New Ideas in Psychology 28 (2010) 64–78

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Questioning Chase and Simon's (1973) "Perception in Chess": The "experience recognition" hypothesis

Alexandre Linhares [a,b,*], Anna Elizabeth T.A. Freitas [b]

[a] The Club of Rome, Lagerhausstrasse 9, 8400 Winterthur, Switzerland
[b] Getulio Vargas Foundation/EBAPE, Praia de Botafogo 190 office # 509, Rio de Janeiro 22250-900, Brazil

ARTICLE INFO

Article history:
Available online 12 August 2009

Keywords:
Cognitive models
Perception
Memory
Cognitive psychology
Experimental psychology
Analogy
Experience Recognition

ABSTRACT

Pattern recognition lies at the heart of the cognitive science endeavor. In this paper, we provide some criticism of this notion, using studies of chess as an example. The game of chess is, as significant evidence shows, a game of abstractions: pressures; force; open files and ranks; time; tightness of defense; old strategies rapidly adapted to new situations. These ideas do not arise on current computational models, which apply brute force by rote-memorization. In this paper we assess the computational models of CHREST and CHUMP, and argue that chess chunks must contain semantic information. This argument leads to a new and contrasting claim, as we propose that key conclusions of Chase and Simon's (1973) influential study stemmed from a *non-sequitur*. In the concluding section, we propose a shift in philosophy, from "pattern recognition" to a framework of "experience recognition".

1. Introduction

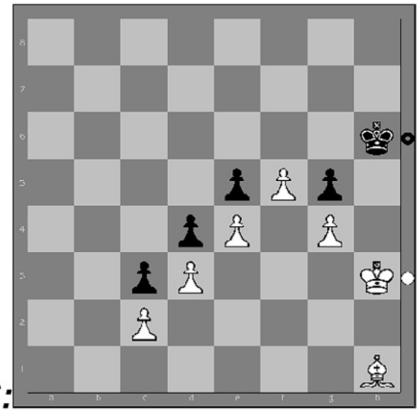
Pattern recognition is perceived as a crucial cognitive ability, and studied in a large number of contexts. In this paper, we propose a shift of perspective: that the most promising avenues towards understanding cognition lie in what we will call "experience recognition". We would like to invite readers to follow our reasoning by carefully studying an influential article in cognitive psychology: Chase and Simon's (1973) "Perception in Chess".

* Corresponding author. Getulio Vargas Foundation/EBAPE, Praia de Botafogo 190 office #509, Rio de Janeiro 22250-900, Brazil. Tel.: +55 21 9996 2505.

E-mail address: linhares@clubofrome.org.br (A. Linhares).

0732-118X/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.newideapsych.2009.07.008

6:



10:

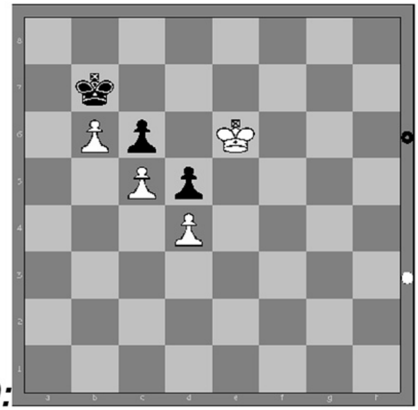


Fig. 1. Positions 6 and 10. White to move and win.

By most accounts, Chase and Simon's (1973) is an extraordinary paper. The experiments showing that chess masters could easily reconstruct complex positions after a brief presentation had already been conducted by de Groot (1965), but Chase and Simon also showed that, in random positions, the masters' reconstruction superiority over beginners nearly vanished. The conclusion that chess masters have from 50.000 to 100.000 chunks encoding combinations of pieces has almost become a cliché in the literature of expertise. Similar recall experiments, in different domains, followed with similar results. That paper spawned the whole field of *situation awareness* (Banbury & Tremblay, 2004; Endsley & Garland, 2000).

But was it correct? The paper had two parts. In the first part, they provided the recall experiments, while the second part studied the content, size, and number of chess chunks. In this paper, we claim that the second part of Chase and Simon (1973) is flawed and that the results presented therein do not hold. Specifically, we question the conclusions concerning the content, size, and number of chess chunks obtained in that paper.

In order to properly develop them, we will look into contemporary computational models of chess expertise directly spawned by that work, and trace their underlying problems back to Chase and Simon (1973). We hence urge the reader to backtrack from present work, and follow our critical review of the modern computational models of chess expertise. We shall start with a summary of work recently conducted.

Consider the positions[1] in Fig. 1. Linhares and Brum (2007, 2009) present psychological experiments showing that, in the eyes of a chess expert, the strategic situations in positions 6 and 10 are remarkably similar. Yet, to "the eyes" of a computer program, these positions are remarkably different, with different piece sets, arrangements, and search tree size. Position 10 is 'solvable' within a brief computation, while position 6, which demands a computation of 20+ plies of the rapidly expanding game tree before any improvement in the evaluation function can be achieved, bears a substantial computational burden.

We shall see that the same effect (perception of similarity at an abstract level) is found in positions in which exchanges take place, and necessarily demand 'looking ahead'. We cannot expand on such results here, except for pointing out the study's final conclusion: experts found positions sharing the same sets of abstract roles as more similar than positions displaying high levels of surface features, such as pieces-on-square combinations. The opposite effect was found in beginners. This goes against current theory.

Linhares (2005) postulates a computational architecture which may explain the semantic information processing involved in chess cognition. Linhares (submitted for publication) presents

[1] In position 6, though the bishop is of no value, the black king cannot defend simultaneously from the passed pawn and the white king's upcoming attack. Therefore, a promotion lies in store for white. Position 10 should be obvious.

a first implementation of such architecture which shows that, from surface information, such as pieces-on-squares, it is possible to rapidly compute topological distances to other pieces, trajectories, interceptions, and abstract roles. It is argued that the abstract roles created account for the master's ability to perceive positions highly different in surface information (pieces-on-squares, search tree size, number of pieces, types of pieces) as similar at a 'strategic level'.

