協力ゲーム Hanabi を用いた 他者行動からの自己状態推定手法の検討

Estimation of Own State by Opponent's Behavior in Cooperative Game Hanabi

大澤 博隆 [1]

Hirotaka Osawa[1]

[1] 筑波大学
[1]University of Tsukuba

Abstract: We used a card game called Hanabi as an evaluation task of imitating human reflective intelligence with artificial intelligence. Hanabi is a cooperative card game with incomplete information. A player cooperates with an opponent in building several card sets constructed with the same color and ordered numbers. We compared human play with random strategy, rational strategy, rational strategy with opponent's viewpoint, and rational strategy with feedbacks from simulated opponent's viewpoints. The results indicate that the strategy with feedbacks from simulated opponent's viewpoints achieves more score than that with the rational only strategy.

1. はじめに

他者の思考を読むという社会的知能は人間の持つ固有の特徴の一つである。こうした社会的知能を解決する知能は、人工知能や認知科学といった分野における研究対象となり得る。Bryne らは、社会における他者の意図の読み込みが人間の知能の進化の主要因になったと考えている[1]。

社会的な知能課題の中で最も難しい物の一つは、他者の行動から自分自身の状態を推定することである。このような熟考的な（reflective な）思考、他人の行動を鑑とする振る舞いは生物学的であり、心理学的な課題である。たとえば、人間の声は他人に対し空気を介して伝わるが、自分に対しては骨伝導によって伝わる[2]。そのため、各個人は自分自身の声を聞くことは出来ない。しかしながら、他者の反応を観察することで、自分の声をより効果を持つように修正することは可能である。心理学分野では、このような自分から見えない情報は"blind spot"と呼ばれる [3].

本研究ではこのような reflective な他者からの自己推定課題の例として、Hanabi と呼ばれる協力ゲームを用い、この課題を解くことで他者からの自己推定の知能がどのように振る舞うか検討する。Hanabi は協力的なカードゲームである。このゲームは AI 研究で使われてきた他のカードゲームと比較して、３つの特徴を持っている。第一に、このゲームは対戦ゲームではなく、協力ゲームであり、全エージェントが協力して得点を競う。全てのプレイヤーは、１~５までのカードの列で表される５色の花火を協力して作り上げる。そして、この花火の大きさが得点となる。第二に、このゲームではプレイヤーは自分のカードは見えないかわり、自分のカード以外の全てのエージェントのカードを知ることができる。全ての情報を俯瞰できるエージェントがいないため、どれか一人のエージェントがリーダーとなり行動を決定する、スターモデルがとれない。第三に、このゲームではエージェント同士のコミュニケーションが制約されている。各エージェントは他のプレイヤーのカードの数もしくは色を教えるため、情報伝達用の資源を消費しなければならない。その他のコミュニケーション手段は用意されない。このような条件下であるため、Hanabi を解く AI プログラムは自然言語処理など、一般的なコミュニケーションのための処理を要求されない。以上のような独特の特徴を評価され、Hanabi はドイツのゲーム大賞を受賞した[4].

筆者はこの Hanabi を解くための人工知能エージェントを実装した。本エージェントは他者の視点とその行動をシミュレートできる。これによって、他者の視点の再現がどのように得点に結びつくかを検討する。

本論文の構成は以下のとおりである。２章では、AI 分野で検討されてきた不完全情報ゲームについて概観し、Hanabi がどのような新しい挑戦をこの分野にもたらすか検討する。３章では、Hanabi のゲームルールを説明する。我々は本課題で、特に２人プレイヤーの Hanabi ゲームについて検討を行った。４章では、Hanabi をプレイさせるために実装したいくつかの戦略について説明を行う。５章では、シミュレーションの具体的な評価プロセスついて説明し、この議論を６章で行う。７章では、本研究の貢献と将来課題について述べ、８章で本研究を結論付ける。

