Chapter 10

Representation, Reduction, and Interdisciplinarity in the Sciences of Memory

John Sutton

Introduction: Memory and Interdisciplinary

Memory is studied at a bewildering number of levels, in a daunting range of disciplines, and with a vast array of methods. Is there any sense at all in which memory theorists — from neurobiologists to narrative theorists, from the developmental to the postcolonial, from the computational to the cross-cultural — are studying the same phenomena? This exploratory review paper sketches the bare outline of a positive framework for understanding current work on memory, both within the various cognitive sciences and across the gulfs between the cognitive and the social sciences.

The project is, obviously, hopelessly ambitious. How could the concepts, models, or practices of such glaringly incompatible activities as clinical neuropsychology and media theory, or developmental psychology and Holocaust studies, ever be imported into neighbouring discursive universes? More to the point, why would anyone bother? Those who work at the "subpersonal" level, trying to understand neural processes or to construct better models of various memory systems, display either respectful neutrality towards social science, or active disdain towards its perceived anti-naturalism. And those who study "social" levels may think of psychology either as important but irrelevant, or as irretrievably marred by individualism. Sensible advice to leave the frameworks distinct can be

Representation in Mind
New Approaches to Mental Representation
Copyright © 2004 by Elsevier Ltd.
All rights of reproduction in any form reserved.
ISBN: 0-08-044394-X

met, I suggest, by arguing not for a unitary view of memory, but for an integrated framework within which different memory-related phenomena might be understood. The skeletal structure of such an integrated framework might already be in place, implicit in overlapping concepts and theories across the levels at which memory is studied.

This paper is merely a provocative preliminary vision of this structure. It should elicit either principled defences of the idea that certain disciplines and levels of explanation in the study of memory should remain insulated and autonomous, or help in beginning to forge the elusive connections I seek. My claims are four, claims to be understood more as guiding hypotheses for further work than as results. Firstly, significant problems, as yet neglected by both practitioners and commentators, about reduction and interdisciplinarity arise even within the relatively restricted list of disciplines comprising the cognitive sciences of memory. Secondly, suitably weak approaches to reduction will allow for substantive intertheoretic and interdisciplinary contact across some of those sciences. Thirdly, even within accepted subdisciplines of cognitive psychology, key concepts and frameworks increasingly point to the relevance of factors outside the individual. And finally, if these three claims go through, there is already a growing body of relevant research on "culture-and-cognition" which shows ways to link certain (broadly connectionist/dynamicist) theories of individual memory with the study of external memory systems in technology and society.

Few commentators have been equipped to analyse the oddly disconnected parallel courses of the recent explosions of work on memory in the cognitive and the social sciences respectively. Even a listing of the modern disciplines of memory reveals the difficulty of the task.[1] Some set of capacities labelled "memory" is a key focus of research, funding, institutional and pedagogical energy, and writing in (for example) molecular neurobiology, computational neuroscience, cognitive neuropsychology and neuropsychiatry, comparative psychology and ethology, cognitive psychology of many stripes, developmental psychology, personality psychology, social psychology, psychoanalysis, psychiatry, sociology, anthropology, philosophy, history, political theory, media studies, museumology, Holocaust studies, postcolonial studies, and literary theory.

Researchers whose attention edges out from their home discipline are thus often driven to express doubts about the unity of the phenomena in question. Susan Engel, for example, notes in the preface to her book *Context is Everything* that "in

[1] This listing of disciplines, and the claims associated, needs to be defended in secure empirical, bibliographical, and sociological studies. For the moment, the impressionistic survey must suffice. The hierarchical order of the list is intended merely to catch an imprecise and general sense of current work, rather than to beg any questions about interdisciplinary relations.

recent years the topic of memory has become so popular, it seems both ubiquitous and yet oddly invisible. People glide seamlessly from a discussion of childhood recollections to national memories, as if they were part of the same phenomenon" (1999: viii).[2] Despite Engel's sensible caution, her balanced and elegant book stretches from developmental and personality psychology to include careful discussions of literary autobiography and historical memory. But I also want to persuade psychologists who, unlike Engel, are sceptical about the relevance of social studies of memory to their domain that there are substantive and rarely asked questions here about the scope of the cognitive sciences. Can we use a common commitment to mental representations, or to an information-processing framework within cognitive science, to mark off its boundaries and to exclude the social sciences of memory? This would require confidence that some such commitment can be specified clearly and non-trivially, and that it would indeed encompass all of the desired subdisciplines of cognitive science. But even then such a framework might, I'll suggest, itself be extended to deal with cultural and historical memory phenomena.

Most episodes of human remembering have multiple causes, describable in different vocabularies at different levels, and which are not restricted to the past events or experiences remembered. So, in part, questions about interdisciplinarity are questions about which causes matter for specific explanatory purposes. Yet often, when writers on memory do acknowledge legitimate issues about how different levels and disciplines might or might not knit together, they are not sensitive to the difficulty of specifying such purposes, and of carrying out particular programs of translation, reduction, or interdisciplinary theory construction. Hype abounds. The distinguished neuroscientists Larry Squire and Eric Kandel write that "memory promises to be the first mental faculty to be understandable in a language that makes a bridge from molecules to mind, that is, from molecules to cells, to brain systems, and to behavior" (1999: 3).[3] They believe that we're on the verge of a truly multidisciplinary, jointly social and natural, science of memory: "the molecular and cognitive study of memory represents only the most recent attempt, historically, to bridge the sciences, which are traditionally concerned with nature

---

[2] Engel's intention, in one of the best recent syntheses of current psychological perspectives, is to sketch theoretical connections linking "the magnificent but bewildering array of studies and insights that have emerged over the last 10 years," in order to investigate lines along which they might be contrasted with each other. Her attempt "to construct a framework for thinking about memory" touches on more of my wild list of disciplines than any other comparable text, and is particularly strong on developmental studies, for which see section on Constructive Remembering below.

[3] For a more cautious popular assessment of the high-level relevance of the neurobiology of memory see Rose (1992, Chapter 13).

and the physical world, and the humanities, which are traditionally concerned with the nature of human experience, and to use this bridge for the improvement of mentally and neurologically ill patients and for the general betterment of humankind" (1999: 215). And to take just one example from the other end of the disciplinary spectrum, the media theorist Barbie Zelizer writes in a review paper that "in pace with the constitution of the social sciences themselves, . . . the study of collective memory has virtually erased disciplinary boundaries" (1995: 216).

Obviously, it's not going to be easy. Introducing a pathbreaking collection of interdisciplinary essays on memory distortion, Fischbach and Coyle demand that "we must crawl into the [black] box and try to understand the brain as it functions in the context of human affairs. This requires that the advances in neuroscience be paralleled by equally creative thinking at the level of psychology and the behavior of societies" (1995: xi). Faced with this requirement, philosophical pessimism may seem appropriate. In a detailed examination of the pitfalls of interdisciplinary theory-construction, Patricia Kitcher uses a historical case study to list "subtle and not so subtle dangers" in moving too fast between disciplines and discourses. Just as did Freud, she argues, so contemporary cognitive scientists often exhibit too much faith in the resources of a neighbouring discipline, disregarding the seriousness of its internal problems; or they take the coherence and harmony of two theories as conclusive evidence for the truth of both (Kitcher 1992: 159–161, 172–174, 180–183). Kitcher's "purpose is not to argue against the interdisciplinary perspective," but she believes that we need to achieve "a more critical understanding of this necessary and exciting project — and so to proceed more cautiously" (1992: 5, 219). Well, perhaps: recent gatherings of memory researchers from different backgrounds may have aided dissemination of specialist work, but don't seem to have left the participants with the desired glowing sense of impending breakthroughs in interdisciplinary theory-construction.[4] But a start must be made somewhere, and occasionally a messy preference for proliferation over prudence in difficult domains may pay off.

