Eur. Phys. J. B (2011)
DOI: 10.1140/epjb/e2011-10905-8

THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

The aggregate complexity of decisions in the game of Go

M.S. Harré[1,a], T. Bossomaier[2], A. Gillett[1], and A. Snyder[1]

[1] The Centre for the Mind, The University of Sydney, 2006 Sydney, Australia
[2] Centre for Research in Complex Systems, Charles Sturt University, Australia

Received 21 November 2010 / Received in final form 2 March 2011
Published online 24 March 2011 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2011

Abstract. Artificial intelligence (AI) research is fast approaching, or perhaps has already reached, a bottleneck whereby further advancement towards practical human-like reasoning in complex tasks needs further quantified input from large studies of human decision-making. Previous studies in psychology, for example, often rely on relatively small cohorts and very specific tasks. These studies have strongly influenced some of the core notions in AI research such as the reinforcement learning and the exploration versus exploitation paradigms. With the goal of contributing to this direction in AI developments we present our findings on the evolution towards world-class decision-making across large cohorts of subjects in the formidable game of Go. Some of these findings directly support previous work on how experts develop their skills but we also report on several previously unknown aspects of the development of expertise that suggests new avenues for AI research to explore. In particular, at the level of play that has so far eluded current AI systems for Go, we are able to quantify the lack of 'predictability' of experts and how this changes with their level of skill.

1 Introduction

This work uses very large databases of professional and amateur players of the game of Go in order to understand the properties of the choices made by populations of players of a known rank. We take a large and tactically well studied area of the Go board and empirically derive a complete game tree of every choice made by players according to their rank. Sorting our results according to this rank, from lowest to highest amateurs and then lowest to highest professionals, provides a very fine grained data-set of changes in behavioural patterns across large populations as they acquire exceptionally high levels of skill in one of the most formidable popular games played today.

The underlying principle of this work is to move the analysis of complex decision tasks away from the detailed local analysis of strongly interacting elements and further towards the domain of weakly interacting contextual elements of a situation. Previous work on Go has successfully shown the utility of seeing the board in terms of the individual pieces (called stones) that have strong local interactions [1]. This technique was used to estimate territory and as a possible foundation on which a decision model could be based. A similar approach views the Go board as a 'conditional random field' [1,2], a technique that is able to relax the strong independence assumptions of hidden markov models and stochastic grammars [3]. Other directions have considered local patterns of stones for decision making [4,5], the representation of the board as a graph [6] and the formal analysis of endgame positions in terms of independent subgames [7]. This work is intended to inform the next generation of AI systems in regard to the complexity of decisions within the context of learning better play through an understanding of the changing contextual dependency of decisions. This perspective implies that whilst we study populations of players and their choices what we have in mind is a single AI that is able to make choices that are consistent with players of a certain skill.

The paper is laid out in the following manner. We first introduce the game of Go along with its basic principles and how we constructed the game trees of decisions from databases. Then the necessary tools of information theory are introduced and described. The game trees are then analysed in terms of information theory and our principal findings are presented. Finally we discuss the consequences of these findings in the context of other research.

2 The game of Go

The game of Go is more than 2 000 years old and still holds a significant cultural place in many asian countries. Despite a vast array of information and analysis that is available to players on almost every aspect of Go strategy, possibly even surpassing that of Chess, the rules of the game are deceptively simple. There are two players each of which plays with either black or white stones on a board

[a] e-mail: mike@centreforthemind.com

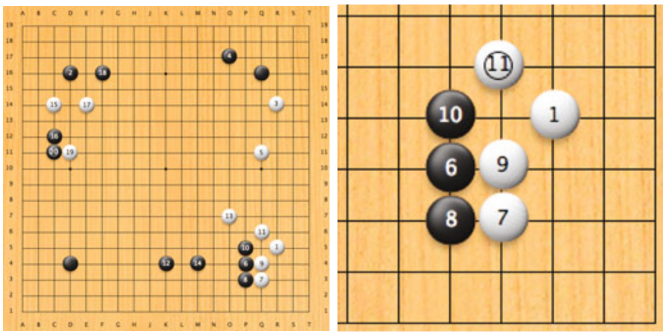

Fig. 1. (Color online) Left: the first 20 moves in a game of Go. Right: Stones played in a 7×7 region in the lower right corner, the numbers record the order in which they were played (moves 2 to 5 were played elsewhere on the board).

laid out with a 19×19 grid. Each player takes it in turn placing one of their stones on one of the vacant intersection points of the grid lines, see Figure 1 for an example game. Once a stone has been placed on the board it cannot be moved except if it is 'captured' by the other player, in which case it is removed from the board. The idea is to surround as large an area as possible with your stones using contiguous stones[1]. A territory is said to belong to a player once the region is encompassed by contiguous stones. There are different sets of scoring rules that result in the same outcome (who won or lost) but might differ in the number of points scored.

The ranks of Go players start with 30 kyu and increase to the first significant rating transition where they go from 1 kyu to 1 Dan. Rank then increases from 1 Dan through to 8 Dan. These ranks, from 30 kyu through to 8 Dan, are all amateur ranks. There are also professional ranks that go from 1 Dan through to 9 Dan. Typically a dedicated player of Go might be able to achieve the rank of 1 Dan amateur after a few years of diligent play. The very top amateur ranks might take decades to achieve and many people never attain them. The professional ranks are awarded after promising young players, typically younger than 18 years old, are conferred their professional status by a professional Go association. There are very few professionals in the world, in the order of hundreds at any one time and their dedication to the game is comparable to world class athletes.

