Memory Concepts

Yadin Dudai, Henry L. Roediger, III and Endel Tulving

A new science of memory is being shaped in front of our very eyes. It rises on the shoulders of giants: psychology, neurobiology and brain research, computational science, philosophy. Each of these parental disciplines contributes a distinctive vocabulary of terms and acronyms, all embedded to some degree or another in zeitgeists and conceptual frameworks. For the practitioners of the science of memory to be able to properly exploit, and benefit from, the rich multidisciplinarity of methods and findings, they must understand the language and *modus operandi* of their colleagues in other subsdisciplines. Such understanding is a *sine qua non* of the success of the venture.

There does exist some commonality in the languages spoken by memory science's parental disciplines already, of course. But this accord often fails to extend to the level where it would matter most, the level of concepts. The terms may be shared, but the concepts that the terms designate are not. Even some of our most basic terms, such as 'memory' or 'memory systems', stand for rather different entities in different subdisciplines of the science of memory, and sometimes even within a given single one.

The project embodied in this volume is an attempt to tackle the problem of the less-than-perfect communication across the discipline boundaries in the new science of memory. We believe that with proper attention, communication and interdisciplinary understanding can be improved. We also believe that a direct confrontation of the issue at what we regard as the most fundamental level of knowledge and analysis - the conceptual level - is the best approach. The discussion of concepts in the contemporary science of memory is underdeveloped. Some exceptions notwithstanding (e.g. Tulving 2000, Dudai 2002), most practicing students of memory seem to shy away from spelling out and

debating the concepts that form, or should form, the foundations of their own science. This project, we hope, will help them overcome their shyness.

What are concepts: The concept of 'concept' is the focus of intense controversies in philosophy and the sciences of the mind (e.g., Laurence and Margolis 1999, Murphy 2002). We are aware of these long-lasting fundamental controversies, as well as of our inability to contribute to their resolution. However, the irony concealed in the attempt to begin a discussion of concepts without even attempting to conceptualize 'concept', did not escape our notice. We could have adopted Bunge's minimalism (1967): concepts are units of thought. Given the scientific context of our discussion, a more formal attempt to delineate the attributes of CONCEPT[1] is desirable and useful, though also potentially controversial and provocative. We think that the most effective way to proceed is to spell out not only what concepts *are*, but also what they *are not*.

Concepts are:

1. Mental representations that encode sets of attributes that describe real or imaginary classes of items, processes or relations.
2. Always linked to other concepts.
3. Essential elements in mental models and theories.
4. Products of mental models and theories.
5. Ultimately expressed in language.

Concepts are not:

1. Entities with spatial coordinates.
2. The terms in language that are used to express them.
3. Items that can be unveiled in experiments (though data can support or refute their validity).
4. Methods or procedures.

[1] Throughout the text we will refer to concepts in SMALL CAPS.

Concepts, conceived in 'mentalese', must be translated into language for the purpose of concretization and communication. This translation is expected to pose an obstacle on the path to shared understanding of nature. Further, scientific disciplines construct languages of some degree of remoteness from natural languages, in which case friction may emerge between the concepts as expressed in the scientific language and the potentially-relevant terms in the natural language. The science of the mind is particularly predisposed to such friction, for, every human mind harbors an intuitive science of the mind even without being aware of it (e.g. see on the role of metaphor in generating and constraining memory models, Roediger 1980). It is possible that this intuitive science of the mind had been shaped by evolution to satisfy autonoetic explanations of certain mental acts, and therefore intuition could make a legitimate guide to some facets of brain output. But science is not allowed to take this assumption for granted. The intimacy of the mind with itself might actually complicate the proper conceptualization of the science of the mind when the analysis shifts to hidden levels of brain and mind.

