ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research xxx (2011) xxx–xxx

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction

Action editor: John Barnden

Alexandre Linhares [a,c,*], Anna Elizabeth T.A. Freitas [a], Alexandre Mendes [b], Jarbas S. Silva [a]

[a] The Getulio Vargas Foundation, Praia de Botafogo 190/509, Rio de Janeiro, Brazil
[b] School of Electrical Engineering and Computer Science, The University of Newcastle, University Drive 2308, Callaghan, New South Wales, Australia
[c] The Club of Rome, Rämistrasse 18 – 8001, Zurich, Switzerland

Received 5 August 2010; accepted 8 December 2010

Abstract

We question Chase and Simon's (1973) study concerning the content of the chess chunks, and we conduct a new variation of the classic chess reconstruction experiments, analyzing 25 types of possible reconstruction errors of grandmasters, masters, and beginners. The differences between the errors conducted in poor, intermediate, and strategically perfect reconstructions provide insights concerning the encoding of experts. The results obtained shed clear light into the debate concerning the importance of abstract thought (i.e., forward search) vs. perceptual processes (i.e., pattern recognition). We claim that a clear solution to this debate is ultimately unfeasible, as our experiments demonstrate high entanglement of perception and reasoning. Our results provide additional evidence that analogy is central to strategic thought in chess.

Keywords: Decision-making; Perception; Combinatorial games; Strategic thought; Cognitive psychology

1. Introduction

Most models of decision-making and strategic thinking are based on game-theoretical ideas. Yet, the human mind works in different ways from those postulated by game theory (or by expected utility theory). While game theory has provided us with valuable insights into numerous scientific fields, the assumption that humans think and act 'rationally' as postulated does not stand empirical enquiry (Camerer, 2003; Zarri, 2009).

Consider the combinatorial game of chess. How does one go from perceiving isolated pieces scattered across a board to a coherent, abstract, vision of how the strategic situation will unfold? Wherein lies the boundary between perception and reasoning?

As is well-known, one can always explore the decision tree, as postulated by game theory (and successfully conducted by AI programs). There is, however, evidence that humans do not expand the game tree extensively, and psychological experiments have shown that pattern recognition plays a crucial role in chess thought (de Groot, 1965; de Groot & Gobet, 1996; Gobet, 1998; Linhares, 2005).

There is an ongoing debate concerning whether humans play through pattern recognition, forward search, or which of these processes 'dominates' the other (Chabris & Hearst, 2005). In this paper we provide new evidence concerning this issue: the question may be empirically irresolvable, as pattern recognition and forward search processes may be deeply entangled. Some models have posed that patterns (i.e., chunks) encode bindings between pieces and the squares they are located, with the possibility of "template"

* Corresponding author at: The Getulio Vargas Foundation, Praia de Botafogo 190/509, Rio de Janeiro, Brazil. Tel.: +55 21 9996 2505.
E-mail address: linhares@clubofrome.org.br (A. Linhares).

1389-0417/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2010.12.006

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

slots: a template slot would hold which particular piece lies at square, say, A4, or, alternately, in which square, say, the white king may be found (Gobet & Simon, 2000; de Groot & Gobet, 1996; Gobet & Jackson, 2002).[1]

It is, however, possible, that chunks encode more abstract bindings, as others have pointed out since de Groot (1965). For example, chunks may encode the fact that the black king is under check (not only its position on the board). These more abstract bindings can include information at many levels of abstraction (Linhares, 2005, submitted for publication; Linhares, Chada, & Aranha, in press), such as: (i) the squares that a piece is able to move to in the upcoming plies, or alternatively, (ii) the fact that a piece is blocked from moving at all, or (iii) a number of defense (or attack) relations held by a certain piece, or even more abstractly, the fact that a piece holds an *abstract role* (the piece has a particular capability in the particular context). We devise manipulations of the classic reconstruction experiments, in which subjects are asked to reproduce positions after brief exposures, to probe some of these "higher-level" bindings.

Our data will show that some of these higher-level bindings are clearly encoded by experts, sometimes *at the cost of* lower-level bindings. The patterns most recognized by experts and disregarded by others are concentrated on abstract relations rather than on superficial features of a situation. By encoding these more abstract features of a situation, experts can perceive similarities between situations that, on the surface, might appear to have no resemblance (Hofstadter, 2001; Hofstadter & FARG, 1995; French 1995). If the essence of a position, *as encoded by experts*, consists in large part of these higher-level bindings, this may help explain the power of generalization from previous experience to situations unforeseen, and our data may provide additional support to the hypothesis that analogies are central to chess perception and thought, perhaps through abstract roles (Linhares, 2005, submitted for publication; Linhares & Brum, 2007, 2009). Our results also point to a number of seemingly counter-intuitive findings.

In the next section, the recent literature of chess cognition is discussed. Section 3 documents materials and methods; Section 4 presents our results; and Section 5 provides a discussion of the findings.

2. Current debates on chess expertise

Chase and Simon (1973)—following de Groot's (1965) footsteps—crafted a crucial study concerning perception in chess: they asked players to reconstruct positions after brief glances at the board, which showed that experts performed much better than novices in reconstructing the boards after a brief display. Chase and Simon also created a new condition with *random* board positions, in which the experts' reconstruction ability dramatically dropped. Chase and Simon concluded that experts saw meaningful chunks in real positions, and since random positions were meaningless, no such chunks (or very few of them) were available for precise reconstruction. This conclusion has been verified in numerous experiments, including other domains (e.g., see reviews in Gobet, 1998; Linhares, 2005; Linhares & Freitas, 2010).

Chase and Simon, however, attempted to go further and study the *content of the chess chunks* held by masters. Chase and Simon's (1973) method for studying the content of chunks has been criticized in Linhares and Freitas (2010), a critique briefly summarized here.

2.1. Chase and Simon's inaccuracy

Chase and Simon believed time could precisely separate chunks, counting, for example, the seconds that players took in placing the pieces on the board. Pieces placed subsequently (small time-intervals) were classified in the same chunk; whereas other pieces would not belong to the chunk—and from that belief their chunk encoding followed.

Their data was surprising: chess masters and beginners did not have any measurable difference. Correlations in their 'perception task' were thus: "Master vs. Class A = 0.93; Master vs. Class B = 0.95, and Class A vs. Class B = 0.92" (p. 65). Correlations on their 'memory task' were similar: "Master vs. Class A = 0.91, Master vs. Class B = 0.95, and Class A vs. Class B = 0.95" (p. 70). Chase and Simon's tasks clearly did not differentiate between masters and beginners. Yet they proclaimed that "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). Except for the study's initial part on board reconstructions, there is no data in Chase and Simon (1973) that separates masters from beginners:

> "*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.*" Linhares and Freitas (2010)

A constant never explains a variable. Because the times measured were strikingly similar between a beginner and a master player, there could be no causality between the high

[1] While we do agree with the general framework of this research programme (which postulates that perception lies at the core of chess, advocates that only computational models can provide a clear, unambiguous refutable theory of expertise, and, of course, includes the theoretical notions of chunks), we do not agree with the specific chunk encodings previously published in the literature (Linhares & Brum, 2009; Linhares & Freitas, 2010). We see this work as *complementary* to computational models such as CHREST, rather than as a *dismissal* of that body of work.

number of chunks held by a master, the low number of chunks held by a beginner, and the resulting measured times. If chunks influenced the time measurements, these measurements would need to be different between subjects. Yet, they were not.

Instead of concluding that the time measurements were not causally affected by chunks, and hence the time-delineated data could not be used to study such chunks, they went on to study the content, number, and size of chunks. Pieces that were included in a reconstruction before a time limit would belong to the same chunk, and others pieces would belong to other chunks. Linhares and Freitas (2010) provide the full argument concerning the shortcomings of the second part of Chase and Simon (1973), and in this work the problem is approached from a different manipulation.

Our new manipulation of the reconstruction experiments is based on three ideas: (i) error differentials, instead of time-intervals, (ii) strategic scenarios of endgame situations and (iii) the reconstruction quality, not the player level.

Error differentials, instead of time-intervals: Our focus here, and our claims, are concentrated on the error differentials of reconstructions. For each type of error studied, we are looking at the differentials in reconstruction quality level. If, for a particular type of error, there are large differences between reconstructions, then those that do not commit the error have learned to quickly encode that type of information (i.e., that type of information is part of a chunk). If, for another type of error, there are few differences between high-quality reconstructions and low-quality reconstructions, then that type of information is not crucial to distinguish between different levels of reconstructions.