A considerable body of work on the psychology of games such as chess has been built up over the years, see [8,9] for overviews. These studies have covered skill and expertise in general [10–12], memory [13,14], perception [15], search [16], learning [17] and visual cues [18,19]. In contrast very few studies have been carried out for Go but recent work has included the comparison of fMRI imaging for Go and chess [20,21]. Another study by Reitman [22] examined the psychological phenomenon of 'chunking' [23] in Go showing that there are very specific but overlapping aggregate spatial structures in Go. Similarly, Zobrist [24], Epstein et al. [25] and Bouzy [26] have all argued that spatial structures in Go are important to the reasoning used by Go players and have implemented perceptual feature modelling in artificial intelligence systems for Go. In terms of developmental psychology, Ping [27] recently demonstrated the positive effects of Go playing on the development of attention in children. In a more extensive study of Go players ranging from 18 through to 78 years of age, Masunaga and Horn [28] have shown that age related cognitive decline can be mitigated through continuing expertise development in Go players.

By far the most extensive body of academic work for the game of Go comes from the AI community, for an overview of earlier work see [29]. Since the computer program "Deep Blue" defeated Kasporov in 1997 [30] much of the interest in chess as a leading edge problem in AI research has dissipated. In its place Go has emerged as a key testbed for research in this area [31]. In particular, considerable advances have been made using novel algorithms on very fast computers, such as UCT monte-carlo implemented by the program MoGo [32,33]. This technique has been extended to include pattern based heuristics to better refine the search space [34,35] and to spatially organise strategies in Go [24,26] and more general game environments [25]. Most of these results are based on the intuition that spatial structure and patterns are relevant to the way in which players search for good moves, a notion supported within the psychological literature (see for example [36]). An aspect of this spatial organisation which is missing in the AI literature is how local organisation is impacted by the global context of the rest of the board. This work contributes in part to understanding this aspect of the development of strong play in Go.

In order to generate sufficient examples of moves, even from large databases of games, the search space necessarily needs to be reduced from the whole board down to some subsection of the board. In this work we focus on the 7×7 corner regions of the board where well studied patterns of moves, called *Joseki*, are played[2]. Studying the move trees in this area provides an insight into how these well understood sequences of moves change with skill (all of the players in our data-set would be expected to know something of Josekis). As a consequence of the small board region and the use of information theory as our analysis tool some interesting conclusions can be drawn regarding the influence and information content of the rest of the board (Sect. 6.2 discusses these aspects).

3 Game trees for Go

The purpose of an empirical decision tree is to represent a significant proportion of the moves made by human players during realistic game play. In order to build

[1] Stones are contiguous if they are on adjacent vertices, i.e. the vertices immediately to north, west, south or east from a given stone, stones that are diagonally adjacent are not contiguous.

[2] Extensive literature, references and discussions can be found at `http://senseis.xmp.net/`

such decision trees we collected game records for approximately 160 000 amateur[3] and professional[4] players such that the two players in each game had the same rank. Then a 7×7 corner section of the board was selected as there are many well known and well studied moves made in this region. Once symmetries had been accounted for, this region makes up more than 4/9 of the total board area and constitutes a significant amount of the spatial structure of the game, particularly in the beginning and middle of the game.

Within this region all first moves made were recorded, there were an average of 15 different first moves made across all ranks (max. = 25, min. = 8). The frequency of each move was then used to construct a probability distribution over the move choices. For each first move made, all subsequent second moves played within the region that were of the alternative colour were recorded and their frequency of occurrence was used to construct a probability distribution over all observed second moves. These probabilities were normalised for every first move, so for 20 first moves there are 20 normalised distributions for the subsequent second moves. This process was continued for the first six moves played within the 7×7 region across all ranks.

There is a subtlety in that, once multiple stones have been played, the order in which these stones were played is irrelevant, what is important is in which position they appear on the board. A case where the order is important is for searching ahead, where the order of planned moves might be strategically relevant. In this work we are interested only in what moves are made given a certain set of stones on the board, not the sequence of moves by which the patterns were arrived at. So there are potentially multiple paths to get to the same stone configurations, these different paths were accounted for in our statistics. Within our data we specifically aggregated the statistics for sequences of paths that resulted in the same pattern of stones on the board, resulting in 'path independent' stone configurations.

4 Information theory and learning

Information theory is used as it provides a unified framework within which learning, complexity, uncertainty and information storage all have well defined and consistent meanings as measured in bits[5]. First we introduce Shannon's entropy measure for a given probability distribution. For a random variable x that may result in n possible outcomes $x_i \in \{x_1, \dots x_n\}$ a probability distribution over outcomes is $p(x = x_i) \equiv p(x_i)$ and the amount of information associated with the distribution (in bits) is given by the entropy [37]:

$$H(p) = -\sum_i p(x_i) \log_2(p(x_i)). \tag{1}$$

Hereafter we drop the subscript 2 on the logs. The entropy is the expected value of $-\log(p(x_i))$, the amount of information associated with the single outcome x_i. $H(p)$ can be thought of in two complementary ways: as the amount of uncertainty an observer has in the outcome of a random event before the event occurs or alternatively as the average amount of information an observer gains by having witnessed the outcome. For a probability distribution p over n (finite) discrete elements the entropy is bounded above and below: $0 \leq H(p) \leq H(p(x_u))$ where $p(x_u)$ is the uniform distribution. $H(p) = 0$ tells us that one and only one outcome ever occurs so there is zero uncertainty in the outcome before it occurs and as the outcome is a certainty no information is gained by having observed that outcome. The entropy of the uniform distribution, $H(p(x_u))$, is the distribution whereby each element is not statistically differentiable from any other element. In this case the uncertainty in what will occur next is a maximum and therefore the information gained having observed an outcome is a maximum as well.

The entropy is also the minimum amount of information required to loss-lessly compress a probabilistic set of elements [37,38]. In particular, when storing or transmitting data, all regularities in the data (i.e. all lack of uniformity) can be accounted for *without loss* by using the number of bits given by the entropy. This led to an alternative interpretation of entropy by Grünwald in the context of learning: "[I]f one has learned something of interest, one has implicitly compressed the data." ([39], p. 595). 'Interest' here means 'the regularities in the data'. That is to say, the lower the entropy the more regularities that have implicitly been extracted from the data.