Core concepts: The aforementioned treatment of concepts does not differentiate among concepts in a language on the basis of importance, frequency, complexity, ontogeny, kinship, etc. Clearly when one wishes to discuss the conceptual framework of a discipline, not all concepts are to be treated equal. Many concepts are expressed as lexical items in natural languages, e.g. SON, COLOR. Others are common to many scholastic disciplines, e.g. CONCEPT, DIMENSION. Still others are endowed with a specific meaning in a given discipline, e.g. CONSOLIDATION, TRANSFER. Our goal here is to identify, delineate and discuss that set of concepts that define the core of the discipline, in this specific case, the science of memory. These are concepts that delineate the domain of knowledge and are necessary for understanding its subject matter, objectives, findings, and models. We define these concepts as the core concepts of the discipline. We do not claim that the core concepts that we have selected comprise a unique, let alone definitive set (and see below). Neither do we claim that mastering the core concepts is a sufficient condition for knowledge; many concepts in natural language and in other disciplines, combined with large bodies of data and theory and their interrelations, are required as well. Knowledge

of the core concepts, however, is promoted by us as a necessary condition for productive scholarship.

Why concepts: Some scholars believe that bothering with concepts is non-essential and even counterproductive. Their major arguments are briefly cited below. We think that this attitude is a privilege that memory scientists can not afford. Further, pragmatism, commonly encountered in those who advance the concept that conceptualization is armchair luxury, actually demands proper attention to concepts. The arguments in favor of concepts are listed below as well.

For concepts:

a. The Common-Ground argument: Neglect of differences in the meaning and usage of concepts could hamper communication, generate confusion, blur research goals, and waste mental and physical resources. This is what Socrates' urges Meno to note:

> *"Socrates: ... Does anyone know what a part of a virtue is, without knowing the whole?*
>
> *"Meno: I suppose not.*
>
> *"Socrates: No, and if you remember, when I replied to you about shape just now, I believe we rejected the type of answer that employs terms which are still in question and not yet agreed upon.*
>
> *"Meno: We did, and rightly.*
>
> *"Socrates: Then please do the same."*

b. The Differentiation argument: Concepts are important in the differentiation of a discipline, shaping its identity, delineating its subject matter and guiding its research programs. This is particularly important in young disciplines, and in those that tend to be data-driven, like much of contemporary neuroscience. The contribution of concepts, for example NATURAL SELECTION and EVOLUTION, to the evolution of modern biology provides an apt example (Mayr 1982). Another striking example is the role of concepts in defining classical physics, modern physics, and the transition between them (Holton 1973). It is noteworthy that although 'interdisciplinarity' has attained the status of a cultural mantra in contemporary culture, the truth is that differentiation into disciplines is critical in

ensuring professionalism and in-depth knowledge. It is multidisciplinarity, rather than interdisciplinarity, that is the engine of modern science; therefore, the concepts of each contributing discipline must be mastered before multidisciplinarity consolidates. And last but not least, as a fringe benefit, conceptualization could assist in tracing and highlighting the historical roots of the evolving discipline, which sometimes may prevent one from reinventing the wheel rather than improving upon it.

c. The Facilitation argument: Conceptualization facilitates the ongoing functionality of the discipline. It does so in multiple ways. It can induce order and coherence, even if only temporary ones, in what otherwise is a rather confusing body of data. It can guide to productive experiments and models. It sometimes provides the opportunity for thought-experiments that might filter out superfluous experiments and faulty theories.

d. The Interlevel argument: The science of memory engages extensive inter-level analysis (Dudai 1992). In this type of analysis, a major goal is to produce correspondence rules ('translation rules', 'bridge laws'), which enable shifts from the terminology of one level of analysis to another, and permit formulation of systematic relationships between findings at one level to those at the other. This becomes difficult if the concepts at each level are not defined and their intercorrespondence clarified. An example from the science of memory illustrates the issue: what is the relevance of MEMORY in cellular neuroscience to MEMORY in cognitive neuroscience? Without addressing this, the relationships of findings at the cellular to the cognitive level, and vice versa, remain fuzzy at best.

e. The Real-thing argument: Methods, techniques, fashions, models, all come and go. Core concepts are with us to stay. Physics will always have ENERGY and the science of memory LEARNING. Familiarization with core concepts hence promotes identification of the invariant, elemental problems in the field.