Strategic scenarios of endgame situations: We concentrate focus on endgame positions, in which the strategic scenarios to unfold are clear (e.g., a mate-in-three). With full knowledge of which pieces play significant roles in each position, we can analyze precisely the reconstructions that committed non-important errors and the reconstructions that committed strategic errors. Sometimes a misplacement of a piece could alter the strategic situation, and sometimes the piece would not be involved at all in the strategic situation, and its misplacement had no effect.

Finally, instead of looking at a *player skill level*, we look at a *reconstruction quality*. Knowledge of reconstruction experiments at the player level is well-known. It is well disseminated that grandmasters reconstruct with great accuracy, and that beginners have enormous difficulty. Of course, in our experiments, reconstruction quality and player skill were associated, as in all previous studies. However, our focus here is on the reconstruction quality, or rather the differentials between high-quality reconstructions and lower-quality ones. The information registered in high-quality reconstructions, but lost in lower-quality ones provides a window into the content of the chess chunk, because experts must have learned to rapidly register such information.

2.2. *The debate between pattern recognition vs. reasoning in chess*

There is an open debate concerning whether reasoning (i.e., forward search) or perception (i.e., pattern recognition) is the most important process underlying chess expertise. Holding (1992) proposes that chess play is mostly conducted through forward search, and even mentions computer programs as a cognitively plausible model.

Gobet and Simon (1998a) argue against that view, pointing to literature and experiments that place pattern recognition at center stage: “No search, no application of an evaluation function is needed for a first approximation [of a good move], and this automatic process is sufficient to play a reasonable game in speed chess or simultaneous games against weaker players” (p. 207).

Chabris and Hearst (2003, 2005), on the other hand, have used computer programs to find out blunders in games (in normal, rapid, and blind chess), and concluded that (i) there was a large gain in performance between the rapid and the normal game conditions; and that (ii) there was no significant gain between the blind and the rapid conditions. Chabris and Hearst (2005) concluded against “theoretical claims about the relative values of pattern recognition and forward search in chess expertise, especially a claim by Gobet and Simon (1996) that pattern recognition is the more important process” (Chabris & Hearst, 2005, p. 1). Chabris and Hearst (2003) theorized that such claims seem to be untestable, yet, the debate stands (Gobet 2003): Jeremic, Vukmirovic, and Radojicic (2010), for example, dispute Chabris and Hearst’s claims concerning the number of blunders in blindfold and rapid conditions—yet they duly mention that the disparity can be due to the computer technology to evaluate blunders, which, of course, means that further research is needed. A new computer program could turn the tables either way.

From a broader perspective, our study will provide additional evidence against cognitive models and theories concerning, or presupposing, “perception modules” encapsulated from “reasoning and inference modules”: though numerous authors have argued that perception and thought seem to be deeply entangled (Chalmers, French, & Hofstadter, 1992; French, 1997, 2008; Hofstadter & FARG, 1995; Linhares, 2000, 2008, submitted for publication), this remains under debate in the chess literature.

We may now proceed to the design of experiments.

3. Materials and methods

3.1. *Data collection*

What type of information is encoded in a chess player’s chunks? Chase and Simon (1973) and subsequent studies focused mainly in depicting the success rate of such reconstructions, for instance, by plotting the number of correctly positioned pieces against the skill of the players. In this study, however, we adopt a novel perspective: *Differences*

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

between the mistakes made by different categories of players indicate which attributes players learn to register as the years of chess play accumulate. What type of information is registered by high-skilled players, and missed by low skilled ones? What are the features of strategically perfect reconstructions, as compared to inferior reconstructions?

Whatever structures are unique to a grandmaster's encoding should not appear on a lower-skilled player's reconstruction (and mistakes should arise specifically due to the lack of these structures). Which would be the most common errors for beginners and for experts? Consider the following examples:

(i) A king, the most important piece of the board, was not registered, i.e., it was omitted in the reconstruction.
(ii) A rook was omitted.
(iii) A defense, i.e., a piece A protecting a piece B of the same color, was omitted in the reconstruction.
(iv) A queen was misplaced, i.e., put in a different square as compared to the original position.

It is not a trivial task, *a priori*, to determine which of these types of errors, in isolation, indicate a superior reconstruction. The method presented in this study allows us to determine the importance of such attributes, i.e., how the strategic vision of the players is structured in high-skilled players, by contrasting the few mistakes they make with the many mistakes made by those without the proper knowledge. (We find out, for example, that queen misplacement is not one of the most differentiating characteristics between high-quality reconstructions and low-quality reconstructions.)

3.2. Participants

Chase and Simon (1973) studied three subjects: a master, a class A player, and a class B player. Here we study six players with higher and more disparate levels of expertise. Two of the players were categorized as *International Grandmasters* (players **Q** and **W**), according to the rules from the World Chess Federation (FIDE rating > 2400); two other players (**E** and **R**) were categorized as FIDE *Masters* (rating > 2200); and two others (**T** and **Y**) were advanced beginners (1500 < rating < 1700). Finally, there was a 7th player, who made comments on the reconstructions, and was also an International Grandmaster.

3.3. Chess positions

A total of 10 positions were used in this study; white to move and win (at least historically) in all matches (Table 1). They represent the state of the match in the last movement, i.e., just before the mate or black's resignation; and were taken from matches with players that had an ELO rating higher than or equal to 2200. In addition to that, four random positions were included to allow the reconstruction of the results from Chase and Simon (1973) and from Gobet and Simon (2000) as a check on our procedure.

The positions were selected in a way such that, after a quick analysis, the strategic situation of the match could be understood. This is the reason why only final positions were chosen. In addition to the imminent mate/resignation, generally the number of pieces is relatively small (min = 12, max = 19, median = 15, mean = 15.6).

Each player had two attempts to reproduce each position. After observing each position on a computer display for 5 s, they would try to reproduce it using a real chess board. After the first trial, all pieces were removed from the chess board, and the process was repeated for a second time. We are following on Chase and Simon's (1973) footsteps, in which players would have a small number of reconstruction attempts. For the best players, two solution attempts usually lead to near perfect reconstruction—giving us a baseline for comparison with the errors of lower-quality reconstructions. The types of errors not committed in strategically perfect reconstructions reflect the types of information expert players have learned to encode. In this analysis we consider the second, final, reconstruction only. Players took around 3 min for both reconstruction attempts. In Table 2 we present the attributes that were used to evaluate the reconstructions.

Examples of attributes consist of piece misplacements, piece omission, different type of piece, and the possible lost chess relations (e.g., if a queen is mistakenly reconstructed

Table 1
Games from which the positions used in this study were taken. ECO stands for "Encyclopedia of Chess Openings", and is the type of opening for each position used.

White	ELO	Black	ELO	Venue	Date	Opening	ECO
Anand Viswanathan	2715	Jussupow Artur M	2665	Candidate match	01/94	Steinitz defence deferred	C75
Ivanchuk Vassily	2735	Jussupow Artur M	2625	Candidate match qf	08/91	Queen's gambit declined	D35
Jussupow Artur M	2610	Spraggett Kevin B	2575	Candidate match qf	01/89	Tarrasch defence	D34
Seirawan Yasser	2570	Tal Mikhail N	2565	Candidate tournament	10/85	Queen's gambit declined	D46
Gurieli Nino	2370	Kachiani-Gersinska Ketino	2415	Candidate tournament w	12/97	Panno variation	E63
Maric Alisa	2460	Peng Zhaoqin	2400	Candidate tournament w	12/97	King's Indian defence	E61
Gurieli Nino	2370	Cramling Pia	2520	Candidate tournament w	12/97	Exchange variation FR	C01
Wang Pin	2345	Gaprindashvili Nona T	2435	Candidate tournament w	10/92	Austrian attack	B09
Guedes Armando	2285	De Asis Dirceu Viana Jr	2380	Capablanca M	05/97	Queen's gambit declined	D37
Harikrishna Pendyala	2200	Belmonte D	2200	World Junior Ch boys	11/98	Flohr-Mikenas variation	A19

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

Table 2
Attributes used in the study to evaluate the reconstructions. Superficial attributes (**s**) bind a particular square of the board with a particular type of piece (e.g., white knight on B4). Other attributes are considered higher-level, more abstract, bindings as a threshold (**a**).