Kitcher sets up her study of Freud's interdisciplinary science of mind by distinguishing three plausible methods for assessing the progress and success of such grand joint ventures. We could "consider the logical form of interdisciplinary

---

[4] I'm aware of five such conferences and series in the last few years: two in Cambridge, MA, on memory distortion in 1994 and on memory and belief in 1997 (see Schacter 1995a; Schacter & Scarry 2000); the Darwin College lectures in Cambridge, England in 1996 (see Fara & Patterson 1998); a C.N.R.S. "Cross-Disciplinary Encounters" series on memory in Paris in 2000 (see http://www.cnrs.fr/cw/en/pres/compress/memoire/sommaire.htm cited at 22 July 2003) and a University of London series on memory in science in 2000. Each of these series included at least one session on neural bases of memory, and at least one session on memory in sociology or history.

theories and arguments"; try to assess current interdisciplinary work in cognitive science directly; or abstract away from current research to gain a better sense of "the potential strengths and weaknesses of interdisciplinary methodology by examining a historical case" (Kitcher 1993: 4). I agree entirely with her view of the utility of what might be called a historical cognitive science.[5] But my method here is the direct description of interdisciplinary approaches in cognitive science. I don't dispute Kitcher's worry about this method, that "it is perhaps somewhat early to judge the effects of interdisciplinary integration on cognitive science," but I don't see the harm in trying: and, by restricting attention to memory, I hope also to open up the possibility of comparing its potential susceptibility to interdisciplinary analysis with that of other cognitive domains. For example, interesting comparative analyses might be made between the case of memory and the study of colour vision, on the one hand, and of dreaming, on the other. My hunch is that the current sciences of memory look less successfully interdisciplinary than the case of colour vision, but more successfully so than those of dreaming. If this is the case, it could be due either to the peculiar history and developmental course of the relevant sciences, or to something peculiar to the particular phenomena in question. Like the present project, this would be work in the philosophy of cognitive science, seeking a synoptic vision by trying to immerse in some detailed areas of some of the contributing sciences in order to motivate debate, and enlisting the help of experts to point out the gaps in such motivated appropriation of their languages.[6]

Notes on Reduction and Interdisciplinarity

There already are partial, imperfect, but promising interdisciplinary research programs within the cognitive sciences of memory. What kind of interdisciplinarity

---

[5] In *Philosophy and Memory Traces: Descartes to connectionism* (Sutton 1998), I tried to do philosophy of cognitive science and historical cognitive science simultaneously, by offering a combined history and defence of the notion of distributed representation in models of memory. The idea was to undermine different patterns of resistance to interdisciplinarity, in the form both of humanists' resistance to cognitive science, and of scientists' resistance to culture and history. But the project was almost entirely underlabouring, and the positive vision of a cultural cognitive science of memory was hardly even sketched there.

[6] So, as should be clear by now, this is not a project in the philosophy of memory as traditionally practiced. The abstract epistemological questions about the very existence of memory traces, about representations and realism, and more recently about self-knowledge and externalism which have characterised Anglophone philosophical discussions of memory have their own importance, but are of little interest to most cognitive scientists. For introductions see Warnock (1987) and Sutton (2002).

do I have in mind? Again, I agree with Kitcher: in difficult and immature domains like this, "it is often extremely difficult to reduce a methodological approach [that of interdisciplinarity] to a set of logical relations" (Kitcher 1992: 4). So I'm not trying to decide in just what respects the cognitive sciences of memory, despite harnessing the vast apparatus of Kuhnian "normal science," are nevertheless paradoxically pre-paradigmatic.[7] Instead I argue that the apparently universalizing urge to connect quite different languages in the sciences of memory does not commit me to an implausibly strong kind of reductionism, or to any classical, microreductive unity-of-science doctrine. By simply mentioning the availability of a range of alternative, weaker approaches to interdisciplinary theory-construction and to reduction in cognitive science, I hope to ward off the immediate concern that my quest will inevitably impose an artificial unity on complex and highly diverse phenomena. After a general discussion of reduction, I'll look at a particular strand of the sciences of memory which is used by Valerie Gray Hardcastle to exemplify interdisciplinary theory-construction.

Interdependent Phenomena and Patchy Reduction

John Bickle takes some pleasure, introducing his persuasive defence of a "new-wave" reductionism, in quoting Jaegwon Kim's remark that being a reductionist nowadays "is a bit like being a logical positivist or a member of the Old Left: an aura of doctrinaire naivete hangs over such a person" (Bickle 1998: 1, quoting Kim 1989). Fighting over the word "reduction" is often, and perhaps should only be, a political activity. It's not at all obvious whether the particular interdisciplinary research on memory which I discuss should be seen as "reductionist" or not. My perspective here is intended to be broad enough to encompass both avowed "new-wave" reductionists like Cliff Hooker, Patricia and Paul Churchland, and John Bickle, and those who claim to reject reduction while encouraging detailed engagement in interdisciplinary programs, like Hardcastle and Kitcher. We can see why the retention or rejection of the term may not matter (at least for current purposes) by examining Kitcher's criteria for interdisciplinarity (1992: 6–7).[8]

---

[7] On cognitive science in general as a case study in the philosophy of science, see the rigorous notion of "research frameworks" developed by Barbara von Eckardt in her remarkable book *What is Cognitive Science?* (1993: 13–56, 345–396. Von Eckardt has recently suggested that cognitive science is "still a mere babe in the woods of science" (1999: 221).

[8] Where Kitcher's three options for assessing interdisciplinary theory-construction, discussed in the first section above, were merely methodological, the three kinds of interdependence she envisages here are metaphysical.

She argues that the idea of mutual influence across apparently separate disciplines requires, minimally, an "assumption that the phenomena dealt with by the different disciplines are in some way interdependent." Such interdependence, for Kitcher, can come in at least three varieties: reduction, which she characterizes classically as involving derivation of reduced laws from reducing laws plus biconditional or bridge laws; supervenience, by which "the properties captured in one science alter if and only if there is some change in the properties characterized by a second science"; or a still weaker kind of genuine interdependence. This last variety of interdependence between phenomena holds "just in case some changes in the properties mentioned by one science alter a significant range of properties mentioned by the second" (1992: 6–7). Yet it is around something like this last kind of weaker, local relation between theories that the "new-wave" reductionists build their more liberal view.[9]

Whatever is implied by these preferred accounts of intertheoretic relations, whether characterized as interdependence or reduction, they are definitely not committed to two extreme claims sometimes ascribed to any substantive reductionism. They do not require the complete purging or elimination of all higher-level terms and concepts from lower-level, "reducing" explanations. And they do not imply that ultimately only the vocabulary of fundamental physics will be legitimate.[10] Neurobiological theories retain terms from "higher-level" theories: Kandel's account of associative learning in the sea slug *Aplysia*, for instance, employs ineliminably psychological terms such as habituation and sensitization. But this doesn't entail that such theories are mere implementations, rather than genuine reductions.[11]

"New-wave" reductionists can agree that the causal generalizations of theories like Kandel's are, as Schaffner puts it, "typically *not* framed in purely biochemical terminology," but still argue that some such theories genuinely describe relations of interdependence between phenomena at different levels. Realistic interesting intertheoretic relations will allow at most for partial reductions. Genuine explanations will produce "many weblike and bushy connections" across levels, with causal sequences described at many different levels of aggregation. In biological

---

[9] The key source for this tradition is Hooker (1981). See also, in addition to the sources cited by Bickle *et al.* (1998).

[10] Both of these positions are ascribed to "radical" reductionists by Gold & Stoljar (1999). For a detailed response see my commentary (Sutton 1999). See also the treatment of long-term potentiation in Stoljar & Gold (1998).

[11] As Gold and Stoljar suggest in a detailed discussion of Kandel: but they don't deal with alternative interpretations of Kandel's work by "new-wave" reductionists, instead assuming an implausibly narrow classical picture of reduction. See Schaffner (1992) and Bickle (1995, 1998: Chapter 5).

cases especially, the generalizability of local reductions may be limited, with some perhaps being specific to the particular system under investigation. So if there is reduction, it is "bound to be patchy" (Schaffner 1992: 337). So the serious problem posed to classical microreductive views by the possibility of multirealizability (by which many different lower-level, reducing terms or laws might in different contexts realize the same higher-level terms or laws) does not bite here: interesting reductions are often going to be heterogeneous (Bickle 1998: Chaps 1 and 4). It's just because of this lower-level heterogeneity that many reductions are revisionary, with the explanations of the higher-level theory either merely approximating the reducing explanations, or being fragmented into many different lower-level explanations (Bickle 1998: 200–201). This is how we have to get what Paul Churchland calls "objective knowledge of a highly idiosyncratic reality" (1996: 306). So Bickle sees Kandel's account of associative learning, with its intentionalist vocabulary of surprise, expectation, and predictability, as genuinely reductive just because of the proposed combinatorial dynamics by which "a few fundamental processes of cellular plasticity, internal to individual neurons, are theorized to occur in a variety of sequences and combinations to produce complex forms of behavioral plasticity" (Bickle 1995: 268).

