Chunks, Schemata, and Retrieval Structures: Past and Current Computational Models

Fernand Gobet[1]*, Peter C. R. Lane[2] and Martyn Lloyd-Kelly[1]

[1] Department of Psychological Sciences, University of Liverpool, Liverpool, UK, [2] School of Computer Science, College Lane, University of Hertfordshire, Hatfield, UK

A recurring question in psychology and cognitive science concerns the expression of theories that are internally consistent and testable. Natural language is unsatisfactory, as theoretical concepts and mechanisms are not stated with sufficient precision (e.g., Newell et al., 1958; Newell and Simon, 1972; Farrell and Lewandowsky, 2010; Jones et al., 2014). Formal and, in particular, computational models avoid the problems of vagueness and under-specification by defining the processes and cognitive mechanisms that occur during a task. They additionally make quantitative and testable predictions, not only about the link between input and output, but also about fine-grained measures such as response times and eye movements. Further, such models can perform complex tasks and, when simulating learning, can use the statistical structure of the environment to help explain behavior.

This Opinion article briefly reviews the extent to which computational modeling has been used to develop theories accounting for the learning and use of chunks, schemata, and retrieval structures. We use the following definitions. A *chunk* is a "meaningful unit of information built from smaller pieces of information" (Gobet and Lane, 2012, p. 541), with the qualification that this information should be of the same kind. A *schema* is "a cognitive structure for representing and retrieving classes of typical situations for which a similar response is required of the learner" (Lane et al., 2000, p. 776). Finally, a retrieval structure is "a set of retrieval cues [that] are organized in a stable structure" (Ericsson and Kintsch, 1995, p. 216). We should point out that there exist plenty of definitions for these terms, which is actually an issue for progress in our understanding. For example, Richman et al. (1991) consider that a retrieval structure is a schema in long-term memory. Even fuzzier is the concept of a "chunk." For example, a chunk is a unit of *declarative* memory in ACT-R (Anderson et al., 2004) and a unit of *procedural* memory in Soar (Newell, 1990), with none of the two meanings corresponding to the definition provided above. For a discussion of the multiple meanings of this term, see Gobet et al. (in revision).

THE PERILS OF USING VERBAL THEORIES

To illustrate the weaknesses of verbal theories, let us consider how Ericsson and Kintsch (1995) applied their long-term working memory (LT-WM) theory to explain the results of Saariluoma's (1989) dictation task. Here, the experimenter dictated the content of a chess position, piece-by-piece, at a rapid rate (typically, one piece every 2 s), and then participants had to reconstruct the position. Saariluoma found that chess experts memorized the board positions taken from games fairly well (more than 80%), but obtained weaker performance with random positions (no more than 60%). Ericsson and Kintsch (1995) explained these results by proposing that strong chess players have acquired a hierarchical retrieval structure corresponding to the 64 squares of a chess board. This structure has two functions: first, it connects individual pieces to their respective

squares, and, second, it allows pieces to be related to each other, representing a position as an integrated hierarchical structure. Thanks to this structure, it is possible to encode information rapidly into long-term memory, and to link it with patterns and schemata.

At first sight, this explanation looks plausible. However, when LT-WM was implemented as a computational model, Gobet (2000) found that recall ranged from 10 to 100% for both game and random positions, depending on parameter settings—essentially all possible outcomes. Being less specified with respect to mechanisms and parameters, LT-WM predicts even more possible outcomes, and is thus non-refutable, a serious problem for a scientific theory.

Of course, it could be the case that Gobet (2000) made incorrect theoretical choices in implementing LT-WM, in the sense that his model did not correspond to what Ericsson and Kintsch (1995) had in mind. Specifically, there exists an indefinite number of possible models that satisfy the verbal description they provided, and some may provide better results than the model implemented by Gobet (2000). However, this is precisely the point made by the authors highlighting the advantages of computational modeling: not enough constraints are provided by verbal theories, and thus too much freedom is left in the way they can be interpreted.

