i.e–f, 12.iii.d, and 16.iv.a–b). By the same token, they welcomed the playfulness in self-presentation that was made possible by the Internet (13.vi.e).

Connectionism was explicitly favoured by some postmodernist writers. Several claimed that Jacques Derrida’s notions of deconstruction and *différance* were largely homologous with PDP ideas, even suggesting that his approach was scientifically authorized by them (Globus 1992; Canfield 1993). The feminist philosopher Elizabeth Wilson rejected that last paradoxical claim but she, too, dressed connectionism in deconstructionist clothes (1998, esp. 14, 24–30, 196 ff.). She saw PDP as providing an opportunity “not merely to rethink cognition, but also to rethink our [i.e. counter-culturalists’] reflexive [self-aware, not automatic] critical [i.e. postmodernist] recoil from neurological theories of the psyche” (Wilson 1998: 14). Connectionism couldn’t solve the philosophical/political questions she and like-minded colleagues were

interested in—but, despite its provenance in (normally suspect) biology and cybernetics, it merited their attention:

Can we [i.e. postmodernists, and feminists in particular] think the subtlety of neurology and cognition on their own terms? Can we read the internal machinations of traditional empiricism in ways that do not return us to the routinized accusations [from the counter-culture, against biological science] of essentialism, reductionism, and political stasis? Specifically, does connectionism offer a political reading of psyche, cognition, and biology not despite its neurocomputational inclinations, but *because* of them? (Wilson 1998: 14)

Her answer was *Yes!* Connectionism, she argued (and she might have added “epigenesis”), offered feminists a way to accept scientific findings about the body without being trapped in a simple-minded biological determinism. On the contrary, it attributed “a fundamental mobility” to the mind/brain (p. 203).

A-Life, in particular, drew interest from some previously hostile sources. The feminist Sarah Kember approvingly described A-Life as “a discipline which developed precisely at the end of the cold war and which rejected the militarist top-down command and control and the masculinist instrumental principles of AI” (Kember 2003, p. vii). Explicitly contrasting cultural practices and self-images influenced by A-Life with “the previous race of cold-war cyborgs” (i.e. GOFAI-based models), she declared: “Posthuman identity, informed by the discourses of artificial life, centres symbolically on the humanization of HAL . . .” (p. 116)—where the emotionless HAL of *2001: A Space Odyssey* was said to exemplify the “failure” of the AI project. (Whether AI, or even GOFAI, has indeed failed is another question: see 13.vii.b.)

Kember also said that A-Life is “a cultural discourse [describing] posthuman life”, where

The posthuman is cyborgian in the sense of its enmeshment, at all levels of materiality and metaphor, with information, communication and biotechnologies and with other non-human actors. (Kember 2003, p. vii)

The pervasive “information” and “communication” she had in mind included telecommunications in general, but especially the Internet. And the “other non-human actors” ranged from semi-autonomous software agents (13.iii.d), through robot surgeons and automated mechanics (13.vi.b), to computerized companions (13.vi.d) and humans-as-avatars (13.vi.e).

But Kember wasn’t attacking this millennial form of man–machine identity. Unlike her predecessor Haraway, whose critique of GOFAI-based “cyborg” culture had *pre-dated* the rise of A-Life, she didn’t see the post-human cyborg as a largely destructive “product of cold war AI”. On the contrary, she saw the cultural legacy of situated AI and A-Life as liberating, in its stress on autonomy and emergent organization.

She even believed that it offered “the resolution of the science wars”, because it enabled one “to become independent of the distinction between nature and culture which forms [their] ‘epistem-onto-logical’ ground” (p. 216). On that view, the battles raging around the Legend were fated not to be won or lost, but to die away. (They haven’t died yet: see Blackburn 2005.)

Philosophically, the flight from centralization and the abstract was supported not only by feminism but also by the phenomenologists’ notion of “situated” intelligence

and “embodiment” (16.vii.a). However, it seeped into the intellectual air being breathed by people who’d never read a word of phenomenology—and who had scant sympathy for feminism.

Even US army generals were affected, despite their scorn for the explicit pronouncements of the counter-culture (P. N. Edwards 1996: 72, 111). They protested against the centralization brought about by computer-based approaches to military matters, as a result of which officers on the ground were, literally, losing control. Their strategic, and sometimes tactical, decisions were pre-empted by game-theoretic simulations, and their logistic decisions taken over by cost–benefit analyses devised by statisticians. Besides threatening their status and self-esteem, this Pentagon-driven trend favoured formal theory over hands-on application. That is, it substituted often unrealistic abstractions for discriminating responses to specific situations in the real world. (What’s more, official reports of what actually happened in the Vietnam war were hugely unreliable as a result: P. N. Edwards 1996: 137–40.)

As for the counter-culture’s attitude to computer technology, that somersaulted too. Roszak’s complaints about “distortive assumptions” had been explicitly aimed at Norbert Wiener, despite the cyberneticist’s attempts to humanize his approach (Wiener 1950). And with the advance of computing—and digital computers—in the late 1960s and 1970s, which depended crucially on military funding (11.i), counter-culturalists had become even more disaffected. By the early 1980s, however, these cultural doomsayers were being explicitly countered by people whose views vied with them for popularity.

Alvin Toffler, for instance, published a widely serialized best-seller, which achieved twelve printings and twelve translations within two years, that challenged “the chic pessimism that is so prevalent today” (1980: 2). Remarking that “Despair—salable and self-indulgent—has dominated the culture for a decade or more”, he argued at length that this attitude was “unwarranted”. The reason for optimism, he said, lay largely in the computer-based technologies, including new forms of communication, to which Roszak had been so hostile.

As we’ve seen, certain applications of personal computers were viewed by late 1980s counter-culturalists as philosophically liberating. The terminology of “emergence”, and/or “life”, made New Age souls—and journalists—even more receptive (15.x.a). Visual, performance, and installation artists, for instance, were quick to respond (Whitelaw 2004).