Interdisciplinary Theory in the Sciences of Memory

To engage in even this much metaphysics is to go beyond the interests of many cognitive scientists. As Endel Tulving, a leading cognitive psychologist of memory, says, "our research community as a whole does not perceive much value in conceptual analysis; there is no promise of social reinforcement for any single individual who might be attracted to the enterprise" (2000: 34). Tulving of course is, from his position of eminence, recommending new attention to conceptual clarity: but interdisciplinarity may in practice be furthered more by the pursuit of particular empirical cases. In this spirit, I outline a discussion of the psychological distinction between explicit and implicit memory by the philosopher Valerie Gray Hardcastle, who argues that it exemplifies (historically and conceptually) the mutual interdependence between disciplines which she sets at the heart of the cognitive scientific enterprise (Hardcastle 1996: Chapter 6).[12]

[12] The example is motivated by Hardcastle's desire to display particular connections between psychology and neuroscience which do not entail the strict explanatory dependence of the former on the latter. For this reason she thinks "we need to stop arguing over the reductionism/antireductionism issue" (1996: 104). Carl Craver is developing an independent account of robust intertheoretic strategies in neuroscience: see Craver & Darden (2001) and Craver (2002).

Hardcastle's case study aims to describe a "strongly interdisciplinary approach," which carves out an explanatory framework not just by relying on evidence from more than one area, but by actively accepting the underlying assumptions of those areas. If successful, the case implicitly answers concerns like Kitcher's about the overhasty importing of alien methods and concepts: "we may use another discipline for collateral support, inspiration, and to help set the parameters of inquiry, but we cannot simply borrow data wholesale from other theories over the same state space" (Hardcastle 1996: 106, 110). The distinction in question, at a first pass, divides explicit or aware remembering from nonconscious influences of past experience on perception, thought, and action. The idea that there are two independent systems in play here is a substantive hypothesis.[13] My concern here is not with its plausibility, but with rehearsing Hardcastle's claim that it is genuinely interdisciplinary.

Initial evidence for the explicit/implicit distinction came from the study of amnesia in clinical neuropsychology. The patient H. M. and others proved able to learn new skills (such as solving simple puzzles), and to remember how to engage in certain procedures, even while explicitly denying having experienced them before, and showing no conscious awareness of even recent events (Hardcastle 1996: 112; compare Schacter 1996: 137–142; Squire & Kandel 1999: 11–16; and, on H. M., Hilts 1995). Secondly, various strands of developmental inquiry suggest the existence in some domains of facilitated processing before, and independently of, explicit knowledge: face recognition, for example, seems to operate automatically from birth in a context-bound system which is later either replaced or augmented by the capacity to achieve explicit learning of specific faces (Hardcastle 1996: 106–111, with references). In turn, cognitive psychologists have designed tests to tap explicit and implicit memory independently. Whereas performance in explicit recognition of previously-presented words, for example, diminishes severely over a week, the ability successfully to complete fragments of words (such as adding "LE" to "TAB__" after previously being shown "TABLE" in a list) declines much less rapidly. The suggestion is that since this facilitation of performance due to previous exposure varies independently of explicit recognition, it is the work of a distinct implicit priming system (Hardcastle 1996: 114).

Hardcastle goes on to describe the varying relation of each of these three traditions (in neuropsychology, developmental psychology, and cognitive psychology) to current neurobiological attempts to localize memory systems in animals,

---

[13] I'm not here discussing the general question of the logic and criteria for postulating and identifying separate memory systems. For a concise account see Schacter & Tulving (1994).

and to recent neuroimaging studies. There are, as she notes, specific puzzles about each strand of evidence here. To what extent, for example, are the implicit memory systems modality-specific? And just how complex and apparently "cognitive" can implicit memory be (compare Schacter 1996: 187–191)? Is it an essential property of implicit systems that their representations can't be manipulated and controlled? Again, rather than evaluating her own preferred synthesis of the evidence,[14] I address Hardcastle's moral: an interdisciplinary theory will include both a general set of principles which apply to phenomena across several different objects of study (different animals, for example), and a series of specific models from which these general principles were derived. The former may be too general to drive specific predictions for any experimental procedures; the latter may be too particular to generalize across many organisms (Hardcastle 1996: 132–133, 138). This is obviously far from a classical reductionist vision:

> As a true discipline still in the making, cognitive science is an amalgam of different disciplines, with each bringing incompatible research techniques and traditions to bear on a common set of interests. We have no reason to expect that the different domains will ultimately fuse into one; each could retain its autonomy as they address a set of central concerns, linked instead by a fuzzy set of entity attributes across a family of models and certain core framework principles (Hardcastle 1996: 135).

The power of Hardcastle's account lies in this unusual combination. On the one hand she accepts the ongoing multiplicity of disciplinary perspectives; but on the other hand she argues that memory scientists can, and often do, use these different perspectives to address just *the same* event space, in compiling *single* explanations which are themselves multi-levelled. When the objects and events under investigation are both complex and idiosyncratic, the best explanations may *have* to be partial or patchy in order to connect to a more general theory which links different but related phenomena. Whether this is seen, with Hardcastle, as a failure of the old reductionist urge to find "a single overarching account"

[14] Hardcastle argues for a revised terminology for the two systems, suggesting "structural episodic memory" for the implicit, task-specific responses, and "semantic episodic memory" for explicit, cross-modal memory (1996: 131–132). This is controversial: one might for example be uneasy about including implicit memory as a variety of episodic memory. But this doesn't affect the metatheoretical point about interdisciplinarity.

(1996: 139), or, with Bickle and the Churchlands, as a triumph of new-wave reductionism, is unimportant: in either case, we have at least the possibility of seeing different (sub)disciplines as dealing with interdependent phenomena.

Constructive Remembering: Source Memory and Development

I have suggested that issues about reduction and interdisciplinarity arise, and are already being fruitfully addressed, within the cognitive sciences of memory. In this section I make the further case that key lines of research within those sciences point to the relevance of factors outside the individual, and of explanatory connections between cognitive and social psychology. My examples are from two strands of an emerging consensus about constructive remembering in cognitive and developmental psychology, in work firstly on source monitoring and source amnesia, and secondly on the development of autobiographical memory. In both cases, larger-scale historical or cultural dynamics may be relevant in understanding both distortion and accuracy in individual memory in both adults and children. Cognitive psychology can't neglect the interface between brain and world.

In the context of the distinction between explicit and implicit memory discussed in the previous section, the focus here is primarily on certain forms of explicit or declarative memory. I'm looking, specifically, at autobiographical episodic memories, explicit experiential memories of past events and episodes in a personal history. Consensus on a definition of episodic memory, sometimes called "personal memory" by philosophers, is surprisingly hard to find: we can work for now with William Brewer's account of "recollective memory" as "a reliving or reviving of my own past phenomenal experience, with the additional knowledge that I've had that experience before" (1996).[15] It's in episodic memory that we achieve a form of "mental time travel," in which we're oriented to events as occurring at particular past times, events which we sometimes knit into autobiographical narratives (Suddendorf & Corballis 1997; Tulving 1983, 1993, 1999). While Brewer's definition leaves open a number of key questions which I

[15] One initial respect in which this seems too strong is in the requirement that I *know* I've had the experience before. Most cognitive psychologists abstract away from the folk use of "remember" as a "success-word," and study the varieties of false memory, thus including cases in which I mistakenly *believe* I've had the experience before.

bypass here,[16] we can focus on clear core cases of episodic remembering, to pick out certain aspects of current cognitive-psychological approaches.

Constructive Remembering

"A variety of conditions exist," declares Daniel Schacter, "in which subjectively compelling memories are grossly inaccurate" (1995b: 22). Many cognitive psychologists now accept, at least in principle, that explanations of even relatively simple memory phenomena will invoke multiple causes. This recognition is of course hard to put into practice. But the recovered-vs-false-memory crisis of the 1990s encouraged academic psychologists to shelve disputes between "ecological" and "lab" approaches to memory, and to forge a swift consensus on the multicausal and constructive nature of remembering, in opposition to extreme views on the possibility of total and exact recovery of long-repressed memories.[17]

There are two related aspects of this relatively recent consensus in cognitive psychology (for reviews see Roediger 1996; Roediger & McDermott 2000; Schacter 1999). Firstly, attention is focussed increasingly on the context of recollection, rather than solely on encoding or on the nature of encoded traces. Realistic representational theories of memory do not assume that recall is fully determined by the nature of the stored representation. On Tulving's notion of "synergistic ecphory," traces (whatever they may be) are "merely potential contributors to recollection": the engram is not the memory, and instead "the [current] cue and

---

[16] Two issues of considerable interest concern time in memory, and the relations (both developmental and conceptual) between episodic memory and autobiographical memory. John Campbell argues that episodic memory presupposes a particular and objective conception of time as linear, which grounds our ability to locate autobiographical episodes as at particular past times (Campbell 1994: Chapter 2, 1997; Hoerl 1999). Not all autobiographical memories are episodic, since I can (non-experientially) remember facts about my life (such as my date and place of birth), using only semantic memory. But the converse question, whether all episodic memories are autobiographical, remains open. My hunch is that any putative consensus on the nature of episodic memory will show it to be, both conceptually and sociologically, an intrinsically interdisciplinary construct, with, in particular, the methods and results of cognitive ethological research on animal memory needing to be integrated: see for example Campbell (1994: 37–41, 64–71), Tomasello (1999: 124–125) and Dennett (2000).