There is also a strong connection between the entropy and the Kolmogorov complexity of a distribution [38] and entropy can often be taken as a proxy for the Kolmogorov complexity [40][6]. We consider two separate and distinct distributions p_1 and p_2 over two different sets of elements of the same size, i.e. the distributions and their supports have nothing in common except the number elements. If the entropies have the relationship $H(p_1) < H(p_2)$ then distribution p_2 might be considered more 'complex' as this distribution has extracted less patterned information from its underlying set of elements than p_1 has from its underlying set of elements. While this provides a certain useful characterisation of complexity it has been noted that such measures only highlight the difference between deterministic and random distributions [41]. However for any stochastic decision process (as opposed to trying to find an objective measure of a system's structure, for example), it is readily seen that it is more difficult (and thereby arguably more 'complex') to choose from a perfectly uniform distribution of options (as all choices are statistically

[3] Amateur game records were collected from the Internet Go Server: www.pandanet.co.jp

[4] Professional game records are from the GoGoD collection: www.gogod.co.uk

[5] We base all of our calculations on log base 2, so measurements are always bits.

[6] This is useful as the Kolmogorov complexity is not a computable function.

indistinguishable from each other) than it is for a distribution where one option occurs 100% of the time (where one choice is perfectly distinguishable from all others). For an overview of the use of complexity measures and their relationship to deterministic systems, see [42].

Mutual information is an extension of entropy whereby the goal is to measure the *joint information* shared by two distributions. In this work we measure how much information is gained about the next choice of move given the current (local) state of the board. The first i moves already played on the board, of which there are k unique variations, are denoted by the set $\{x_1, \ldots, x_i\}^j$ for $i \in \{1, \ldots, 5\}$ and the index term $j \in \{1, \ldots, k\}$. The (marginal) probability that the jth unique pattern of i stones on the board occurs is $p(\{x_1, \ldots, x_i\}^j)$. The (marginal) probability that the lth unique move x^l_{i+1} ever occurs at move $i+1$ is $p(x^l_{i+1})$ and the (joint) probability that the sequence $\{x_1, \ldots, x_i\}^j$ is followed by x^l_{i+1} is $p(\{x_1, \ldots, x_i\}^k, x^l_{i+1})$. x^l_{i+1} represents *all* moves observed for move $i+1$, irrespective of the stones already played. In this sense $H(p(x_i))$ is the unconditional entropy of the probability of all variations of the ith move. The mutual information between the stones already on the board and the next stone placed on the board is [43]:

$$I(\{x_1, \ldots, x_i\}; x_{i+1}) = \sum_{k,l} p(\{x_1, \ldots, x_i\}^k, x^l_{i+1}) \log \left(\frac{p(\{x_1, \ldots, x_i\}^k, x^l_{i+1})}{p(\{x_1, \ldots, x_i\}^k) p(x^l_{i+1})} \right). \quad (2)$$

For example, in the case of how well move 1 predicts move 2, this equation reduces to the much simpler form:

$$I(x_1; x_2) = \sum_{k,l} p(x^k_1, x^l_2) \log \left(\frac{p(x^k_1, x^l_2)}{p(x^k_1) p(x^l_2)} \right). \quad (3)$$

This explicitly calculates how predictable move two is based on move one: if $I(x_1; x_2) = 0$ then move two is independent of move one and if $I(x_1; x_2) = H(p(x_1))$ then move two is entirely decided by the choice of move one. Generally the entropy measures a distribution's uniformity whereas mutual information measures the relative dependency of two distributions. In the case of independent distributions $p(x^k_1, x^l_2) = p(x^k_1)p(x^l_2)$ we have $I(x_1; x_2) = 0$. On the other hand, equation (3) can be rewritten as $I(x_1; x_2) = H(p(x_1)) + H(p(x_2)) - H(p(x_1), p(x_2))$. The $H(p(x_i))$ terms are simply entropies as in equation (1). The

$$H(p(x_1), p(x_2)) = -\sum_{i,j} p(x^i_1, x^j_2) \log(p(x^i_1, x^j_2))$$

term is the joint entropy. For any random variables x and y, $H(p(x), p(y)) \geq H(p(x))$ where equality holds if and only if the outcome y is a deterministic function of x.

5 Results

In this section we present the principal findings of our work. First the entropy and then the mutual information results are outlined. In the last section we discuss these results and place them in the context of previous work.

5.1 Entropy

Figure 2 plots the cumulative average entropies for the first six moves by player ranks with the best fit linear trend added for the amateurs and professionals. The linear trends can be plotted as a function of a 'continuous' rank index[7] ρ with equations (a for amateur, p for professional): $H_a(\rho) = -0.1806\rho + 7.2322$ and $H_p(\rho) = 0.0070\rho + 5.7222$. There is a distinct downward linear trend for the amateur players as their rank increases. This is not the case for the professionals though; the average cumulative move entropy across all player ranks is almost flat. Note that reliable statistics were not achievable for some of the moves in our data, specifically the senior amateurs and junior professionals, due to scarce data. So in the case of cumulative plots (Figs. 2 and 3) these data points were omitted as they would not make sense. On the other hand it is possible to include some of the data points for some moves in the other non-cumulative plots (Figs. 4 and 5 are not cumulative) and doing so enables a better estimation of the inflection points.

In order to understand the components of these trends better, Figure 3 plots the individual entropies for each move. Note that the first three moves have distinguishably higher average entropies than the second three for both the amateurs and the professionals. The best linear approximations have been included. For the amateurs, the gradient of the linear trend for first three moves is -0.0133 ($r^2 = 0.441$) and for the second three moves is -0.0469 ($r^2 = 0.890$). For the professionals, the gradient of the linear trend for the first three moves is -0.0097 ($r^2 = 0.204$) and for the second three is 0.012 ($r^2 = 0.213$). Some ranks were excluded as accurate move entropies could not be calculated due to a paucity of data, and so cumulative entropies would make no sense, see [44] for the source of our error analysis.