However, modern computational models, such as CHREST or CHUMP (Gobet, 1997, 1998; Gobet & Jackson, 2002; Gobet & Simon, 1996), define human cognition in chess by placing high emphasis on sets of pieces-on-squares, and use mechanisms such as discrimination nets to retrieve chunks from LTM. In this paper we assess these models from the viewpoint that chess chunks must contain semantic information, an argument which leads to the claims concerning the nature of chess chunks originally made by Chase and Simon (1973).

2. Models of chess expertise: a look at CHREST

Judging from the number of citations it has received in ISI indexed literature, CHREST is the most influential theory of chess expertise[2]. Unlike verbal theories, CHREST is an 'evolving computer implementation' of the hypothesis underlying its theoretical foundations. It is both (i) rigorous in the definition of terms, such as chunks, templates, the processes underlying chess mastery—which are usually loosely defined in SEEK (Holding, 1992) and other 'verbal' theories; and also (ii) affords to be tested in careful simulations and compared to actual human data.

Before we analyze CHREST, it may be interesting to comment on a previous experiment in cognitive linguistics. Sachs (1967) showed that subjects were able to recall the meaning of a paragraph while rapidly forgetting the specific wording involved. One paragraph used in her experiment was:

> *In Holland a man named Lippershey was an eyeglass maker. One day his children were playing with some lenses, and they discovered that things seemed very close if two lenses were held about a foot apart. Lippershey began experimenting, and his "spyglass" attracted much attention. He sent a letter about it to Galileo, the great Italian scientist. Galileo built his own instrument, took it out on the first clear night, and was amazed to find the empty dark spaces in the sky filled with stars!*

Sachs then asked which phase was found in the text:

1. Galileo, the great Italian scientist, sent him a letter about it.
2. He sent Galileo, the great Italian scientist, a letter about it.
3. He sent a letter about it to Galileo, the great Italian scientist.

Subjects easily perceived that #1 could not be in the text, but most thought #2 was the correct sentence. Subjects were able to retain the essence of the message, but could not retain the specific wording used. The key point is that memory is more prone to find semantic similarity than surface similarity.

The goal of CHREST is to be a "psychological model of human chess expertise" (de Groot & Gobet, 1996, p. 215), in fact, a "unified model" (de Groot & Gobet, 1996), which includes the cognition behind learning, perception, and memory. It is a symbolic system such as the well known SOAR or EPAM. The system implements a visual scanning of a position in a visual field (which qualitatively matches humans' eye saccades), learns by storing new patterns in LTM, recognizes patterns already stored in LTM, enabling it to construct a representation of the board in STM, and finally reconstructs the position with the content embedded in the chunks (referred to by pointers in STM). The matching to human data, either experts or novices, either after short or long presentation times, either with regular or with random positions, is nothing short of remarkable.

This work states that "it is so hard to become a master […because…] several nets need to be constructed, with many connections within them (redundancy) and between them (production and

[2] For example, as of January 2005, the recent references concerning CHREST cited herein had over 200 citations, and if we include the 1973 papers from Chase and Simon, the number of citations grows to over 1000.

semantic links)" (de Groot & Gobet, 1996, p. 246). Since these 'several nets' are fundamental in understanding how CHREST, and thus the underlying theory, operates, we must delve a bit into their mechanisms (the same applies to CHUMP, below).

3. Discrimination nets

At the core of discrimination nets lies a process of branching tests. In CHREST, whenever the visual field scans a new piece, say <Bb3>, then the system will look into LTM for a potential chunk candidate. Let us suppose it finds an initial branch, or node, with that information. This node may be connected to other nodes, such as <Pa2>, <Pa3>, and so on. Each of these nodes (which in this hypothetical example may already have been found in previous positions and stored in LTM) will act as a test for a potential chunk. The meaning of such a branch is: "You have found a White Bishop on square b3. Do you see now a White Pawn on a2?" Each of these tests will guide eye saccades and offer branching points for the system to identify a potential chunk (eventually involving a variety of pieces). ***The tests always include a specific piece-position pairing*** (or POS, for piece-on-square).

Discrimination nets are usually decision trees, enabled to evolve by two specific mechanisms: *familiarization* and *discrimination*. Familiarization concerns the cases in which the chunk is "recognized as being compatible with the chunk the net sorted it to" (de Groot & Gobet, 1996, p. 226); discrimination concerns the cases in which a chunk is "sorted to a node and is recognized as being different from the image of the node, then a new test is added" (de Groot & Gobet, 1996). The image of the node is simply the representation of a previously known chunk. This leads us to some first rigorous definitions offered by CHREST, that on the nature of chunks and the content of STM:

"As in Chase and Simon (1973), chunks are defined as sequences of pieces having corrected latencies of less than 2s between successive pieces." (Gobet & Simon, 2000b, p. 661). We will return to this term, latencies, soon. "A chunk is encoded as *a list of the pieces on their squares* (POS), sorted in arbitrary order, for example: <Kg2, Re1, Pf2, Pg3, Ph4>. A chunk can consist of a single POS (for example, <Kg2>)" (Gobet & Simon, 2000, p. 671). What experts do is thus to "…recognize chunks [from LTM], and place a pointer to them in STM. These chunks, each of which contain several elements that novices see as units, allow experts to recall information well beyond what non-experts can recall" (Gobet & Simon, 2000, p. 652).

The content of STM is a queue of pointers to LTM nodes, in traditional first-in, first-out fashion. That is, potential chunks are evaluated and discarded rapidly after new candidates appear. There is one special chunk, referred to as the *hypothesis*, which gets the privilege of never leaving STM—this is due to the fact that it is the largest chunk found in LTM so far.

4. Simulations: learning and performance phases

How does the system learn? According to the authors, "during the learning phase, the program scans a database of several thousands of chess positions taken from masters' games. It fixates squares with simulated eye movements, and learns chunks using the discrimination and familiarization processes. Templates and similarity links are also created at this time." (Gobet & Simon, 2000b, p. 672)

Templates are "…simply large chunks describing patterns that are met frequently in Master's practice, especially common opening variations, which evolve into [templates]" (Gobet & Simon, 2000b, p. 679). Templates have a *core*, (i.e., a CHREST chunk), but they are more flexible than chunks, encoding a series of *slots* with additional information. There are *square-slots* that may encode pieces surrounding the core, and *piece-slots*, which encode the squares occupied by pieces that usually are linked to the core (one such example is provided in Figs. 2 and 3).