[17] This speculation on the sociology of recent cognitive psychology needs careful historical support. For the earlier battles between "real-life" and cognitivist opponents see for example Middleton & Edwards (1990), and special issues of *American Psychologist 46* (1991) and *The Psychologist 5* (1992). A later, ecumenical treatment is Koriat & Goldsmith (1996). Ian Hacking's brilliant *Rewriting the Soul* (1995) was perhaps too early to catch the effects of the recovered memory movement on academic theories of memory.

the engram conspire" in any act of episodic remembering (Schacter 1996: 56–71, 105; 1982: 181–189; Tulving 1983: 12–14). So, as Susan Engel puts it, "one creates the memory at the moment one needs it, rather than merely pulling out an intact item, image, or story," so that specific features of any context of recall may be direct causes of the content and format of the memory-as-retrieved (1999: 6).

Secondly, cognitive psychologists have come to accept more flexible and dynamic pictures of long-term "storage." This is partly due to the influence of connectionist or Parallel Distributed Processing (PDP) models, but also to a rich array of studies on misinformation, bias, and the role of schemas in memory, some of which I discuss below. This internal plasticity is one of the most curious and characteristic features of human memory, and one which clearly differentiates our cognitive systems from current digital computers. It's pretty useful for the contents of our files to remain exactly the same from the moment I close them at night to the moment I open them again in the morning. But various kinds of reorganization and realignment often happen to the information retained in my brain over the same period. In us, many memories do not naturally sit still in cold storage.

This consensus about the constructive nature of remembering needn't be unrealistically overdescribed. It's not that accuracy and reliability in memory are suddenly shown by science to be impossible or unlikely. Rather, the assumption is that understanding of the mechanisms of distortion will also illuminate the processes operating in veridical remembering (Mitchell & Johnson 2000: 179–180). Neither "accuracy" or "reliability" are transparent notions in this context, and "truth" in memory, though not forever inaccessible, is neither a single nor a simple thing. Verbatim recall and other forms of exact reproduction are rarely necessary for success in remembering (Rubin 1995).

Constructive Memory and Source Monitoring

The "source monitoring" framework developed by Marcia Johnson and her colleagues exemplifies a method of "experimental phenomenology" which takes subjective judgements and feelings as core explananda for cognitive psychological theory. "Source monitoring" is the process of recalling when, where, and how some information was acquired (Johnson *et al.* 1993; Johnson & Raye 2000; Mitchell & Johnson 2000). Johnson argues that there is typically no tag or label on a memory to specify its source. Instead, activated contents must be "evaluated and attributed to particular sources through decision processes performed during remembering" (Johnson *et al.* 1993: 3). Our subjective experience of autobiographical remembering, it's suggested, depends on our source attributions: when some content doesn't

have the right set of qualitative characteristics, it's likely to be experienced not as something I remember through personal experience, but just as something I know (Rajaram & Roediger 1997).[18] These attributions are not determined by properties of the "memory records" alone, but are influenced also by a range of aspects of the current context, including biases, motivations, current agendas, attention, stress, metamemory assumptions, apparent compatibility with other knowledge, and social setting. Johnson suggests, further, that the ease of attribution of a content to a specific source is one "compelling reason for experiencing it as belonging to our personal past," thus offering another promising line of enquiry into the multiple roots of autobiographical remembering (Johnson *et al*. 1993: 21).

As a theoretical framework subsuming a variety of experimental results on memory distortion, misinformation, eyewitness testimony and suggestibility, the source monitoring research is driven by evidence from a range of (sub)disciplines. As well as cognitive psychology, it draws on cognitive neuroscience (both clinical studies of frontal lobe lesions, and neuroimaging investigations), and on personality, developmental, and social psychology. In each case, the project is to study conditions under which source monitoring breaks down, in confusions of (for example) reality and fiction, internal and external sources, trustworthy and unreliable sources, or perception and imagination. In each case, contents become unglued from their origins and misattributed to another source or context. There are highly dramatic instances, as when the psychologist Donald Thomson was accused of rape when in fact the victim had been watching him being interviewed on TV prior to the rape; but related processes operate in our mundane tendencies falsely to "remember" seeing a thematically-related word such as "sweet" in a presented list of words like "candy," "sugar," and "taste" in which "sweet" did not in fact appear (Schacter 1999: 188–190, with references).[19] Source forgetting drives the generation of plausible, but in such cases incorrect, attributions.

There is no sharp line between the normal, fallible but functional operation of source monitoring processes, and more pathological and confabulatory contaminations and intrusions. Some forms of source monitoring are harder for children, and develop at different stages; some forms are more difficult for older adults. Some individual difference factors (including imaging ability, hypnotizability, and dissociative tendencies) may be related to particular kinds of

[18] There are various intermediate possibilities, in which subjects feel they are neither straightforwardly "remembering," nor just "knowing," yet they somehow have access to the information in question.

[19] Schacter also includes ordinary "cryptomnesia" and unintentional plagiarism, in which we are remembering but do not subjectively feel that we are, because we misattribute to our own imagination a thought or idea which in fact derives from a specific prior experience.

source confusions.[20] But misattribution is not always unadaptive: indeed, normal subjects apparently produce false recognitions in certain contexts *more* often than do amnesics with medial temporal lobe damage (Schacter 1999: 190). If correct, this intriguing evidence suggests that accurate source monitoring is often less important than the kind of generalization on the basis of similarity and theme which (in normal subjects) may "give rise to distortions as an inherent byproduct" (McClelland 1995: 84).[21] Susan Engel may overstate a little in writing that "it is the norm rather than the exception to be unable to distinguish between what happened, what you feel about what happened, and what others may have said about what happened" (1999: 16): but the source monitoring framework at least suggests that understanding the conditions in which such confusions occur will require attention to a large range of affective, motivational, temperamental, and social contextual factors which influence the mapping of information to source.

The idea that cognitive psychology studies the individual mind, leaving social processes to be treated by the social sciences, thus looks untenable. The creation of a false memory may often be explained neither by an overconfident therapist alone, nor by a purely internal process. As Janice Haaken argues, the source monitoring framework requires us to discuss the recovered memory crisis without either overconfident individualism, or a focus on "memory deceivers on the other side of the fence": instead, we have to trace the subtle sedimentations of culture in some individual memories. "The source of a remembrance," Haaken concludes, "may not be readily or immediately located in discrete events in an individual past but rather may be found in the complex web of converging group experiences" (1998: 111).

Autobiographical Memory Development and Scaffolding

Susan Engel's 12-year-old son, working on a class writing assignment, looked seriously at his mother before asking "Mom, what is my most important memory?" (1999: 24). Autobiographical memory develops in a shared environment, in which its content as well as its expression is influenced by that interpersonal and cultural context. Developmental studies are a rich and flourishing area within the sciences of memory. Despite great variety in methods and theoretical assumptions, most

---

[20] See the special issue of *Applied Cognitive Psychology 12* (1998), "Individual Differences and Memory Distortion."

[21] This point is also relevant to philosophical accounts of the role of memory in maintaining continuity of personal identity over time. Where traditional theories of personal identity focus on current contact in memory with specific episodes in the personal past, it may be, as Marya Schechtman argues, "precisely insofar as our memories smooth over the boundaries between the different moments in our lives . . . that we are able to produce a coherent life history" (1994: 13).

schools of developmental thought are thoroughly interdisciplinary, calling to differing degrees on neuropsychology and social psychology as well as cognitive psychology; and most accept, at least in some explanatory contexts, the significant causal influence over time of the remembering environment. Children learn to remember in company.