5.2 Mutual information

Using equation (2), the mutual information between the distribution of each unique sequence of moves and the distribution of all possible next moves were calculated. The results are plotted in Figure 4. Using the variable ρ for rank as used for the entropies, the best fit quadratic interpolation of the mutual information between successive moves are:

$$I(x_1; x_2, \rho) = -0.0026\rho^2 + 0.0587\rho + 0.7445$$
$$I(\{x_1, x_2\}; x_3, \rho) = -0.0029\rho^2 + 0.0657\rho + 1.0317$$
$$I(\{x_1, \ldots, x_3\}; x_4, \rho) = -0.0037\rho^2 + 0.0765\rho + 2.0085$$
$$I(\{x_1, \ldots, x_4\}; x_5, \rho) = -0.0039\rho^2 + 0.0792\rho + 2.3886$$
$$I(\{x_1, \ldots, x_5\}; x_6, \rho) = -0.0037\rho^2 + 0.0794\rho + 2.6319$$

[7] For this purpose we set the following integer values for ρ (rank): am2q = 1, am1q = 2, ... , pr9d = 19.

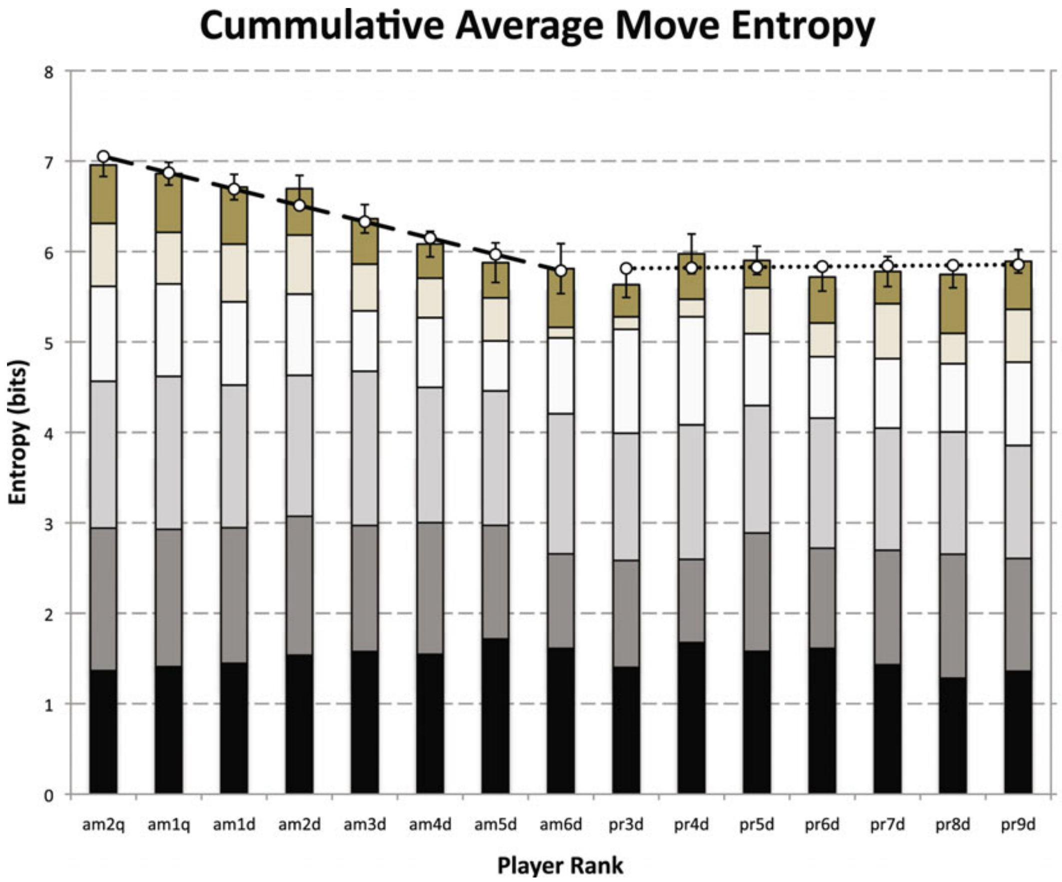

Fig. 2. (Color online) A plot of the cumulative entropies of the decision tree. The linear trend for the amateur is negative ($r^2 = 0.95911$), however the professional linear trend has a low r^2 value (=0.01588, reflecting the near zero gradient). Error bars are $\pm 2\sigma$ and both linear approximations lie within $\pm 2\sigma$ of the observed total entropies for all ranks.

and the respective residual errors (r^2 terms) for these equations are, from top to bottom: 0.65791, 0.73576, 0.68054, 0.81934, 0.7142. Here we have included all ranks of players. In this case the mutual information curves are not cumulative as in the entropies, but we are also interested in where the inflection points are in the quadratics used to fit the data.

Using these curves as continuous approximations to the discrete data in the variable ρ, we want to find the inflection point where the mutual information peaks as ρ increases. To do so we differentiate the five quadratics listed above with respect to ρ and solve for where this differential is zero for each of the curves. The mean inflection point in the mutual information is at $\rho = 10.77$ (min. $= 10.15$, max. $= 11.34$), i.e. slightly more than half-way between 8 Dan amateur ($\rho = 10$) and 1 Dan professional ($\rho = 11$). This suggests a fundamental change in the nature of skill development as players turn professional.