A BRIEF REVIEW OF COMPUTER MODELS

Chunks

Although dominated by verbal theories, the literature includes a number of computational models of chunking. Several models have accounted for expert chess memory and perception (Ellis, 1973; Simon and Gilmartin, 1973; De Groot et al., 1996; Gobet and Simon, 1996a,b; Saariluoma and Laine, 2001) and the use of chunks for chess problem solving (Berliner and Campbell, 1984; Gobet and Jansen, 1994). Implicit learning has led to the development of several models (Servan-Schreiber and Anderson, 1990; French et al., 2011; Lane and Gobet, 2012a; Perruchet et al., 2014). Models have also been developed to account for short-term memory experiments (Robinet et al., 2011; Mathy and Feldman, 2012), spelling (Simon and Simon, 1973), alphabet recitation (Klahr et al., 1983), and verbal learning (Feigenbaum and Simon, 1984; Gobet et al., 2001). Mathematical models have been developed to estimate the most efficient size of chunks (Dirlam, 1972), the number of chunks necessary to reach expertise (Simon and Gilmartin, 1973), or the amount of monochrestic (single-use) knowledge held by experts (Chassy and Gobet, 2011). Rabinovich et al. (2014) used non-linear dynamics to develop a model of hierarchical chunking in the brain.

Some of these models have led to important empirical discoveries. For example, modeling with CHREST (Gobet and Simon, 1996a,b) predicted that, contrary to a widely held opinion at the time, chess experts should show superior memory recall over weaker players even for random positions. To explain this counter-intuitive prediction, Gobet and Simon argued that a model with a larger number of chunks was more likely to find, serendipitously, chunks in a random position than a model with few chunks. The prediction was supported by a meta-analysis of the available data as well as the collection of new data. Interestingly, Vicente and Wang (1998) challenged this explanation by arguing that the chess positions used in the literature still contain constraints (e.g., there is only one white King and at most eight black pawns). They argued that, with positions where all pieces have the same likelihood of being selected ("truly random positions"), the skill effect should disappear. By contrast, simulations with CHREST predicted that, again, there should be a reliable skill effect, albeit smaller than with standard random positions, because the likelihood to find a chunk fortuitously is smaller. The model's predictions were upheld (Gobet and Waters, 2003): not only did the empirical data show a skill effect, but they also were close to the absolute values predicted by the model. This result is theoretically important, because the other mechanisms proposed to explain experts' superiority with game positions (e.g., high-level knowledge, schemata, or retrieval structures) fail to explain their superiority with random and fully random positions (Gobet, 2015). It thus provides direct support for chunking mechanisms. In addition, and importantly, the fact that new predictions were made by a computational model and supported empirically is important, since this rebuts a common criticism that computational modeling only serves to fit data.

Schemata

Unfortunately, there has been little modeling work with respect to schemata (Lane et al., 2000). CHREST is perhaps unique in explaining the processes underpinning the learning and use of schemata (called *templates*) in a variety of domains. Simulations have been carried out in board games (Gobet and Simon, 2000; Gobet, 2009; Bossomaier et al., 2012), diagrammatic reasoning (Lane et al., 2001), implicit learning (Lane and Gobet, 2012a), categorisation (Lane and Gobet, 2014), and agent modeling (Lloyd-Kelly et al., 2014, 2015). When a broader definition of a schema is used, work with neural networks (St. John and McClelland, 1990) and Soar (Laird et al., 1987) address issues related to schemata; Kintsch (1992) proposes some simulations with respect to language.

Retrieval Structures

Overall, the least modeling research has been carried out on retrieval structures. To our knowledge, the only models having tackled this issue used *chunking networks*. Richman et al. (1995) used EPAM-IV to simulate the growth of expertise in the digit-span task, accounting for how an individual was able to memorize up to 106 digits dictated every second. The retrieval structure used in their simulations consisted of digits at the bottom level, chunks (e.g., running times) at the second level, super-groups combining chunks at the third level, and super-group clusters combining super-groups at the fourth level. Gobet (2013) used CHREST to simulate how a chess master was able to memorize several briefly presented chess boards simultaneously. In this model, the retrieval structure consisted of the list of chess world champions. In both cases, the retrieval structures were directly based on the strategies used by the human participants.