Children start talking about the past "almost as soon as they begin talking," but the form of their references to past events takes some years to develop (Nelson & Fivush 2000: 286).[22] At early stages, adults provide much of both the structure and the content of young children's references to the past, providing "scaffolding" for the children's memories. Where initially children use generic event memories implicitly, like scripts, as a basis on which to understand routines and generate expectations, they gradually develop the ability spontaneously to refer to specific past episodes with rich phenomenal content. Although there are a number of different theories of how these changes unfold, with the relative roles of language, temporal representation, theory of mind and metarepresentational capacities, and self-schemata being as yet uncertain, joint reminiscing is a key part of the process (Howe & Courage 1997; Perner & Ruffman 1995; for integrative discussion see Welch-Ross 1995).

Children gradually develop perspectival temporal frameworks in which to locate memories of idiosyncratic events. Memory sharing practices, often initiated by adults, encourage the idea of different perspectives on the same once-occupied time (McCormack & Hoerl 1999). In developing this temporal perspective-switching, children start to take memories as objects for negotiation, shared attention, and discussion. Realization of the existence of discrepancies between versions of the past goes along with the development of some kind of self-schema, as children begin to collect stories into some kind of personal history. The ability to view one's life retrospectively is sophisticated, and follows adult guidance in simpler conversations about the past. On the social interactionist picture of the development of memory, defended by Robyn Fivush, Katherine Nelson, and others, parental and cultural models or strategies for the recounting of past events act as initial scaffolding on which children start to hang their own memories. They then internalize the forms and narrative conventions appropriate to their context, so that "early reminiscing begins as an interpersonal process and only becomes intrapersonal over time" (Engel 1999: 27; see for example Fivush 1991, 1994; Nelson 1993). The point is not that children can't remember in solitude, nor that they remember only what they talk about, but that both shared

[22] There are of course also a range of memory phenomena before language: for surveys of the methods used to study memory in infants, and current thinking about the results, see Mandler & McDonough (1997) and Rovee-Collier & Hayne (2000).

and inner reminiscing may alter the form and the content of autobiographical memories. Variations in narrative practices may then reappear in the subjective idiosyncrasies of early remembering. Cultural variations in the nature and contexts of talk about the past, and intracultural variation in the motivations for and richness of remembered narratives, have been convincingly demonstrated.[23]

It is not clear whether these variations have longer-term effects, nor how narrative practices interact with the simultaneous development of self-schemata and of a theory of mind. But my comment on this work is simpler: I want to insist on the possibility of strong interaction between individual and shared reminiscence without assuming that the developing internal representations are a straightforward projection of the shared narratives. Fivush and her colleagues occasionally write as if the format of autobiographical memory is itself linguistic or language-like, as if children simply incorporate the forms and contents of local external narratives: following Vygotsky, Fivush argues that "the narrative forms that children are learning to organize their recounting of past experiences are also used for organizing their internal representations of past experiences" (1994: 138). This is possible, but the argument slides too quickly: we can accept that what Peggy Miller calls the "distribution of storytelling rights" in a culture or in a family may strongly influence the uses and the contents of individual memories (Miller *et al.* 1990), without having to assume that either the format or the organization of those individual memories is literally linguistic or narrative. In the final section of the paper, I suggest that this same situation crops up more generally in the sciences of memory. We want to allow a variety of ways in which individual and cultural processes and representations can interact, complementing or conflicting with each other, while retaining asymmetry between those internal and external representations. Social norms and cultural narratives are not simply downloaded into the mind, and yet each mind is more intimately tangled with such norms and narratives than traditional individualistic cognitive science has often allowed.

Cognition and Culture: Collective Memory and External Memory

Traces Inside and Outside the Mind

The work I've described in cognitive and developmental psychology is as yet unintegrated with the central projects of cognitive science, in that it proceeds

[23] See for example the studies of "elaborative" and "pragmatic" reminiscing styles by Reese *et al.* (1993) and Welch-Ross (1997); and of psychological effects of cross-cultural linguistic styles by Mullen & Yi (1995) and MacDonald *et al.* (2000).

without substantive assumptions about the nature of mental representation and computation. But the experimental evidence gathered in these subdisciplines must be compatible with a theory of mental representation: the constraint works both ways. So if I'm right that such central parts of the cognitive psychology of memory as these already point to the relevance of factors outside the individual, there are two tasks for the wildly optimistic interdisciplinarian to pursue. Firstly, different approaches to mental representation within cognitive science must be related to, and tested against, this data and these cognitive-psychological frameworks. Secondly, we must find vocabularies and methods for connecting the cognitive and the social sciences of memory. In this final section, I address this second task: this requires dropping neutrality between cognitive-scientific theories, and applying to the case of memory some recent work in cognitive anthropology and embodied cognition.

The aim is to find a general framework for the sciences of memory in which the concepts of social or "collective" memory, and of "external memory," will be integral parts of cognitive science, rather than social constructionist myths or humanistic curiosities. If memories are not stored independently, permanently, and explicitly within the individual mind (but are, for example, superpositionally retained as dispositions of the connection weights of neural networks), then the relatively unstable individual memory needs support from more stable external scaffolding or props. One theoretical goal might be a broad metaphysics of traces inside and outside the mind (perhaps along the ambitious lines suggested by Leyton 1992). Such a picture would not collapse the distinction between internal and external representation and processing, but would provide a framework for investigating whether and how our interaction with different forms of external information systems might in turn affect the format and processing of individual memories. Culture and technology are products of cognition and action, but in the human case, as Merlin Donald argues, such products in turn "have direct effects upon individual cognition" (1991: 10).

Collective Memory

Some explanations in the social sciences will refer, among multiple causes, to (appropriately flexible) internal processes of schematization or reconstruction. And some explanations in the cognitive sciences will refer to the transformations of external representations. The resources of situated cognitive science can be put to work in forging fruitful notions of collective and external memory. There's widespread scepticism about the very idea of "collective memory." Naturalists are uneasy about the taint of Jungian archetypes, or morphic resonance. And recently

even sociologists and historians who work on what seem to be social memory phenomena have tried to disclaim the notion. The historian Pieter Lagrou (2000) suggests that the term "memory" amalgamates "memory proper, which is an inalienably individual capacity, and "memory" as a metaphor, in an anthropomorphism that is often not conscious, for the entire set of current representations in a community." James Young, working on memories of the Holocaust, prefers to use the term "collected memory" instead of "collective memory," because "societies cannot remember in any other way than through their constituents' memories" (1993: xi).[24] In discussing the work of the great sociological theorist of collective memory, Maurice Halbwachs, Fentress and Wickham worry that his concept of collective consciousness was "curiously disconnected from the actual thought processes of any particular person," leaving later sociological accounts with the danger of treating the individual as "a sort of automaton, passively obeying the interiorized collective will" (1992: ix–x). This widespread embarrassment is understandable among sociologists and historians seeking explanatory models which are both flexible and naturalistic. But it isn't necessary.

Although not my aim here, it's possible to find a more subtle, and more plausible, account of the relations between individual and collective representations in Halbwachs' own work.[25] It's true that he was highly critical of the individualism of psychological theory between the wars, but his positive, anti-individualist views do not rest on mystery. For Halbwachs, briefly, "there is no point in seeking where memories are preserved in my brain or in some nook of my mind to which I alone have access: for they are recalled to me externally" (1925/1992: 38).[26] The people and groups around me normally "give me the means to reconstruct them." Collective frameworks of memory are not the simple product of isolated individual memories, constructed after the fact by combinations of separate reminiscences, but are rather, in part, their source, the instruments used in the particular acts of recall. There's a sharp contrast, argues Halbwachs in an intriguing chapter on dreams, between remembering and the actual "state of isolation" of the dreamer,

---

[24] Compare the odd locution of Jacques Le Goff (1992: 53): "at a metaphorical but important level, . . . the absence, or voluntary or involuntary loss, of collective memory among peoples and nations can cause serious problems of collective identity. The connections between different forms of memory may also be not metaphorical but real in character."

[25] Another compatible project here would be the application to memory of recent work in social ontology, which seeks to naturalize notions like "joint action" and "mutual knowledge." See for example Tuomela (1995). But my argument about memory doesn't depend on the details of any particular such project.

[26] The view that Halbwachs simply neglects psychology is unfortunately and erroneously supported by this translation, which simply omits the bulk of the early chapters of Halbwachs' work, which cover dreams, language, constructive memory, and the localization of memories.

who isn't capable directly of reliance on the frameworks of collective memory: "it is not in memory but in the dream that the mind is most removed from society" (1925/1992: 42).[27] So in ordinary remembering, which is either actually or potentially shared, it is the public scaffolding of various forms, in the physical, symbolic, and mnemonic environment, which triggers the specific form and content of individual memory. The contemporary sociologist Paul Connerton puts the point strongly: "it's not because thoughts are similar that we can evoke them; it is rather because the same group is interested in those memories, and is able to evoke them, that they are assembled together in our minds" (1989: 37).