From Figure 4 it is not possible to see to what extent the plotted mutual information reflects the theoretical maximum possible values. Recall that mutual information is strictly bounded: $0 \leq I(\{x_1 \dots x_i\}; x_{i+1}) \leq H(p(\{x_1 \dots x_i\}))$, and it is of interest to what extent Figure 4 reflects this range. In order to do so, Figure 5 plots $\frac{I(\{x_1 \dots x_i\}; x_{i+1})}{H(p(\{x_1 \dots x_i\}))} \times 100\% =$ percentage of the theoretical maximum predictability for Figure 4.

6 Discussion

Developers of future AI systems, if their systems are to emulate the cognitive abilities of humans in complex tasks, need to be as well informed as possible regarding the nature and evolution of human behaviour as they acquire the skill AI systems hope to emulate. This study has sought to quantify some of the behavioural attributes of complex decision tasks in terms of information theory, an approach which enables measurements of both human and artificial systems such that direct comparisons might be made. Currently 'brute force' techniques, such as monte-carlo algorithms and their recent variations [34,35,45], have been able to play Go to a strong amateur level. It is a little below this strong amateur level at which this study begins and includes the very best players in the world, highlighting some of the subtleties with which AI algorithms need to contend in order to perform beyond their current levels. This section discusses the important conclusions that may be drawn from this work and places them within the context of previous results.

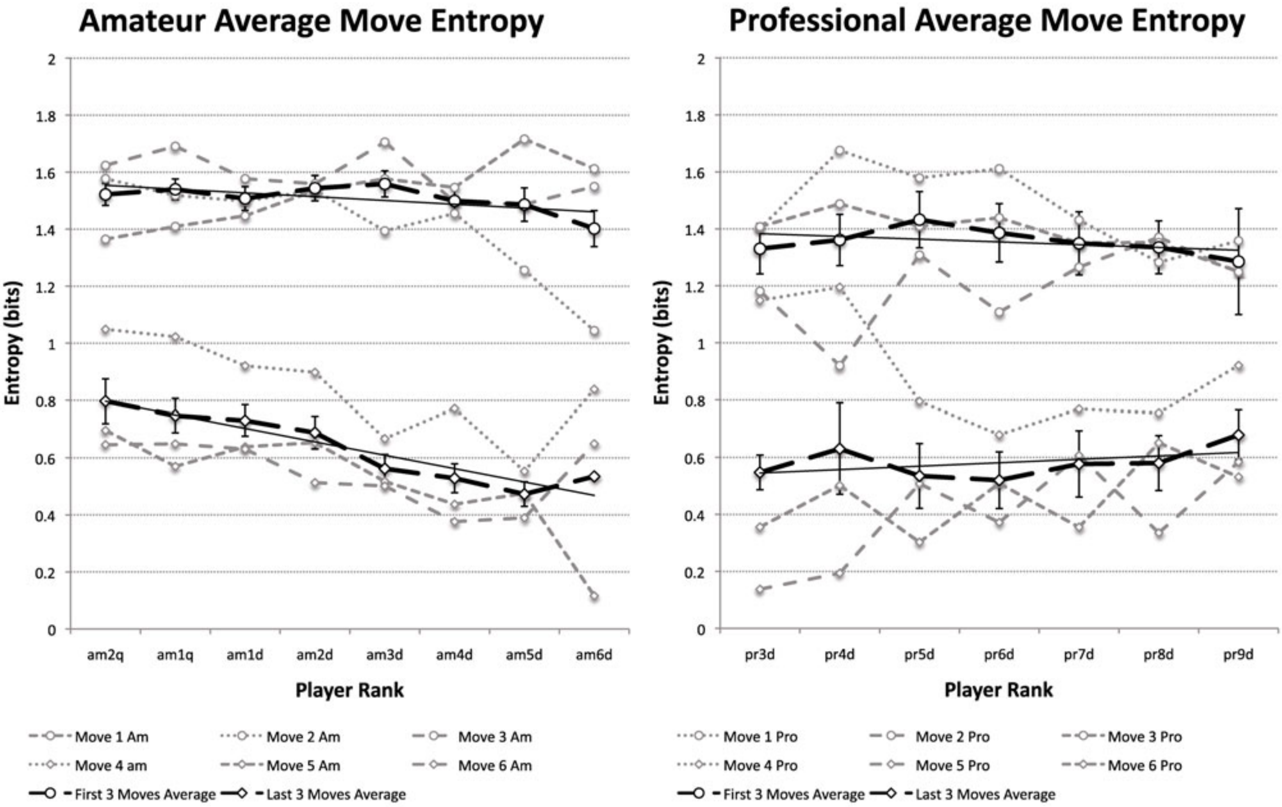

Fig. 3. The component parts of the entropies of Figure 2 and the linear trends for the first three moves and last three moves. These curves, like Figure 2, are averages across all branches at a given move number. For example there is only one branch for move one, but there is an average of 15 branches for each rank at move two that have been averaged in order to get the plotted values.