CONCLUSION

While a fair number of models have investigated chunking, very little computational work has been devoted to schemata and retrieval structures, although some informal theories exist. This is regrettable, since schemata and retrieval structures are key structures of the human mind, not least because they link together various kinds of knowledge. Why are there so few models? Possible answers include the lack of methodology for developing models (but see Lane and Gobet, 2012b), the technical skills and time required, the difficulty of deciding what constitutes a good model (Roberts and Pashler, 2000), the poor specification of current theories, and also the difficulty in finding suitable data to model. Naively, one could argue that modeling is not necessary, as data from neuroscience will answer all the key questions in the long term. However, this is unlikely to be the case. Data from neuroscience are actually inconsistent and confusing (Uttal, 2011; Guida et al., 2012) and computational models are necessary for making sense of them (Gobet, 2014)!

A striking result of this review is the number and breadth of coverage of models for the chunking network family, whose main representatives are EPAM and CHREST. This should not really come as a surprise and reflects the influence of Herbert Simon, who was one of the first to advocate chunking as a key mechanism of human cognition (e.g., Feigenbaum and Simon, 1962; Simon, 1974) and also carried out influential empirical research on retrieval structures and schemata (Larkin et al., 1980; Gobet and Simon, 1996c). He also studied topics, most notably expertise in chess, for which chunking comes as a natural explanation (e.g.,

Simon and Chase, 1973). Finally, chunks, schemata, and retrieval structures dovetail with his hypothesis that human cognition can be described as a physical symbol system (Simon and Newell, 1976), where symbols are discrete units.

Beyond historical reasons, chunking network models offer several theoretical advantages, including: (a) they provide learning mechanisms; (b) they are organized hierarchically; (c) they include time parameters, making it possible to make precise predictions; (d) they are efficient and scale up (learning hundreds of thousands of chunks can be done in a few minutes; (e) they provide an architecture seamlessly implementing chunks, schemata and retrieval structures; and (f) they can be (and have been) used to account for phenomena in different psychological provinces, including perception, memory, problem solving and decision making, and language.

Without the use of formal models, in particular computational models, progress in understanding chunks, retrieval structures, and schemata will be slow, if possible at all. By highlighting the strengths of computational modeling and the weaknesses of verbal theorizing, we hope to have encouraged other researchers to develop computational models accounting for each of these structures and how they work together.

FUNDING

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. L. (2004). An integrated theory of the mind. *Psychol. Rev.* 111, 1036–1060. doi: 10.1037/0033-295X.111.4.1036

Berliner, H. J., and Campbell, M. (1984). Using chunking to solve chess pawn endgames. *Artif. Intell.* 23, 97–120. doi: 10.1016/0004-3702(84)90006-7

Bossomaier, T., Traish, J., Gobet, F., and Lane, P. C. R. (2012). "Neuro-cognitive model of move location in the game of Go," in *2012 International Joint Conference on Neural Networks* (Brisbane, QLD).

Chassy, P., and Gobet, F. (2011). Measuring chess experts' single-use sequence knowledge: an archival study of departure from 'theoretical' openings. *PLoS ONE* 6:e26692. doi: 10.1371/journal.pone.0026692

De Groot, A. D., Gobet, F., and Jongman, R. W. (1996). *Perception and Memory in Chess: Heuristics of the Professional Eye*. Assen: Van Gorcum.

Dirlam, D. K. (1972). Most efficient chunk sizes. *Cogn. Psychol.* 3, 355–359. doi: 10.1016/0010-0285(72)90012-6

Ellis, S. H. (1973). *Structure and Experience in the Matching and Reproduction of Chess Patterns*. Doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.* 102, 211–245. doi: 10.1037/0033-295X.102.2.211

Farrell, S., and Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Curr. Dir. Psychol. Sci.* 19, 329–335. doi: 10.1177/0963721410386677

Feigenbaum, E. A., and Simon, H. A. (1962). A theory of the serial position effect. *Br. J. Psychol.* 53, 307–320. doi: 10.1111/j.2044-8295.1962.tb00836.x

Feigenbaum, E. A., and Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cogn. Sci.* 8, 305–336. doi: 10.1207/s15516709cog0804_1

French, R. M., Addyman, C., and Mareschal, D. (2011). TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychol. Rev.* 118, 614. doi: 10.1037/a0025255

Gobet, F. (2000). "Long-term working memory: a computational implementation for chess expertise," in *Proceedings of the Third International Conference on Cognitive Modelling,* eds N. Taatgen and J. Aasman (Netherlands: Universal Press), 150–157.