After learning, in the 'performance' phase, simulations are conducted in order to reconstruct positions briefly shown. There are variations on the level of expertise (e.g., beginners, masters, etc), on the presentation times, and on regular vs. random positions. The system displays an incredible fit with human data, making mistakes when humans also are expected to. But CHREST does not select moves, and since humans learn chess by selecting moves, it was a natural step for this theory to expand into a move-selecting, game-playing, system, named CHUMP.

CORE:

Pc4 Pe4 Pf3 Pg2 Ph2 Be3 Nc3
pc6 pd6 pf7 pg6 nc5 bg7

SLOTS:

SQUARE SLOTS:

c5: —
d4: —
g1: <whiteking>
c3: —
e5: —
e3: —
f7: —

PIECE SLOTS:

white rook : <e1>
white bishop : <f1>
white knight : —
black bishop : —
black knight : <d7>
black pawn : <b7>

Fig. 2. (After Gobet & Simon, 2000). An example CHREST template after 300,000 chunks are stored into LTM—a discrimination net size argued to account for a master players' level of expertise.

4.1. A look at CHUMP

CHUMP (Gobet & Jansen, 1994) is a system of the same family and theoretical foundations, with the objective of selecting a move (as opposed to reconstructing a briefly shown position). It "grows two [discrimination] nets, one for patterns of pieces, as in CHREST, and the other for moves and sequences

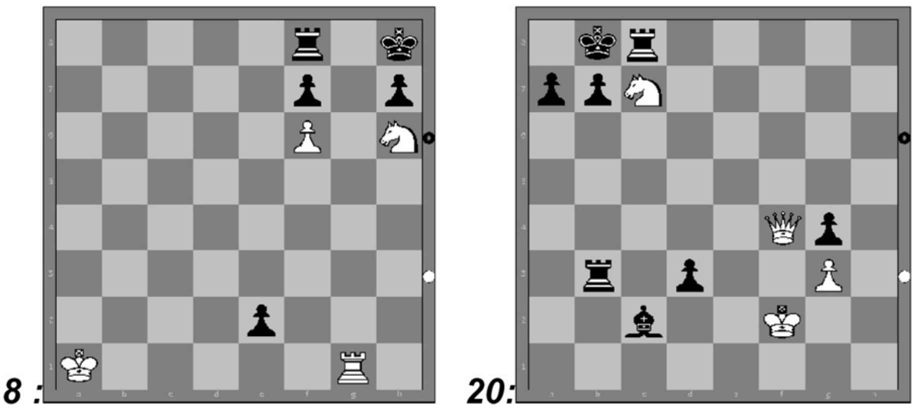


Fig. 3. Experience recognition and analogy-making on chess.

of moves. In addition, the two nets are connected by associative links" (de Groot & Gobet, 1996, p. 245), which enable the system to propose moves based on the chunks previously stored in LTM:

"For example, given the pattern <Pf2, Pg3, Ph3, Bg2> and given that move played in the game, say 'Bg2xd5', is accessible to a node in the *move-net*, CHUMP creates a connection between the two nodes. In the future, the pattern <Pf2, Pg3, Ph3, Bg2> will serve as a condition to the action 'Bg2xd5'." (de Groot & Gobet, 1996, p. 245) Hence, the fundamental mechanism underlying the whole system is also composed of additional discrimination nets, this time involving moves and sequences of moves. de Groot and Gobet (1996, p. 245) speculate about a future implementation of a semantic memory, also based on discrimination nets: "The same mechanism may be used to implement a rather complex semantic memory. In addition to nets for patterns of pieces and for (sequences of) moves, nets could be created for openings, plans, heuristics, tactical concepts, positional concepts, etc" (de Groot & Gobet, 1996).

One of the key underlying assumptions of the system stems from the fact that chess masters are incredibly fast in selecting a good move. Because this has been seen in numerous experiments, not to mention in fast tournaments or in simultaneous games against multiple opponents, the designers have focused on a system that can very rapidly and efficiently obtain moves:

"It seems plausible, for example, that beginners and weak players will focus on abstract, non-located relations between pieces, like a 'Knight attacks a piece that defends a Pawn'. [...] Such knowledge is general, but its price is that it takes time to interpret it and to apply it to a given position. As expertise grows, we expect that players will tie abstract, declarative knowledge to specific instances. The resulting compiled, variable-free knowledge is of course *limited in its application* but is very rapid to access and very reliable. This is the type of productions implemented in CHUMP." (de Groot & Gobet, 1996, p. 246, emphasis mine). This argument is built to defend the storage of pieces on squares.

Perhaps the most profound problem with these models is that they deeply lack the flexibility present in human cognition. Take a new look at position 10 in Fig. 1, and imagine all the pieces shifting right (or down, or both) one (or two or three) square(s). Because in this "new" position the pieces now reside in different squares, according to the chunking theories, completely new chunks would be brought from LTM. To rephrase, the chunking theories—which are based on mechanisms as rigid as discrimination nets (i.e., decision trees)—"see" a completely different position with a new set of chunks when we shift the pieces. But is this shifted position actually different from the original in any significant sense? No: experts have reported to us that this shifting of position 10 "means nothing, it's the very same position" (Linhares & Brum, 2007). Models based on discrimination nets do not exhibit the flexibility involved in human perception. Every piece may have moved, but in essence, the position is the same—and so is the strategy for play.

CHUMP has been criticized by others: "[...] one important objection to simple chunking theories of chess skill involves the production system link between chunk recognition and move selection. How does recognizing relatively small chunks or patterns lead to the choice of specific moves [...]?" (Chabris & Hearst, 2003, p. 644; see also McGregor & Howes, 2002). Since the chunks recognized in current theories are perceived mostly in isolation, and the context in which they are found determines a good or bad move, it is hard to accept the notion that they could be directly associated with strategies for play. A global, abstract, perspective of the situation is needed. It is clear that perception of a specific chunk cannot lead directly to any move, because, for each chunk, there is a large number of positions in which it appears. Each position presents distinct move possibilities. Chabris and Hearst (2003) then suggest that "other higher level processes" (such as associative factors, extensive search, evaluation, visualization, and perhaps "some additional as yet unidentified higher-conceptual or representational theories"), must intervene between knowledge of patterns and selection of moves. Bishops are generally stronger than pawns, but as we have seen in position 6, that is not always the case. Any chunk that includes that bishop cannot simultaneously be linked to a strategy in which the bishop is a key piece and other strategies in which the bishop is irrelevant. In the following discussion, we demonstrate that a global view of the situation is needed, even in the simple case of position 6.