Granted, computers remained something of an embarrassment—and they still are. One art critic recently defended his approval of interactive art like this:

[These new aesthetic theories] propose personal and social growth through technically mediated, collaborative interaction. They can be interpreted as aesthetic models for reordering cultural values and recreating the world. As much as these theories depend on the same technologies that support global capitalism, they stand in stark contrast to the profit-motivated logic that increasingly transforms the complexion of social relations and cultural identity into a mirror-reflection of base economic principles. (E. A. Shanker 2003: 6)

In short, the computer hasn’t been all-conquering. Theories resting on computer technologies still arouse philosophical/sociopolitical suspicion in certain quarters.

So far, we’ve focused on how the counter-culture responded to cognitive science. But what about influences in the opposite direction?

The theoretical developments mentioned above weren't primarily caused by the sociopolitical values of the counter-culture. Indeed, the intellectual seeds had been sown long before its rise, and nurtured out of public view for twenty years (4.viii, 12.i–ii and iv–v, and 15.iii–vi). However, some of the relevant concepts may have occurred to the scientists concerned partly as a result of it. For new ideas are often hugely overdetermined, in the sense that *many* influences and associations play a part in their generation (Boden 1990a: 186–98, 244–8).

For example, consider what Stephen Toulmin has called the postmodernist “revaluation of the concrete”, and the closely associated distaste for top-down hierarchical control—even in hard-headed business management (Toulmin 1999, ch. 5). This was reflected in a number of ways within cognitive science:

- \* by connectionist work on distributed control and bottom-up emergence in anthropology, AI, and psychology (8.iii and 12);
- \* by GOFAI work on ‘agents’ (13.iii.d–e), interactive interfaces (13.v–vi), and even AI programming languages (10.vi.b);
- \* by A-Life and situated robotics (13.iii.b–d, 15.vii and viii.a–b);
- \* and by some aspects of philosophy (12.x).
- \* It was explicitly endorsed by Papert, in relation to various late-century trends in psychology (Turkle and Papert 1990).
- \* Arguably, it was even indicated by the new willingness of laboratory neurophysiologists to hobnob with practising physicians (see 14.i.a).

Irrespective of whether these late-century scientific ideas were part-caused by sociopolitical influences, such influences helped determine whether they met a receptive audience. They were more readily accepted, even *within* the field, because of this particular *Zeitgeist*.

In intellectual history in general, remarks of that type are common. Indeed, the pioneering, and punctilious, historian of psychology Edwin Boring (1957) often cited the *Zeitgeist* in his work. On the very first page of his *magnum opus*, he said:

Discovery and its acceptance are [limited] by the habits of thought that pertain to the culture of any region and period, that is to say, by the *Zeitgeist*: an idea too strange or preposterous to be thought in one period of western civilization may be readily accepted as true only a century or two later. Slow change is the rule—at least for the basic ideas. On the other hand, the more superficial fashions as to what is important, what is worth doing and talking about, change much more rapidly . . . (Boring 1957: 3)

Later chapters featured the *Zeitgeist* too (his index lists a dozen entries, some pointing to several pages).

Boring is sometimes mocked as a result, by critics who feel that he wheeled in this Hegelian notion almost as an extra character in his script. (For instance, he spoke of individual scientists becoming “the means by which the *Zeitgeist* prevails”: p. 23.) Perhaps he was tempted by the animistic flavour of the term, which literally means “the spirit of the times”, concepts/assumptions that inform virtually all aspects of a given culture. But whichever word one chooses to use, the point is that—as a pervasive intellectual background, informing virtually all areas of life—the *Zeitgeist* is a real phenomenon. In the second half of the twentieth century, then, counter-cultural ideas and values had a significant effect on the history of cognitive science.

In sum, although my narrative of cognitive science is predominantly ‘internalist’, dealing with the data and theoretical ideas as such, it’s partly ‘externalist’ too. We’ll see repeatedly that social and personal factors played a role within the scientific community (and a fortiori in the public media: P. N. Edwards 1996). Sometimes, these influences were explicitly acknowledged by the scientists concerned (e.g. Resnick 1994: 6–19). More often, they were implicit—betrayed by the choice of theoretical terminology and/or illustrative metaphor (e.g. AI workers’ descriptions of heterarchical programming: see 10.iv.a). But in either case, they were there. Since the Legend—though highly attractive to many scientists—is false, that’s only to be expected.

### e. Hardly hero worship

The fourth warning concerns the Romantic myth of the creative prodigy. The thinkers I’ve chosen to discuss weren’t intellectual heroes solely responsible for the ideas I attribute to them. Sometimes, it’s even doubtful whether they *could* have had the idea independently. Herbert Simon, for instance, has said as much about the three-man origin of what’s often (wrongly: 10.i.b) called the first AI program:

We [three] were in closest communication during the whole period, through long association had developed an extraordinary capacity to communicate even our subtleties to each other, and *the whole product must be regarded as joint and inseparable. I am firmly convinced that none of us alone had much chance of accomplishing [it].* (quoted in McCorduck 1979: 139; italics added)

Quite apart from other *individuals*, the writers discussed in this narrative were influenced and/or encouraged by the social context at the time. Indeed, one historian of science has said that the “great man” view of history might well be replaced by a “great opportunities” view, “with the emphasis on the socially given possibilities rather than on the people who exploit them” (Fleck 1982: 217). In AI, for instance, socially driven changes in funding policies have offered encouragement and discouragement alike (Chapters 11.i and v.b, and 12.iii.e and vii.b).

Occasionally, of course, people’s self-serving rhetoric suggests the contrary. I’ve not come across any example so extreme as James Watson’s *The Double Helix*, whose unfairness to several other DNA pioneers, including Rosalind Franklin, led Harvard University Press to drop its plans to publish it (Maddox 2002: 312). But a few famous names in cognitive science are guilty to a lesser degree (see 14.v.d). Certainly, some individuals were exceptionally fertile thinkers. Turing, McCulloch, von Neumann, Marr, and Simon are examples—which is why each of them features at length in *several* chapters. But even geniuses aren’t lone geniuses.

An important idea is rarely, if ever, due to only one mind. Indeed, it often arises near-simultaneously in several. Given that creativity involves either novel associations of familiar (and shared) ideas, or the exploration and transformation of structured conceptual spaces acquired from one’s culture, this is only to be expected (Boden 1990a, 1994b).