Blunt claims like these no doubt strike many readers as not simply anti-individualist, but as anti-cognitivist: how can there be a cognitive science of memory which includes or even allows for such a displacement of explanatory relevance from the individual mind/brain to the natural or cultural world? The assumption to be questioned is that cognitivist, information-processing commitments automatically rule out notions of external or collective memory.[28] I'm going to make this case, and the stronger case that certain cognitive theories specifically *require* attention to external and collective memory, by direct engagement with the implications for memory science of recent work on the "embedded" and "extended" mind. But first I make two preliminary point about current movements in the social sciences of memory.

Collective Memory and Schema Theory

Firstly, there is ample naturalism in history, anthropology, and sociology, in studies of memory which are compatible with, even if not actively constrained by, current work in cognitive science. For example, Michael Schudson has classified forms of collective memory, arguing for distinctions between three kinds: socially mediated individual memories, cultural forms and artifacts which hold and interpret the past for social mediation, and individual memories which are in turn constructed from the cultural forms (1995; compare Zelizer 1995; Assmann 1997: 1–22 on "mnemohistory"; and the review by Olick & Robbins 1998). The main current concern for a naturalist about this social scientific work should not, I suggest,

---

[27] The point of the contrast is to argue that memories do not endure unchanged in an unconscious state, for later autonomous extraction and manipulation by either the dream or the will: Halbwachs doesn't deny the role of memory in dreams, which are "made of fragments of memory mutilated and mixed up with others," but underlines the absence of interpersonal contact and comparison in dreaming.

[28] Though not applied to memory, this general point is the core of Ronald McClamrock's book *Existential Cognition* (1995).

be the danger of seeing collective memory as floating free of individuals, but rather the temptation to overlinguisticize the form of the internal representations which construct and are permeated by collective memories. Halbwachs sometimes wrote like this, arguing for instance that "one cannot think about the events of one's past without discoursing upon them" (1925/1992: 53), and the powerful influence of Russian psychology on Anglophone developmental theory often has the same result (Bakhurst 1990). But this is to project too quickly the format of external, expressed memories back inwards onto internal memory. But this linguistic-constructivist conception of mental representation, which *is* at odds with the post-connectionist cognitive science I describe below, is not necessary: my memories can be called forth socially, moulded and formed by external influences, without having themselves to be, in their internal aspect, linguaform.

Secondly, the history of the concept of "schema" instructively reveals psychologists and cognitive anthropologists struggling to find a vocabulary for relations between internal and external memories which neither collapses the distinction nor sees the internal as simply the reflection of the social. When Bartlett imported the term into the psychology of memory from neurophysiology, he worried about its implications of stasis:

> I strongly dislike the term 'schema'. It is at once too definite and too sketchy . . . . It suggests some persistent, but fragmentary 'form of arrangement', and it does not indicate what is very essential to the notion, that the organised mass results of past changes . . . are actively *doing* something all the time (Bartlett 1932: 201).[29]

So a "schema" is not, in one sense, a definite cognitive structure at all, and Bartlett makes the same point in relation to the notion, closely related in his system, of a memory trace:

> Though we may still talk of traces, there is no reason in the world for regarding these as made complete, stored up somewhere, and then re-excited at some much later moment. The traces that our evidence allows us to speak of are interest-determined, interest-carried traces. They live with our interests and with them they change (1932: 211–212).[30]

---

[29] See also Bartlett's Chapter 18 on Halbwachs and collective memory. For Bartlett, a schema has both a conservative and a creative aspect, tending both to homogenize or conventionalize the new, and to support. See now the essays in Saito (2000), Iran-Nejad & Winsler (2000) and Kashima (2000).

[30] For an extended defence of such a dynamic conception of traces, see Sutton (1998: 277–316).

The concept of a schema need not imply a settled structure in order for it to play its required explanatory roles. As an enduring but modifiable set of tendencies or dispositions, a schema may usefully be invoked to explain, for example, the way a story may be normalized in the remembering and retelling, with the schema driving easy inferences to uncertain or untold parts of the story. Cognitive-psychological accounts of the schema were implemented in connectionist models in the 1980s, with the history of past processing being "stored" in the (enduring but modifiable) matrix of connection weights of the neural network, and thus influencing the processing of new and related input (Rumelhart *et al.* 1986). And more recently, cognitive anthropologists have found this tradition a useful way of modelling both the "centripetal" forces of cultural reproduction and the competing "centrifugal" processes of variation and inconsistency: Claudia Strauss and Naomi Quinn, for example, take the connectionist version of schema theory to show how cultural learning produces responses which are permeated by tradition and yet not rigidly repetitive. Remembering occurs on the spot, in a context, and yet can be guided (without being determined) by cultural norms: because connectionism emphatically rejects a linguistic model of internal memory, it's easier to see that the traces culture leaves on and in individual brains and bodies are not downloaded *copies* of any specific cultural instructions, but rather flexible and particular action-oriented responses (Strauss & Quinn 1997: Chapter 3).[31] The dynamics of intrapersonal thoughts, feelings, and motives, conclude Strauss and Quinn, may be quite different from those of extrapersonal messages and practices, even if we accept that the boundaries between the two realms are permeable (1997: 8 and *passim*). This is a key point which it's not easy to hold onto in the excitement of the new embedded cognition movements, and to which I'll return in conclusion.

Distributed Cognition and External Memory

It's no accident that memory is at the heart of recent work on "the extended mind." On top of the connectionist focus on the plasticity of superpositionally stored memory traces, theorists like Andy Clark, Mark Rowlands, and Edwin Hutchins examine dynamic forms of interplay between such internal representations and the

[31] See also Strauss & Quinn (1997: 44–47) on how to reinterpret Bourdieu's notion of the *habitus*, as a set of embodied dispositions driving regulated improvisation, into a naturalistic cultural cognitive science. Compare D'Andrade (1995); and Barry Smith's account of Hayek's notion of partial and dynamic cognitive maps, which act not as passive memory stores but as active memory-competence, in Smith (1997).

(natural and social) environment.[32] Linked through what Clark calls "continuous reciprocal causation," brain and world are often engaged in an ongoing interactive dance from which adaptive action results (Clark 1997: 163–166). Tim van Gelder explains the motivation for seeing genuinely cognitive processing as seeping out of the skull:

> Since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three. Interaction between the inner and the outer is . . . a matter of coupling, such that both sets of processes continually influence each other's direction of change (van Gelder 1995: 373).

This idea of the "extended mind" has led Clark *et al*. to argue that memory contents, as well as memory processes, sometimes spread or leak out of the brain and are left in the world (Clark & Chalmers 1998). Just as our unique problem-solving abilities depend most importantly "on our abilities to *dissipate* reasoning" by building "designer environments" (Clark 1997: 180, 191), so our unique abilities to access, manage, and manipulate large bodies of information depend more on the technological and cultural symbolic networks we've constructed to plug ourselves in to (Donald 1991: 269–360; Rowlands 1999). I don't want here either to defend this startling idea against individualist objections, or to suggest that it's compatible only with these connectionist/dynamicist accounts of internal representation. This "active externalism," unlike traditional views about wide content and social externalism, does not violate the general considerations in favour of narrow content mentioned by Frank Jackson in this volume, because we agree with Jackson that knowledge of the world or the past depends on traces left in our region of the world or of time: what the active externalist denies is just that the division between the regions accessible to us, and those not, *must* coincide with the edges of the individual body.[33] The idea of extended memory systems leaves

---

[32] The best single treatment is Clark (1997). See also Hutchins (1995: Chapter 9) and Rowlands (1999: Chapter 6); for an introduction see now Clark (2001: Chapters 7–8), and for the role of philosophy of biology in understanding embodied cognition see Griffiths & Stotz (2000).

[33] See Section 2.4 of Jackson's paper: the region in which traces are accessible to cognition may extend beyond the individual body, as in Jackson's example of a sealed spaceship, or contract within the body. Jackson's notion of an impermeable barrier between self and world needs better motivation. For other objections to the very idea of the extended mind, see Adams & Aizawa (2001). The extended mind thesis admits of stronger metaphysical, and weaker methodological readings: although I'm sympathetic to both, as long as the stronger is understood correctly, most of the argument in this section relies only on the weaker point.

room for, or positively invites, the correlative extension of the cognitive sciences of memory to include historical, technological, and cross-cultural investigations.

