Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

The Artificial Memory of Mr. Polly: Memory Simulation in Databases and the Emergence of Knowledge

Gisela Susanne Bahr (biomedical engineer), CARL (Cognition Applied Research Lab) Florida Institute of Technology, Melbourne, Florida U.S.A. Email: <gsbahr@outlook.com>.

William H. Allen (research scientist), Harris Institute for Information Assurance (HIAI), Florida Institute of Technology, Melbourne, Florida U.S.A. Email: <wallen@fit.edu>.

Philip J. Bernhard (educator, research scientist), Department of Computer Science and Cyber Security, Florida Institute of Technology, Melbourne, Florida U.S.A. Email: <pbernhar@cs.fit.edu>.

Stephen Wood (ocean engineer), Underwater Technologies and Robotics, Ocean, Marine and Environmental Systems, Florida Institute of Technology, Melbourne, Florida U.S.A. Email: <gwood07@harris.com>.

Manuscript received 4 February 2016

ABSTRACT: Human memory may be characterized by five dimensions: 1) large capacity; 2) associativity; 3) diversity of memory systems; 4) change over time; and 5) the unified memory experience. The organization and multidimensionality underlying memory can be represented with set theory. This offers a new mathematical perspective, which is the foundation for the cognitive memory architecture Ardemia. The authors present a relational database implementation of Ardemia that supports the artificial memory of Mr. Polly, the main character in H.G. Well's novel *The History of Mr. Polly*. In addition to the implementation of Mr. Polly's artificial memory using TimeGlue, his memory is probed with a collection of everyday memory queries that are related to temporal and schema knowledge. The investigation of Mr. Polly's knowledge suggests an alternative representation of schemas; rather than fixed structures or explicit associations, it is possible to model schemas as the results of the interaction between existing knowledge and remembering.

Human Long-term Memory is an example of naturally occurring big data. It is vast, probably in the order of Exabytes and can last up to a lifetime [1]. Given the rapid development of technology and the capabilities of big data storage and querying, we asked the question whether we could create a system of artificial memory that resembled the long-term memory of a person. Such a system would be instrumental for conducting research on memory on realistic time and capacity scales. For example, artificial memory could be created, cloned and systematically manipulated to investigate memory phenomena: This may include the study of forgetting mechanism, simulating how and what we forget, or an investigation of the emergence of time and schema knowledge. In addition the use of artificial memory as a simulation environment and research tool, future applications may create personal artificial memory in the form of mobile or wearable devices. They may assist users as a cognitive multimedia support systems for reviewing

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

and searching of memory data, privately and as needed in life-logging or quantified-self applications.
Given that this was a first attempt of creating a human-like memory system with relational database technologies, it seemed advisable to avoid the ethical complications and privacy concerns of working with the memories of a real person. Instead we investigated a fictitious character, Mr. Alfred Polly, the main character in the novel *The History of Mr. Polly* by H.G. Wells (Fig. 1). Over 100 years after the original publication, we brought Mr. Polly's memory to life with the goal to investigate his everyday memory.

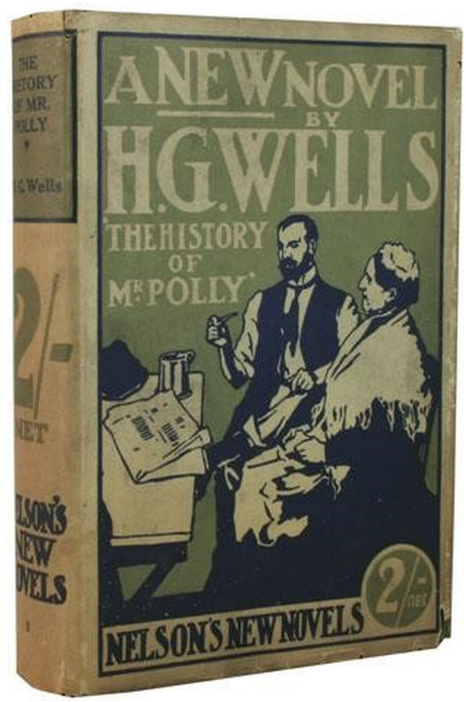

Fig. 1. The first edition of H.G. Wells' *The History of Mr. Polly* published by Nelson, 1910. Public domain.