It’s made even more likely by the fact that science has been an increasingly communal enterprise since the mid-seventeenth century (2.ii.b–c). Today, that’s often flagged by multi-authorship: most of the scientific papers cited in the References have more than one author (one lists twenty-five). But even the most prolific credit listings are likely to

omit the names of people who were, in fact, part of the communicative network that made the discovery possible.

One of the most important techniques in connectionist AI, for instance—namely, backpropagation—was independently discovered by at least four people, and prefigured by several others (12.vi.d). And the group who actually got the credit for it were just that: a *group*, who collaborated for several years to improve one member's initial idea before publishing it.

Sometimes, this has bizarre, even comic, consequences. The Nobel Prize committee, when considering Ivan Pavlov in 1903, were much discomfited by his frequent declarations in his *Lectures on the Work of the Main Digestive Glands* (1897/1902) that his discovery of the conditioned reflex was “the deed of the entire laboratory”. It's clear from the discussions recorded in the Nobel archives (Todes 2001) that Pavlov's modesty nearly prevented his winning the prize.

But perhaps “modesty” is the wrong word here? For Pavlov was (rightly) proud of his role as a pioneering—and visionary—laboratory manager. The more that scientific research depends on complex equipment and/or multidisciplinary cooperation, the more important this role becomes. Indeed, J. Robert Oppenheimer became world-famous as the scientific manager of the Manhattan Project, not as its leading researcher. Of course, he had enough scientific knowledge and imagination to have a good nose for new ideas—even (when heading the Princeton Institute after the war) some in the nascent cognitive psychology (Bruner 1983: 96, 121). So did Sydney Brenner. When Brenner was director of the Molecular Biology Unit in Cambridge, he offered an unused cubbyhole to the young Marr—who had scant interest in molecular biology, but was exploring highly abstract ideas about the brain (see 14.v.b).

In that particular case, there's no ambiguity: Brenner provided the space, not the intellectual content. In general, however, assigning individual responsibility for creative ideas isn't straightforward. And the more that people are interacting, the more this is true.

(Discovery and invention are in much the same boat, so far as group influences are concerned. Seymour Cray, the charismatic inventor of the supercomputer, couldn't have designed ‘his’ machine without a rich network of technical and social relationships—even including his child's willingness to accept help with her algebra from invisible “elves”, working their magic long after her bedtime: MacKenzie and Elzen 1991.)

This fact lies behind the many horror stories about supervisors stealing their research students' ideas. No names, no pack-drill—except to pay tribute to the psychologist Edward Tolman. He opened his major book by thanking his research students not merely for doing most of the experimental work, but for having many of the theoretical ideas—“which ideas I have often no doubt quite shamelessly appropriated as if they were my own” (Tolman 1932, preface). To be sure, they might not have had them without Tolman's prompting. And they might not have been able to develop them as coherently as he did. Nevertheless, the ideas recounted in Tolman's book weren't just *Tolman's*. His own view was that “If it had not been for those students, this book could not have been written.”

Finally, people may be unable to recall just who said what to whom—and even just who wrote what down. For instance, George Mandler was named in two different

reports of a 1959 conference on language learning as the author of a provocative “manifesto” attacking both behaviourist and early-cognitive approaches (Mandler 2002a: 348, 350). (Specifically, it criticized the “glib invocation of ‘schemas,’ structures,’ and ‘organization’” and the “mere postulation” of new mechanisms and processes.) Mandler now—over forty years later—admits to being *one of three* initiators, and “probably” one of three authors. However, another member of the trio says that he himself didn’t take part in the actual writing (which took place one evening), even though he’d contributed some of the ideas.

This isn’t a case, for once, of an unpleasant priority dispute: the person who was given the public credit is loath to take it undeserved. Rather, the people concerned simply can’t remember.

And what if they could? Even then, their memories couldn’t necessarily be taken at face value. For as Mandler points out (2002a: 348), memory is construction, not recall (see Chapter 5.ii.b). If—as is usual—they were motivated to *claim* priority, instead of shrugging it off, the constructive process might have been biased accordingly. The same applies, of course, to all the other personal recollections quoted in this history.

## f. Discovering discoveries

The next three warnings concern judgements of originality—of ideas, as well as people. The attribution of an original idea to one person or group rather than another, and even its recognition as important, or as a ‘discovery’, are complex social processes (Brannigan 1981). These judgements aren’t subject to cut and dried criteria. They involve social negotiation and rivalry, as well as historical enquiry and theoretical argument. Let’s consider *discovery* first.

*Discovery* is a highly loaded term (Sturm and Gigerenzer 2006). An idea deemed by some people as a discovery may not be so regarded by others—even by those whom one might expect to appreciate it.

A famous example is Charles Darwin’s paper on natural selection, which—alongside one by Alfred Wallace on the same topic—was read at the Linnaean Society on 1 July 1858. (Whether this counts as ‘simultaneous’ discovery is doubtful: Darwin had been working on this idea for many years before Wallace came up with it; however, both had been inspired by Thomas Malthus’s *Essay on the Principles of Population*.) This first public presentation of the theory of evolution failed to impress the Linnaean’s president, Thomas Bell. His Presidential Address some ten months later declared:

The year which has passed has not, indeed, been marked by any of those striking discoveries which at once revolutionize, so to speak, the department of science on which they bear.

Only those little words “at once” can save Bell, today, from ridicule. As Janet Browne (2002: 42) has remarked, his verdict, though “accurate enough in the short term”, has become known as “one of the most unfortunate misjudgments in the history of science”.

Some cognitive scientists—though by no means all—would say much the same of Minsky and Papert’s unrelenting dismissal of connectionist AI (Chapter 12.iii). And, almost exactly 100 years after Bell’s misjudgement in London, a similar blindness occurred at MIT (C. G. Gross 2002: 85). In 1959 Jerome Lettvin, the lead author of what

would become perhaps the most famous paper in neuroscience (see 14.iii.a), wasn't invited to a two-week local seminar closely related to his (MIT-based) research. In the event, a visitor from England (Horace Barlow) arranged an invitation for him, and a second—much less famous—paper by Lettvin's team was tagged on at the end of the official Proceedings (Lettvin *et al.* 1961).