Attribute name	Description
Omitted kings	(s) Counts the number of omitted kings
Omitted queens	(s) Counts the number of omitted queens
Omitted rooks	(s) Counts the number of omitted rooks
Omitted knights	(s) Counts the number of omitted knights
Omitted dark bishop	(s) Counts the number of omitted bishops in black squares
Omitted light bishop	(s) Counts the number of omitted bishops in white squares
Omitted pawns	(s) Counts the number of omitted pawns
Misplaced kings	(s) Counts the number of misplaced kings
Misplaced queens	(s) Counts the number of misplaced queens
Misplaced rooks	(s) Counts the number of misplaced rooks
Misplaced knights	(s) Counts the number of misplaced knights
Misplaced dark bishop	(s) Counts the number of misplaced bishops in black squares
Misplaced light bishop	(s) Counts the number of misplaced bishops in white squares
Misplaced pawns	(s) Counts the number of misplaced pawns
Lost attacks	(a) Counts the number of lost attacks, as a result of omitted, wrongly located, or blocked pieces
Lost defenses	(a) Counts the number of lost defenses, as a result of omitted, wrongly located, or blocked pieces
Misplaced attack	(a) Counts the number of attacks created, as a result of omitted, wrongly located, or blocked pieces
Misplaced defenses	(a) Counts the number of defenses created, as a result of omitted, wrongly located, or blocked pieces
Different color	(s) Counts the number of pieces with a different piece color in the reconstructions
Different piece	(s) Counts the number of pieces of different types in the reconstructions
Number of moved pieces	(a) Counts the number of pieces moved in the reconstructions
Moved piece total Euclidean distance	(a) Sum of the length of the movements in the reconstruction, given by the Manhattan distance (i.e. number of squares between the original and reproduced locations of each piece)
Moved piece total topological distance	(a) Sum of movements in the reconstruction, given by the topological distance (i.e. following the movement rules of the piece). If a movement is *not possible*, e.g. a pawn moved to the square adjacent to it, a penalty of six movements is given (no penalty would distort the topological distance between the pieces)
Geometrical structures 3 pieces	(a) Counts the number of chains of pieces with exactly three elements, which were not registered in the reconstruction
Geometrical structures 4 pieces	(a) Counts the number of chains of pieces with exactly four elements, which were not registered in the reconstruction

as a rook, any attack/defense on the diagonals will be lost). This leads us to a distinction between superficial attributes and abstract attributes. Superficial attributes bind a specific piece to a specific square (say, a white rook occupies square a1). Superficial attributes are those that do not necessarily have a direct relation with the strategic situation of the position (the piece combination might be shifted, or an attacking piece might be substituted to another piece type, while playing the same role). Abstract attributes, on the other hand, are those that are a consequence of specific combinations of pieces, and are independent of the particular set of pieces present in the position (for example, a piece p_1 attacks/defends piece p_2, without consideration to location or even the types of pieces).

3.4. Data standardization

Attributes were standardized through the equation below:

$$N(R_P(A)) = \left(\frac{Max_P(A) - A_{R_P}}{Max_P(A)}\right),$$

where

- R_P is the reconstruction R of position P;
- A_{R_P} is the number of errors of attribute A in reconstruction R of position P;
- $Max_P(A)$ is the maximum number of errors for attribute A given all reconstructions of position P. $Max_P(A) > 0$ for all positions P, in which error type A can arise (i.e., a position without knights cannot contain errors such as *Misplaced_knights*).
- $N(R_P(A))$ is the relative value, within the interval [0, 1], of attribute A from reconstruction R of position P. A value of 1 indicates a perfect reconstruction for attribute A; a value of 0 indicates that reconstruction R is the worst one for attribute A, position P; intermediate values indicate relative errors between a perfect reconstruction and the worst one, for each attribute and for each position (note that for every position there were always perfect reconstructions).

The reader may thus notice that, after this standardization, higher values represent better reconstructions. For example, a value of 1 in attribute *Omitted rooks* means that board reconstruction was the one with the least number of such omissions. A value of zero means that the board reconstruction had the highest number of such errors.

4. Results

Which attributes differentiate reconstructions into different classes? Clustering through expectation–maximization (Dempster, Laird, & Rubin, 1977) generated three distinct

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

classes of reconstructions (number of clusters obtained through cross-validation; log likelihood 1.62177):

(i) *Strategically perfect reconstructions*: Reconstructions in which the strategic scenario is not altered (reconstruction mistakes are allowed, but they do not change the strategic situation). Because these are end-game scenarios in which we have full knowledge of what is about to happen, reconstruction errors that close the upcoming avenue of events (or that open other avenues) are considered to change the strategic situation.
(ii) *Intermediate reconstructions*: Reconstructions in which most of the information was preserved, but the strategic relations were lost.
(iii) *Poor reconstructions*: Reconstructions in which not only the strategic relations are lost, but also several pieces and other relations.

These three classes were assigned labels, namely *class 0 = poor reconstructions*; *class 1 = intermediate reconstructions*; *class 2 = strategically perfect reconstructions*. Three experiments were conducted. The first contrasts the three classes simultaneously. The second experiment contrasts class 2 (strategically perfect reconstructions) against classes 0 and 1 grouped together. Finally, the third experiment differentiates poor reconstructions (class 0) from intermediate and perfect reconstructions put together (classes 1 and 2—note that most of the strategically perfect reconstructions are associated to players **Q** and **E**).

Two types of methods were used to analyze the players' reconstructions. In each experiment we look for both *evaluation of attributes* and *classification*:

(i) *Evaluation of degree of differentiation for each attribute*: This was done using two methods: χ^2 and the *class-information entropy* test. χ^2 computes, for each attribute, the sum, across all classes, of the squared difference between the sum of the actual scores over all reconstructions in the class and the expected sum of such scores under the null hypothesis of no ability for this attribute to differentiate the classes. Class-information entropy, on the other hand, discretizes the attribute values using thresholds, and tries to minimize the sum, across all classes, of the noise that is associated with discretized values of the attribute. Both methods generate a value that indicates the 'success' of the attribute in separating examples from each class. A value of zero indicates that the attribute separates the classes with the "accuracy" of random choice—that is, the attribute does not convey meaningful information to separate the classes. For details, we refer readers to Witten and Frank (2005), Kvam and Vidakovic (2007), and Hall and Holmes (2003). Evaluation tests allow us to order the attributes by their ability to differentiate between reconstructions: what kinds of errors most differentiate high-quality reconstructions from poor reconstructions? Because high-quality reconstructions avoid some types of errors, found in lower-quality reconstructions, this is the type of information that more skilled players must be encoding. Because these tests did not provide sufficient information regarding the particular attribute threshold values that separate high-quality reconstructions from lower-quality ones, we also provide classification tests.
(ii) *Classification*: Our second test involves the use of classifiers such as J48 and *Naïve Bayes*. We have also tested *classification through linear regression* classifier, but the results were inferior and are not reported here. Classification tests tell us the particular threshold values that the attributes carry in order to separate reconstructions as high-quality, low-quality, or intermediate quality. These tests do not necessarily provide information regarding the order of importance of attributes.

4.1. Decision trees using the J48 algorithm

A decision tree was used for the classification of the examples, which in this study correspond to the reconstructions. Internal nodes of the tree contain attribute evaluations, which guide the classification of the example. Leaf nodes contain labels, which correspond to the class assigned to examples classified (correctly/incorrectly) in that node. J48 generates a decision tree that is used for the classification of the examples, which in this study correspond to the reconstructions.

Several algorithms are available to generate decision-tree-based classification models. Among the most traditional, we must cite ID3, for Iterative Dichotomiser 3 (Quinlan, 1986); C4.5 (Quinlan, 1993), an extension of ID3 that uses the concept of class entropy, allowing for missing values and discrete variables; and finally J48 (Witten & Frank, 2005), an extension of C4.5 with parameters that control the size of the tree; the elimination of branches with no bearing on classification accuracy; a minimum number of examples classified in each leaf node; among other characteristics. We have employed the algorithms from the *Waikato Environment for Knowledge Analysis* (Witten & Frank, 2005).