Against concepts:

a. The Mental-Prison argument: Defining concepts may impose our cognitive limitations on nature. Burke (1757) defined it succinctly: *"For when we define, we seem in danger of circumscribing nature within the bounds of our own notions."*

b. The Ignorance argument: This is related to *a* above, implying that since we are ignorant about most of nature, we could prematurely adopt misnomers or erroneous concepts, only to propagate further ignorance.
c. The Fuzziness argument: Since in reality very few concepts attain the decisiveness demanded from a unique definition of the entity in question, and since often over time exceptions or contradictions are found to interpretations of 'established' concepts, the entire process of conceptualization might be just a useless intellectual game.
d. The Pragmatics argument: The history of science includes many cases in which signal discoveries were made in the absence of conceptualization, often serendipitously. If this pragmatism works well, why bother?

All in all, our view is that the *'for'* far outweigh the *'against'* arguments, though the latter, especially the ones referring to cognitive limitations and ignorance, should not be belittled, let alone neglected.

The concepts of the science of memory: The set of concepts covered in this volume is the outcome of an iterative exchange among us. We also sought input along the way from invitees to this project. In brief, the concepts in this set are: *a.* First and foremost, concepts as defined above, not methods, experimental paradigms, findings or research objectives. *b*. Deemed by us to express basic ideas in the science of memory that are obligatory for appreciating the discipline. *c.* Elementary or first-order, as opposed to complex or higher-order (e.g. TRANSFER VS. POSITIVE TRANSFER, NEGATIVE TRANSFER, ANALOGICAL TRANSFER; note that WORKING MEMORY and MEMORY SYSTEMS are considered here elementary though they are complex in natural language). In addition: *d.* Even if used in natural language or in folk psychology, they require precise definition, or acquire specific meaning in the science of memory. *e.* Even if currently used predominantly only in some subdisciplines of the science of memory, their discussion by other subdisciplines is highly valuable for the development of knowledge in the discipline as a whole. All in all, we therefore propose to consider this set as core concepts.

On selection criteria: A few additional considerations will further explain why were the 16 key words in the current project selected to designate the core concepts. Why MEMORY and LEARNING, CODING & REPRESENTATION, PLASTICITY, CONTEXT, ENCODING, WORKING MEMORY, CONSOLIDATION, PERSISTENCE, RETRIEVAL, REMEMBERING, TRANSFER, INHIBITION, FORGETTING, MEMORY SYSTEMS, PHYLOGENESIS & EVOLUTION - but not any of the many other terms that dot the pages of learned journals and books, and make their multi-colored appearances on countless power-point presentations at meetings and conferences? Where are *amnesia, amygdala, artificial neural networks, compensation, classical and instrumental conditioning, CREB, emotional memory, experimental extinction, false memory, functional neuroimaging, gene expression, habit, habituation, hippocampus, imprinting, learning set, levels of processing, limbic system, long-term potentiation, medial temporal lobe, metamemory, priming, prospective memory, punishment, reward, sensitization, sensory memory, skill, and state-dependent learning*, to name just a few?

Many popular memory terms received short-shrift because, although at first sight they may appear to refer to concepts, they mostly represent categories that correspond to what 'concepts are not' (see above). These sorts of terms were classified as 'non concepts' for the present purposes. Some of them are labels of methods or procedures, some refer to tasks, some to physical objects, and still others designate empirical findings, models, and other sorts of things that one can intuit as not quite right. For example, *conditioning*, with its many subcategories, did not make the grade because it is easy to think of it primarily as a method, or procedure, or a task. The same is true of *functional neuroimaging*, and *priming*. *Amygdala* and *hippocampus*, and all other anatomical terms, were set aside, because they refer to physical objects, entities with spatiotemporal coordinates. Finally, terms such as *false memory*, *imprinting*, and *state-dependent learning* failed the grade, because they connote particular kinds of experimental findings more than they point to concept.