Sometimes, someone makes a discovery they don't count as a discovery—not because they don't realize its interest (although this happens too), but because they can't explain it. Lettvin still hasn't published a remarkable finding of the late 1950s, concerning a discrimination made by the frog's retina which seems to be far too complex for a retina to make (see 14.iv.a). Because it's a mystery in theoretical terms, Lettvin has never reported it officially. In his eyes, then, he's noticed something but discovered nothing.

In other cases, an idea is used to great effect in a specific context, but without anyone's recognizing its wider implications. One example is the Watt governor (Chapter 4.v.a). This was copied in countless machines of the steam age, but it was seen as a mechanical trick not a theoretical principle. Its importance as an example of a general type of control mechanism wasn't recognized for ninety years—and even so, it wasn't fully appreciated for another seventy. Indeed, James Watt wasn't the first to use such a mechanism: it's been found in some ancient Greek automata and fourteenth-century clocks. One might say, then, that feedback was *invented* long before it was *discovered*.

Some judgements about what counts as a discovery are grounded in explicitly heroic assumptions. So people may say: “If so-and-so thought of it, it must be important”, or “If so-and-so was involved, he must have been the leader” (very rarely “she”, of course). Sometimes they don't even realize that they're doing this: the name of the Nobel prizewinner Simon was often wrongly put first in references to papers co-authored by his younger colleague Newell (see Chapter 6.iii.b).

By contrast, some individuals are treated as anti-heroes. One example is the French mathematician Louis de Branges, recently described by a reviewer as personally “cranky”, but “not a crank” (Sabbagh 2004). His latest work, in which he claims to have solved a famous mathematical problem (the Riemann hypothesis), is being systematically ignored by his peers, and science journalists are being told not to bother with it—by people who themselves haven't actually read it. (Apparently, it is fiendishly difficult even for professional mathematicians, and would require “a team” of experts working for at least six months.) De Branges is known to have done fine work in the past, but he's very unpopular (there's some suggestion that he may have Asperger's syndrome: see Chapter 7.vi.f). In short:

It may be that a possible solution of one of the most important problems in mathematics is never investigated because no one likes the solution's author . . . The entire mathematical profession is turning its back on what could be the most important development in the last hundred years of mathematics. (Sabbagh 2004)

I can't think of such an extreme case in cognitive science. But it's certainly true that personal animosities often hinder, and sometimes even prevent, proper consideration of new ideas. For example, AI people in the late 1960s failed to take proper account of Dreyfus's criticisms, because they resented his savage attack on them—not to mention his technical ignorance (11.ii.b).

Such heroic/anti-heroic assumptions are often buttressed by social snobberies of various kinds. These include the superior trust accorded to the word of a “gentleman” (Shapin 1994), which was crucial to the emergence of scientific communities in the seventeenth century (Chapter 2.ii.b).

On the negative side, they include suspicion of uneducated ‘country bumpkins’ (such as the champion of the neurone theory, Santiago Ramón y Cajal: 2.viii.c), and systematic undervaluing of the contribution of technicians (Shapin 1989; Schaffer 1994). For example, Richard Gregory’s influential work on visual illusions (6.ii.d) owed much to the ingenuity of his technician Stephen Salter (later, a professor of engineering who invented an ingenious way of harnessing wave power). Gregory is generous enough to acknowledge this, in print as well as conversation. Many others wouldn’t.

Group loyalties enter the picture, too. Judgements of originality and/or value can be strongly influenced, for instance, by chauvinistic nationalism. In his inaugural address in January 1996, President Clinton called the electronic computer an American invention—to the chagrin of compatriots of Turing, Max Newman, Thomas Flowers, Frederick Williams, and Maurice Wilkes (3.v.b–d).

Such judgements can be skewed also by the ‘not-invented-here’ syndrome. This systematically distorts the reminiscences and the bibliographies of workers in at least one leading AI laboratory, and at least one leading department of linguistics (Chapter 9.viii.a and ix.a).

It even prevented proper recognition being given to the only working AI program to be presented in the final days of the famous Dartmouth Summer School in 1956 (Chapters 6.iv.b and 10.i.b). Although it electrified a few people present there, the program wasn’t taken up as the paradigm for AI. Far from it. One of the originators could afford to be ‘philosophical’ about this years afterwards, but it had rankled at the time:

[The new field of AI] was going off into different directions. They [i.e. Minsky and John McCarthy] didn’t want to hear from us, and we sure didn’t want to hear from them: we had something to *show* them! . . . In a way, it was ironic because we already had done the first example of what they were after; and second, they didn’t pay much attention to it. But that’s not unusual. The “Not Invented Here” sign is up almost everywhere, you know. (H. A. Simon, interview in Crevier 1993: 49)

As for *which* groups are involved in evaluating a discovery, that can rest largely on chance. Frank Rosenblatt’s work on “perceptrons” became very widely known, very quickly, because he’d been communicating with physicists as well as psychologists (12.ii.e). This was more accidental than intellectual. He’d been working alongside physicists (in his university’s Aeronautical Laboratory) because he’d been borrowing their computer: the psychologists didn’t have one.

An idea may be hailed as a discovery largely because the context in which it’s put forward makes it highly significant. For example, Goethe is commonly credited with having discovered a particular similarity between the bones of the rabbit’s jaw and ours. In fact, someone else had noticed it before him. Goethe, however, placed this anatomical fact in the context of an ambitious philosophy of the unity of nature. The rabbit’s jawbone was of interest because he related *all* vertebrate skulls to a single archetype, and because he saw morphological archetypes in the structure of flowers as well as skulls (2.vi.e).

### g. So what's new?

To view something as a discovery is to see it both as *valuable* and as *new*. The Goethe example (above) shows that an idea may be more or less valued depending on its intellectual context. Judgement, even negotiation, must enter in. But what about the newness? One might think that novelty is a more cut-and-dried matter. In fact, however, it's a minefield for the unwary.

There are three very different reasons for this. One is obvious, even boring: namely, unavoidable ignorance. No one can be *sure* of knowing about every previous thought that's relevant to a given topic, nor even of having read everything that's been published about it (especially if some of the texts exist only in a language one doesn't read).