This article has two main parts in addition to the introduction and the conclusions. The first part presents an overview of the memory simulation architecture and environment Ardemia [2], with focus on two unique aspects that are related to temporal processing. The second part of this article presents the artificial memory system of Mr. Polly and an investigation of his everyday memory for time-related information and schema knowledge.

Part 1: Ardemia

The multidimensionality of long-term memory can be conceptualized with set theory and has been implemented in a long-term memory architecture, called Ardemia. This part of the article, presents Ardemia's conceptual foundations in set theory, including the Time Universe of Memories, and a relevant implementation aspect, TimeGlue. We chose to highlight time-related features of the architecture, because of their importance for the representation for temporal and schema memories presented in part 2.

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

Five Dimensions of Long-term Memory
Human long-term memory can be described with five dimensions:

- **Capacity**, which refers to the vast storage volume of human long term memory
- **Associativity**, which refers to the intrinsic connectivity of memories and memory systems
- **Diversity**, which refers to multiple, related long term memory systems
- **Unified memory experience**, which refers to the seamless integration of the memory, systems during recall
- **Change over time**, which refers to the fact that human long term memory changes constantly as new data are added and older data or memories are modified or forgotten

This next section presents how this multidimensionality can be conceptualized with set theory and modeled as the Time Universe of Personal Memory.

Human Memories as Sets in the Time Universe
Set theory was first formalized by German mathematician Georg Cantor. It is concerned with the investigation of sets and their properties. Cantor's defines,

> By set we mean any collection M into a whole of definite, distinct objects m (m are the "elements" of M) of our perception [Anschauung,] or of our thought [3].

Paraphrased a set represents a *unified whole* that is constructed from individual elements that are conceivable by the human mind. The set concept can be readily adapted to define memory. Like a set, "the whole of a person's long term memory is a collection of conceivable definite, distinct experiences, facts and knowledge."

How memory is organized has been a matter of empirical investigation and it appears that memory is composed of multiple overlapping systems as seen in Fig. 2. These systems can be viewed as sets of related memories that are implicit, explicit, episodic, semantic, sensory, and procedural in nature. The domain of these memory sets is the whole of a person's memory in its current state. This can also be called a person's memory universe. In addition to the personal memory universe, Fig. 2 captures that the systems (sets) are related. Set intersections indicate, for example, that skill memory includes sensory (modality) memories, such as the unique feel of one's usual keyboard while playing the piano or typing. Within the context of a person's memory universe, the organization of the memory systems as sets and their relatedness can be formalized as propositions of their functional associativity. Some examples are,

- *Implicit and explicit memories* are mutually exclusive. Hence, their set representations do not intersect.
- *Episodic memories* are remembrances of self that we can remember. Hence, the episodic set intersects with the explicit set.

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

- *Semantic memories* consist of knowledge that can be recalled implicitly or explicitly. Hence the semantic set intersects with implicit and explicit sets.
- *Sensory memories* overlap in part with all memories but are not necessarily contained by them.

