

Strategic planning through topological inference in the game of Go. A non-linear oscillator network model

Abstract. The ancient oriental game of Go is the last of the great perfect knowledge games wherein strong human players still outperform AI players, but the gap is narrowing. Superior strategic skill is keeping humans several steps ahead even as computers catch up tactically. Taking the perspective that large scale topological perception is part of the key to strategic excellence, we present a new method for generating strategic plans in the game of Go. Treating stones in play as continuous perturbations of a dynamical system composed of coupled non-linear oscillators, we exploit transitory synchrony in the wave patterns thus generated to make topological inferences about possible territories later in the game. Our system is shown, using professional game records, to focus search on appropriate areas of the space of continuations, and to generate similar results to some existing computer Go 'static mapping' standards as a special case.

Wave models offer a new approach to whole board integration which we can use to form multiple direct representations of relatively likely, winning territorial topographies, or 'strategic maps', given a board position. We anticipate that these strategic maps will be very useful to computer Go players and can help to inform us about how humans achieve such skilled topological perception in Go. By framing tactical search and pattern analysis with specific strategic hypotheses, large scale topological inference will help to direct search towards winning continuations.

1 Introduction

In the years since computer chess players have reached levels of performance competitive with the very best players in the world, the ancient oriental game of Go has increasingly been recognised as a new grand challenge to AI researchers. The large board and simple rules give rise to a game tree of tremendous depth and breadth. There are in the region of $361!$ different games of Go. The uniform nature of the pieces and board - each player simply places black or white stones on the intersections of a 19×19 matrix - forces heuristic evaluation to consider relational rather than intrinsic properties of game states.

This combination of problems has led research to explore avenues beyond the classic mini-max game tree search [11] paradigm that has been so successful in other perfect knowledge games. One interesting facet of Go is its strongly visual-spatial nature. Once played, stones in Go do not move around the board (although they may be captured and removed if surrounded), leaving something like a historical trace of the games progress as play goes on. Thanks to this visual continuity, Go has often been anecdotally compared to the incremental growth of an image, mutually 'painted' by the two competing players. Since the early days of computer Go, it has been suggested that the visual-spatial perceptual processes natural to the human nervous system play a strong role in enabling human expertise in Go. This

visual analogy led to the development of 'influence' [14] and 'territory' [2] estimating algorithms based on morphological operators (erosion and dilation) often used in image processing. More recently, significant progress has been made with the application of Bayesian methods to local pattern recognition [9].

In the last few years, a new generation of computer Go players based on Monte Carlo Tree Search ('MCTS') [6] together with the construction of 'Upper Confidence Trees' ('UCT' - a variant on Upper Confidence techniques developed from multi-armed bandit games [1] and adapted to tree search [7]) have achieved a great leap forward in performance. These systems, combined with some of the pattern matching techniques mentioned above, are now competitive with professional players in 9×9 Go, the smallest board size in regular use. On the full size (19×19) board, however, strong human players are still significantly ahead of the best computers. This suggests that scaling is a problem for MCTS-UCT approaches. The game tree grows exponentially with the size of the board, so this is not surprising. This new paradigm has dominated the field over the past few years, and together with increasing computational resources will undoubtedly continue to produce improvements in computer Go. The opportunity cost, however, will be a reduction in active work on strategic planning, which we consider to be the great remaining challenge for AI in Go. Comments from professional players suggest that the new generation of MCTS computer players are very strong locally and tactically, but can be outplayed strategically.

We introduce here a dynamical system for whole board strategic inference in Go. The next section briefly describes our system. This work extends existing visual-spatial approaches, and Section 2 details results showing that measurements taken from our system can produce similar outputs to the Zobrist [14] and Bouzy [2] algorithms. In Section 3 we show our system extends this work, by producing multiple inferences about potential topological continuations consistent with the position information available (according to the parameter set governing the system). Section 4 discusses the implications of our results and how future work may develop the findings reported here.