4.2. Classification via Naïve Bayes

Despite the assumption of independence, Naïve Bayes has been surprisingly successful in several machine learning applications by employing Bayes' Theorem; i.e., given hypothesis H and evidence E:

$$P[H|E] = \frac{P[E|H]P[H]}{P[E]}$$

where $P[H|E]$ means the probability of hypothesis H given the evidence E, i.e. updating the *a posteriori* probability

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

given a prior event. The method allows us to establish an estimation of the probability distributions of each attribute, given a class of reconstructions. The method is well-known and additional information is widely available (see, for instance, Witten & Frank, 2005), and we may now proceed to the aforementioned experiments.

4.3. Experiment #1: Differentiation of classes 0, 1 and 2

4.3.1. Attribute ranking

Table 3 shows the results of χ^2 and of the class-information entropy tests. Both methods returned the same set of 20 attributes as relevant, with slightly different rankings. The five irrelevant attributes were *Misplaced dark bishop*, *Misplaced queens*, *Omitted dark bishop*, *Misplaced knights* and *Omitted light bishop*. This is a result worthy of note: the misplacing of a queen, the most powerful piece of the game, is *not* a defining attribute between excellent reconstructions and poor ones. This is counter-intuitive, to say the least, and is one of our findings to be discussed further on.

4.3.2. Discussion

Note that both χ^2 and the class-information entropy tests yield similar results. Among the most important attributes to differentiate the three classes, there are both superficial and abstract ones.

Consider a comparison between the six most important attributes and the six least important ones, according to χ^2. The most important attributes are *Lost attacks*, *Moved piece total topological distance*, *Lost defenses*, *Number of moved pieces*, *Moved piece total Euclidean distance* and *Omitted pawns*. In contrast, the six least important attributes are *Omitted knights*, *Misplaced dark bishop*, *Misplaced queens*, *Omitted dark bishop*, *Misplaced knights* and *Omitted light bishop*. Only *Omitted knights* is shown in Table 3; the other five attributes have χ^2 values equal to zero and were omitted.

Notice also that the most important attributes are all abstract, except *Omitted pawns*. These attributes are not concerned with a particular type of piece. On the other hand, the least important attributes are all connected to omission or location errors of some particular type of piece. The current theory, in models such as CHREST, CHUMP, following the original EPAM (Gobet, 1998), suggest that abstract errors should be of low relevance, as these models presuppose that the key-structures of representations follow a precise binding of particular-piece-on-particular-square (e.g., a black knight on square A3).

For a better visualization of these results, Fig. 1 shows a heatmap with the value of all 25 attributes, i.e. the 20 attributes present in Table 3, plus the five irrelevant ones, for the 60 reconstructions (six players, 10 positions). The attributes correspond to the rows, and the reconstructions to the columns. The class that each example belongs to is represented by a color code, namely green = poor reconstructions (values close to 0); blue = intermediate reconstructions (intermediate values); red = strategically perfect reconstructions (values close to 1). The classes are shown as an extra row below the heatmap.

The heatmap is generated by a combinatorial optimization algorithm that permutes columns and rows, in order to minimize energy (as measured by incongruence with neighboring cells). Thus, it provides a "bird's-eye" view of our data. Note, for example, that strategically perfect reconstructions are associated to the skilled players, by looking at the columns mostly filled with red spots (which mean no or few relative mistakes concerning each particular attribute). By looking at rows, one can readily see the

Table 3
Results from Experiment #1: attributes that differentiate the three classes, according to a χ^2 and the class-information entropy.

χ^2		Class-information entropy	
Value	Attribute	Value	Attribute
72.73	Lost attacks	0.944	Lost attacks
37.92	Moved piece total topological distance	0.515	Moved piece total topological distance
34.84	Lost defenses	0.504	Lost defenses
34.38	Number of moved pieces	0.496	Omitted pawns
34.38	Moved piece total Euclidean distance	0.455	Moved piece total Euclidean distance
32.84	Omitted pawns	0.455	Number of moved pieces
24.45	Different color	0.342	Geometrical structures 3 pieces
24.08	Misplaced defenses	0.341	Misplaced defenses
22.69	Geometrical structures 3 pieces	0.311	Omitted rooks
20.40	Misplaced pawns	0.310	Different color
19.83	Omitted rooks	0.285	Different piece
17.78	Misplaced kings	0.283	Misplaced kings
17.30	Different piece	0.267	Misplaced pawns
14.17	Misplaced attack	0.201	Geometrical structures 4 pieces
13.58	Misplaced rooks	0.186	Misplaced attack
12.81	Geometrical structures 4 pieces	0.182	Misplaced rooks
11.03	Omitted queens	0.177	Omitted queens
10.72	Omitted kings	0.152	Omitted kings
7.96	Misplaced light bishop	0.116	Misplaced light bishop
6.79	Omitted knights	0.111	Omitted knights

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

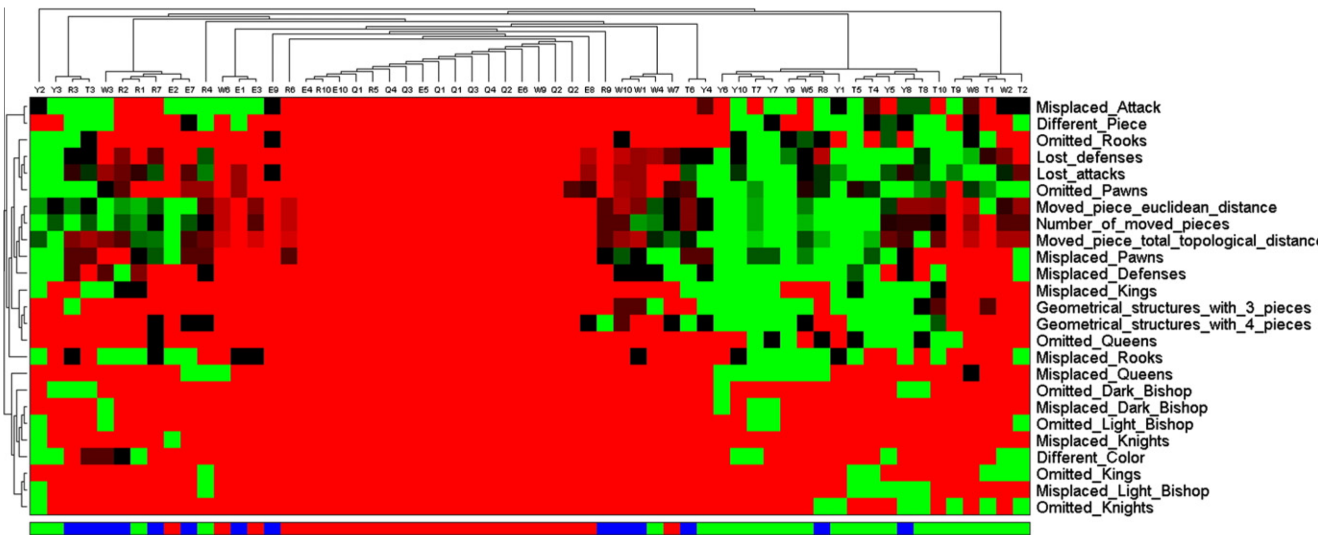


Fig. 1. Heatmap generated using Biominer's hierarchical clustering and showing the relations between all 25 attributes and the 60 reconstructions (Moscato, Mendes, & Berretta, 2007). In the heatmap, the red color is associated to zero or few errors for the attribute in the reconstruction; green is associated to large numbers of errors; and black is associated to intermediate numbers of errors. The row below the heatmap indicates the class of the example; i.e. green = poor reconstructions; blue = intermediate reconstructions; red = strategically perfect reconstructions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Results from Experiment #1: summarized results for the classifiers.

	Naïve Bayes		J48	
Correctly classified examples	49	81.67%	**52**	**86.67%**
Incorrectly classified examples	11	18.33%	**8**	**13.33%**
Kappa statistic	0.7197		0.7946	
Mean absolute error	0.1246		0.1044	
Root mean squared error	0.3345		0.2881	
Relative absolute error	28.56%		23.93%	
Root relative squared error	71.54%		61.62%	

disparities in the distribution of errors, for each particular attribute. Some types of errors (attributes) have an almost uniform distribution, and hence do not differentiate high-quality from low-quality reconstructions (e.g., Misplaced_knights). On the other hand, other types of errors provide large variation across reconstructions (e.g., Lost_attacks). These high-variation attributes provide valuable information concerning chess chunks encodings.