The idea that "external memory" or external representation in general is neither merely metaphorical, nor a straightforward expression of more fundamental mental representations, can seem to rest on the claim that the external representations — information in notebooks, for example, which meets certain criteria of accessibility and reliability — are functionally isomorphic to mental representations.[34] But whatever the merits of this claim, it sits oddly with connectionism: if internal representations are "stored" only in an implicit and distributed fashion, to be reconstructed only in a context, they don't look very like, and have quite different causal properties to, the fixed and context-independent symbols written in a notebook (the point is forcefully made by O'Brien 1998). But the case for the extended mind doesn't rest on such a case for functional isomorphism between the inner and the outer. Rather, as Clark makes clear in other moods, the core idea is that quite disparate internal and external elements are simultaneously coopted into integrated larger cognitive systems. The external media on which we rely so much as cognitive scaffolding are for Clark "best seen as alien but *complementary* to the brain's style of storage and computation. The brain need not waste its time *replicating* such capacities. Rather, it must learn to interface with the external media in ways that maximally exploit their peculiar virtues" (1997: 220).

But different external media for the storage, transmission, and transformation of information have their own peculiar virtues. The various kinds of memory scaffolding which humans have used, from knots, rhymes, codes, and sketches to artificial memory techniques, photographs, books, and computers each have different properties. The resources of the historian and social scientist must again be included in cognitive science.[35] Merlin Donald carefully catalogues the differences between (internal) engrams and the (external) "exograms" which humans have produced since the Upper Palaeolithic era: "unlike the constantly-moving and fading contents of biological working memory, the contents of this externally-driven processor can be frozen in time, reviewed, refined, and reformatted" (Donald 1991: 308–319; 2000). While the enduring and expandable nature of many external symbol systems has indeed altered the informational environment in which brains

---

[34] This is suggested, notably, by the celebrated case of Inga and Otto described by Clark & Chalmers (1998), where the key argument is that if external information is used in ways which would be seen as cognitive if performed internally, there's then no principled reason not to see it as genuinely cognitive. In consequence, Adams & Aizawa (2001) treat the entire argument for the extended mind as resting on such functional isomorphisms.

[35] I've tried to put this into practice in one case study in Sutton (2000).

develop, media theorists remind us that not all such systems are designed to hold information permanently in context- or medium-independent fashions, and that not all those systems which are designed to do so actually succeed (see for example Klein 1997; Tofts & McKeich 1997; and in art history, Kwint 1999).[36]

Yet the communal, transmissable bodies of knowledge which do pervade human culture are not merely convenient holding-pens for information too unwieldy to store internally. Instead, they are essential tools for thinking which often alter the cognitive processes in which brains, bodies, and external symbol systems are all entangled. As Clark in particular argues, the internalizing of relatively context-free representational formats brings new cognitive possibilities, and burdens:

> By 'freezing' our own thoughts in the memorable, context-resistant, modality-transcending format of a sentence, we create a special kind of mental object — an object that is amenable to scrutiny from multiple cognitive angles, is not doomed to alter or change every time we are exposed to new inputs or information, and fixes the ideas at a high level of abstraction from the idiosyncratic details of their proximal origins in sensory input (Clark 1997: 210).

By thus assimilating particular kinds of props and pivots over the course of cognitive development, we regulate our own minds, imposing an approximation of rigidity and inflexibility on our own mental representations.

In this final section, I've taken a swift tour through some ways of thinking about collective memory and external memory which might be naturalistically acceptable. I've programmatically suggested that the post-connectionist notion of the extended mind, carefully developed, shows us how to treat external memory within a general information-processing framework. Whether this turns into a universalizing cognitive science of society, or a contextualist social cognitive science, depends on the care with which particular projects are pursued. I don't want to appear an ebullient optimist, hyping up the current state of interdisciplinary endeavours. Such schemes in the sciences of memory are as yet remorselessly difficult, and (despite actual local interdisciplinary contacts) there is a general sense that those sciences are disconnected from each other. I hope that this provides sufficient motivation for trying to show how such different memory researchers — from neurobiology to narrative theory, from the developmental to the postcolonial, from the computational to the cross-cultural, might one day be able to talk to each other.

---

[36] Again, Michael Leyton's provocative theory of cognition (1992) as the recovery of the past from the shape, in particular from the asymmetries, of objects in the present, may offer a unified framework.

Acknowledgments

I gave earlier versions of this paper in 2000 in a seminar series on "Memory in Science" at the London School of Economics, and at the Representation in Mind conference at Sydney University. Thanks to both audiences for helpful comments, especially to Hasok Chang, Wayne Christensen, Ian Gold, Carl Hoefer, Cliff Hooker, Stephen Jacyna, Doris McIlwain, Daniel Stoljar, Will Sutton, Maria Trochatos, Andrew Warwick, and Elizabeth Wilson.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*, 43–64.

Assmann, J. (1997). *Moses the Egyptian*. Harvard University Press.

Bakhurst, D. (1990). *Social memory in Soviet thought*. In: Middleton & Edwards (1990) (pp. 203–226).

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bickle, J. (1995). Psychoneural reduction of the genuinely cognitive: Some accomplished facts. *Philosophical Psychology*, *8*, 265–285.

Bickle, J. (1998). *Psychoneural reduction: The new wave*. MIT Press.

Brewer, W. (1996). What is recollective memory? In: D. Rubin (Ed.), *Remembering our past* (pp. 19–66). Cambridge University Press.

Campbell, J. (1994). *Past, space, and self*. MIT Press.

Campbell, J. (1997). The structure of time in autobiographical memory. *European Journal of Philosophy*, *5*, 105–118.

Churchland, P. M. (1996). Flanagan on moral knowledge. In: R. McCauley (Ed.), *The churchlands and their critics*. Blackwell.

Churchland, P. M., & Churchland, P. S. (1998). Intertheoretic reduction: A neuroscientist's field guide. In: *On the contrary: Critical essays 1987–1997*. MIT Press.

Clark, A. (1997). *Being there: Putting and brain, body, world together again*. MIT Press.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.

Connerton, P. (1989). *How societies remember*. Cambridge University Press.

Craver, C. F. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science*, *69*(Supplement), 83–97.

Craver, C., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In: P. Machamer, R. Grush, & P. McLaughlin (Eds), *Theory and method in neuroscience* (pp. 112–137). Pittsburgh University Press.

D'Andrade, R. (1995). *The development of cognitive anthropology*. Cambridge University Press.

Dennett, D. (2000). Making tools for thinking. In: D. Sperber (Ed.), *Metarepresentations*. Oxford University Press.

Donald, M. (1991). *Origins of the modern mind*. Harvard University Press.

Donald, M. (2000). The central role of culture in cognitive evolution: A reflection on the myth of the "isolated mind". In: L. Nucci, G. Saxe, & E. Turiel (Eds), *Culture, thought, and development* (pp. 19–38). Lawrence Erlbaum.

Engel, S. (1999). *Context is everything: The nature of memory*. W. H. Freeman & Co.

Fara, P., & Patterson, K. (Eds) (1998). *Memory*. Cambridge University Press.

Fentress, J., & Wickham, C. (1992). *Social memory*. Blackwell.

Fischbach, G. D., & Coyle, J. T. (1995). *Preface*. In: Schacter (1995a) (pp. ix–xi).

Fivush, R. (1991). The social construction of personal narratives. *Merrill-Palmer Quarterly*, *37*, 59–81.

Fivush, R. (1994). Constructing narrative, emotion, and self in parent-child conversations about the past. In: U. Neisser, & R. Fivush (Eds), *The remembering self* (pp. 136–157). Cambridge University Press.

Gold, I., & Stoljar, D. (1999). A neuron doctrine in the philosophy of neuroscience. *Behavioral and Brain Sciences*, *22*, 809–869.

Griffiths, P. E., & Stotz, K. (2000). How the mind grows: A developmental perspective on the biology of cognition. *Synthese*, *122*, 29–51.

Haaken, J. (1998). *Pillar of salt: Gender, memory, and the perils of looking back*. Rutgers University Press.

Hacking, I. (1995). *Rewriting the soul: Multiple personality and the sciences of memory*. Princeton University Press.

Halbwachs, M. (1925/1992). The social frameworks of memory. In: L. A. Cose (Ed.), *Halbwachs, on collective memory*. Chicago University Press.

Hardcastle, V. G. (1996). *How to build a theory in cognitive science*. SUNY Press.

Hilts, P. J. (1995). *Memory's ghost*. Simon & Schuster.

Hoerl, C. (1999). Memory, amnesia, and the past. *Mind and Language*, *14*, 227–251.

Hooker, C. (1981). Towards a general theory of reduction. *Dialogue*, *20*, 38–59, 201–236, 496–529.