5. How could CHUMP deal with position 6?

How could CHUMP handle position 6? It should be obvious to any player of the game that white has a winning advantage and that the only moves that make sense are white king to either g2 or g3. The question we now face is: how could CHUMP find out those moves? This should be an easy problem: there are only six moves available for white in this position, and two would be correct (in the sense that they could form the starting point for a winning strategy). But, given the specified mechanisms, it is far from easy. Position 6 leads us to two key ideas: CHUMP needs a global view, representing the situation with one chunk only. And the probability that the system would have acquired this specific chunk is vanishingly small.

Why CHUMP must find one chunk only: As pointed out in Gobet and Simon (2000), in the "learning phase the program scans a database of several thousands of chess positions". This enables CHREST to build a discrimination net with 300,000 nodes—which is argued to resemble the knowledge of a master player. There are numerous ways to chunk the pieces, as seen above; however, since CHREST at a "master level" creates large chunks with numerous pieces (Fig. 3), it is possible that the system would build one single chunk or template to account for position 6, or perhaps two or three.

In order for CHUMP to solve this problem, it would have to find appropriate chunks in its discrimination net, and potential moves associated to such chunks in the move discrimination net. Depending on the particular chunking selected, and on the move discrimination net, the system would have reasons to play the following wrong moves:

- An unblocked advanced pawn obviously suggests the attempt to promote it, so the system would immediately have a reason for the f6 move; in other words, the system should have in its move discrimination net numerous moves of this kind;
- The bishop is currently defending only the e4 pawn, but any of its two possible moves would let it defend two pieces, so, once again, the system would have a reason for moving the bishop — and numerous moves of this kind should appear in the move discrimination net.

Though CHUMP does not look for exact position matches, and only chunks, there is nothing to be gained by looking at isolated patterns. *Any framing in separate chunks (or templates) will not give the necessary cues for the global perspective of this position*: first, white king goes to a2, and then abruptly changes direction, threatening the e5 pawn. Which postulated (isolated) chunks could lead to (directly associated) production-system moves following such a zig-zag-like trajectory? If the system found more than one chunk/template on this board, how could it restrain movement of the passed pawn until the white king was found in the appropriate square, numerous plies afterwards? A single chunk is needed to prevent the move of the passed pawn during such a large number of plies. Otherwise, in all probability, in at least one ply a production to move the passed pawn should be triggered. So let us consider, for the sake of argument, the best-case scenario that the system effectively managed to create a global description of the position by finding that chunk. Can it now, in principle, solve this position?

Why the probability of having acquired that chunk is vanishingly small: Let us consider how CHUMP could find a reason for the correct white king moves. The only possibility stems from having accumulated chunks with these specific piece-square pairings, and then finding associated king moves in the move-discrimination net. If the position is not scanned, the chunk cannot be found. Now, CHREST at a 'master' level of position reconstruction scans "thousands of chess positions" in its learning phase. Since it must be carved into one global chunk, could this position, or perhaps one following it, be in that set? We have conducted a search on a database of 504,550 full games. If we assume a (rather pessimistic) average number of 30 moves per game in the database, then we have over 15 million positions available to include in a discrimination net of moves and move sequences—a gain of several orders of magnitude from CHREST's 'master' level experiments. In order to *raise the chances* of finding this position, we considered a series of subsequent ones, by varying the white king from its original square to c4, an obvious point in its trajectory (and therefore easier to find in the database). We also varied the black king from its original squares to f7, e7, and d6, also obvious possibilities. We varied the bishop's position. We finally took out the bishop from the position. In all these searches, not a single game included this position or the numerous variations used — in total over a *hundred million positions* taken from real games were scanned. So the conclusion we are led to is that it is extremely unlikely that CHUMP, looking *at a mere thousands of positions*, would be capable of justifying a correct sequence of white king moves.

To summarize our analysis, CHUMP could not have the required discrimination nets for the white king moves unless its 'learning phase' was increased by several orders of magnitude. But if a mere *thousands of positions* suffice to reach 'master level' in CHREST, it would be highly questionable to make this increase in magnitude in CHUMP while retaining the current 'master level' architecture of CHREST. There seems to be something lacking in the CHUMP architecture and we propose that the main deficiency is the lack of semantics obtained by accessing abstract roles. Another point worth of mention is that, because there are no piece-on-squares matches, neither CHREST nor CHUMP could account for the abstract similarity between positions 6 and 10, or 8 and 20.

The major problem involved here stems from the conclusions obtained in the Chase and Simon (1973) study, to which we should turn our attention next.

5.1. Chase and Simon (1973)

The Chase and Simon (1973) study is divided in two parts. The first part (p. 55–61) showed that, when chess masters looked at a board for 5 s, they could reproduce it with enormous accuracy, while beginners could not recall more than a few pieces. This difference could not be explained by masters' greater memory, for, in randomized positions, the effect disappeared, with masters and beginners able to reproduce only a few pieces of the board. This was not only a clear reproduction of the de Groot (1965) results, but brought a theory of chunks to distinguish masters from beginners.

The second part of the paper asked what the nature of the chunks is like: What is the content of a chess chunk? What is the size of a chunk? How many chunks do masters have? To probe these, Chase and Simon devised two tasks, a 'perception task', and a 'memory task'. These tasks looked at master subject, a Class A subject, and a beginner, Class B subject. To make our argument clearer, let us refer exclusively to the master and to the beginner. This will not alter the arguments that follow in any way, and may simplify understanding of our disagreement with that study.