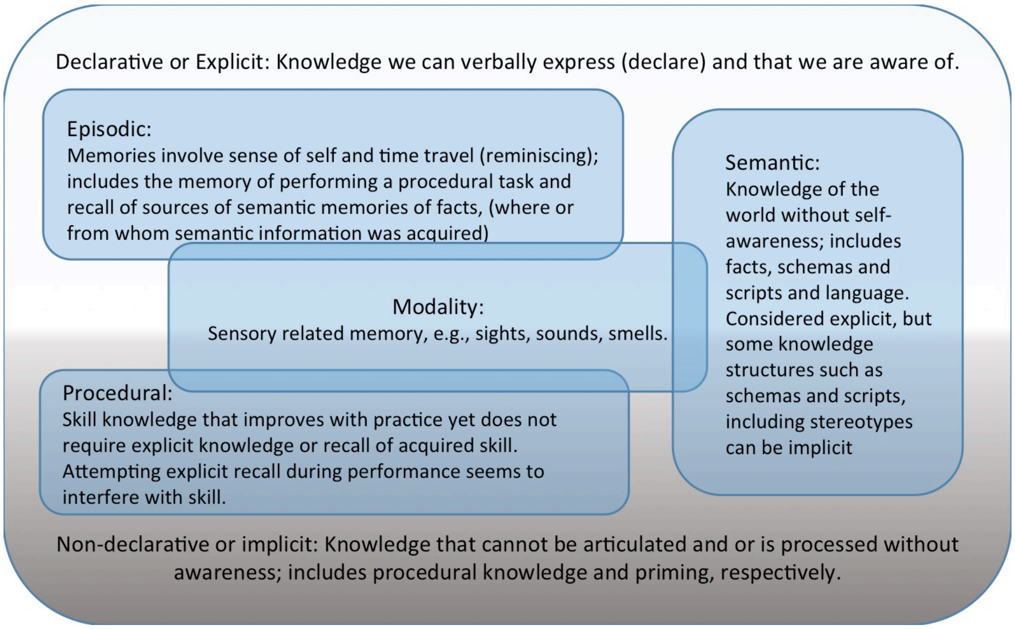

Fig. 2. Long-Term Memory Systems as related sets in the Personal Memory Universe. (© Gisela S. Bahr)

It is evident that the associations between elements and sets are function based rather than explicitly drawn (as in a graph or a network for example). Therefore the number of possible combinations, the expressible associativity, is only limited by the functional relationships and operation performed on the data. A full functional analyses is beyond the scope of this article but a relevant observation is that sets can be related to each other and that based on these relationships, elements can be drawn from different sets to form a new set. Such constructed sets (called relations) are similar to remembering in such a way that recall is the process of reassembling data from multiple systems (sets) into a coherent memory (a new set) without apparent effort or awareness of these processes.

The Time Universe of Memories

Modeling human memory systems as interrelated sets embodies the dimensions large capacity (a person's memory universe), diversity of systems (multiple sets), associativity (functions) and the unified memory experience (relations), but not change over time. Simply put, the current state of the memory universe is mapped to a fixed point in time and cannot be modified as a result of this mapping. If we imagine time as a continuous line, then adding or changing memories requires two things, (1) the expansion of the personal memories to include multiple states and (2) the mapping of these states to different points on the timeline. As such, a time-sensitive memory

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

system may be constructed from multiple versions of the personal universe. These distinct versions exist in parallel and map to corresponding points on a temporal axis; collectively, the versions of person's universe form a new and greater context, which is called Time Universe of Memories. In this Time Universe, personal memories are no longer static but express dynamic, temporal phenomena like forgetting, reorganization and knowledge acquisition. Hence the dynamic nature of memory emerges in the Time Universe by animating the various states of personal memory along the time axis.

Viewing Human Long-Term Memory from a set-theoretical perspective creates a new mathematical approach for the modeling and simulation of memory. This approach can be related to systems designed for storing and managing big data that are based on set theory. The next section presents an implementation aspect using an appropriate, set-theory related database approach that is based on the relational model of data by E.F. Codd [4].

TimeGlue

To create an artificial memory architecture, the majority of the memory dimensions can be readily modeled and implemented in a relational database management systems, such as MySQL or MSSQL. For example, large capacity is an inherent feature of databases; associativity is defined by functions and results from operations on existing data; systems diversity is represented by different datasets (tables) and datatypes; and the unified memory experience emerges during query processing as the construction of new relations from existing data. Likewise, change over time is an integral aspect of modern database solutions and multiple approaches exist to managing data over time. A review of over 20 temporal models led us to the design of a temporal variable called TimeGlue. (For a list of modeling approaches see Appendix A.)