2. 関連研究との違い

2.1. 人狼ゲームの発展

ゲームをプレイするエージェントの作成は、人工知能研究におけるランドマーク課題の一つである[5]。完全情報ゲームとして、チェッカー、オセロ、チェス、将棋、囲碁といった課題が取り組まれてきた[6][7]。これらのゲームでは、全ての情報は両方のプレイヤーから観測可能であり、エージェントは勝利のために、必ずしも他者の意図を推測する必要はない。

これに対し、カードゲームなどのゲームには、他者の情報が観測不可能な不完全情報ゲームが存在する[8]。これらのゲームも研究対象となっている。ポーカーはその中でもよく知られた例であり、いくつかの理論的な分析や大会が行われている[9][10]。この他に、ブリッジや Do Zi Zhu(闘地主)といったゲームに関する研究が知られている[11] [12]。

2.1. Hanabi の特徴

これらのゲームと比較し、Hanabi はエージェント研究に貢献する３つの特色を持っている。

第一に、このゲームは協力ゲームである。Hanabi に参加する全てのプレイヤーはお互いに協力してカードを積み上げ、花火を作ることが求められる。このような条件はマルチエージェントシステムにおける協調問題によく似ている。

第二に、全てのプレイヤーは自分以外のプレイヤーのカードを観測可能である。これは客観的な視点を持ったプレイヤーが存在しないことを意味しており、リーダー不在の状況での協調を求められる課題である。これも、マルチエージェントシステムにおける協調課題によく似ている。

第三に、Hanabi ではプレイヤー間のコミュニケーションが厳しく制約されている。プレイヤーは他のプレイヤーの色、もしくは数字を教えることができるだけであり、その教示には情報カウンターと呼ばれる資源を消費する必要がある。このように制限された条件のため、Hanabi の解法には自然言語処理からの意味理解が必要とされない。また、ゲーム理論におけるチープトークと呼ばれる利得を伴わない情報交換がない、という過程を意味している[14]。Hanabi をエージェント課題として用いることで、言語に依存しない形の社会的知能、意図の読み合いを再現することが可能である。

3. 制約条件

3.1. Hanabi のルール

Hanabi は 2 から 5 人のプレイヤーによって行われるゲームである。本研究では、2 人のプレイヤーによるゲームのみを扱う。Hanabi は 5 色(白、赤、青、黄、緑)、50 枚のカードを使用する。1 色につき、10 枚のカードが存在する。1 のカードは 3 枚、2 から 4 までのカードは 2 枚、5 のカードは 1 枚存在する。本ゲームのゴールは、1 から 5 までの数字のカードが重なった、5 つの異なる色の山（花火）を作ることである。

ゲーム開始時に、各プレイヤーは 5 枚のカードを配られる。残りの 40 枚のカードは山札として積まれる。また、2 人のプレイヤーは 8 枚の情報カウンターを共有する。2 人のプレイヤーが交互にターンを重ね、協力して花火を作成する。

各プレイヤーは、ゲームの各ターンにおいて 3 つの行動が許されている。1 つめの行動、情報提示である。この行動の場合には、各プレイヤーは相手の持っているカードの色か数字を教えることができる。情報を教える場合には、情報カウンターを一つ消費する。例えば、相手のカードが赤 1、緑 2、緑 3、白 2、白 4 である場合、相手に緑の色を教える場合には、2 番目のカードと 3 番目のカードが緑である、と教えることが可能である。また、数字 2 を教える場合には、2 番目と 4 番目のカードが 2 であると教えることが可能である。どちらか一方しか教えない、ということはできない。また、で情報カウンターが存在しない場合には、相手に情報を教えることはできない。

2 つめの行動は、カードの破棄である。プレイヤーは自分が持っている 5 枚のカードのうち、必要ないと考えるカードを捨て、その代わりに新しいカードを山札から補充し、さらに共有する情報カウンターを戻すことができる。破棄したカードは自分を含めた全員に公開される。一度破棄したカードは、2 度と使うことはできない。例えば、1 の数字のカードは同じ色のものが 3 枚あるため、どれかを捨てても、他 2 枚のどちらかのカードを使うことで、花火を完成させることができる。しかしながら、5 のカードは各色 1 枚ずつしかないため、このカードを捨ててしまうと、この色の花火が完成することはない。また、情報カウンターが既に 8 個ある場合には、カードを捨てることはできない。