In these perception and memory tasks, Chase and Simon collected data consisting of 'interpiece interval times' while subjects reconstructed the boards, such as times taken "within glances at the board, and in between glances". The results were unequivocal: the data was ***exactly the same for masters and beginners*** (see figs. 3 and 4 of that paper). They pointed this out clearly:

[Perception task, p. 65] "The first thing to notice is that the data are quite similar for all subjects. The latencies show the same systematic trends, and, for the probabilities, the product moment correlation between subjects are quite high: Master vs. Class A = .93; Master vs. Class B = .95, and Class A vs. Class B = .92. The same is true for the between glance data [...] Thus, the same kinds and degrees of relatedness between successive pieces holds for subjects of very different skills." They were to claim later that successive pieces formed chunks.

[Memory task, p. 70] "Again the pattern of latencies and probabilities look the same for all subjects, and the correlations are about the same as in the perception of data: Master vs. Class A = .91, Master vs. Class B = .95, and Class A vs. Class B = .95".

Herein lies our problem. Chase and Simon had shown in the initial pages that the master had chunks and that the beginner lacked them; this accounted for the master's ability to reconstruct the board to an average of 81% correct vs. 33% in one task, and 62% correct on average vs. 18% on another task.

Now, in the subsequent part, they had collected a large amount of data, consisting of interpiece interval times—which, they claimed, could establish the size of chunks; with lower times leading to higher chances of pieces being included in the same chunk. The data were unequivocal: the master player exhibited behavior similar to the beginner to an extraordinary extreme, and Chase and Simon clearly pointed that out. Consider, for instance, figs. 3 and 4 of that study: those figures display frequency distributions of the times involved, broken down by subject. There is no distinguishing feature between the subjects. The master, the class A player, and the class B player randomly alternate positions between the highest frequency, lowest frequency, and the middle spot across interpiece intervals. The correlation of .95 should speak by itself. If a master has more chunks, or different content in those chunks, it certainly does not show up in the data.

Our only conclusion from the data is that, whatever difference exists between the contents of chunks of the master and the beginner, it cannot be in any way inferred from that dataset. It should be clear that no information concerning the 'content of the chess chunk' can be obtained there, because the master has chunks, the beginner lacks them, and *in this task their behavior is indistinguishable*. The task is *not* affected, in any way, by the content of the chunks. If the content of the chunks does not interfere with the results, yielding different results for master and beginner, this must lead to the conclusion that no inference on the content of the chunks can be made by looking at the collected data. In other words, the selected experiment is not probing the content of chunks—otherwise results from a dedicated master would be different from those of an amateur beginner.

Yet, Chase and Simon draw the exact opposite conclusion: "These probabilities are informative about the underlying structures that the subjects are perceiving" (p. 68). This is, to say the least, questionable. Master or beginner, the probabilities were undistinguishable from each other, with the correlation between master and a beginner at .95. The probabilities cannot be informative if there are no differences between those that have the "underlying structures" (i.e., chunks) and those that lack them.

After that point, they move on to characterize chunks: "Our data gives us an operational method of characterizing chunks, which we will apply to the middle-game memory experiments of subject M" (p. 78). This is also immensely problematic. One can only wonder why the master would be necessary. Because performance of the master in the tasks is simply indistinguishable from a beginner, the same results would be gathered from a beginner subject— and perhaps even from a child. If "the same kinds and degrees of relatedness between successive pieces holds for subjects of very different skills" (p. 65), then how can the chunks of masters and beginners be any different?

That is their main conclusion concerning the *content* of chess chunks. Concerning chunk *size*, Chase and Simon find a small difference between the sizes of chunks for the memory experiment, as the average "number of pieces per chunk was 2.5, 2.1, and 1.9" (p. 76), for the master, the class B player, and the beginner. There are, however, a number of problems here. First, a beginner has very limited knowledge of chess, and a master has dedicated at least a full decade of intense study to deserve that title. Can a difference of less than a third in chunk size explain the difference between decade of intense study and a year of learning? Chunking theory applies do other domains also. Would it be acceptable to say that a Yanomami, with a single year of study, could learn Hindi to the same level of difference (1/3) from someone who masters Hindi? It is reasonable to say that a mere 1/3rd of chunk size separates a master of Chinese ideograms from a European with a mere year of training?

If the reader is still unconvinced, their previous sentence has that difference vanishing. In the perception task, Chase and Simon present the following average chunk sizes (in number of pieces) for master, Class B, and Class C player: "2.0, 2.8, and 2.0" (p.75). They hence conclude, remarkably, that chunk sizes are not that important to encode knowledge: "it appears that the chunks are about the same size in both tasks, but that chess skill is reflected in the speed with which chunks are perceived in the perception task and the size of the chunks in the memory task" (p. 76). It is, however, as impossible to differentiate the trained master from the random beginner in the perception task times as it is in the memory task: "The distribution of time intervals for the two tasks are not dissimilar" (p. 63). A glance at figs. 3 and 4 of Chase and Simon shows that it is impossible to distinguish the master from the beginner according to time, so the inclusion of time in one task and the exclusion of time in the other task seems to be a rather arbitrary way to explain the data. We are not arguing against chunking theory *per se*, but about this particular dataset and conclusions.

Ambitious conclusions have been drawn over that dataset, such as the estimated size of chunks, the number of chunks a master player has, and the content of the chunks (pieces based on proximity, piece color, etc). This eventually led to the view that chunks contain pieces-on-squares in detriment of information of semantic value, and deeply influenced research on computational models.

As we have seen, those conclusions are invalid. The argument, reaffirmed, has three points: (i) Performance on the tasks is indistinguishable between a master and a beginner. So, (ii) the content of the chess chunk (or the number or size of them) does not alter the data collected on the tasks, otherwise performance would be distinguishable. Therefore, (iii) any inference based on that data concerning the nature of the chunk is not valid.

This paper is not making a minor claim about some obscure technical error: *all the results* in the 20 pages following page 61 are claimed to be invalid. Chess is a game of abstractions: pressures; force; open files and ranks; time; tightness of defense; old strategies rapidly adapted to new situations. These ideas do not arise on current computational models, which apply brute force by rote-memorization.

If, in a given experiment, a beginner is able to produce the same dataset of Bobby Fischer, there will be nothing in that dataset to illuminate the nature of chess genius.

5.2. "Experience recognition" and analogy as the core activity in cognition

Though we have provided a methodological critique of a well-known study in this paper, we believe that the same methodological critique is valid to a larger body of literature. In this concluding section, we would like to introduce the term "*experience recognition*". Experience Recognition is a term that summarizes our viewpoint and can readily be contrasted with "pattern recognition".