TimeGlue is instrumental for the modeling of the temporal axis of the time universe. It was inspired by Pavlov's contiguity principle, which states that perceptions and events occurring at the same time are linked in memory [5]. As such TimeGlue represents the time of encoding in natural memory and was operationalized as a timestamp that attaches itself to every memory in every system at the time of encoding (transaction time).

TimeGlue supports the construction of a temporal axis, because timestamps do not repeat and can be ordered to form a linear axis. The relationship between TimeGlue and each memory datum is unique (one-to-one) within each system and are shared across systems. Sharing a timestamp across systems creates an emergent phenomenon, which is the linking of systems based on transaction time. This is instrumental during simulated remembering, because TimeGlue permits the synchronization and coordination of simulated memory systems.

In addition to temporal ordering and associativity, TimeGlue supports logical time, which is constructed from inferences based on existing data. Logical time can involve temporal relationships (first, last, before, after or "around some date"), intervals, frequencies, etc.

The next part of this article presents the implications of the temporal modeling for the probing of an artificial memory. The focus is on temporal memory and schema emergence.

doi: 10.1162/LEON_a_01441

Part 2: Mr. Polly's Artificial Memory

This part consists of a brief introduction to Mr. Polly's memory and the first attempt to probe a person's artificial memory for temporal and schema knowledge using a series of everyday memory probes.

The person whose memory was implemented in Ardemia belongs to the main character of a novel by H.G. Wells, Mr. Alfred Polly. We sampled a period during his early twenties while he worked in a small shop in Cambridge, England, selling men's clothing and shoes. Wells describes him as a young man who is less interested in his job and more concerned with adventure books and the opposite sex.

For the primary data population, memories were extracted from the book. They involved memories of people, events and mundane tasks, such as getting up, eating meals, talking to colleagues, shopping, etc. Data also included plausible emotions that could be reasonably inferred from Mr. Polly's personality. Secondary sources supplied historical and weather data.

Asking Mr. Polly Questions

To ask Mr. Polly questions, it was necessary to probe his artificial memory. Because the number of possible questions is only limited by one's imagination, we arbitrarily selected twelve probes into temporal and schema knowledge.

Unlike temporal knowledge, which seems to be an intuitive property of memory, the concept of cognitive schemas has not been discussed until Sir Frederic C. Bartlett first defined the term in his theory of remembering in 1932 [6]. His ideas about knowledge abstracted from prior experiences were largely forgotten until Marvin Minsky referred to schemas as the basis for intelligent behavior [7]. To date cognitive schemas and scripts are considered one of the hallmarks of cognition and memory organization [8].

One might think of a schema as a mechanism for data reduction, which allows us abstract generic information and use it as a blueprint for the past and expectation of the future. Details are forgotten and typical patterns are recalled instead. What remains a topic of investigation is how schemas are represented in memory.

Given the intrinsic link between time and emergence of typicality, we focused on a selection of memory probes that involve temporal inferences and schema knowledge. As seen in Table 1, Questions 1–4 tap into Mr. Polly's logical time relating to events in episodic memory and Questions 5–12 inquire about facts that are patterns he may have acquired while working or socializing. These questions are deliberately reflective of everyday tasks, events and remembering. They were coded with structured query language (SQL) [9].

The replies to the questions (query output) are also given in Table 1. Mr. Polly's answers are brief, factual and honest. It is apparent that the response language is choppy and in third rather than first person. Future versions of Mr. Polly's artificial memory will address these limitations.