2 Outline of the model

Strategy in 19×19 Go is sufficiently complex for computer players to run into the 'frame problem' [13] to a significant extent - there are too many relationships present to all be attended to, but how do we know which ones to disregard until we have attended to them and identified them as unimportant? We consider this proliferation of relations to be one of the main reasons strategic Go is so difficult for computers. Wheeler [13], among others, has suggested that the way to avoid the 'frame problem' is with a dynamical system in which the causal contribution of each systemic component partially

determines, and is partially determined by, the causal contributions of large numbers of other systemic components. This is precisely our approach here. By combining spatial (rather than game-tree) representation with dynamical lookahead, we hope to overcome some of the limitations faced by static mapping and dynamical game-tree based approaches.

It is difficult to communicate the dynamical behaviour of the surface using static images. One may imagine a surface which expresses the influence of the white and black stones as peaks and troughs in constantly shifting patterns of waves which emanate from the energetic inputs of the stones in play. Areas of strong influence may be sufficiently constrained by the stones in the area to be constantly held up or down. The coupled oscillators provide a two dimensional medium through which the influence of black and white stones can propagate as waves. The interaction of these waves produces transitory patterns of peaks and troughs which are used to represent potential white vs black territories. The effect is not unlike the subjective experience of imagining how black and white groups might appear, grow and connect up when one is playing Go.

2.1 Computational architecture

The system is based on a two dimensional, locally coupled network of 19x19 Hopf oscillators [8], which correspond to the 19x19 intersections on the Go board. Oscillator networks have been applied to a range of other scene analysis tasks [12]. At present, each oscillator is coupled to its von Neumann neighbourhood, via a coupling constant ε . Other coupling schemes may well be interesting, but are not explored here. Stones on the board are coupled to the system as continuous energetic perturbations applied to the oscillators at empty points neighbouring the stone (in Go terms, the stones 'liberties'), via a coupling constant κ . This means that a stones energetic contribution, and the distribution of that contribution, depends on its liberties. The influence of the stones present constrains the trajectory of the dynamical surface. Edge and corner nodes are self-coupled to produce the effect of reflection at boundaries. There is no 'learning', in that the system does not undergo any long term changes in structure.

A single Hopf oscillator is a dynamical system of the form;

$$\dot{x} = \gamma(\mu - r^2)x - \omega y \quad (1)$$

$$\dot{y} = \gamma(\mu - r^2)y + \omega x \quad (2)$$

$$r^2 = x^2 + y^2 \quad (3)$$

Where μ controls the amplitude of the oscillations, and γ controls the speed of recovery after perturbation. A single Hopf oscillator has an unstable equilibrium at 0 and a stable harmonic limit cycle determined by the parameters μ (amplitude) and ω (frequency). The surface in our model consists of 19x19 such oscillators, each coupled to their von Neumann neighbourhood (the coupling term $F(n)$) by a constant ε , and to the perturbation matrix $F(p)$, which is determined by the Go stones on the board by a constant κ giving;

$$\dot{x}_{ij} = \gamma(\mu - r_{ij}^2)x_{ij} - \omega y_{ij} + \varepsilon F(n_{ij}) + \kappa F(p_{ij}) \quad (4)$$

$$\dot{y}_{ij} = \gamma(\mu - r_{ij}^2)y_{ij} + \omega x_{ij} \quad (5)$$

with

$$F(n_{ij}) = x_{i+1j} + x_{i-1j} + x_{ij+1} + x_{ij-1} \quad (6)$$

Constants are set to $\gamma = 8$, $\omega = 0.13$, $\mu = 1$, $\kappa = 0.1$, $\varepsilon = 0.08$. Somewhat arbitrarily, we set ω to 0.13 here. $F(p)$ is defined by the