Though not exactly the same, our proposal of experience recognition is close to the interactionist literature (See, for instance, Beer, 2003; Bickhard, 2009; Pfeifer & Scheier, 1999; Thelen & Smith, 1996), which has argued for the study of the interactions between the environment and an agent's predispositions (with the full complexity of the agent's particular body and mind and goals, etc). Our view is also close to Klein (1999).

In *pattern recognition* journals, conferences, and books, there is generally an emphasis on the pattern (i.e., the raw data, such as a waveform of music, a bidimensional image in vision systems, or a chess position in the aforementioned models). We propose that this emphasis on the data *outside of any understanding* is highly counterproductive (Linhares, 2000; Smith, 1996). As put succinctly—by an anonymous referee—"*pattern is in the mind of the beholder*". Our emphasis is on the *experience* that a human or a model has gone through, embedded in its long-term memory. In order to review this idea, let us consider Hofstadter's (2001) proposal of analogy as the core of cognition.

Chess has been seen as a "problem-solving", "inference-making" task, up to this point. We argue that, for high-level players, analogies to previous experiences provide the "atoms of abstract thought" (see also the discussion on Indurkhya, 2006, 2007). Consider positions 8 and 20 of Linhares and Brum (2007), in Fig. 3. These positions share no pieces on squares, yet they have a remarkable resemblance in strategic terms. For readers unfamiliar with chess, the solutions are: In position 8, a variant of a position taken from (Charness, Reingold, Pomplun, & Stampe, 2001), white moves rook to g8 check, black rook captures white rook, white knight captures black pawn at f7 checkmate. In position 20 white moves knight to a6 check (by knight and by queen), black king escapes to a8, white moves queen to b8 check, black rook captures white queen, white knight returns to c7 checkmate. These variations of "Philidor's mate" display high strategic similarity and no similarity at a surface appearance level.

We thus propose that analogy-making is a central process in chess cognition; and hypothesize that this idea extends to other domains. Let us ground these ideas with some concrete examples: perception of letters (Fig. 4), perception of letter strings (Fig. 5), Bongard problems (Fig. 6), and Physics problems

Fig. 4. Uppercase Bs and 13s formed with the same raw data.

ABC→ABD IJK → IJL	ABC→ABD XYZ→WYZ
ABC→ABD IIIJJJKKK→IIIJJJLLL	ABC→ABD RBBOOO→RBBOOOO

Fig. 5. Some interesting problems dealt with by the Copycat and Metacat projects (Hofstadter & FARG, 1995; Marshall, 1999; Mitchell, 1993; Mitchell & Hofstadter, 1990)..

(Fig. 7). Hopefully these examples will convey the full significance of the flexibility of perception and the abstract roles underlying it.

In Fig. 4, we have the very same surface structure forming either an uppercase “B”, or the number 13. How can the same surface structure be so easily perceived as two different things? What is ambiguity? And how can countless potential surface ways to render a letter lead humans to instantly perceive “the same thing”? What is generalization? Can a unified theory explain both the ambiguity inherent in the ‘B’s and ‘13’s and the generalization among different typefaces? An emerging solution to these puzzles is that abstract roles are crucial to letter perception (Hofstadter & FARG, 1995; McGraw, 1995; Rehling, 2001). In the terminology of Rehling (2001), the ‘B’s are chunked as a “long left-post” alongside two “right-bowls”. The ‘13’s are formed by two chunks, each with its own role—the ‘1’s are chunked as a single “long post”, and the ‘3’s are separately chunked as “two right-bowls”. In this theory—as in our chess theory—what stays fixed, what anchors similarity, is a set of abstract roles, and each abstract role can be found in countless distinctive surface displays (Rehling, 2001). This leads to our proposal that chunks are formed by sets of abstract roles, each of which is detached from specific implementations.

In the game of chess, abstract roles arise from a piece’s spheres of influence (Linhares, 2008). In perception of letters, such as those in Fig. 4, abstract roles arise by fast eye saccades. Our experience of saccading eyes in particular angles and turns is one of the most important information to be recognized. The raw data is (almost) of no consequence, *if it is able to trigger the same pattern of eye saccades* (i.e., the same experience). The crucial encoding is not image-like, but more like a set of vectors representing eye saccades. And the result is crucially different is one expects to recognize a character or a number.

The problems in Fig. 5 ask the solver to come up with answers to questions such as “if ABC goes to ABD, then IJK goes to what?”—usually represented by “ABC → ABD: IJK → ?”(Mitchell, 1993; Chalmers et al, 1992; see also French, 1995). In the perception of these letter strings, abstract roles once again anchor deep similarity. In the upper left case, things are very simple: ABC and IJK play the roles of “letter staircases”, so the transformation is simply to do a “double step” at the rightmost spot. But situations in this domain may be less trivial. In the lower left case, the abstract role played by C is mapped to the group of letters KKK, so the “double step” is now applied to the whole group. In the upper right case, the abstract role of the A is simply “first letter of the alphabet”, which maps in an opposite way to the Z, playing the role of “last letter of the alphabet”. This suggests to the system that now the role of a “staircase from first letter with double step at the right end” can map to “staircase

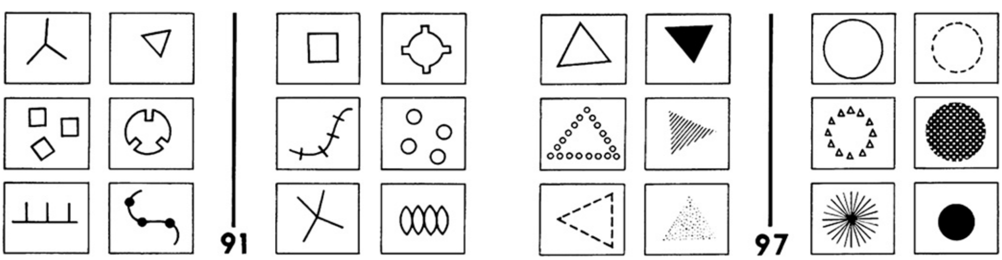


Fig. 6. Bongard problems 91 and 97.