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

Table 1. Temporal and Schema Queries: Q&A with Mr. Polly's Control Memory

Q1: What have you been dreaming about recently?	sheep, world war 2, India, work
Q2: How did you feel when he first met Sam?	friendly
Q3: Where did you last meet Sam?	teashop
Q4: Who did you meet first? Sam or Lance?	Sam
Q5: How many times per week do you usually pick up your laundry from Mrs.Pipps?	1
Q6: Where did you usually see Ruby?	in his bed
Follow-up question: Sorry to be indiscreet but exactly did you do?	daydreams about Ruby, dreams about Ruby, thinks about Ruby
Q7: Mr. Polly, What item do you sell the most?	Pants
Q8: Mr. Polly, What do you sell the least?	Cufflinks
Q9: Mr. Polly, Do you less gloves or less mufflers?	Gloves
Question10: Mr. Polly, In the springtime do you sell more hats or hankerchiefs?	hankerchief
Q11: What you usually dream about at night?	Sheep, Ruby, work
Q12: What you usually do in the mornings?	wakes up, eats, sleeps

The results of the memory probes indicate that Mr. Polly's artificial memory is capable of producing answers that are not explicitly coded into artificial memory but instead emerge as the result of query processing. It appears that logical time reveals temporal knowledge and schema-like knowledge that can be inferred from instances that have occurred over time. Hence Mr. Polly's memory is not encumbered with the explicit storage of derived patterns but is capable of generating inferences about time and schemas *on the fly*, as the result of query processing.

Conclusion

Schank makes the point that intelligence and higher-order cognition are often recall and memory in disguise [10]. For instance, when answering questions and in conversation, humans rarely create innovative answers, but instead seem to rely on existing knowledge of previous events, general facts, personal preferences, as well as typical patterns and expectations. To store such knowledge, it has been suggested that schemas could be represented by data containers, similar to frames [11] or number of connections [12] that are considered part of memory; the simulations of Mr. Polly's recall suggests an alternative model: schemas could be the results of a data-process interaction.

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

Schemas as abstractions from accumulated experience, created by processes that retrieve and organize existing memories into new knowledge, have a provocative implication for human cognition: Similar to artificial memories in Ardemia, schema-like abstractions, such stereotypes, may be cognitively constructed *on the fly*. Therefore stereotypes and what appear to be rigid preconceptions are potentially malleable cognitive phenomena. As such they can be influenced and possibly changed by acquiring new memories.

The investigation of Mr. Polly's artificial memory has revealed that artificial memory systems are capable of providing intelligent answers, demonstrating schema and temporal knowledge that is not explicitly stored or coded into memory. Therefore simulated remembering is not limited to retrieval of prior knowledge or stored data, but integrates retrieval and computation in response to a specific question. The artificial memory presented here was implemented in Ardemia, a set-theoretically based, relational memory architectures that embodies the five dimensions of Human Long-term Memory: capacity, associativity, diversity, change over time and the unified memory experience.

References and Notes

1. G.S. Bahr and S.L. Wood, "The big data between your ears: Human inspired heuristics for forgetting in databases," in 2015 IEEE Int. Conf. Multimedia and Expo Workshops (ICMEW), pp.1–6, 29 June 2015–3 July 2015.

2. The name "Ardemia" is pronounced (Ar-Dee-Me-Ah) and the phonetic spelling of the acronym RDMA (Relational Data Memory Architecture). The data model and implementation details can be found at <http://www.artificialmemory.org/> or at or <www.gsbahr.com>.

3. J.W. Dauben, *Georg Cantor* (Princeton Univ. Press, 1970), German quote on p.170, translated by G.S. Bahr.

4. E.F. Codd, "A Relational Model of Data for Large Shared Data Banks," *Comm. ACM* **13**, no. 6, pp. 377–387 (June 1970).

5. The simultaneity of events is an inherently subjective experience and based on point of view. This observation is in line with Einstein's special relativity (A. Einstein, "Zur Elektrodynamik bewegter Körper," *Annalen der Physik* **322**, no.10, pp. 891–921 [1905]) as discussed in *Time Machines: Time Travel in Physics, Metaphysics, and Sci. Fiction* by P.J. Nahin (American Institute of Physics, 1993). As a result of the subjective simultaneity, every memory formed by association through TimeGlue is unique and compatible only with its owner and his or her viewpoint.