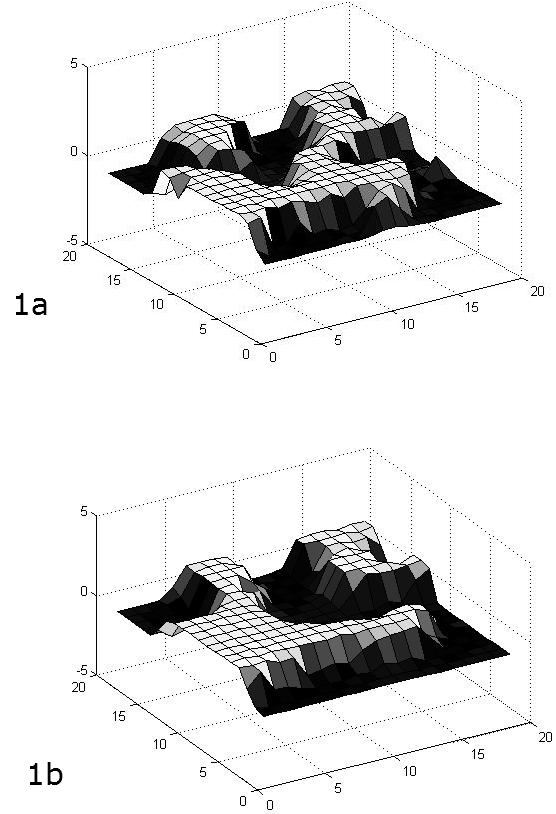


Figure 1. Snapshots of travelling waves on the oscillatory surface

stones on the board. Representing the board as a matrix, all white stones contribute +1, and all black stones -1, to each of their empty 4-neighbours ('liberties'). Slightly higher or lower ω gives shorter or longer lengths to the surface waves. Large changes in ω result in qualitatively different behaviour including unwanted phase locking and a tendency towards static equilibria. μ is set to 1. The absolute level of μ is not so important as its level relative to κ , as this determines the ability of stones played to constrain the intrinsic oscillatory dynamics of the surface. In Go terms, this translates to the question 'How much influence does a stone have?'. If μ is too high relative to κ , stones barely effect the surface at all. If κ is too high relative to μ , the perturbation from the stones overpowers the oscillatory tendency of the surface, resulting in a static state enforced by the stones. The relationship between κ and ε is also important. Here we use $\kappa = 0.1$ and $\varepsilon = 0.08$, meaning that stones have a slightly greater effect than surface levels. γ is kept constant here. Small changes in γ have very little effect. In the course of further work, these parameters will be a target for optimisation via professional game statistics and play against other computer systems. At this early stage of development we simply choose settings which produce interesting and desired behaviour by hand, accepting that these are imperfect.

See Figure 1 for snapshots of travelling waves on the surface given a board position from a real professional Go game. Transitory patterns of synchrony, caught in these snapshots as black and white areas, provide the basis for a series of topological hypotheses about territory distributions given different continuations from the current

position. We call these 'strategic hypotheses' or 'strategic maps'.

3 Comparison to existing visual-spatial models

3.1 Zobrist 'Influence' Maps

Zobrist, one of the first Go programmers, made extensive use of the visual analogy, particularly in the design of his 'influence function' [14]. Influence is a concept in Go describing the relative dominance of each colour in particular areas of the board. Figures 2a and b show two representations of a position taken from a real professional game at move 40. 2c shows the result of the application of the Zobrist algorithm. We can achieve a similar result by averaging (taking the mean) over time the amplitudes of the oscillators that comprise the oscillatory surface. Figure 2d shows the result of running our system on the same game. The darker an area, the more blacks influence, vice versa for white. It is clear that our system produces an influence map which is qualitatively similar to the Zobrist map. There are some differences though, as would be expected. The distance between the vectors ($\rho = 2c - 2d$) is 109.99 or about 15 percent of the maximum possible 722, which is close to the mean over 1000 professional positions of 111.5, showing that the example is typical.