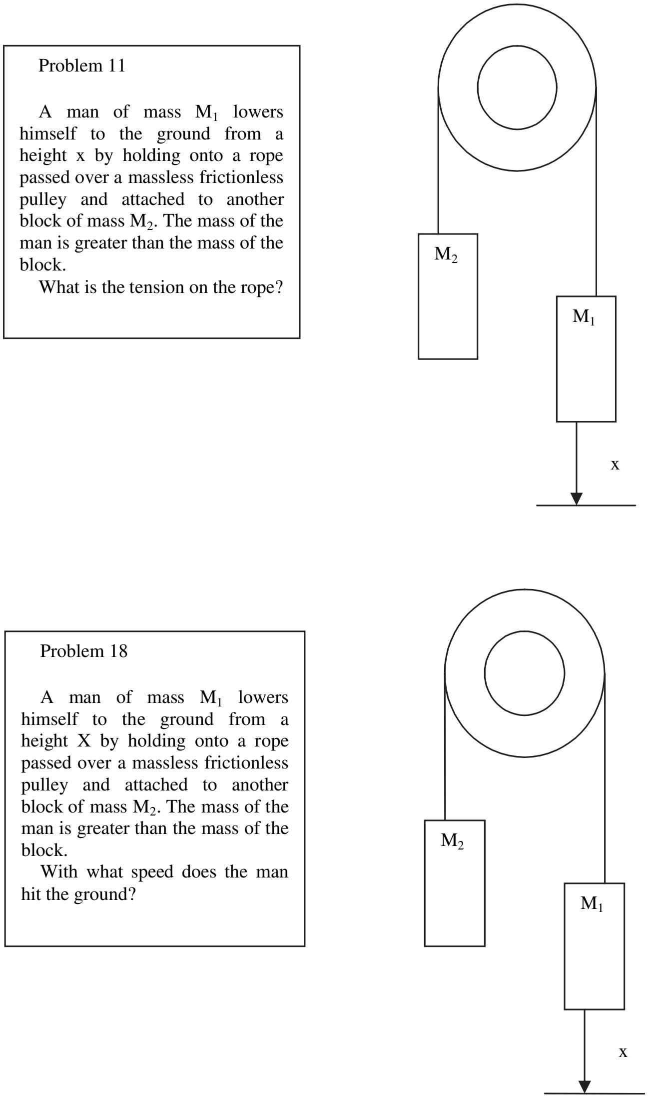


Fig. 7. Experience recognition in Physics (after Chi et al., 1981).

from last letter with double step at the left end", in a strikingly symmetrical structure. Of course, numerous other variations are also possible; each of them carrying its own level of aesthetic resemblance and symmetry to the original. Finally, the roles played by ABC in the lower right case are simply "first-second-third", which maps to the roles played by the lengths of groups of letters, such as "first number, second number, third number". So the transformation to the "fourth letter" slips to "group of length 4", and the string of Os grows by one. Abstract roles enable the perception of similarity between extremely different surface structures.

Each Bongard problem (BP) in Fig. 6 asks us to find a distinction between two classes of visual figures (Bongard, 1970; Foundalis, 2006; Hofstadter, 1979; Linhares, 2000). The six boxes on the left side of a BP share some similarity that differs from the six boxes on the right side. Once again, the similarity between extremely different figures can be perceived only if one is capable of abstracting from the raw patterns presented and find an abstract role that the elements in each box play in a given context. It is intriguing how the flexibility of human perception can handle this type of problem by employing a complex interaction between subconscious and conscious processing. For example, the solution to problem #91 is "three vs. four", and it tends to emerge relatively rapidly for humans. But for that solution to emerge, the reader should notice that a subconscious decision has been made to interpret each square in the left side as "one chunk", while on the right side the square is chunked as "four line segments". Strict interpretations of the raw data would lead to the "consistent" views of "three squares on the left; one on the right", or to "12 line segments on the left, four on the right" (Linhares, 2000). Because human perception is highly flexible, the "inconsistent view" decision is made subconsciously, and is not even perceived after the problem has been solved.

Or consider Physics. Experienced physicists see physics problems not by their surface features, but by the underlying laws of nature involved in each of them. Chi, Feltovich, and Glaser (1981) used a set of problems such as those in Fig. 7 to demonstrate that beginners classify problems according to surface similarity. The problem structure is exactly the same, with only the question being varied—in an important sense: problem 11 is a force problem, problem 18 is an energy problem. Experienced scientists saw the huge divide between these problems, while first-year undergraduates thought they were very similar.

We hope these examples may convey to the reader how this emerging theory may chart a largely unexplored territory in cognitive psychology. If perception is based on analogy, which is based on abstract roles perceived instantly and mostly unconsciously, then analogy and perception of abstract roles may be central to human cognition. And we propose that this pervasiveness of analogy stems from our necessity to continuously match the contents of our short-term memory with those of long-term memory, in a process of experience recognition.

Experience recognition also seems to arise in groups and in organizations. Consider for example, a population of consumers. In the 80s, the largest European carmaker decided to invest heavily in the U.S. market, by introducing a large range of cars that were bestsellers in Europe. Before this decision, the carmaker had 63% of the imported car market in the U.S. But the carmaker was called Volkswagen. The American consumer's experience with Volkswagen consisted of the Beetle, an inexpensive and odd-looking car first sold in 1938. American consumers rejected the idea of a large, well-built, modern-looking, powerful and expensive Volkswagen. To make matters worse, the company decided to withdraw the Beetle from the market—and its share of the imported car market in the U.S. dramatically fell from 63% to less than 4%. The *exact same cars* were being sold in Europe and in the U.S—the only difference was in the consumers' experiences of what a "Volkswagen" means. A twenty thousand dollar Volkswagen seemed, to Americans, like a practical joke. Similarly, "the new Honda" to an American consumer meant a new car model; to the Japanese, it meant a new motorcycle (Ries & Trout, 1993).

When Iranian Ayatollah Ruhollah Khomeini declared a death sentence to writer Salman Rushdie, the Catholic Church did not stand for the principle of "Thou shall not kill"(Hofstadter & FARG, 1995). It recognized its experience of trying to censor "The last temptation of Christ", a film, and sided with the Iranians. *L'Osservatore Romano*, a key Vatican publication, condemned Rushdie's book as "blasphemous". The Head of the French Congregation, Cardinal Decourtray, called it an "insult to God"; Cardinal O'Connor from New York made it clear that it was crucial to "let Moslems know we disapprove of attacks on their religion."