4.4. Classification using Naïve Bayes and J48

The results for the classifiers are shown in Table 4. These values were generated using 10-fold cross-validation, i.e., the 60 reconstructions were divided into 10 disjoint sets with six elements each. Then, a classifier is created with 54 reconstructions as a training set and tested in the remaining six reconstructions. The process is repeated 10 times, each of which with a different test set. At the end, the classifiers are combined into an average classifier, which corresponds to the reported one (note that, because of the way the classifier is generated, it is possible, at least in principle, that a test on new data may not be so favorable as the results on Table 4—however, 10-fold cross-validation usually minimizes this sort of problem).

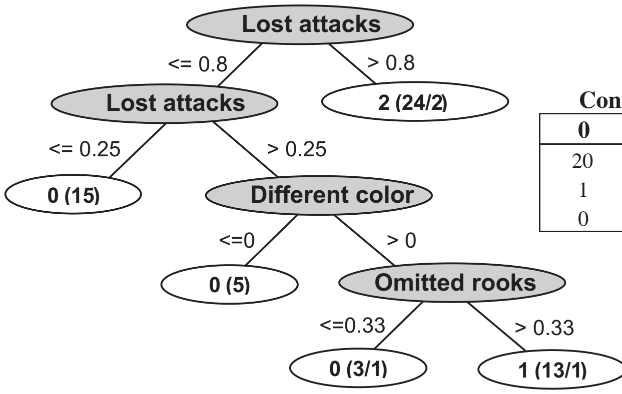
As can be seen in Table 4, J48 was more successful than Naïve Bayes. J48 correctly classified 86.67% of the reconstructions, presenting (i) the lowest absolute error, (ii) the lowest relative error, and (iii) the highest *Kappa statistic*, which indicates a significantly better classification performance. The decision tree generated by J48 is shown in Fig. 2.

4.4.1. Discussion

Fig. 2 depicts a decision tree with nine nodes in total, five of them being leaf nodes. Each internal node contains an attribute and the branches indicate the threshold values for the associated attributes. Each leaf node has a class label (0, 1 or 2) followed by one or two values between parentheses. The first number indicates the total number of reconstructions classified in that node, and the second indicates the number of incorrect classifications. If there is only one number between parentheses, all reconstructions were correctly classified.

In the chess context, these results indicate that the preservation of attacks seems to be the most differentiating characteristic of high-quality reconstructions. Note that *Lost attacks* > 0.8 immediately takes the reconstruction to class 2, i.e. strategically perfect (24 classifications—with two incorrect ones). Note that in the second node the attribute *Lost attacks* is used again. A value lower than or equal to 0.25 immediately leads to the classification of the reconstruction as poor (all 15 reconstructions classified in this node are correct). This is a significant finding: Failure to rapidly recognize the *strategically significant* attacks seems

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006



Confusion matrix

0	1	2	← Classified as
20	2	1	**Class 0**
1	10	3	**Class 1**
0	1	22	**Class 2**

Fig. 2. Decision tree for Experiment #1 created through J48 and the confusion matrix for the classification of reconstructions. If the value of Lost attacks is higher than 0.8 (low relative errors), the reconstruction is classified as excellent (class 2, 24 reconstructions correctly classified, two reconstructions incorrectly classified). On the other hand, if this value is lower than or equal to 0.25, the reconstruction is classified as of lowest quality (class 0, 15 reconstructions correctly classified). The third branchpoint tests whether the maximum number of different color errors was achieved, which leads to five correctly classified instances of class 0. The last branchpoint puts a threshold value of 0.33 concerning the Omitted rooks type of error. A value lower or equal to this threshold leads to class 0 (three correctly classified reconstructions, and one incorrectly classified). A value higher than 0.33 leads to the class 1 of intermediate reconstructions (13 correct classifications, one incorrect). Notice that the number of incorrectly classified reconstructions in the tree and in the confusion matrix is different. The tree depicts the average model obtained using 10-fold cross-validation and the classification of the 60 reconstructions. On the other hand, the confusion matrix shows the median of the classification errors for each one of the 10 classifiers.

Table 5
Results from Experiment #2: attributes that differentiate class 2 from the others, χ^2 and the class-information entropy tests.

χ^2		Class-information entropy	
Value	Attribute	Value	Attribute
48.13	Lost attacks	0.685	Lost attacks
37.81	Moved piece total topological distance	0.511	Moved piece total topological distance
34.26	Number of moved pieces	0.452	Number of moved pieces
34.26	Moved piece total Euclidean distance	0.452	Moved piece total Euclidean distance
28.90	Lost defenses	0.406	Lost defenses
24.86	Omitted pawns	0.394	Omitted pawns
23.08	Misplaced defenses	0.327	Misplaced defenses
20.38	Misplaced pawns	0.284	Geometrical structures 3 pieces
17.28	Omitted rooks	0.284	Different piece
17.28	Different piece	0.284	Omitted rooks
17.28	Geometrical structures 3 pieces	0.267	Misplaced pawns
15.98	Misplaced kings	0.265	Misplaced kings
13.93	Misplaced attack	0.183	Misplaced attack
12.56	Misplaced rooks	0.177	Geometrical structures 4 pieces
10.32	Geometrical structures 4 pieces	0.171	Misplaced rooks
9.32	Omitted queens	0.161	Omitted queens
8.37	Different color	0.146	Different color
5.73	Omitted knights	0.102	Omitted knights
4.93	Misplaced light bishop	0.088	Misplaced light bishop
4.14	Omitted dark bishop	0.075	Omitted dark bishop
4.14	Omitted kings	0.075	Omitted kings
3.39	Omitted light bishop	0.061	Omitted light bishop
2.66	Misplaced dark bishop	0.049	Misplaced dark bishop

to be the most decisive skill of all those probed in this study in what concerns the quality of reconstructions.

After 22 (out of 23) strategically perfect reconstructions done, and 15 (out of 23) correctly classified as poor, there are still eight poor reconstructions to be differentiated from the intermediate ones. In this case, the decision tree starts to use more superficial attributes, e.g. *Different color* and *Omitted rooks*. These attributes have a simple expected consequence, for any mistake degrades relative reconstruction quality.

Table 6
Results from Experiment #2: summarized results for the classifiers.

	Naïve Bayes		J48	
Correctly classified examples	**55**	**91.66%**	53	88.33%
Incorrectly classified examples	**5**	**8.33%**	7	11.66%
Kappa statistic	0.8193		0.7512	
Mean absolute error	0.0557		0.0805	
Root mean squared error	0.2353		0.2525	
Relative absolute error	17.23%		24.93%	
Root relative squared error	59.08%		63.40%	

Table 7
Results from Experiment #2: mean and standard deviation for the probability distributions associated to the attributes with the best precision values for the Naïve Bayes classifier; for classes 0/1 and class 2.

	Classes 0/1 mean	Classes 0/1 stdev	Class 2 mean	Class 2 stdev
Moved piece total topological distance	0.4788	0.3603	0.9130	0.2435
Moved piece total euclidean distance	0.4022	0.3533	0.9130	0.2274
Number of moved pieces	0.3825	0.3358	0.9064	0.2313
Lost defences	0.4615	0.3877	0.9799	0.0689
Lost attacks	0.3808	0.3246	0.9921	0.0371
Omitted pawns	0.4108	0.3889	0.9522	0.1247

Table 8
Results from Experiment #3: attributes that differentiate class 0 from the other two, according to χ^2 and the class-information entropy tests.

χ^2		Class-information entropy	
Value	Attribute	Value	Attribute
39.26	Lost attacks	0.592	Lost attacks
30.93	Lost defenses	0.400	Lost defenses
28.21	Omitted pawns	0.363	Omitted pawns
19.76	Geometrical structures 3 pieces	0.246	Geometrical structures 3 pieces
19.40	Moved piece total topological distance	0.239	Moved piece total topological distance
18.05	Omitted rooks	0.229	Omitted rooks
16.68	Moved piece total Euclidean distance	0.226	Moved piece total Euclidean distance
15.39	Misplaced defenses	0.209	Different color
15.23	Number of moved pieces	0.207	Number of moved pieces
14.85	Different color	0.189	Misplaced defenses
14.70	Misplaced pawns	0.178	Misplaced pawns
12.50	Misplaced kings	0.152	Omitted kings
10.72	Omitted kings	0.150	Misplaced kings
10.45	Geometrical structures 4 pieces	0.125	Geometrical structures 4 pieces
8.53	Omitted queens	0.101	Omitted queens
7.52	Misplaced light bishop	0.092	Misplaced light bishop

4.5. Experiment #2: Differentiation of class 2 against classes 0 and 1, simultaneously

4.5.1. Attribute ranking

As in experiment #1, both tests obtained similar results, presenting the same sets of relevant/irrelevant attributes, but in slightly different rankings. Table 5 shows the relevant attributes; the irrelevant ones were *Misplaced queens* and *Misplaced knights*.