Moreover, publication is sometimes long delayed, and/or long unread, so that later work is mistakenly believed to have been the pioneer. One example is Paul Werbos's algorithm defining what was later called "backprop" (12.iv.d). Not only did this lie buried for many years in a computer manual (unread by psychologists), but wider publication was at first deliberately suppressed by a governmental committee, because other results in the same report were politically embarrassing. It was only later that Werbos's priority could be recognized. Another example is Konrad Zuse's work on computers, which remained unpublished for many years—and wasn't translated from the German for several years after that (3.v.a, 10.v.f). In short, judgements of historical novelty can only be provisional. (Maybe they should carry a government health warning?)

The second difficulty in assigning priority is more interesting: no creative idea is entirely new. It's either a novel combination of familiar ideas, or an exploration or transformation of a culturally accepted style of thought (Boden 1990a, chs. 3–5). Even in the latter case, where the new idea can be so surprising as to seem *impossible*, there will be some intelligible relation with the previous way of thinking. It may take a long time, and considerable persuasion, for someone/everyone to recognize this in a particular instance (such as the holographic theory of memory, or the travelling-depolarization theory of nervous conduction: 12.v.c and 2.viii.e). But the point stands, nevertheless. It follows that there are differing *degrees* and varying *types* of novelty, never absolute newness.

It's the third reason which is especially relevant here, however. Namely, people are human—often, all too human. Priority claims can be grounded in various kinds of moral frailty: from deliberate deception, through uncritical self-deception, to lazy (i.e. avoidable) ignorance. Instances of each of these can be found in cognitive science.

For a start, there's the unfortunate habit of representing other people's work as one's own. Occasionally, this boils down to outright *theft*. Two examples, both utterly trivial in the grand scheme of things but pretty annoying to me: Many years ago, I was sent a book for review that was 75 per cent lifted, largely word-for-word, from parts of one of my own—cited only twice. (I felt unable to point this out, and merely said that the book was "highly derivative"; but another reviewer, Donald Michie, did so—thanks, Donald!) And recently, while googling on the Web, I found some "Detailed Comments" on heterarchy (10.iv.a) comprising pp. 125–42 of the same book—which wasn't even mentioned. (On regoogling in April 2005, I could no longer find this entry; it must have been removed.) The limit of this person's originality was to make a few cosmetic changes, including one—adding the words "I'm afraid" to a critical remark

of mine—specifically intended to strengthen the impression that he was the author of the stolen comments.

I don't know of any non-trivial examples quite as shameless as this, although I've been told (in confidence) of several that come very close. But it's pertinent to note that the pseudonymous Father Hacker includes the theft—and judicious renaming—of old ideas as a key item of advice in his spoof 'Guide for the Young AI-Researcher' (*AISB Quarterly* 2003; see Chapter 13.vii.a).

More common than such shameless theft is the deliberate suggestion that one's new ideas are more original, and/or more independent, than they actually are.

An exceptionally dishonourable example was directed against a cognitive scientist—let's call him Dr X—who worked in the hard sciences before turning to matters of the mind. During that period, he'd had idea A about topic B. When he was preparing a report for publication, his colleague Dr Y—who was thinking along the same lines, on idea A'—persuaded him to delay publication until his own thought was more advanced, so that they could co-author an even stronger paper. Dr X agreed, and waited for Dr Y to contact him. In vain: soon afterwards, Dr Y published alone—and later received a Nobel Prize for this work.

Within cognitive science as such, there are a number of cases of people exaggerating, even deliberately misrepresenting, their own originality. Sometimes, the earlier author isn't even mentioned—in which case only evidence from informal conversations (e.g. between the two individuals involved) can support the charge of deliberate plagiarism. Since this evidence is (a) unpublished and (b) often given in confidence, I haven't specified any such charges in my story: but I'm sorry to say that I could have done.

More commonly, the original author *is* mentioned—but in a misleading way. As one colleague remarked to me: "The art of plagiarism is to make a marginal and inaccurate citation of earlier work, but to give the impression that it is new and yours." For example, Marr (who has been accused several times of exaggerating his priority: 14.v.d) used this ploy. His early work on stereopsis has been described as "a minor variation" of a model published by someone else, who was cited dismissively by Marr in a low-profile footnote (J. A. Anderson and Rosenfeld 1998: 231–2). "Low-profile" not just because it was a footnote, but because it came almost at the end of a long list of footnotes. A generous acknowledgement this was not.

A more general way of exaggerating the novelty of one's own work is to describe it not as an exciting new development within an *existing* field, but as something quite different. Many pioneers mentioned in later chapters were guilty of this, in that they ignored/denied the close links—historical, sociological, methodological, and even philosophical—between their work and earlier forms of cognitive science.

Consider, for instance, a 'historical' paper by the situated roboticist Rodney Brooks (1991*b*). In the opening paragraphs he mentioned "traditional" AI, implying (correctly) that his new approach was part of the AI enterprise as a whole. In the main body of the text, however, he repeatedly used the term "Artificial Intelligence" to mean only symbolic AI—implying that his approach was even more new, even more revolutionary, than it was (cf. 13.iii.b). Other examples of this self-glorifying rhetorical strategy include the frequent attempts to distance connectionism from AI as a whole (Boden 1991), and to exclude A-Life from "cognitive science" because it doesn't share the Cartesian assumptions of the traditionalists (15.viii.b–c).

Even more common than deliberately overplaying one's originality is the *ignorant* suggestion that one's ideas are new. George Miller was bitterly accused by Newell and Simon of not giving them proper recognition as the originators of many ideas in *Plans and the Structure of Behavior* (Chapter 6.iv.c). However, as Miller said later, "they were old familiar ideas; the fact that they had thought of it for themselves didn't mean that nobody ever thought of it before" (1986: 213). In the event, he redrafted the text and added zillions of footnotes "so they would no longer claim that those were their ideas". They had some excuse, since they hadn't been trained as psychologists. Nevertheless, and despite Simon's knowledge of Gestalt research, they were evidencing the shameful ignorance of past work that pervaded American psychology from the 1920s to early 1960s (see Chapter 5.i.b).

A more excusable example concerns logic programming (Chapter 10.v.f). Its first occurrence (so far as is known) was due to Zuse, in 1946. But his ideas weren't taken up by the early computer scientists. Moreover, the description of his Plankalkul language wasn't published until 1972 (nor translated from the German until four years later). So "the early computer scientists", who were based in England and the USA (3.v.b–e), didn't know about them. The people who "pioneered" logic programming in the early 1960s may therefore be forgiven for believing that they were the very first to think along these lines.