Figure 2

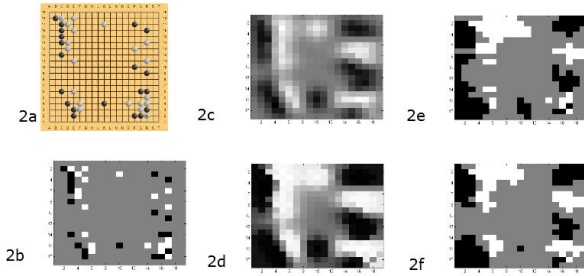


Figure 2. Comparison with existing spatial mapping algorithms

3.2 Bouzy 'Territory' Maps

The ultimate aim in Go is to establish more 'territory' than ones opponent. Territory is operationally defined by mutual agreement and inference rather than by a specific rule. Essentially, one's territory is an area of the board over which one has so much dominance that the other player can not play there without having their stones surrounded and eaten by ones own stones. If the other player doubts that an area is truly territory in this sense, they may play there and try to establish a live group. Territory and influence are related, but territory is in general more difficult to achieve - a player may have influence in an area, but not gain definite territory there.

Bouzy extended Zobrist's ideas to territory estimation. See [2] for details of the Bouzy algorithm. We designate an intersection as 'territory' if the amplitude of the oscillator at that intersection never falls below some threshold (in absolute terms). We set the threshold to 0.8 here, with the amplitudes μ of the oscillators set to 1. So the oscillatory tendency of the point must be heavily constrained by nearby stones to count as territory. Figure 22e shows the results of running the Bouzy algorithm. Non-zero intersections are considered to be territory. 2f shows the output of our system. Clearly the two analyses produce similar results. Here ρ is 53, or about 7.5 percent. This is

close to the mean ρ of 55.203 over 1000 professional games, showing that the example is typical. Our aim in this section is simply to establish by comparison to existing standards a certain level of relevance for the behaviour of our system. The Zobrist and Bouzy algorithms are not held to be perfect evaluations, and so perfect agreement is not to be expected or even necessarily desired. The system also scores finished games correctly when dead stones are removed by hand, as does the Bouzy algorithm.

The Zobrist algorithm consists of four consecutive dilations. The Bouzy algorithm consists of a series (usually 5) of consecutive dilations, followed by a larger number (usually 21) erosions. The similarity of these static mapping approaches to our results reflects an underlying algorithmic similarity between the convolutions used for dilation and erosion and the local coupling scheme for the oscillators. When surface levels are averaged over time, many oscillations can cancel out and the effect is similar to convolution.

4 Strategic planning through topological inference

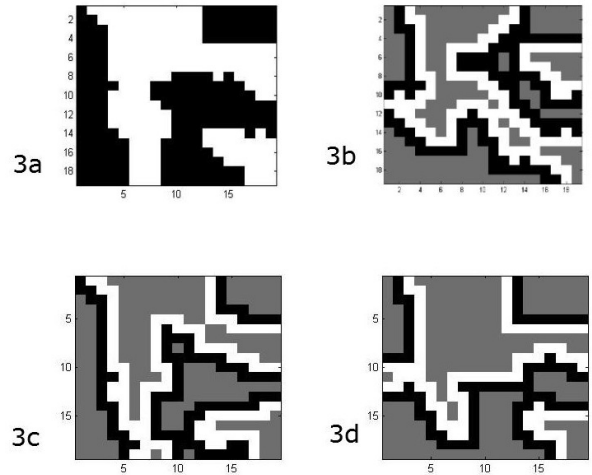


Figure 3. Some strategies suggested by our system for the position in Figure 2

4.1 Methods

Next we show how our system extends the capabilities of topological mapping in Go. Using the same game position as in Figure 2, we generate a range of strategic maps. Figure 3a-d shows 4 different strategies suggested by our system as suitable to the topology in this game. For 3b-d we have edged the maps. This form of display may be more appealing to Go players sensibilities than the solid blocks. Each point on the board is assigned to be black or white, giving us a scoreable board state¹. Figures 3a-b are winning topologies for black, 3c-d for white. Comparing 3a-d with Figure 2, it is clear that the prototopology suggested by the original position strongly constrains the topology of the resultant strategies, although considerable variation is still possible. Thus the trajectory of the dynamical surface explores a series of variations on a 'topological theme'.