Both χ^2 and *class-information entropy* resulted in similar attribute rankings. There are 23 relevant and two irrelevant attributes, and the seven most relevant were the same for both: *Lost attacks*, *Moved piece total topological distance*, *Number of moved pieces*, *Moved piece total Euclidean distance*, *Lost defenses*, *Omitted pawns* and *Misplaced defenses*. Note that these seven attributes, except *Omitted pawns*, are all abstract, i.e., they are more concerned to the relations between the pieces than the more superficial information of the position.

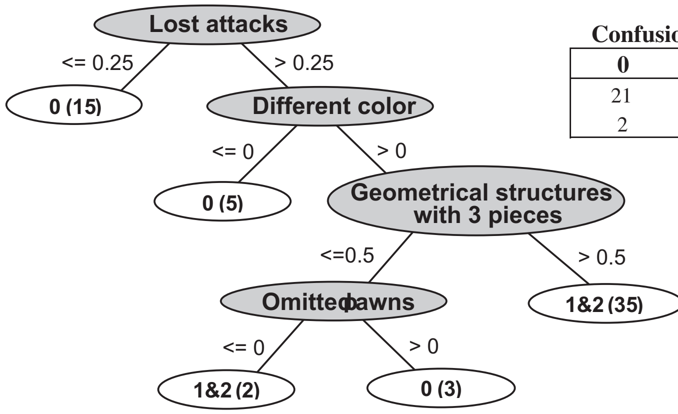
4.5.2. Classification

The same classifiers were tested, but results were different in this case (Table 6). The results indicate that the Naïve Bayes classifier was the best one, with 91.66% of the reconstructions correctly classified. In addition, the *Kappa statistic* was the highest, and both absolute and relative errors were the lowest.

Table 9
Results from Experiment #3: summarized results for the classifiers.

	Naïve Bayes		J48	
Correctly classified examples	51	85%	**56**	**93.33%**
Incorrectly classified examples	9	15%	**4**	**6.66%**
Kappa statistic	0.6953		0.8590	
Mean absolute error	0.0958		0.0499	
Root mean squared error	0.3032		0.2130	
Relative absolute error	29.67%		15.44%	
Root relative squared error	76.13%		53.47%	

Because visualization of the Naïve Bayes classifier is not as intuitive as a decision tree, let us look at its results in more detail. For each class and attribute, there are values associated to probability distributions, such as mean and standard deviation, a *weighted sum* of the attribute/class, and a precision level (the lower the value, the better is the Naïve Bayes result). The six attributes with precision level lower or equal to 0.1 were *Moved piece total topological distance* (precision = 0.0476); *Moved piece total euclidean distance* (precision = 0.0588); *Number of moved pieces* (precision = 0.0769); *Lost defences* (precision = 0.0769); *Lost attacks* (precision = 0.0909); and *Omitted pawns*



Confusion matrix

0	1 & 2	← Classified as
21	2	**Class 0**
2	35	**Classes 1 & 2**

Fig. 3. Decision tree for Experiment #3 created by J48 and the confusion matrix for the classification of the reconstructions. As in Fig. 2, the number of wrongly classified reconstructions is different in the tree and in the confusion matrix.

(precision = 0.1). These are the attributes with the greatest differences in the probability distributions for classes 0/1 and class 2. In Table 7, we show mean and standard deviation values for the probability distributions of these attributes.

These results indicate a sizeable difference for these attributes between strategically perfect reconstructions and all others. For instance, considering the means, in the case of class 2, the worst case is 0.9064, for the attribute measuring number of moved pieces, against 0.3825 for classes 0/1. Note the importance of this difference given the standard deviations. This table summarizes some of the main elements that are used (and learned over the years) by grandmasters. Here, we need again to emphasize that these results oppose the current theory of expertise, as will be discussed on the conclusion.

4.6. *Experiment #3: Differentiation of class 0 against classes 1 and 2, simultaneously*

4.6.1. *Attribute ranking*

Table 8 displays ranks differentiating class 0 (poor reconstructions) from classes 1 and 2 put together.

4.6.2. *Discussion*

Again, both methods obtain similar results; same set of attributes, with slightly different rankings. The main attributes for both methods were *Lost attacks*, *Lost defences*, *Omitted pawns*, *Geometrical structures 3 pieces*, *Moved piece total topological distance*, *Omitted rooks* and *Moved piece total Euclidean distance*. In this case, the number of non-abstract attributes in considerable – *Omitted pawns*, *Omitted rooks* and *Geometrical structures with 3 pieces*. This indicates that *in order to identify poor reconstructions*, attributes associated to superficial structures become relevant.

4.6.3. *Classification*

The classification results in Table 9 show a better performance of the J48 classifier, which classified correctly 56 reconstructions, against only four misclassifications. Next, in Fig. 3, we present the decision tree obtained by the method.

4.6.4. *Discussion*

The decision tree in Fig. 3 indicates that when there is a large number of *Lost attacks* (i.e., leading to a relative value of *Lost attacks* $\leqslant 0.25$), the reconstruction is classified as poor (15 in total). For the other 45 reconstructions, the next test to be applied uses the attribute *Different color*, which refers to errors related to different colors of the same piece type. If this value is zero (i.e. worst value possible relative to all reconstructions), the reconstruction is classified as poor again (5 are classified in this node). The remaining 40 reconstructions are now tested for *Geometrical structures with 3 pieces* (i.e. chains of pawns with 3 pieces or two pawns and another piece). If the associated value is higher than 0.5, the model classifies them as intermediate/strategically perfect (35 in total). Otherwise, a last attribute, *Omitted pawns*, is used. If no pawns were omitted, the classification is poor. This hints at the interesting possibility of *attentional shifts*, in which this type of error is made in intermediate reconstructions but not in poor ones (a *U-shaped learning curve*). This is also counter-intuitive and will be discussed in the conclusion.

As a final confirmatory experiment, we tested whether the null hypothesis of no difference between the three classes could be rejected at the $p < 0.05$ level, through an ANOVA using attributes as independent variables, and the classes as dependent variables. There was no significant effect of reconstruction categories in just four out of 25 dependent variables: number of omitted dark bishops; number of omitted light bishops; number of misplaced knights; number of misplaced dark bishops. In the other hand, all other attributes are significant at the $p < 0.05$ level, according to Table 11.

5. Conclusion

Cicero, at a crossroads, once proclaimed that "*We must not say every mistake is a foolish one*". That is our approach

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

to understanding human perception and thought under strategic scenarios: there is skill hidden behind mistakes.[2] We have redesigned the chess reconstruction experiments—i.e., tasks in which players of varying skill levels reconstruct positions after two 5-s observations. The innovation here is that we have thrown attention to the *mistakes* obtained in the reconstructions. We have ranked the errors and applied classifiers to them, enabling us to precisely pinpoint which types of attributes distinguish high-quality reconstructions from poor reconstructions. Because it is well-known that player level is associated with ability to reconstruct the positions, our approach contrasted the few errors of FIDE grandmasters with those of beginners (and also masters), we believe that this error differential information throws important light into the chunks and structures encoded by experts.

In our study, the data gathered from novices and from experts is unlike that of Chase and Simon (1973) and subsequent studies. Chase and Simon found strikingly similar timings between masters and class B players, and used this timing data to analyze the content of chess chunks. Because there was no difference at all in the timing data, we claim that chunks cannot bear any causality relation to the timings observed on their memory or perception tasks. And if the timings are not causally affected by chunks, no useful information concerning the content, number, or size of the chess chunk can be obtained through use of the timing data (for the full argument see Linhares and Freitas (2010); for background see also Gobet and Simon (1998b) for a review of critiques different from ours).