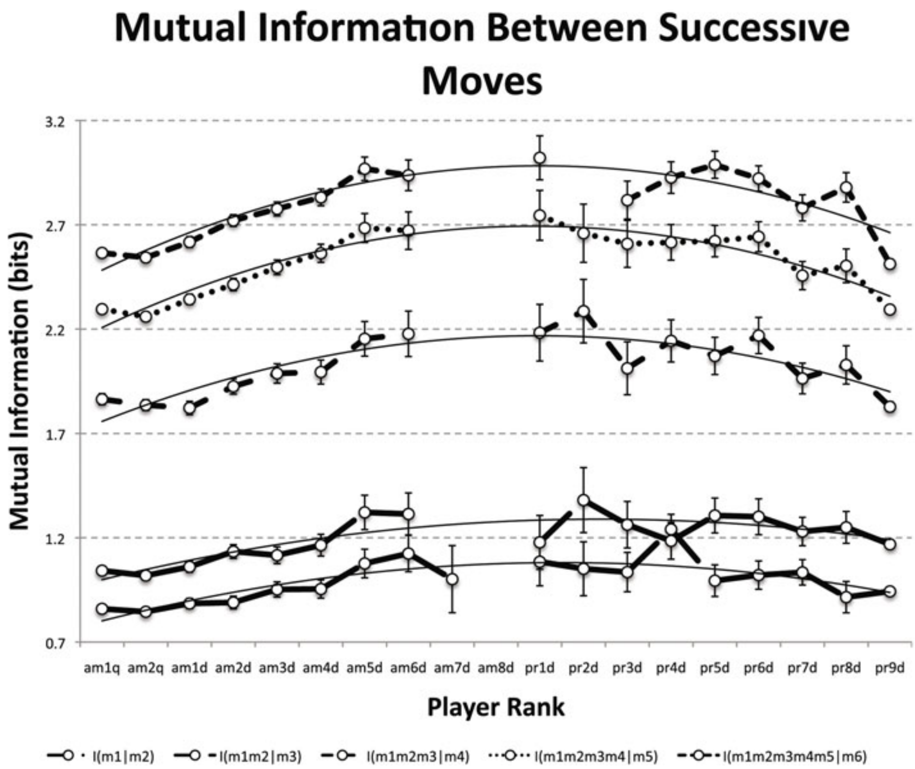

Fig. 4. Mutual information between successive moves.

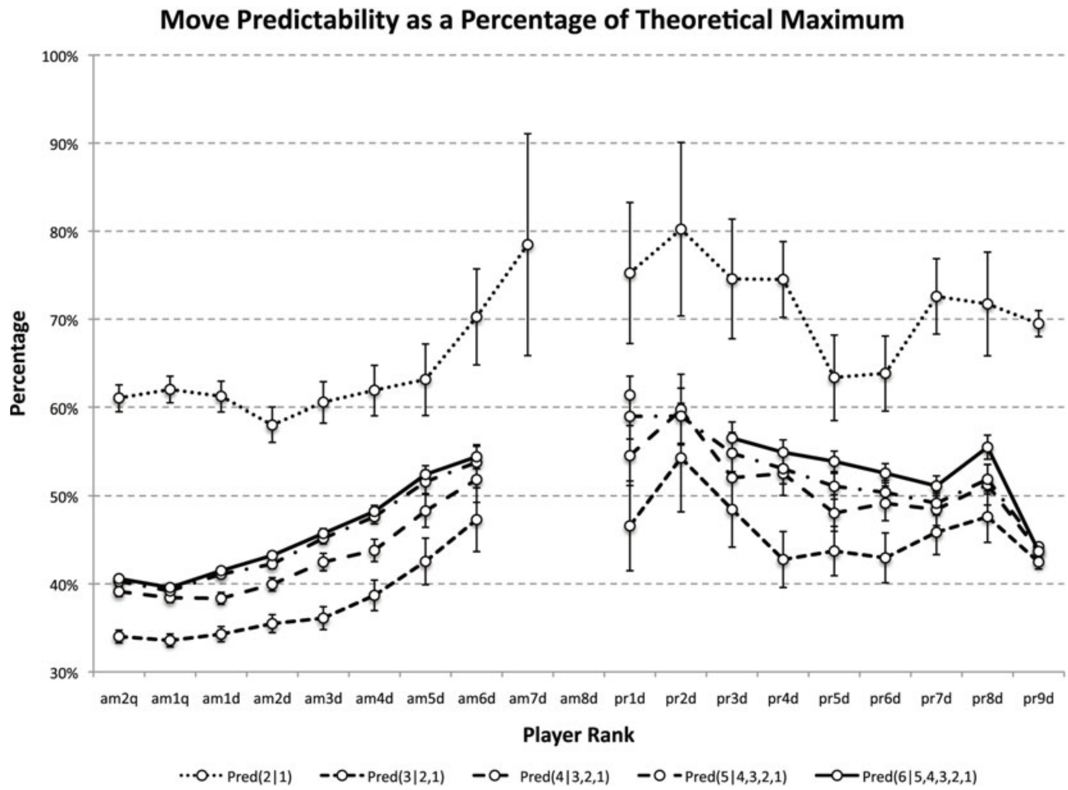

Fig. 5. The 'predictability' of each move as a percentage of the theoretical maximum of mutual information, e.g. $pred(3|2,1)$ is the average predictability of move 3 across all possible variations of moves 1 and 2. From bottom to top, the order of the curves is: $pred(3,|2,1)$, $pred(4|3,2,1)$, $pred(5|4,3,2,1)$, $pred(6|5,4,3,2,1)$, $pred(2|1)$.

6.1 Learning in a complex task environment

A prominent theoretical framework used for learning the best choice in a given task is reinforcement learning [46]. In this approach a record of past rewards for given actions is kept and used to inform future decisions. This approach has not only been successful within the the AI community, but it has recently shown considerable explanatory power in neuroscience [47,48]. Within this paradigm the notion of exploration versus exploitation is used extensively: exploration is achieved by choosing more randomly between the available options and exploitation is achieved by choosing based on historically more rewarding options.

A modification of this approach uses a meta-parameter in order to control the degree to which of these two strategies are favoured: exploration of the alternatives or exploiting the historically better options [49,50]. In such models where a meta-parameter is a controllable variable, typically once an exploration phase has informed a reinforcement learning algorithm of the better options, then an exploitation phase is adopted where knowledge gained in the exploration of the space is used to make better decisions, consequently improving performance. Alternatively it has been suggested that learning is the process by which regularities are extracted from data, and consequently learning is data compression [39]. From this point of view entropy is a measure of how much has been learned by the players.

In these terms, we see in Figure 2 a decreasing entropy for amateur players as rank increases. This strongly suggests that the uncertainty in move selection is decreasing because players are taking advantage of the better choices of moves as they learned through past experiences of the game. Alternatively, the players are able to 'compress' the data by extracting more information from past data and have thereby learned more from the data. This tendency is more prominent in the last three moves than the first three, Figure 3, possibly because the first three moves are learned more quickly than the next three moves. This effect has mostly vanished in the cumulative entropies by the time the players become professionals, although curiously this seems to have been caused by a balancing of a slight increase in entropy for the first three moves and slight decrease in entropy for the next three moves.