Now, we fully realize that the distinction we are making here between concepts and 'non-concepts' is rather easy to criticize. After all, many concepts are closely tied to things such as methods and findings. *Encoding*, *consolidation*, and *transfer*, which appear in our list of final 16, are concepts by definition, but all of them also point to particular methods, and

tasks, and findings. Yet, *conditioning* and *priming*, which are not on the list, also imply particular methods and tasks, and particular kinds of empirical findings. Therefore, it is easy to ask why they should be classified differently. While admitting to the inevitability of a certain degree of peremptoriness in drawing up the rules of this new game - conceptual analysis of the science of memory - we hope that additional considerations will help.

One such consideration consisted in the idea that the referent of a term should explain something, rather than only describe it. We know that this putative distinction will open another hornet's nest, because countless learned sages of the past have not been able to agree on the reality or usefulness of the distinction. Nevertheless, it did turn out to help us in our deliberations. When we could not see much explanatory surplus meaning in the definition of a term, and decided that it 'just described' something, we skipped it. This is why terms such as *Alzheimer's disease*, *amnesia*, *habit*, *reward*, and *skill* did not make it to the final team.

A third difference between our concepts and 'non-concepts', or perhaps just a corollary of the second, had to do with the requirement that a concept do a reasonably good job of answering certain kinds of questions that crop up in our science all the time, even if implicitly. These are questions such as 'What (exactly) do you mean when you use the term X?' 'What is X?' and 'What does X do?' Because we are discussing concepts, we note that the 'do' in the last question refers to the (explanatory) role that a given concept is meant to play in a larger scheme of things, such as a model or 'theory'. We could imagine plausible answers to these questions asked about the 16 core concepts, whereas such images would be more difficult to come up with in situations where X refers to venerable expressions such as *association*, or *skill*.

The aforementioned considerations did eliminate a number of potential candidates for core concepts, but certainly not all. There are lots of viable candidates left. *Attention*, *binding*, *motivation*, *reinforcement* and many others quickly come to mind as examples. These will have to wait their turn. Some of these, we trust, will be covered in the course of the discussion of the core concepts as defined here; others may get their due when the science of memory project moves on to methods, facts, and theories; and still others may triumph in

the end as dark horses of the present set. Lastly, we are well aware of the fact that idiosyncrasies are not eliminated by group effort. Paradoxically, they sometimes become reinforced via unwarranted consensual self-contentment. It is quite possible that other teams will come up with somewhat different lists, even if they use similar selection rules. The modest claim we wish to make about the present set is that each term captures an important idea about learning and memory, and that these 16 core terms represent a good starting place.

On the exposition of concepts in this volume: The order in which the set of core concepts appears in this volume is not a given. This order may be construed as reflecting a bias toward an ontogenetic agenda, where 'ontogeny' implies the unfolding of the life-history of a platonic memory item. Actually, this coupling of order-of-presentation to order-of-development is mostly a marriage of convenience. After introducing the two meta-concepts of MEMORY and LEARNING, the list proceeds to concepts referring to memory formation (CODING & REPRESENTATION, PLASTICITY, CONTEXT, ENCODING, WORKING MEMORY), maturation and use (CONSOLIDATION, PERSISTENCE, RETRIEVAL, REMEMBERING, TRANSFER), and lack of availability or loss (INHIBITION, FORGETTING). The list concludes with evolutionary issues (MEMORY SYSTEMS, PHYLOGENESIS & EVOLUTION). We do expect debates about this order even if the biography of a memory item is accepted as a convenient guiding principle. For example, psychologists and neuroscientists may differ on the preferred position of WORKING MEMORY in the list, depending on whether they consider it to be primarily related to the formation or to the use, respectively, of a memory item.