We, however, analyzed the (few or none) errors grandmasters commit, and we contrasted these to the (intermediate) errors that masters commit and to the (numerous) errors that beginners commit. Unlike the data stemming from Chase and Simon (1973), our data clearly differentiates high-quality reconstructions from poor reconstructions. And we claim that the errors committed in poor reconstructions and not committed in high-quality ones shed light on the types of structures encoded in chunks: experts must have learned to rapidly encode such features and do not commit the corresponding errors. Table 10 shows the most important differentiating attributes we have found in each of the three experiments (χ^2 *test).*

Note that *Lost attacks* is overwhelmingly the defining factor differentiating high-quality reconstructions from poor ones (note also that this is in line with the findings of McGregor and Howes (2002)). *Lost defenses* and *Omitted pawns* also arise prominently here (and *Omitted rooks* in Experiment #3). Global measures such as *Moved piece total topological/Euclidean distance* and *Number of moved pieces* also arise. The importance of global measures is intuitively understood and should also be expected by alternative theories. But why not other attributes? Why isn't the misplacement of a queen, the most powerful piece on the board, a significant differentiating factor? In the next section we discuss some counter-intuitive findings.

[2] Recent studies on the Einstellung effect, by Bilalic, Mcleod, and Gobet (2008a, 2008b) bring a refreshing approach to this idea.

5.1. Counter-intuitive results

There are some results that we would like to elaborate on, and provide an interpretation for wider discussion: the results concerning the *Misplaced queen* attribute, some attentional shifts displayed, and the significance of the *Lost attacks* and *Lost defenses* attributes.

5.1.1. Misplaced queen

That this attribute is not a differentiating factor between reconstructions is highly counter-intuitive. There should be a number of reasons for this. First, as can be seen in the heatmap of Fig. 1, this attribute has relatively low variation across the reconstructions. Precisely because the queen is important, most reconstructions from players of different levels present it correctly. If beginners are able to register the queens' positions, and they are to some extent, then the attribute should be less important as a differentiating factor. Moreover, consider the strategic situation: a queen attacks (and defends) through diagonals, through ranks and through files. Therefore, in what concerns attack and defense relations, a queen can "substitute" for rooks, bishops, pawns, *while preserving* the attack or defense relation. A misplaced queen, if "substituting" for a piece on a significant relation in the strategic situation, would *not* alter the unfolding strategic scenario—if the same abstract role of a piece can be placed by a queen, it is possible that players may register the abstract role and not register the type of piece. This is in line with the view of analogy at center stage in chess (Linhares, 2005, submitted for publication; Linhares & Brum, 2007, 2009), or, as Hofstadter puts it, "slippability" (Hofstadter & FARG, 1995). This will be seen in a reconstruction example below.

5.1.2. Attentional shifts

Here is an interesting result: for some kinds of configurations, especially pawn chains, beginners sometimes perform *better* than masters (i.e., masters may commit some mistakes that a beginner did not commit). And grandmasters will generally *not* commit such mistakes. What we have perceived over the experiment is that the lowest-quality reconstructions (given by beginners) may correctly register a pawn chain, all the while missing many more important pieces in other areas. Masters, on the other hand, tend to rapidly focus on the most important areas of the board, and, in the limited time allotted, miss the specific pawn chain configurations. Finally, grandmasters could easily register both the most important chess relations *and* pawn chains, usually exhibiting few errors (one subject misplaced a single pawn in all 10 reconstructions—a pawn that was playing no role whatsoever in the strategic situation). That this does occur occasionally seems to provide evidence for attentional shifts of masters towards the most important

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

Table 10
Most important attributes given by χ^2 in the experiments.

Experiment #1	Experiment #2	Experiment #3
Lost attacks	Lost attacks	Lost attacks
Moved piece total topological distance	Moved piece total topological distance	Lost defenses
Lost defenses	Number of moved pieces	Omitted pawns
Number of moved pieces	Moved piece total Euclidean distance	Geometrical structures 3 pieces
Moved piece total Euclidean distance	Lost defenses	Moved piece total topological distance
Omitted pawns	Omitted pawns	Omitted rooks
Different color	Misplaced defenses	Moved piece total Euclidean distance

Table 11
ANOVA with the classes as independent variables and attributes as dependent variables.

Dependent variable	MSE	F (2, 60)	p
Omitted_Kings	0.483	6.203	0.004
Omitted_Queens	0.614	6.338	0.003
Omitted_Rooks	1.534	14.793	0.000
Omitted_Knights	0.392	3.635	0.033
Omitted_Dark_Bishop	0.217	2.492	0.092
Omitted_Light_Bishop	0.175	2.359	0.104
Omitted_Pawns	2.513	28.247	0.000
Misplaced_Kings	1.465	10.700	0.000
Misplaced_Queens	0.469	3.574	0.034
Misplaced_Rooks	1.121	7.256	0.002
Misplaced_Knights	0.010	0.302	0.740
Misplaced_Dark_Bishop	0.098	1.580	0.215
Misplaced_Light_Bishop	0.410	4.357	0.017
Misplaced_Pawns	1.461	11.888	0.000
Lost_attacks	3.419	81.942	0.000
Lost_defenses	2.743	39.800	0.000
Misplaced_Attack	0.886	5.334	0.008
Misplaced_Defenses	1.964	15.452	0.000
Different_Color	0.740	7.548	0.001
Different_Piece	1.355	10.124	0.000
Number_of_moved_pieces	2.002	21.740	0.000
Moved_piece_euclidean_distance	1.918	19.531	.000
Moved_piece_total_topological_distance	1.772	18.772	0.000
Geometrical_structures_with_3_pieces	1.932	16.559	0.000
Geometrical_structures_with_4_pieces	1.135	8.625	0.001

relations *in detriment to* configurations such as pawn chains. This is something we cannot prove, but we do conjecture it is true. It does seem that these attentional shifts are a cost placed for recalling the most important areas of the board, for those subjects that are able to do it but cannot recall the entire board. And, given that de Groot and Gobet (1996, chap. 6), for instance, have shown that the eye saccades of high-skilled players rapidly converge on the "most relevant areas of the board", it is certainly plausible that, if unable to reconstruct the board in its entirety, higher-skilled players may reconstruct its more relevant parts.

5.1.3. Lost attacks and lost defenses

By a large factor, of all (non-global) attributes studied here, *Lost attacks* and *Lost defenses* are the most important differentiating factors between high-quality reconstructions and low-quality ones. This suggests that possible chess moves are encoded in the expert's chunks (i.e., beyond the "superficial features"), and will be illustrated in the following, final section. It also suggests a natural extension to modern computational chunking theories: instead of encoding only bindings between pieces and squares, and template slots for open squares or piece location, perhaps chunks could encode the fact that a piece p_1 holds a particular relation to a piece p_2. Current computational models encode low-level information, and, recognized chunks evoke productions with possible movements. Our data suggests that chunks encode not only bottom-up, superficial information, but also top-down hypotheses concerning what may happen next. With this encoding, whenever a computational model preserved an attack while committing a reconstruction error (say, by placing an attacking queen instead of a rook), the reconstructed position would reflect the human-like, psychologically plausible behavior found here.

5.2. Entanglement of perception and reasoning

In the chess literature there is an intense debate concerning pattern recognition and forward thinking (Chabris & Hearst, 2003, 2005; Gobet & Simon, 1996). Our results point to a hypothesis of entanglement between these processes. It seems that, embedded in the chunks recognized by grandmasters, are the *subsequent movement implications*, such as attacks and defenses. This particular point has been argued previously, for instance, by de Groot (1965), or Saariluoma and Kalakoski (1998)—but we propose a qualifier below: superficial information may be lost in the process. If chunks encode top-down hypotheses of what is about to happen, it may be impossible to clearly set apart forward search from pattern recognition (examples of models illustrating bottom-up and top-down information include Hofstadter and FARG (1995) and Sun and Zhang (2004)). The Capyblanca project (Linhares, submitted for publication) shows that high entanglement of perception and higher cognition, of bottom-up (data driven) and top-down (hypothesis-driven) processes, is theoretically feasible from a computational modeling standpoint.