Howe, M. L., & Courage, M. L. (1997). The emergence and development of autobiographical memory. *Psychological Review*, *104*, 499–523.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Iran-Nejad, A., & Winsler, A. (2000). Bartlett's schema theory and modern accounts of learning and remembering. *Journal of Mind and Behavior*, *21*, 5–36.

Jackson, F. (2002). *Representation and experience* (this volume).

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Johnson, M. K., & Raye, C. L. (2000). *Cognitive and brain mechanisms of false memories and beliefs*. In: Schacter & Scarry (2000) (pp. 35–86).

Kashima, Y. (2000). Recovering Bartlett's social psychology of cultural dynamics. *European Journal of Social Psychology*, *30*, 383–403.

Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, *63*, 31–47.

Kitcher, P. (1992). *Freud's dream: A complete interdisciplinary science of mind*. MIT Press.
Klein, N. M. (1997). *The history of forgetting*. Verso Books.
Koriat, A., & Goldsmith, M. (1996). Memory metaphors and the real-life/laboratory controversy. *Behavioral and Brain Sciences*, *19*, 167–228.
Kwint, M. (1999). *Introduction: The physical past*. In: M. Kwint, C. Breward, & J. Aynsley (Eds), *Material memories* (pp. 1–16). Berg.
Lagrou, P. (2000). History and memory: The example of the two world wars. *CNRS Cross-Disciplinary Encounters*, at http://www.cnrs.fr/cw/en/pres/compress/memoire/lagrou.htm cited at 22 July 2003.
Le Goff, J. (1992). *History and memory*. Columbia University Press.
Leyton, M. (1992). *Symmetry, causality, mind*. MIT Press.
McClamrock, R. (1995). *Existential cognition*. Chicago University Press.
McClelland, J. L. (1995). Constructive memory and memory distortions. In: Schacter (1995a) (pp. 69–90).
McCormack, T., & Hoerl, C. (1999). Memory and temporal perspective. *Developmental Review*, *19*, 154–182.
MacDonald, S., Uesiliana, K., & Hayne, H. (2000). Cross-cultural and gender differences in childhood amnesia. *Memory*, *8*, 365–376.
Mandler, J. M., & McDonough, L. (1997). Nonverbal recall. In: N. L. Stein *et al*. (Eds), *Memory for everyday and emotional events* (pp. 141–164). Lawrence Erlbaum.
Middleton, D., & Edwards, D. (Eds) (1990). *Collective remembering*. Sage.
Miller, P. J., Potts, R., Fung, H., Hoogstra, L., & Mintz, J. (1990). Narrative practices and the social construction of self in childhood. *American Ethnologist*, *17*, 292–311.
Mitchell, K. J., & Johnson, M. K. (2000). *Source monitoring: Attributing mental experiences*. In: Tulving & Craik (2000) (pp. 179–195).
Mullen, M. K., & Yi, S. (1995). The cultural context of talk about the past: Implications for the development of autobiographical memory. *Cognitive Development*, *10*, 407–419.
Nelson, K. (1993). The psychological and social origins of autobiographical memory. *Psychological Science*, *4*, 7–14.
Nelson, K., & Fivush, R. (2000). *Socialization of memory*. In: Tulving & Craik (2000) (pp. 283–295).
O'Brien, G. (1998). Review of Andy Clark, being there. *Metascience*, *7*, 78–83.
Olick, J. K., & Robbins, J. (1998). Social memory studies: From "collective memory" to the historical sociology of mnemonic practices. *Annual Review of Sociology*, *24*, 105–140.
Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, *59*, 516–548.
Rajaram, S., & Roediger, H. L. (1997). Remembering and knowing as states of consciousness during retrieval. In: J. Cohen, & J. Schooler (Eds), *Scientific approaches to consciousness* (pp. 213–240). Lawrence Erlbaum.
Reese, E., Haden, C. A., & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development*, *8*, 403–430.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76–100.
Roediger, H. L., & McDermott, K. B. (2000). Distortions of memory. In: Tulving & Craik (2000) (pp. 149–164).
Rose, S. (1992). *The making of memory: From molecules to mind*. Bantam.
Rovee-Collier, C., & Hayne, H. (2000). Memory in infancy and early childhood. In: Tulving & Craik (2000) (pp. 267–282).
Rowlands, M. (1999). *The body in the mind*. Cambridge University Press.
Rubin, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes*. Oxford University Press.
Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In: J. McClelland, & D. Rumelhart (Eds), *Parallel distributed processing* (Vol. 2, pp. 7–57). MIT Press.
Saito, A. (Ed.) (2000). *Bartlett, culture, and cognition*. Psychology Press.
Schacter, D. L. (1982). *Stranger behind the Engram: Theories of memory and the psychology of science*. Lawrence Erlbaum.
Schacter, D. L. (1995a). *Memory distortion: How minds, brains, and societies reconstruct the past*. Harvard University Press.
Schacter, D. L. (1995b). *Memory distortion: History and current status*. In: Schacter (1995a) (pp. 1–43).
Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past*. Basic Books.
Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, 54*, 182–203.
Schacter, D. L., & Scarry, E. (Eds) (2000). *Memory, brain, and belief*. Harvard University Press.
Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In: D. Schacter, & E. Tulving (Eds), *Memory systems 1994* (pp. 1–38). MIT Press.
Schaffner, K. (1992). Philosophy of medicine. In: M. Salmon, J. Earman, C. Glymour, & J. Lennox (Eds), *Introduction to the philosophy of science* (pp. 323–339). Prentice-Hall.
Schechtman, M. (1994). The truth about memory. *Philosophical Psychology, 7*, 3–18.
Schudson, M. (1995). *Dynamics of distortion in collective memory*. In: Schacter (1995a) (pp. 346–364).
Smith, B. (1997). Hayek and connectionism. In: S. F. Frowen (Ed.), *Hayek: Economist and social philosopher* (pp. 9–29). Macmillan.
Squire, L. R., & Kandel, E. R. (1999). *Memory: From mind to molecules*. Scientific American Library.
Stoljar, D., & Gold, I. (1998). On biological and cognitive neuroscience. *Mind and Language, 13*, 110–131.
Strauss, C., & Quinn, N. (1997). *A cognitive theory of cultural meaning*. Cambridge University Press.
Suddendorf, T., & Corballis, M. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social and General Psychology Monographs, 123*, 133–167.
Sutton, J. (1998). *Philosophy and memory traces: Descartes to connectionism*. Cambridge University Press.

Sutton, J. (1999). The churchlands' neuron doctrine: Both cognitive and reductionist. *Behavioral and Brain Sciences*, *22*, 850–851.

Sutton, J. (2000). Body, mind, and order: Local memory and the control of mental representations in medieval and renaissance sciences of self. In: G. Freeland, & A. Corones (Eds), *1543 and all that: Word and image in the proto-scientific revolution* (pp. 117–150). Kluwer.

Sutton, J. (2002). Memory: Philosophical issues. In: L. Nadel (Ed.), *Encyclopedia of cognitive science*. Macmillan.

Tofts, D., & McKeich, M. (1997). *Memory trade: A prehistory of cyberculture*. Interface Books.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science*, *2*, 67–70.

Tulving, E. (1999). Episodic vs. semantic memory. In: F. Keil, & R. Wilson (Eds), *The MIT encyclopedia of the cognitive sciences* (pp. 278–280). MIT Press.

Tulving, E. (2000). *Concepts of memory*. In: Tulving & Craik (2000) (pp. 33–43).

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

Tuomela, R. (1995). *The importance of us*. Stanford University Press.

van Gelder, T. (1995). What could cognition be, if not computation? *Journal of Philosophy*, *92*, 345–381.

von Eckardt, B. (1993). *What is cognitive science?* MIT Press.

von Eckardt, B. (1999). Critical notice of Hardcastle, *How to build a theory in cognitive science. Philosophy and Phenomenological Research*, *59*, 221–224.

Warnock, M. (1987). *Memory.* Faber.

Welch-Ross, M. (1995). An integrative model of the development of autobiographical memory. *Developmental Review*, *15*, 338–365.

Welch-Ross, M. (1997). Mother-child participation in conversations about the past: Relations to preschoolers' theory of mind. *Developmental Psychology*, *33*, 618–629.

Young, J. (1993). *The texture of memory: Holocaust, memorials, and meaning*. Yale University Press.

Zelizer, B. (1995). Reading the past against the grain: The shape of memory studies. *Critical Studies in Mass Communication*, *12*, 204–239.

Uncited references

Churchland & Churchland (1998), Jackson (2002) and Tulving & Craik (2000).