On concepts in an emerging discipline: Pluralism is the ambrosia of science. This appreciation is already connoted by our remark above that other teams might come up with somewhat different lists of core concepts. Even in a mature, paradigmatic science, pluralism impels revaluation of paradigms and ultimately paradigm shifts. Intellectual polyphony is particularly essential in younger disciplines, while bodies of core concepts, research programs, working models and heuristic theories are still being shaped on the basis of the contribution of the source disciplines. A tension is therefore generated between the wish to define and cultivate the identity of the new discipline and the need to

remain tuned to what other disciplines have to offer. We are aware of this tension. It is definitely not the intention of this volume to artificially impose a unified view of what each of the core concepts of the science of memory means. Rather, the idea is to clearly present what the different concepts mean to experts from different backgrounds, hoping that this will result in better communication and mutual understanding, less fuzziness, and possibly improved research programs, models and theories. This intention is reflected in the structure of the enterprise: for each concept, at least two position papers are presented, followed by an integrative essay. The latter is hence not a jury verdict; it is only an interim report on an intellectual dialogue. The position papers and the integrative attempt should be considered as a trigger for further contemplation of what the science of memory is, and a modest contribution to the maturation of our science.

On concepts as concepts: Finally, glancing at the selection, we can not refrain from noting that the list reflects some basic controversies in the sciences of concepts and mind, which in turn reflect on the science of memory. Without exceeding too much the scope of this discussion, or promising answers that we can not deliver, two selected notions are noteworthy. Do the concepts represent natural or artificial types? By natural type we mean a class honored by nature in the absence of obligatory intervention of the human mind. This issue is manifestation of the good-old debate on the role of *a-priori* and *a-posteriori* (valuable concepts per se) in the interaction of our mind with the world. We leave it for the reader to take this complex issue from here, only adding that the terms selected for discussion in this book possibly represent a spectrum of more-natural to more-artificial type of entities. A second, related issue is whether the concepts represent distinctly segregated types of entities in the world, or only prototypes formed in our mind by statistical analysis of properties. The first view is deterministic, the latter statistical. If statistical, we should not aspire to uncover sharp boundaries in taxonomies, e.g. taxonomy of MEMORY SYSTEMS. But the science of memory - not unlike most other scientific disciplines - is still far away from the point where such fundamental problems of classification impose practical limitations on knowledge.

References

Bunge M (1967/1998) Philosophy of Science, Vol. 1, Transaction Publishers, New Bruswick, NJ.

Burke E (1757/1990). Introduction on taste. In: A philosophical enquiry into the origin of our idea of the sublime and the beautiful. Oxford University Press, Oxford.

Dudai Y (1992) Why 'learning' and 'memory' should be redefined (or, an agenda for focused reductionism). Concepts Neurosci. 3, 99-121.

Dudai Y (2002) Memory from A to Z. Keywords, concepts, and beyond. Oxford University Press, Oxford.

Holt G, Brush SG (1973). Introduction to the concepts and theories in physical science. 2nd ed. Addison-Wesley, Reading, MA.

Laurence S, Margolis E (1999) Concepts and cognitive science. In: Concepts. Core readings (eds. Margolis E, Laurence S), pp. 3-81, MIT Press, Boston, MA.

Mayr E (1982) The growth of biological thought. Harvard University Press, Cambridge, MA.

Murphy GL (2002) The big book of concepts. MIT Press, Cambridge, MA.

Plato, *Meno* 79c,d, in: The Collected Dialogues (eds. Hamilton E and Cairns H), Princeton University Press, Princeton, NJ.

Roediger HL (1980) Memory metaphors in cognitive psychology. Mem & Cog 8, 231-246.

Tulving E (2000) Concepts of memory. In: The Oxford Handbook of Memory (eds. Tulving E, Craig FIM), pp. 33-43, Oxford University Press, NY.