In our study, some reconstruction mistakes are *not counted as long as they do not alter the essence of the situation*. For example, consider a bishop attacking a pawn. If, in the reconstruction, the bishop is relocated to another

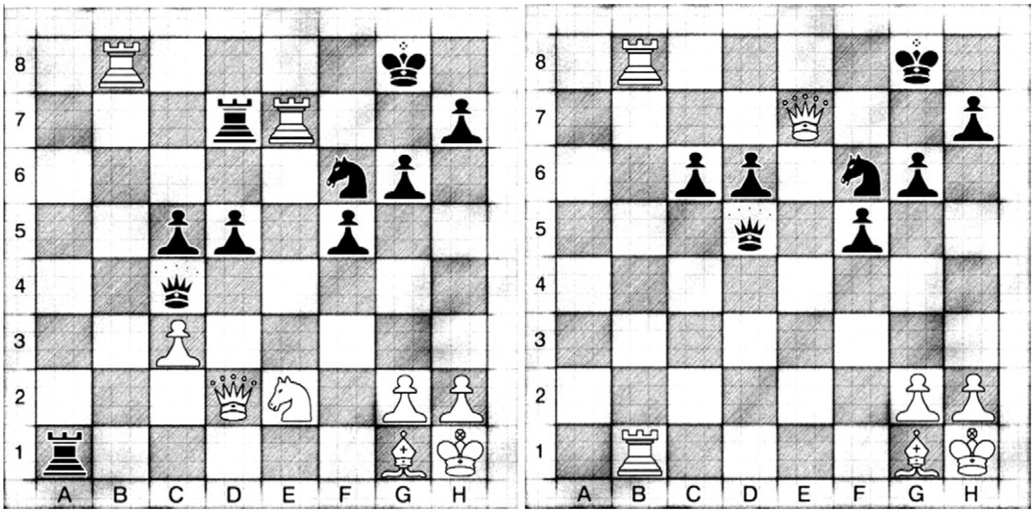


Fig. 4. Left: original position. Right: a player's reconstruction preserves the essence of the situation while committing a large number of errors of various sorts.

square on the same diagonal, but remains attacking the same pawn, then it is not considered as an error under all possible attributes (tested herein). There will be a *Misplaced bishop* error, but not a *Lost attacks* error. If the bishop is changed to a queen, there will be errors of *Piece type* and possibly *Misplacement*, but if the diagonal attack remains, there is no *Lost attack*: the reconstructed position clearly *preserves* the attack, even if it has changed some of its superficial characteristics. This is considered for all the chess relations: all the attacks and defenses leading to the end of the game. Because we are probing endgame positions in which the strategic situation is crystal clear, it is possible to determine *a priori* what is relevant and what is not relevant.

This enables us to see immense beauty in reconstructions such as the one presented in Fig. 4. Notice that this reconstruction has a large number of errors, of many kinds: the original position has 19 pieces, the reconstruction has only 15. Three rooks are misplaced. Two black pawns are misplaced. The white and black queens are also misplaced. One white pawn is omitted, as is one white knight. Ten out of 19 pieces (or 52% of pieces) exhibit errors *at a superficial level.* Under the perspective of pieces-on-squares, this is a dire reconstruction. Current pattern-recognition theories would not see this reconstruction as preserving valuable information.

Yet, deeper probing shows something remarkable. Consider the essence of the position. In the original position, the black king is under threat from the white rooks, and though it can block the check (through r-d8, or k-e8), checkmate is unavoidable. Though 52% of the pieces were either misplaced or simply omitted from the reconstruction, *the essence of the situation remains intact*: the black king continues to be under 'the same' attack (with the white queen replacing the role previously taken by a rook). It is simply remarkable that 52% of pieces can be incorrectly reconstructed with the strategic situation remaining, in effect, intact. This is only possible, we propose, because players encode information at an abstract level, concerning relations and combinations of relations; all the while disregarding, in the brief 10 s displayed, which particular pieces were on each particular squares. These positions are strategically equivalent in the sense of Linhares and Brum (2007, 2009). In the terminology of Linhares (2005, submitted for publication) and Linhares and Brum (2007, 2009) the white queen is said to assume the same "abstract role" of the f7 rook. Relations (and reasoning at an abstract level) were preserved; whereas the underlying pieces were lost. Bindings at a higher-level (p_1 attacks p_2, regardless of piece type or placement) are preserved *at the cost of bindings at a lower-level* (piece type, piece-to-square). The wrong pieces promptly assume the same roles as the original pieces, exhibiting strong "slippability". While 52% of the surface information has mostly vanished, the essence of the situation remains there, pristinely intact. Given that, in the combinatorial nature of the chess microcosm, a single misplaced piece can destroy a large advantage, what are the odds that this could be a coincidence?

Acknowledgements

This work has been generously supported by Grant E-26/110.790/2009 from the FAPERJ Foundation, Grants 470341/2009-2 and 301207/2009-7 from the CNPq of the Brazilian Ministry of Science & Technology, and the PROPESQUISA Program of the Getulio Vargas Foundation. All the collected data is available upon request from Dr. Linhares.

References

Bilalic, M., Mcleod, P., & Gobet, F. (2008a). Inflexibility of experts – Reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology, 56*, 73–102.

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006

Bilalic, M., McLeod, P., & Gobet, F. (2008b). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition, 108*, 652–661.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Chabris, C. F., & Hearst, E. S. (2005). Search, recognition, and visualization in chess: Rebuttal to Gobet's critique of Chabris and Hearst (2003), unpublished manuscript.

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.

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

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

De Groot, A., & Gobet, F. (1996). *Perception and memory in chess*. Assen: Van Gorcum.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological), 39*(1), 1–38.

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

French, R. M. (1997). When coffee cups are like old elephants or why representation modules don't make sense. In A. Riegler & M. Peschl (Eds.), *Proceedings of the international conference new trends in cognitive science* (pp. 158–163). Austrian Society for Cognitive Science.

French, R. M. (2008). Relational priming is to analogy-making as one-ball juggling is to seven-ball juggling. *Behavioral and Brain Sciences, 31*, 386–387.

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

Gobet, F. (2003). Forward search, pattern recognition and visualization in expert chess: A reply to Chabris and Hearst (2003), unpublished manuscript.

Gobet, F., & Jackson, S. (2002). In search of templates. *Cognitive Systems Research, 3*(1), 35–44.

Gobet, F., & Simon, H. A. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master-level chess. *Psychological Science, 7*(1), 52–55.

Gobet, F., & Simon, H. A. (1998a). Pattern recognition makes search possible: Comments on Holding (1992). *Psychological Research, 61*, 204–208.

Gobet, F., & Simon, H. A. (1998b). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*, 225–255.

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

Hall, M., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering, 15*(6), 1437–1447.

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., & Fluid Analogies Research Group (FARG) (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. New York: Basic Books.

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

Jeremic, V., Vukmirovic, D., & Radojicic, Z. (2010). Does playing blindfold chess reduce the quality of game: Comments on Chabris and Hearst (2003). *Cognitive Science, 34*, 1–9.

Kvam, P., & Vidakovic, B. (2007). *Nonparametric statistics with applications to science and engineering*. New York: John Wiley & Sons.

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. *Behavioral and Brain Sciences, 31*, 389–390.

Linhares, A. (2011). The emergence of choice: 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., & Brum, P. (2009). How can experts see the invisible? Reply to Bilalic and Gobet. *Cognitive Science, 33*, 748–751.

Linhares, A., Chada, D. M., & Aranha, C. N. (in press). The emergence of Miller's magic number on a sparse distributed memory. *PLoS ONE*, 5, e15592. doi:10.1371/journal.pone.0015592.

Linhares, A., & Freitas, A. E. T. A. (2010). Questioning Chase and Simon's (1973) "perception in chess": The "experience recognition" hypothesis. *New Ideas in Psychology, 28*(1), 64–78.

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.

Moscato, P., Mendes, A., & Berretta, R. (2007). Benchmarking a memetic algorithm for ordering microarray data. *Biosystems, 88*(1–2), 56–75.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*(1), 81–106.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo: Morgan Kaufmann.

Saariluoma, P., & Kalakoski, V. (1998). Apperception and imagery in blindfold chess. *Memory, 6*(1), 67–90.

Sun, R., & Zhang, X. (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research, 5*, 63–89.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.

Zarri, L. (2009). On social utility payoffs in games: A methodological comparison between behavioural and rational game theory. *Theory and Decision*. doi:10.1007/s11238-009-9146-2 (pre-published online).

Please cite this article in press as: Linhares, A., et al. Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research* (2011), doi:10.1016/j.cogsys.2010.12.006