It might be thought that the decrease in entropy between successive moves, particularly moves 3 and 4, is due to the decrease in options available to the players as the local region of the board fills up with stones. This effect would only be slight when considering 6 stones out of a possible 49 positions, and the very sharp decrease between moves 3 and 4 cannot be explained this way as we should expect entropy to decrease smoothly as the local positions fill with stones. Also note that while the first 3 moves have different entropies from the last 3, there is no consistent order within each group of 3 moves as would be expected if the effect was due to decreasing move options.

6.2 Context dependency of move choices

In order to quantify the relationship between prior moves made and the next move, we measured the mutual information between the probability of a certain sequence of stones being played and the next choice of move (Fig. 4). This measures how predictable the next move is based on the probability of a prior pattern of stones having been placed on the board.

We observed a considerable difference in the degree of predictability from one move to the next. For example the first move typically shares about 0.9 to 1.1 bits of information with the second move, equating to between 60% and 80% of the theoretical maximum possible (Fig. 5). This tells us that there is very little other information, such as the rest of the board, being used to decide what the next move will be, almost all of the uncertainty in the second move is explained by the uncertainty in the first move. This is not consistent across each successive move though. The choice of move three is considerably less predictable based on the uncertainty in the first two moves. The total predictability increases because two stones are more informative than one[8]. However the increase in shared information provided by this second stone is typically only about 0.1 to 0.3 bits (the difference between $I(x_1; x_2)$ and $I(\{x_1, x_2\}; x_3)$). This is shown in Figure 5 as ranging between approximately 35% and 55% of the theoretical maximum (the bottom most curve in this plot), a significant drop compared to how informative the first move is of the second move.

We suggest that these significant and consistent differences between the predictability of one move given the uncertainty in the previous moves is due to the different focus the players have on the local versus the global context of board strategy. In choosing the second move, most of the information used by the players is based on the stone already placed. In choosing the third stone to play, most of the information used by the players is based on the rest of the board, i.e. the non-local aspect of the decision. In this sense the rest of the board acts to induce a degree of uncertainty in which move to make next. This changing uncertainty, and hence changing emphasis on local versus global context, is expressed in how dependent the next move is on the stones that are played locally.

6.3 Expertise acquisition and the ‘Inverted-U’ effect

We make a final comment regarding the striking and consistent nature of the “inverted-U” observed as a function of player rank for the mutual information measures, Figure 4. The inverted-U property of novice-skilled-expert comparisons have previously been reported in the literature [51–54]. Here we have been able to show that, across a significant portion of the range of player skills, the inverted-U property holds for the predictability of the next move in Go. In order to explain this effect we break it into two components, the increasing branch (prior to players turning professional) and the decreasing branch (after players turn professional).

Prior to turning professional, Figure 2 shows a decreasing entropy and Figure 4 shows an increase in predictability. This is consistent with the notion of reinforcement learning with meta-parameters discussed earlier, the entropy decreases due to choosing strategies that have proven to have good outcomes in the past, and which choice will be made is informed by the local pattern of stones thereby increasing the mutual information.

After turning professional this no longer holds, the players have likely minimised the entropy for these moves as far as they possibly can within the local region. However now the predictability of the next move starts to decrease for a fixed level of entropy. This suggests that professional players are now learning to obfuscate their moves by playing less predictably within the local region and using more information from the rest of the board i.e. they are increasing their global strategic awareness in order to play more subtle local variations. Importantly this is achieved without significant variation in the marginal entropies, so there is no change in the exploitation-exploration balance or equivalently there is no variation in the amount of information the players have extracted from the data, the change in behaviour is of a qualitatively different nature to that which has been observed previously for expertise in complex task environments.

6.4 Conclusion

This work has aimed at studying the complexity of the decisions made by many thousands of Go players as their skill progresses towards that of the best players in the world. Using large databases of game records we have been able to show that there are significant behavioural patterns, some of which support previous work in both psychological and artificial intelligence research and others which are entirely new to both fields. Most significantly we have been able to show that the strategic development of player’s behaviour at the very top level of play, the level that has not yet been conquered by brute force search techniques, is not as simple as a trade-off between exploitation-exploration or the compressibility of relevant information in the underlying decision task.

This work was supported by ARC grant number DP0881829 and USAF grant number 094094.

References

1. D. Stern, T. Graepel, D. MacKay, Advances in neural information processing **16**, 33 (2004)

[8] This is not *always* the case, for example in Figure 4 for pr4d the first data point sits higher than the second data point. This is due to there being multiple paths to the same local pattern of stones.