Rather like our "inconsistent perception" of Bongard problem #91, the Vatican had to decide between "thou shall not kill" vs. their own recent experience. As a testament to force of experience recognition in decision-making, the Vatican sided with its experience.

Lawyers usually present precedent decisions that are analogous to their case, but highly dissimilar in specific events. Scientists explain complex concepts by analogy: DNA is like a staircase; DNA is like a will; DNA is like a fingerprint on a crime scene—each analogy will bring forth some aspects of the situation we have experience with (while pushing other aspects into the background). Analogies and experience recognition arise early in human life; as Lakoff and others have pointed out (Lakoff, 1990; Lakoff & Johnson, 1999), young children can instantly understand that a "cold, distant, person" does not carry the literal meaning of either low physical temperature or geographical distance. The encoded experience of having a mother's warm body close to one's own seems sufficient to, later in life, enable the discernment between the literal and the metaphorical.

Politicians argue using analogies to previous events, to previous policies, and to previous politicians; Terms like "Patriot Act", accounting by "Fair Value", or "Tax Relief", are common tools of politicians, for those who will argue against the policies instantly lose the moral ground. Who can argue for "Unfair Value" accounting? Who can argue against a kind of "relief"? From the abstract game of chess to our major political issues, we believe that strategic thought, the intricate, knotty, search for what to do, will, in the end, be seen as the outcome of such analogies.

To summarize: we propose that what our brains compare our situations with are encoded experiences, not to the raw data outside of any understanding. *Internal encoding takes priority over raw external data. Experience recognition takes precedence over pattern recognition.* We are constantly comparing the inflow of information around us to encoded experiences in long-term memory. Numerous studies are conducted with an exclusive focus on the raw data, the "pattern recognition", and no focus on comparing the current processing with previous internal encodings. We believe that approach is counterproductive. We hence propose that emphasis of a substantial body of research be pulled from the perspective of "pattern recognition" towards "experience recognition".

References

Banbury, S., & Tremblay, S. (2004). A cognitive approach to situation awareness. In *Theory and application* (pp. 359). Ashgate Publishing.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior, 11*(4), 209–243.

Bickhard, M. H. (2009). The interactivist model. *Synthese, 166*(3), 547–591.

Bongard, M. M. (1970). *Pattern recognition*. New York: Spartan Books.

Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science, 27*, 637–648.

Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: a critique of artificial intelligence methodology. *Journal of Experimental and Theoretical Artificial Intelligence, 4*(3), 185–211.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: evidence from eye movements. *Memory & Cognition, 29*, 1146–1152.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement*. CRC Publishing. p. 408.

Foundalis, H. (2006). PHAEACO: a cognitive architecture inspired by Bongard's problems. Ph.D. thesis, Indiana University, Bloomington.

French, R. M. (1995). *The subtlety of sameness*. MIT Press.

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking and Reasoning, 3*, 291–313.

Gobet, F. (1998). 'Expert memory: a comparison of four theories. *Cognition, 66*, 115–152.

Gobet, F., & Jackson, S. (2002). In search of templates. *Cognitive Systems Research, 3*, 35–44.

Gobet, F., & Jansen, P. (1994). Towards a chess program based on a model of human memory. In H. J. van den Herik (Ed.), *Advances in computer chess 7*. Maastricht.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: a mechanism for recalling several boards. *Cognitive Psychology, 31*, 1–40.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24*, 651–682.

de Groot, A. D. (1965). *Thought and choice in chess*. New York: Mouton.

de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Assen: Van Gorcum.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

Hofstadter, D. R. (2001). Epilogue: analogy as the core of cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 499–538). MIT Press.

Hofstadter, D. R., & Farg. (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. New York: Basic Books.

Holding. (1992). Theories of chess skill. *Psychological Research, 54*, 10–16.

Indurkhya, B. (2006). Emergent representations, interaction theory and the cognitive force of metaphor. *New Ideas in Psychology, 24*, 133–162.

Indurkhya, B. (2007). Rationality and reasoning with metaphors. *New Ideas in Psychology, 25*, 16–36.

Klein, G. (1999). *Sources of power*. Cambridge: MIT Press.

Lakoff, G. (1990). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

Linhares, A. (2000). A glimpse at the metaphysics of Bongard problems. *Artificial Intelligence, 121*, 251–270.

Linhares, A. (2005). An active symbols theory of chess intuition. *Minds and Machines, 15*, 131–181.

Linhares, A. (2008). Dynamic sets of potentially interchangeable connotations: a theory of mental objects. *The Behavioral and Brain Sciences, 31*, 389–390.

Linhares, A. Decision-making and strategic thinking through analogies, submitted for publication.

Linhares, A., & Brum, P. (2007). Understanding our understanding of strategic scenarios: what role do chunks play? *Cognitive Science, 31*, 989–1007.

Linhares, A., & P. Brum. (2009). How can experts see the invisible? Reply to Bilalic and Gobet. *Cognitive Science*, 33, 748–751.

McGraw, G. (1995). Letter spirit (part one): emergent high-level perception of letters using fluid concepts. PhD thesis, Indiana University, Bloomington.

McGregor, S. J., & Howes, A. (2002). The role of attack and defense semantics in skilled player's memory for chess positions. *Memory and Cognition, 30*, 707–717.

Marshall, J. (1999). Metacat: a self-watching cognitive architecture for analogy-making and high level perception. PhD thesis, Indiana University, Bloomington.

Mitchell, M. (1993). *Analogy-making as perception*. Cambridge: MIT Press.

Mitchell, M., & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy-making. *Physica D, 42*, 322–334.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge: MIT Press.

Rehling, J. A. (2001). Letter spirit (part two): modeling creativity in a visual domain. PhD thesis, Indiana University, Bloomington.

Ries, A., & Trout, J. (1993). *The 22 immutable laws of marketing*. New York: HarperCollins.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse, *Perception and Psychophysics*, 2, 437–442.

Smith, B. C. (1996). *On the origin of objects*. Cambridge: MIT Press.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.