Some cases of sincere-but-mistaken claims to originality rest not on ignorance so much as on *failure to recognize* the similarities between one's own ideas and others'. This failure, in retrospect, can be very surprising. Thus Simon recalled in his autobiography that he and his father—an early servo-engineer who'd designed gun-turret controls for battleships in the First World War—didn't appreciate their shared interests until it was almost too late:

It wasn't until [the end of the Second World War] that I realized that his whole life had been spent in what you might call protocybernetic work, and that it was just a direct ancestor to this whole business. And until the last year or two of his life—he died in 1948—we never had a conversation about this. *He used to tell me about his work, but that was about his work, and I used to tell him about what I was doing, but that was about what I was doing*, and I don't think the thought crossed either of our minds, certainly not until about 1947 or 1948, that these had any relation to each other. And I don't really understand that now. (Simon, quoted in McCorduck 1979: 130; italics added)

Perhaps the explanation was a form of what the Gestalt psychologists had called "functional fixedness" (5.ii.b). Much as someone may be unable to tie two far-separated strings together because they see a nearby pair of pliers only as pliers, not as *a weighty object suitable for use as a pendulum bob*, so Simon (and his father) apparently thought in terms of "his work" and "my work"—labels which prevented them from noticing the conceptual links. Someone with an emotional and/or professional investment in classifying an idea as "my discovery" may be especially likely to fall victim to this type of blindness.

Failing to recognize the similarity can depend on failing to be consciously aware of the other person's idea. Paul McCartney was accused (in July 2003) of having drawn the superb melody, and some of the words, for *Yesterday* from a song released when he was only 11 years old. (The older version was played on BBC Radio, and the likeness

was striking.) McCartney had already reported that he'd been so surprised at waking up with the tune (but no lyrics) "all there" that he'd thought at the time that something like this might have happened. He'd even asked his friends and fellow Beatles if they'd ever heard the tune before, but no one had.

This case of unrecognized memory wasn't identified as such until some fifty years after the original memory was formed. So priority claims, however sincere, can only be provisional. For instance, the neuroscientist Michael Arbib sincerely believed that he'd thought up the name *Rana computatrix* for his computerized frog all by himself, but realized years later that he'd been inspired by a similar name used by someone else (see 14.vii.c). Ideas about names for frogs, computerized or not, are of course relatively trivial. But who can know how many analogous cases concern ideas much more significant than that?

Sometimes, people deliberately suggest that they've done relevant previous work when in fact they haven't. The Nobel prizewinner David Hubel has confessed to a deception of this kind:

[Torstein Wiesel and I had gone to a lecture by Vernon Mountcastle in the late 1950s] in which he had amazed us by reporting on the results of recording from some 900 somatosensory cells, for those days an astronomic number. We knew we could never catch up, *so we catapulted ourselves to respectability* by calling our first cell No. 3000 and numbering subsequent ones from there. When Vernon visited our circus tent we were in the middle of a three-unit recording, cells number 3007, 3008 and 3009. We made sure that we mentioned their identification numbers. (Hubel 1982: 516; italics added)

This was just a youthful prank, of course. But less innocent examples sometimes appear in print. They ride on the fact mentioned above, that by and large scientists trust each other's empirical reports even if they disagree on their theoretical interpretation. ("By and large", because in important cases the reported experiments must be replicated by someone else.)

Another prank, confessed years after it happened, was meant to suggest not just originality but effortless superiority:

Alan Kay, Marvin Minsky, and I got together and did some back-of-the-envelope calculations—we actually killed about five minutes to find an envelope so we could later say we did back-of-the-envelope calculations—on how much knowledge would be required [for an AI system embodying common sense: see 10.viii.c] and how much time it would take. That was a million frames over ten years. (D. B. Lenat, interviewed in Shasha and Lazere 1995: 233)

Pretty harmless, admittedly . . . but some attempts to influence people's judgement by providing misleading information are more blameworthy than this one.

Even if someone honestly cites inspiration by, or similarity to, a previous worker, it's often not straightforward to decide who "discovered" the idea—because it's not clear *what*, exactly, the relevant "new" idea is.

The development of hypertext is an illustration (Chapters 10.i.h and 13.v). Individuals commonly cited as crucial originators include Vannevar Bush, Douglas Engelbart, Joseph Licklider, Theodor (Ted) Nelson, and Alan Kay. And Licklider, a leading influence on library science, was well aware that librarians had identified some of the core problematic issues long before—if not quite as early as Gabriel Naudé (see

Preface). To understand the history of hypertext, then, one must distinguish carefully between statements of:

- \* a general conceptual framework;
- \* abstract organizational principles;
- \* desirable/conceivable end-results;
- \* in-principle-possible technical methods;
- \* feasible computational techniques;
- \* commercially efficient designs;
- \* and specific improvements of pre-existing technologies.

Without doing this, one can't say just what was "new" about an individual's (or a research group's) contribution. The same applies, of course, to many other discoveries besides hypertext.

## h. Rhetoric and publication

The seventh warning concerns not what ideas people come up with, but how they present them. For the history of science is not only about what people thought, but also about what they were *thought* to have thought. And that, in turn, depends on how—and where—they chose to express them.

The identification of discoveries (both as *new* and as *valuable*) can depend heavily on the discoverer's rhetorical skills, or lack of them. We'll see in Chapter 9.vii.b, for instance, that Chomsky's review of Skinner's *Verbal Behavior* became famous largely because of its sparkling, stinging wit. By contrast, several classic papers in cognitive science—even including *the* classic paper (Chapter 4.iii.f)—were all but ignored on first publication, because of their notational and/or mathematical difficulty.

Arguably, one case in point is the early work of the highly creative cognitive scientist Stephen Grossberg. He was perhaps the first to formulate three ideas that are influential today under the names of other people: Hopfield nets, the Marr–Albus model of the cerebellum, and Kohonen self-organizing maps.