¹ If we assume Chinese rules, which are ambivalent to the particular placing of stones within these territories.

As shown in Figure 3, the system can produce multiple options for winning strategies from the same position. Many of these will not follow the actual continuation. There is no existing standard to objectively evaluate the 'correctness' of these 'counter-factual' strategic hypotheses. Go games diverge so quickly that it is difficult to gather data about multiple continuations from the same position past the first few moves. However, we can still approximate a normative function using actual continuations from professional game records. Although the actual continuation is in general not to be considered as an exclusive example of the only correct continuation, it is the only available example of a continuation considered acceptable by professional players from a given position. Assuming that reasonable continuations are not uniformly distributed in the space of possible topologies, we want to find subregions which contain a relatively dense population of good continuations. We consider that the neighbourhood of the actual continuation in professional games is likely to be one such subregion.

Thus we take as the target function thresholded 'influence maps' generated by our system (as described in the previous section) of actual continuations after a certain number of moves, and measure the distance between this target vector and the vector described by a suggested strategy. We use an influence map as the target because we are more interested in change in the broad topology of the board than in the exact position of individual stones.

If the system has focussed on an inappropriate subregion of the space of topological continuations, we would expect strategic maps generated by our system to be (on average over many positions) no closer to the target than randomly generated continuations. If the subregion is appropriate, we would expect consistently smaller distances to be found. Here we take 100 professional games (taken from the GoGod database [www.gogod.co.uk] at move number 40, and follow the actual continuation in 10 steps of 10 moves, from move 50 to move 140. For each step of 10 moves into the game, we run the system for 2000 time steps, generating 2000 measurements of distance between the current surface topology and the target topology. We record the mean distance over the course of the run and also the lowest distance (i.e. best match) achieved. We also generate 2000 random continuations for each game and take the same measurements².

4.2 Results

Figure 4a shows the results of this comparison averaged over 100 games. Both the mean distance and the lowest distance are consistently smaller for the strategy maps produced by our system than for randomised continuations. As we look further into the game (movement along the x axis in Figure 4a), both mean and minimum distance increase towards the random levels. As the interaction between the competing players plans drives the continuation into increasingly unpredictable and complicated lines of play, the available information in the current position is increasingly insufficient to infer the continuation, and our system tends towards the random baseline. Randomly, we may expect to hit the right colour 50 percent of the time by chance. The randomiser achieves a mean level about 5 percent above this thanks to the 40 stones already on the board and occasionally makes an usually lucky guess.

Testing 100,000 randomised topologies we find only 11 such 'lucky guesses' which are closer to the target (at 40 moves ahead) than the mean distance of the systematic trajectory. This suggests

that the systems trajectory is limited to a sub-space somewhere in the region of 0.01 percent of that explored by the randomiser. In reality an even smaller space is probably explored, due to the tendency of the trajectory to repeatedly visit similar areas within this subregion. Yet after, for example, 40 moves further into a game, it is consistently about 30 percent closer to the target than the randomiser.

Figure 4b plots distance from the target for systemic vs randomised individual continuations during a run, with a target vector taken from 40 moves into the same game we have been considering above. Looking at the variation over sequential measurements, the difference in the form of the trajectory is clear. Whereas the randomiser generates at each time step a continuation which is independent of the previous continuation, our systems trajectory passes through a continuous series of gradually changing continuations, 'variations on a theme', which repeatedly passes through subregions of the search space relatively near to the target. This behaviour can be seen in Figure 4b.