2. S. Sanner, T. Graepel, R. Herbrich, T. Minka, Learning CRFs with Hierarchical Features: An Application to Go, in *International Conference on Machine Learning (ICML) Workshop* (2007)
3. J. Lafferty, A. McCallum, F. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Machine Learning – International Workshop then Conference* (2001), pp. 282–289
4. T. Cazenave, Advances in Computer Games **9**, 275 (2001)
5. D. Stern, R. Herbrich, T. Graepel, Bayesian pattern ranking for move prediction in the game of Go, in *Proc. 23rd int. conf. on machine learning* (2006), Vol. 148, pp. 873–880
6. T. Graepel, M. Goutrie, M. Krüger, R. Herbrich, *Learning on graphs in the game of Go, Artificial Neural Networks – ICANN* (2001), pp. 347–352
7. E. Berlekamp, D. Wolfe, *Mathematical Go: Chilling Gets the Last Point*, edited by A.K. Peters (1997)
8. N. Charness, Psychol. Res. **54**, 4 (1992)
9. F. Gobet, A. de Voogt, J. Retschitzki, *Moves in mind: The psychology of board games* (Psychology Pr, 2004)
10. H.A. Simon, W.G. Chase, Am. Sci. **61**, 393 (1973)
11. D. Holding, *The psychology of chess skill* (L. Erlbaum Assoc., Hillsdale, NJ, 1985)
12. D. Holding, Psychol. Res. **54**, 10 (1992)
13. H. Simon, K. Gilmartin, Cogn. Psychol. **5**, 29 (1973)
14. F. Gobet, H. Simon, *Recall of rapidly presented random chess positions is a function of skill* (1996)
15. W.G. Chase, H.A. Simon, Cogn. Psychol. **4**, 55 (1973)
16. N. Charness, Journal of Experimental Psychology: Human Perception and Performance **7**, 467 (1981)
17. A. Cleveland, The American Journal of Psychology **18**, 269 (1907)
18. E. Reingold, N. Charness, M. Pomplun, D. Stampe, Psychol. Sci. **12**, 48 (2001)
19. A. Waters, F. Gobet, G. Leyden, British Journal of Psychology **93**, 557 (2002)
20. M. Atherton, J. Zhuang, W. Bart, X. Hu, S. He, Cognitive Brain Research **16**, 26 (2003)
21. X. Chen, D. Zhang, X. Zhang, Z. Li, X. Meng, S. He, X. Hu, Cognitive Brain Research **16**, 32 (2003)
22. J. Reitman, Cogn. Psychol. **8**, 336 (1976)
23. F. Gobet, P. Lane, S. Croker, P. Cheng, G. Jones, I. Oliver, J. Pine, Trends in Cognitive Sciences **5**, 236 (2001)
24. A. Zobrist, A model of visual organization for the game of Go, in *Proceedings of the May 14-16, 1969, spring joint computer conference* (ACM, 1969), pp. 103–112
25. S. Epstein, J. Gelfand, E. Lock, Constraints **3**, 239 (1998)
26. B. Bouzy, Spatial Reasoning in the game of Go, in *Workshop on Representations and Processes in Vision and Natural Language, ECAI* (Citeseer, 1996), pp. 78–80
27. X. Ping, K. Keqing, Psychological Research, 03 (2009)
28. H. Masunaga, J. Horn, Learning and individual differences **12**, 5 (2000)
29. X. Cai, D. Wunsch, Computer Go: A grand challenge to AI, in *Challenges for Computational Iintelligence* (Springer Berlin, 2007), pp. 443–465
30. M.S. Campbell, A.J. Hoane, F. Hsu, *Search control methods in Deep Blue*, *AAAI Spring Symposium on Search Techniques for Problem Solving under Uncertainty and Incomplete Information*, pp. 19–23 (1999)
31. J. Burmeister, J. Wiles, The challenge of Go as a domain for AI research: A comparison between Go and chess, in *Proceedings of the Third Australian and New Zealand Conference on Intelligent Information Systems* (1995)
32. S. Gelly, D. Silver, Achieving master level play in 9×9 computer go, in *Proceedings of AAAI* (2008), pp. 1537–1540
33. C. Lee, M. Wang, G. Chaslot, J. Hoock, A. Rimmel, O. Teytaud, S. Tsai, S. Hsu, T. Hong, Computational Intelligence and AI in Games, IEEE Transactions on **1**, 73 (2009)
34. T. Cazenave, B. Helmstetter, Combining Tactical Search and Monte-Carlo in the Game of Go, in *Proceedings of the IEEE Symposium on Computational Intelligence in Games* (2005), pp. 171–175
35. S. Gelly, Y. Wang, R. Munos, O. Teytaud, *Modification of UCT with Patterns in Monte-Carlo Go* (2006)
36. F. Gobet, H. Simon, Psychol. Res. **61**, 204 (1998)
37. C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948)
38. T. Cover, J. Thomas, *Elements of Information Theory*, 2nd edn. (Wiley, 2006)
39. P. Grünwald, *The Minimum Description Length Principle* (The MIT Press, 2007)
40. S. Leung-Yan-Cheong, T. Cover, IEEE Trans. Inf. Theory **24**, 331 (2002)
41. J. Crutchfield, K. Young, Phys. Rev. Lett. **63**, 105 (1989)
42. G. Boffetta, M. Cencini, M. Falcioni, A. Vulpiani, Phys. Rep. **356**, 367 (2002)
43. D. Mackay, *Information theory, inference, and learning algorithms* (Cambridge University Press, New York, 2003)
44. M. Roulston, Physica D: Nonlinear Phenomena **125**, 285 (1999)
45. Y. Wang, S. Gelly, Modifications of UCT and sequence-like simulations for Monte-Carlo Go, in *Proceedings of the IEEE Symposium on Computational Intelligence and Games* (2007), pp. 171–182
46. R. Sutton, A. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, 1998)
47. M. Frank, A. Moustafa, H. Haughey, T. Curran, K. Hutchison, Proceedings of the National Academy of Sciences **104**, 16311 (2007)
48. M. Cohen, C. Ranganath, J. Neurosci. **27**, 371 (2007)
49. S. Ishii, W. Yoshida, J. Yoshimoto, Neural Networks **15**, 665 (2002)
50. N. Schweighofer, K. Doya, Neural Networks **16**, 5 (2003)
51. F. Gobet, H. Simon, Mem. Cogn. **24**, 493 (1996)
52. M. Van De Wiel, H. Boshuizen, H. Schmidt, Eur. J. Cogn. Psychol. **12**, 323 (2000)
53. H. Boshuizen, R. Bromme, H. Gruber, *Professional learning: Gaps and transitions on the way from novice to expert* (Kluwer Academic Publishers, 2004)
54. I. Davies, P. Green, M. Rosemann, M. Indulska, S. Gallo, Data and Knowledge Engineering **58**, 358 (2006),