(These things are tricky. Although Kohonen 1982 didn't cite Grossberg directly, he did cite four papers by Christoph von der Malsburg. The earliest of these had clearly acknowledged Grossberg as its prime inspiration: see Chapter 14.vi.b. But without following that particular paper-trail, readers might be unaware of the kinship between Kohonen's ideas and Grossberg's. This is, of course, a general point: the fact that Bloggs didn't cite Squoggins, or even hadn't read Squoggins, *doesn't* mean that he wasn't influenced by Squoggins.)

Grossberg also pioneered many more notions—including back propagation—that are commonly attributed to others, if not actually named after them. As he puts it:

This has, all too often, been the story of my life. It's tragic really, and it's almost broken my heart several times. (J. A. Anderson and Rosenfeld 1998: 179)

Trying to live with so many false [priority] claims has been difficult for me at times. If I try to get credit where it is due, then people who want the credit for themselves often mount a disinformation campaign in which they claim that all that I think about is priority. Because I have been a very productive pioneer, who innovated quite a few ideas and models, that can create quite a chorus of disinformation! If I don't try to get credit for my discoveries, then

I am left with the feeling that eventually most of my ideas may become attributed to other people . . . (pp. 186–7)

Grossberg’s own diagnosis of his “tragic” life history identifies three causes:

The problem is that, [1] although I would often have an idea first, I usually had it too far ahead of its time. Or [2] *I would develop it too mathematically for most readers*. Most of all, [3] I’ve had too many ideas for me to be identified with all of them . . . [From the mid-1960s on] many things that I discovered started getting named after other people. (J. A. Anderson and Rosenfeld 1998: 179; italics added)

The key sense in which his work was “too far ahead of its time” was the unfamiliarity of the mathematics. He was talking about brain and behaviour as a complex dynamical system (as he put it: “nonlinear, nonlocal, and nonstationary” — 1980: 351), a theoretical approach that didn’t become popular in cognitive science until the late 1980s. But it’s the second point that’s of special interest here: the rhetorical style.

His early work was largely unintelligible even to the few psychologists who took the trouble to read it (see Chapters 12.v.g and 14.vi.a). He combined intellectually demanding (and unfamiliar) mathematics with a host of interdisciplinary details, most of which would be unfamiliar to any individual reader. They were there because he was trying to show the unsuspected theoretical *unity* behind hugely diverse data. His writing was unusually voluminous too: 500 pages for his first-year graduate report (1964), and many long and richly cross-referenced journal articles. Faced with this challenge from a youngster they’d never heard of, most people gave up before reaching the end, if they could summon up the courage to start reading at all.

Scientifically speaking, the weakness was theirs—not his. It’s now taken for granted (which it wasn’t then) that the problems he was discussing demand a fair degree of mathematical literacy on the researcher’s part. Specifically, they require non-linear mathematics. (Broadly speaking, this is a type of mathematics in which a very small change can bring about a much larger, and global, change.) In general, indeed, one can’t expect cutting-edge ideas to be easy to understand. Humans being human, however, new scientific claims have to compete for attention. In a cynical mood, one might even say that they have to compete *in the marketplace* for attention. So any rhetorical obstacles put in readers’ way will tend to obscure the science, however worthy it may be. Grossberg wasn’t easy to read.

By contrast, the group who (wrongly) got the credit for discovering back propagation went to exceptional lengths to describe their work in an intelligible way (12.vi.a). If they didn’t emulate the sparkle of Chomsky’s famous review, they did achieve clarity. This doesn’t count as a *scientific* achievement. But, for good or ill, it does—and did—make a difference in bringing scientific ideas into historical prominence.

Sometimes, the rhetorical obstacles are so great that a primary author gains recognition largely through the efforts of a secondary one. For instance, Richard Montague’s theory of semantics was nigh opaque to most cognitive scientists, even including linguists accustomed to dealing with formalism. Without Barbara Partee’s (1975) more accessible introduction, his ideas might have been virtually ignored except by his fellow logicians. In fact, they became hugely influential in theoretical linguistics—and underlay the most important challenge to Chomsky (9.ix.c–d).

If one thinks of rhetoric as *presentational* skill, verbal or otherwise, one can even find examples in robotics. We'll see in Chapter 4.vii.b, for instance, that whereas William Grey Walter's mobile "tortoises" caused a sensation in the early 1950s, his—even more interesting—learning machine did not. The problem, apparently, was that this was built as an ugly metal box sitting boringly on a bench. Even though it was sometimes connected (wired) to a mobile robot, its significance wasn't recognized at the time. Another example is the robot arm built by Minsky when he was a boy: no one paid any attention—until he put the sleeve from a flannel shirt on it (Newquist 1994: 62).

The line between rhetorical and theoretical difference, in turn, is often unclear. I implied (above) that the people who independently 'discovered' backprop, Hopfield nets, or the Marr–Albus rule each discovered *the very same thing*. But this is debatable—which is why I said that Grossberg is "arguably" a case in point, and had "perhaps" discovered many ideas later attributed to others.

In general, it's easier to decide such matters when (as in these examples) the idea is expressed mathematically. Two authors might use an identical mathematical equation, or two equations that can be *proved* equivalent. That's not possible for verbally expressed theories, where the many subtleties of interpretation and association of ideas complicate matters hugely. Indeed, showing that two verbal theories or concepts are equivalent is analogous to translating from one natural language to another—a deeply problematic enterprise (9.iv.b and x).

Even with mathematical examples, however, there can be difficulties. A third person, Shun-Ichi Amari, is sometimes said to have "discovered" Hopfield nets (also known as additive nets). But whereas Amari gave only fourteen lines and a mathematical equation, John Hopfield drew out the implications at length, showing how they could be explored in computer simulations. Largely because it was so much easier for cognitive scientists to understand and be inspired by Hopfield, he got the credit. Yet Amari had defined the identical mathematical function—and Hopfield had cited him (see 12.v.f).

As for Grossberg, he'd defined additive nets even earlier, while he was still a student. But he hadn't given a succinct account of them (although he did so later, as we'll see: 12.v.g). Nor had he presented this new idea alone. He'd combined it with many others in describing the processes within a dynamical brain–behaviour system, so that even mathematical psychologists were unable to appreciate it. The scientific value was greater than that of Hopfield's paper (see 14.vi). But the impact, at the time, was less.