Figure 4

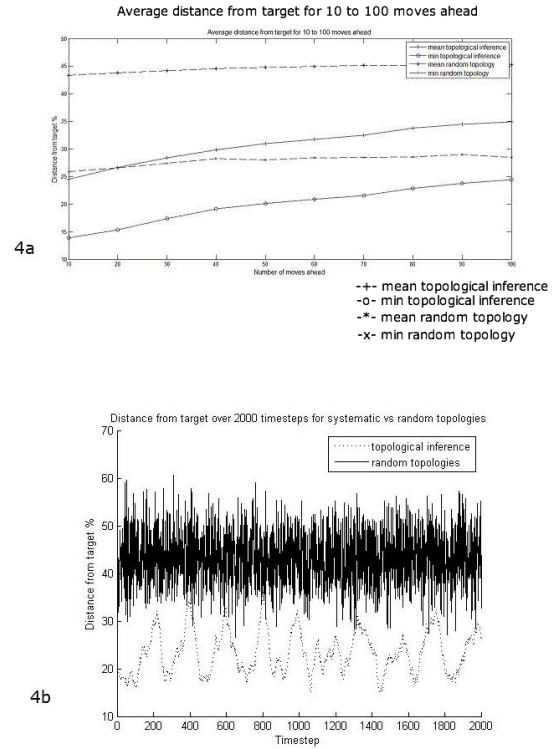


Figure 4. Results of comparing strategies generated by topological inference in our system versus randomised strategies

5 Discussion

We interpret our results as evidence that the trajectory of the dynamical system inhabits a relatively productive subregion of the space of topological continuations. This gives us confidence in the utility of 'counter-factual' strategies generated, but does not mean that the entire trajectory occupies reasonable or useful states. Automating the process of choosing points on the trajectory as strategic plans will be

² We generate random continuations by assigning each empty intersection on the board randomly to black or white, then convolving the board 5 times and thresholding to achieve areas which are similar to territories in Go.

addressed in a separate paper. Here, we have used a semi-automated process which still involves some input from the experimenter.

When involved in a game in progress, the future of the game is undetermined. Given the complexity of the game and the uncertainty related to opponents moves, it is misleading to imagine that one can expect to find, or identify even if found, a Platonic 'perfect' strategy in these circumstances. The reduced region in which the system concentrates its exploration will undoubtedly cut out good strategies that exist in unexplored regions, but there is no reason to believe that human players do not suffer the same limitation when they focus their search. We do not need to identify every possible good strategy. It is at any stage sufficient to find strategies which beat our opponent, who does not have perfect knowledge either. Opponent modelling can be addressed in our system simply by generating winning strategies for the opponent as well. Their moves may then be interpreted as steps expressing their strategic intentions.

The strategic maps produced by the system map directly to the Go board. Assuming Chinese rules, we can count unambiguously the territories of each player in a given strategic hypothesis, and so find a series of winning strategies. Part of the definition of a good strategy is, of course, one in which we win. The other major questions are whether, and how, the strategy is tactically achievable. This question can only be addressed by further work integrating the system with tactical modules and playing against rated opponents. The role of our system in a Go playing AI would thus be to propose a number of winning strategic maps compatible with the current position. Tactical search can then be limited to examining the tactical viability of a limited number of specific strategies. Thus the system can 'frame' local, tactical analysis with specific, global hypotheses. Feedback from tactical modules may then be used to refine strategic hypotheses, providing the opportunity for powerful mutual bootstrapping. We suggest that this interaction of global topological inference with local pattern matching and tree search is the route to deep, directed lookahead along winning continuations in computer Go.