6. F.C. Bartlett, *Remembering: A Study in Experimental and Social Psychology, 2nd ed.* (Cambridge Univ. Press, 1995), originally published 1932.

7. M. Minsky, "A framework for representing knowledge," in *The Psychology of Computer Vision,* P. Winston, ed. (New York: McGraw-Hill, 1975) pp. 211–77.

doi: 10.1162/LEON_a_01441

8. S.P. Marshall, *Schemas in Problem Solving* (Cambridge Univ. Press, 1995). See also Robert J. Sternberg, *Textbook Cognitive Psychology,* 6th Ed. (Wadsworth, 2011) Chapter 8; Robert W. Weisberg and Lauretta M. Reeves, *Cognition: From Memory to Creativity, 1st Ed.* (Wiley, 2013).

9. The query code is available in text format at <http://www.artificialmemory.org/> or <www.gsbahr.com>.

10. R.C. Schank, Tell Me a Story: A New Look at Real and Artificial Memory. Scribner, 1991.

11. Minsky [7].

12. J.L. McClelland, “Retrieving general and specific information from stored knowledge of specifics,” in Proc. Third Annu. Conf. Cognitive Sci. Society, pp. 170–172, 1981.

13. Codd [4].

Leonardo Just Accepted MS.
doi: 10.1162/LEON_a_01441

Appendix A

In the seminal work on the relational model, Codd [13] considered time-dependent data the greater part of information captured in the relational model. Consequently, temporal aspects of relational databases have been addressed and further developed. Major contributors to this advancement of implementing temporal aspects have been Richard Snodgrass, Hugh Darwen and Chris Date, and Victor Vianu. Over 20 relational data models that include temporal components have been proposed of which some, in addition to concept development, include a more formal temporal relational algebra. See Tables 1, 2 and 3, which list these 20 models. The key to understanding the models is their respective treatment of temporal dimensions. For example, two types of time data that have been universally accepted are transaction time and valid time. Transaction time refers to the time that data was added, modified or deleted in the database. Valid time refers to the period or point in time when the datum is relevant. For example, a datum may be valid in the past, future or present. Using these two time variables we can classify temporal data models into three categories, those that are based on transaction time or valid time or both. The latter are called bi-temporal models.

Table 1. Transaction Time Models

Name/Identifier of Model	**Key Temporal Properties**	**Reference**
DATA	Temporal data is implicit and handled as event-stamped tuples; cannot display sequences.	K.A. Kimball, “The Data System,” Master’s thesis, University of Pennsylvania, 1978.
Postgres DM	Similar to DATA but support sequences.	L. Rowe and M. Stonebraker, “The Postgres Papers,” Technical Report UCB/ERL M86/85, University of California, Berkeley, CA, June 1987.
DM/T	Temporal data is implicit with backlog of history of any changes.	C.S. Jensen, L. Mark and N. Roussopoulos, “Incremental Implementation Model for Relational Databases with Transaction Time,” IEEE Trans. Knowledge and Data Eng. **3**, no. 4, pp. 461–473, 1991.