Besides the difficulties in getting recognition after a discovery has been published, there's the problem of getting it published in the first place. Strictly, official publication isn't necessary for an idea to be recognized. Chomsky's mimeographed research notes, like Ludwig Wittgenstein's lectures, were read by some experts long before appearing 'in press' (Chapter 9.vi.a). So were Brian Smith's research notes on the nature of computation (16.ix.e). But publication certainly helps.

If Bertrand Russell and Alfred North Whitehead hadn't had some cash to spare in 1910, cognitive science might have arisen much later than it did. For they had to help subsidize the publication of their epochal *Principia Mathematica* (see 4.iii.b). Work on cellular automata, by contrast, might have burgeoned earlier, had von Neumann's lectures not remained unpublished for many years after his death (15.v.b). Occasionally, wider publication is actively prevented. For example, the pioneering work on computing done at Bletchley Park in the Second World War remained an official secret for over

thirty years (3.v.d). And the ‘true’ discoverer of backprop was unable to circulate his idea more accessibly because of the political sensitivities of a US government department (12.vi.d).

In addition, the source of publication matters. One could publish an idea in *The Beano*, and its scientific value wouldn’t be affected—but its reception by other scientists, and therefore its place in history, would. In the seventeenth century acceptance, or even respectful consideration, required the word of a gentleman. Today’s equivalent is the peer-reviewed journal. And the perceived relevance and quality of the peers make a difference. In the 1950s it was acceptable, even if not the authors’ preferred choice, for an important neurophysiological paper to be published in the *Proceedings of the Institute of Radio Engineers*, because of the influence of cybernetics (see Chapter 14.iii.a). Today, no biological reader would see it. And the people popularly associated with backprop made a point of publishing this idea simultaneously in the high-prestige *Nature* and in a more accessible, easily affordable, but well-respected source: an MIT Press trade book, aimed at students and professionals in all areas of cognitive science.

Last but not least, it follows from what’s just been said that the editors and reviewers of the professional journals have the power to block, or to censor, the ideas sent into them. This can affect even well-known scientists, if they’re presenting ideas not favoured by the editor concerned. Bruner himself has come up against this problem:

The human sciences are about human beings in specific situations, and why should I hide that fact? You know, the editor of the *Journal of Experimental Psychology* typically *deletes anything of that sort* from my papers. That’s probably why I don’t write as much for the *Journal of Experimental Psychology* any more. (quoted in Shore 2004: 157; italics added)

And stories are rife about as-yet-unknown writers being spurned by their chosen journals. Indeed, work which eventually won a Nobel Prize was initially rejected by an “elder statesman” reviewing it for the editor of the *Psychological Review* (Chapter 7.iv.f).

## i. An explanatory can of worms

A new field of enquiry typically spawns new journals (see 6.v.c and 10.ii.b). It’s easy to assume that this is because there are so many discoveries, of an increasingly detailed nature, that there wouldn’t be room for them in the existing journals—even if the editors happened to be interested in them.

That’s true, to be sure. But it’s also true that the existing editors may reject *the general approach* underlying all the discoveries—in which case, they’ll probably refuse to recognize them *as* discoveries. (Similarly, researchers are usually loath to offer jobs to youngsters who disagree with them: Newell was a refreshing exception, as we’ll see in Chapter 10.v.b.) The problem, then, isn’t a mere lack of space, or even a mere lack of interest, but a (supposed) lack of intellectual respectability. New journals are founded, accordingly, which do regard these new ideas as respectable.

Bruner’s current publication difficulties (mentioned above) illustrate this point. For they’re due less to specific, and unacceptably shocking, ideas on his part than to a general disagreement about what counts as explanation in psychology. Having started out in the 1950s as a straightforward, if highly creative, experimentalist (6.ii.a–c), he’s now prepared also to consider *interpretative*, hermeneutic, accounts (8.ii.a). In a

nutshell, this means that he relies on his intuition for “narratives” (a critic might say “gossip”), to understand “human beings in specific situations”. But whether *human beings in specific situations* is, as he claims, a proper subject for a theoretical psychology is problematic.

The eighth caveat, then, is that Bruner’s little story about his problems with the *Journal of Experimental Psychology* has opened a philosophical can of worms that we’ll encounter at various points in later chapters. And there are two species of worm in the can.

On the one hand, it’s highly controversial whether theoretical (as opposed to therapeutic) psychology should hope to deal with *particularities*. Some people, such as Bruner (and his long-time Harvard colleague Gordon Allport: 1942, 1946), argue that to understand specific events in individual human lives we need empathetic interpretation, not naturalistic science (Chapter 7.iii.d). Indeed, some philosophers argue that psychological phenomena *in general* (not just particularities) can’t, in principle, be explained in scientific terms (14.x.c and 16.vi–viii). To the editors of the *Journal of Experimental Psychology*, such a view is an abomination. Naturally, then, they won’t be willing to publish papers expressing it.

On the other hand, it’s also disputed whether showing how certain things—either particularities (e.g. Joe Bloggs’s saying “It wasn’t me!”) or general phenomena (e.g. language)—are *possible* really counts as ‘explaining’ them (7.iii.d). Since that’s precisely what the computational approach enables us to do, cognitive scientists have a stake in this abstract philosophical dispute.

You’ll have noticed that the FAQs listed in Section i.a weren’t about what Joe Bloggs did or didn’t do last Saturday, nor about what he’ll do next week. Rather, they concern *how it’s possible* for him—or anyone else—to do the sorts of things that he does every day of his life. If cognitive science can explain that (and if it really is an ‘explanation’), it will have done what it set out to do.

## 1.iv. Envoi

As Naudé recognized (Preface, preamble), others would have told this tale differently. For as we’ve seen, people disagree about *just what counts* as cognitive science. They also disagree about what message they’d want to send in writing about it. Instead of relaying (as I’ve tried to do) the enormous interest and promise of the field, and its necessary interdisciplinarity, some would pen an account of the futile pursuit of a philosophical illusion (Chapters 11.iv.a and 16.vi–viii).

What’s more, the many factors mentioned in Section iii make historical attribution within cognitive science a difficult, and delicate, matter. Such judgements can turn out in more than one way.

If one worried too much about all that, however, no history could even be attempted. So here goes. Now that we know what we’re talking about, let’s begin the story.