Having access to multiple strategies enables flexible play in the face of uncertainty regarding the opponents moves. We may aim to play moves which keep our strategic options as wide as possible, by contributing to many potential winning strategies. The absence of abstract representations and symbolic computations means that we do not have a complicated framework which must be constructed anew to incorporate new developments in the game as it progresses. Simply coupling the new stone to the system as described above is sufficient for the system to adjust its trajectory to incorporate each new move. This paper introduces a system in the early stages of its development. Future work will concentrate on optimisation of parameters, integration with tactical search, and further possible uses for our system, such as identifying strategic 'regions of interest'. For example, where winning strategy(s) require a black group in the upper right, but there are currently no black stones in that region, we can identify this as a region that requires development. Since the days of Walter Freeman's Mass Action in the Nervous System [5], many researchers have speculated that spatio-temporal dynamics of the cortical surface encode patterns and neural computation above and beyond the learning processes inherent in synaptic transmission. Varela et al [10], Sejnowski et al [4] and Sporns et al [3], revisit these ideas with the much greater information now available on the timing and coherence of neural spike trains. Our results here suggest a possible role for cortical surface waves in the organisation and exploration of topologically organised spaces.

6 CONCLUSION

We have presented a new method for producing strategic maps in computer Go. Our approach extends earlier work on visual-spatial methods in Go, enabling dynamic global exploration through topological inference. We have shown that our system can produce results similar to these existing algorithms as a special case, and that despite constraining this exploratory process to a relatively small subregion of the space of possible topologies, the system consistently follows a trajectory closer to continuations in real professional games than randomised continuations. This provides evidence that this subregion is populated by 'reasonable continuations' (given available information), based on the assumption that real professional continuations are members of this sub-population of 'reasonable continuations' and that this sub-population is over represented in the subspace nearby the actual continuation. The strategic maps generated by our system are intended to be used to direct tactical search towards winning continuations among this sub-population.

REFERENCES

- [1] P. Auer, N. Cesa-Bianchi, and P. Fischer, 'Finite-time analysis of the multiarmed bandit problem', *Machine Learning*, **47**(2/3), 235–256, (2002).
- [2] B. Bouzy, 'A model of visual organisation for the game of go', *International Journal of Pattern Recognition and Artificial Intelligence*, **17**(2), 257–268, (2003).
- [3] E.T. Bullmore and O. Sporns, 'Complex brain networks: graph-theoretical analysis of structural and functional systems', *Nat. Rev. Neuroscience*, **10**, 186–198, (2009).
- [4] D. Durstewitz, J.K. Seamans, and T.J. Sejnowski, 'Neurocomputational models of working memory', *Nature Neuroscience*, **3**, 1184–1191, (2000).
- [5] J. Freeman, W. *Mass Action in the Nervous System*, Academic Press, New York, 1st edn., 1975.
- [6] S. Gelly and Y. Wang, 'Exploration exploitation in go: Uct for monte-carlo go', *Twentieth Annual Conference on Neural Information Processing Systems*, **20**, –, (2006).
- [7] L. Kocsis and C. Szepesvari, 'Bandit based monte-carlo planning', *15th European Conference on Machine Learning*, **15**, 282–293, (2006).
- [8] L. Righetti, J. Buchli, and A. J. Ijspeert, 'Dynamic hebbian learning in adaptive frequency oscillators', *Physica D*, **216**(2), 269–281, (2006).
- [9] D. Stern, R. Herbrich, and T. Graepel, 'Bayesian pattern ranking for move prediction in the game of go', *Proceedings of the 23rd international conference on Machine learning*, **23**, 873–880, (2006).
- [10] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, 'The brainweb: phase synchronization and large-scale integration', *Nat. Rev. Neuroscience*, **2**, 229–239, (2001).
- [11] J von Neumann, 'Zur theorie der gesellschaftsspiele', *Math. Annalen*, **100**, 295–320, (1928).
- [12] D. Wang, 'The time dimension for scene analysis', *IEEE Trans. Neural Networks*, **16**, 1401–1426, (2005).
- [13] M. Wheeler, 'Cognition in context: Phenomenology, situated robotics, and the frame problem', *International Journal of Philosophical Studies*, **16**(3), 323–349, (2008).
- [14] A. Zobrist, 'A model of visual organisation for the game of go', *Proc. AFIPS Spring Joint Computer Conference*, –, 103–111, (1969).