Gobet, F. (2009). Using a cognitive architecture for addressing the question of cognitive universals in cross-cultural psychology: the example of awalé. *J. Cross Cult. Psychol.* 40, 627–648. doi: 10.1177/0022022109335186

Gobet, F. (2013). "Chunks and templates in semantic long-term memory: the importance of specialization," in *Expertise and Skill Acquisition: The impact of William G. Chase,* ed J. J. Staszewski (New York, NY: Psychology Press), 117–146.

Gobet, F. (2014). William R. Uttal: mind and brain: a critical appraisal of cognitive neuroscience. *Minds Mach.* 24, 221–226. doi: 10.1007/s11023-013-9335-1

Gobet, F. (2015). *Understanding Expertise: A Multidisciplinary Approach*. London: Palgrave.

Gobet, F., and Jansen, P. J. (1994). "Towards a chess program based on a model of human memory," in *Advances in Computer Chess* 7, eds H. J. van den Herik, I. S. Herschberg, and J. W. H. M. Uiterwijk (Maastricht: University of Limburg Press), 35–60.

Gobet, F., and Lane, P. C. R. (2012). "Chunking mechanisms and learning," in *Encyclopedia of the Sciences of Learning,* ed N. M. Seel (New York, NY: Springer), 541–544.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends Cogn. Sci.* 5, 236–243. doi: 10.1016/S1364-6613(00)01662-4

Gobet, F., and Simon, H. A. (1996a). Recall of random and distorted positions. implications for the theory of expertise. *Mem. Cogn.* 24, 493–503. doi: 10.3758/BF03200937

Gobet, F., and Simon, H. A. (1996b). Recall of rapidly presented random chess positions is a function of skill. *Psychon. Bull. Rev.* 3, 159–163. doi: 10.3758/BF03212414

Gobet, F., and Simon, H. A. (1996c). Templates in chess memory: a mechanism for recalling several boards. *Cogn. Psychol.* 31, 1–40. doi: 10.1006/cogp.1996.0011

Gobet, F., and Simon, H. A. (2000). Five seconds or sixty? presentation time in expert memory. *Cogn. Sci.* 24, 651–682. doi: 10.1207/s15516709cog2404_4

Gobet, F., and Waters, A. J. (2003). The role of constraints in expert memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1082–1094. doi: 10.1037/0278-7393.29.6.1082

Guida, A., Gobet, F., Tardieu, H., and Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain Cogn.* 79, 221–244. doi: 10.1016/j.bandc.2012.01.010

Jones, G., Gobet, F., Freudenthal, D., Watson, S. E., and Pine, J. M. (2014). Why computational models are better than verbal theories: the case of nonword repetition. *Dev. Sci.* 17, 298–310. doi: 10.1111/desc.12111

Kintsch, W. (1992). *Comprehension*. Cambridge, MA: Cambridge University Press.

Klahr, D., Chase, W. G., and Lovelace, E. A. (1983). Structure and process in alphabetic retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 462–477. doi: 10.1037/0278-7393.9.3.462

Laird, J. E., Newell, A., and Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artif. Intell.* 33, 1–64. doi: 10.1016/0004-3702(87) 90050-6

Lane, P. C. R., Cheng, P. C. H., and Gobet, F. (2001). "Learning perceptual chunks for problem decomposition," in *Proceedings of the 23rd Meeting of the Cognitive Science Society* (Mahwah, NJ: Erlbaum), 528–533.

Lane, P. C. R., and Gobet, F. (2012a). "CHREST models of implicit learning and board game interpretation," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin), 148–157.