Table 2. Valid Time Models

Name/Identifier of Model	**Key Temporal Properties**	**Reference**
Time Oriented Database (TOD)	Designed for medical applications. Stores patient's visits, visit intervals and associated data.	R.L. Blum, "Displaying Clinical Data from a Time-Oriented Database," *Comput. Biol. Med*. **11**, no. 4, pp. 197–210, 1981.
Logical Model	Models valid period as activation start and end times.	N. Mahmood, A. Burney and K. Ahsan, "A Logical Temporal Relational Data Model," *JCSI Int. J.Computer Science* **7**, no. 1, pp. 1–9, 2010.
Legol 2.0 Tuples	Temporal data designed to support temporal (historical) ordering and valid time as Start-Stop.	S. Jones, P. Mason and R. Stamper, "Legol 2.0: A Relational Specification Language for Complex Rules," *Information Systems* **4**, no. 4, pp. 293–305, 1979.
Historical Database Model	Temporal data adds two attributes whose respective domains represent indivisible units of time, and Boolean states of valid or not.	J. Clifford and D.S. Warren, "Formal Semantics for Time in Databases," *ACM Trans. Database Systems* **8**, no. 2, pp. 214–254, 1983.
Temporally Oriented Model	Uses snapshots of data and creates sequences indexed by valid time.	G. Ariav, "A Temporally Oriented Data Model," *ACM Trans. Database Systems* **11**, no. 4, pp. 499–527, 1986.
Temporal Relational Model	Valid temporal data involves time invariant valid time (cannot be changed) and time varying data (data that change over time).	S.B. Navathe and R. Ahmed, "TSQL—A Language Interface for History Databases," in *Proc. Conf. Temporal Aspects in Information Systems*, pp. 113–128, France, May 1987.
Sadeghi	Supports Valid time start and stop.	R. Sadeghi, "A Database Query Language for Operations on Historical Data," PhD thesis, Dundee College of Technology, Dundee, Scotland, December 1987.
Sarda	Models valid time as a user defined period, i.e. an attribute in a valid-time separate relation.	N. Sarda, "Algebra and Query Language for a Historical Data Model," *The Computer J.* **33**, no. 1, pp. 11–18, 1990.
Temporal Data Model	Valid time is sequence identified by a time value pair.	A. Segev and A. Shoshani, "Logical Modeling of Temporal Data," in U. Dayal and I. Traiger, eds, *Proc. ACM SIGMOD*, pp. 454–466, San Francisco, CA, May 1987.

Historical Relational Data Model	Valid time can be associated with entire tuple and the individual values of a tuple.	J. Clifford and A. Croker, "The Historical Relational Data Model and Algebra Based on Lifespans," *Proc. Int. Conf. on Data Eng.*, pp. 528–537, Los Angeles, CA, February 1987.
Tansel	Valid time can be supported using time-invariant or varying attributes, that are simple or sets.	A.U. Tansel, "Adding Time Dimension to Relational Model and Extending Relational Algebra," *Information Systems* **11**, no. 4, pp. 343–355, 1986.
Homogenous and Multihomogenous Model	Temporal data includes valid time as elements and as relations; multihomogenous extension support merging of two timestamps into one.	S.K. Gadia and C.S. Yeung, "A Generalized Model for a Relational Temporal Database," *Proc. ACM SIGMOD*, pp. 251–259, Chicago, IL, June 1988.
Temporal Relational Model	Supports valid time with nested timestamps and associated them with attribute values of tuples.	N.A. Lorentzos, "A Formal Extension of the Relational Model for the Representation of Generic Intervals," PhD thesis, Birkbeck College, University of London, 1988.

Table 3. Bitemporal Models

Name/Identifier of Model	**Key Temporal Properties**	**Reference**
Time Relational Model	Temporal data include effective (valid) start and stop time, transaction time start and stop and deletion time.	J. Ben-Zvi, "The Time Relational Model," PhD thesis, Computer Science Department, UCLA, 1982.
Ahn	Temporal data include sequences with individual transaction times and valid time.	R.T. Snodgrass and I. Ahn, "Temporal Databases," *IEEE Computer* **19**, no. 9, pp. 35–42, 1986.
TQuel	Temporal data include insertion time, transaction time of deletion, start and end of valid time.	R.T. Snodgrass, "The Temporal Query Language TQUEL," *ACM Transactions on Database Systems* **12**, no. 2, pp. 247–298, 1987.
Bitemporal Conceptual Data Model	Simplifies previous approaches to attach a set of chronons that represent the validity of the semantically unique tuples within database as well as the query results.	C.S. Jensen, R.T. Snodgrass and M.D. Soo (1995, 2012), "The TSQL2 Data Model," in Snodgrass, R.T. (ed.), The *TSQL2 Temporal Query Language* (Springer, 2012).