Lane, P. C. R., and Gobet, F. (2012b). A theory-driven testing methodology for developing scientific software. *J. Exp. Theor. Artif. Intell.* 24, 421–456. doi: 10.1080/0952813X.2012.695443

Lane, P. C. R., and Gobet, F. (2014). Evolving non-dominated parameter sets for computational models from multiple experiments. *J. Artif. Gen. Intell.* 4, 1–30. doi: 10.2478/jagi-2013-0001

Lane, P. C. R., Gobet, F., and Cheng, P. C. H. (2000). "Learning-based constraints on schemata," in *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society* (Philadelphia, USA: Erlbaum), 776–781.

Larkin, J. H., McDermott, J., Simon, D. P., and Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science* 208, 1335–1342. doi: 10.1126/science.208.4450.1335

Lloyd-Kelly, M., Gobet, F., and Lane, P. C. R. (2015). "The art of balance: problem-solving vs. pattern-recognition," in *7th International Conference on Agents and Artificial Intelligence (ICAART)* (Lisbon).

Lloyd-Kelly, M., Lane, P. C. R., and Gobet, F. (2014). "The effects of bounding rationality on the performance and learning of CHREST agents in Tileworld," in *Proceedings of AI-2014: Thirty-fourth SGAI International Conference on Artificial Intelligence,* eds M. Bramer and M. Petridis (London: Springer), 149–162.

Mathy, F., and Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122, 346–362. doi: 10.1016/j.cognition.2011.11.003

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychol. Rev.* 65, 151–166. doi: 10.1037/h0048495

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., and Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychol.* 149, 1–8. doi: 10.1016/j.actpsy.2014.01.015

Rabinovich, M. I., Varona, P., Tristan, I., and Afraimovich, V. S. (2014). Chunking dynamics: heteroclinics in mind. *Front. Comput. Neurosci.* 8:22. doi: 10.3389/fncom.2014.00022

Richman, H. B., Simon, H. A., and Gobet, F. (1991). "Applying retrieval structures to chess memory," in *Paper Presented at the International Conference on Memory* (Lancaster, PA).

Richman, H. B., Staszewski, J. J., and Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychol. Rev.* 102, 305–330. doi: 10.1037/0033-295X.102.2.305

Roberts, S., and Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychol. Rev.* 107, 358–367. doi: 10.1037/0033-295X.107.2.358

Robinet, V., Lemaire, B., and Gordon, M. B. (2011). MDLChunker: a MDL-based cognitive model of inductive learning. *Cogn. Sci.* 35, 1352–1389. doi: 10.1111/j.1551-6709.2011.01188.x

Saariluoma, P. (1989). Chess players' recall of auditorily presented chess positions. *Eur. J. Cogn. Psychol.* 1, 309–320. doi: 10.1080/09541448908403091

Saariluoma, P., and Laine, T. (2001). Novice construction of chess memory. *Scand. J. Psychol.* 42, 137–146. doi: 10.1111/1467-9450.00223

Servan-Schreiber, E., and Anderson, J. (1990). Learning artificial grammars with competitive chunking. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 592–608. doi: 10.1037/0278-7393.16.4.592

Simon, D. P., and Simon, H. A. (1973). Alternative uses of phonemic information in spelling. *Rev. Educ. Res.* 43, 115–137. doi: 10.3102/00346543043001115

Simon, H. A. (1974). How big is a chunk? *Science* 183, 482–488. doi: 10.1126/science.183.4124.482

Simon, H. A., and Chase, W. G. (1973). Skill in chess. *Am. Sci.* 61, 393–403.

Simon, H. A., and Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cogn. Psychol.* 5, 29–46. doi: 10.1016/0010-0285(73)90024-8

Simon, H. A., and Newell, A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. Assoc. Comput. Mach.* 19, 113–126. doi: 10.1145/360018.360022

St. John, M. F., and McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artif. Intell.* 46, 217–257. doi: 10.1016/0004-3702(90)90008-N

Uttal, W. R. (2011). *Mind and Brain: A Critical Appraisal of Cognitive Neuroscience*. Cambridge, MA: MIT Press.

Vicente, K. J., and Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychol. Rev.* 105, 33–57. doi: 10.1037/0033-295X.105.1.33