3 つめの行動は、プレイである。プレイヤーは自分が持つカードのうち、どれか一つを花火につなげる「プレイ」を行うことができる。もし、自分の出したカードが既存の同色の花火よりも 1 つだけ大きい数字の場合、花火を成長させることができる。例えば、緑の花火が 1 から 3 までの数字で構成されている時、緑 4 のカードを出すことで、緑の花火を 1 から 4 までの数字に成長させることができる。ただし、緑 3 や緑 5 のように、繋がらない数字を出してしまうと、それは失敗となり、出したカードは花火に接続されず破棄され、ゲーム中で二度と使用できなくなる。失敗が 3 回繰り返されると、ゲームはその時点で終了する。また、その色の花火が存在しない時にその色の 1 のカードを出すと、新しく花火を作ることが可能である。プレイ後に、カードを補充するためプレイヤーは山札から新たにカードを 1 枚加える。なおオリジナルのゲームでは、ある色の花火が 1 から 5 までのカードを揃えた場合、成功となり情報カウンターを 1 つ戻すことができる、というボー

ナスがあるが、単純化のため今回はこのルールを採用しない。

3 回プレイが失敗する場合、山札からカードが尽きて、さらに一周した場合、全ての色の Hanabi が完成する場合の、どれか 3 つの場合にゲームが終了する。終了後には、各色の花火のそれぞれのカード枚数を合計し、これが特典となる。最大で各色 5 種×5 枚のカードで、25 点となる。

3.2. 定式化

以下のように表記の定式化を行う。

3.2.1. カード名

それぞれのカードを各色の最初の文字と数字で表す。例えば、赤の 4 のカードは R4、緑の 1 のカードは G1 と表す。情報の足りない場合には、そのカードをアンダーラインの形で表し、可能なカードの集合の表記とする。例えば、R_というカードはそのカードが{R1,R2,R3,R4,R5}のどれかであることを表す。

3.2.1. ゲーム盤の表記

ゲームの山札を集合 P、捨てられたカードを集合 T、各花火を集合 D で表す。P は両プレイヤーから観測不可能なカードの順列集合であり、P={Y3,W1,R2,…}のように表す。T は捨てられたカードの集合であり、両プレイヤーから観測可能なカードの集合となる（T={R3,G2,Y1,…}）。D は各花火の集合を表しており、空集合を含めた順序集合として D={{},{B1},{},{R1,R2},{G1,G2,G3}}のように表される。ゲームの開始時に P は 50 のカードの集合であり、T は 0、D は 5 つの空集合の集合{{},{},{},{},{}}となる。

全てのプレイヤーは各プレイヤー自身の視点を保持する。これを W_{pl} や W_{co} のように表す（pl は player であり、co は cooperator(協力者)を表す）。W は 5 枚のカードの状態の順序集合 C を保持している。各プレイヤーの視点上の自分のカードや相手のカードを記述可能である（C_{pl}={R_,R_,__,__,_3}, C_{co}={R1,R2,W1,W2,G3}など）。

3.2.2. プレイヤの役割

各プレイヤは関数 F で表される。F は入力として D, P, T, W を受け取り、出力として A を返す($A = F_{pl}(D, P, T, W_{pl})$)。

3.2.3: 基本作戦

もし相手がカードに関する情報を与えた場合、そのカードに対する情報は狭まる。例えば、あるカードを指して他方のプレイヤーが赤と指摘した場合、そのカードが赤であると同時に、他のカードは赤でないという情報が入る。このため、可能集合は常に減少する。

3.2.4: プレイ可能カードの定義

もし、あるカードがプレイするのが可能であるという確定的な情報を持つ場合、そのカードをプレイ可能カードと定義する。例えば、場に花火が一つも出ておらず、あるカードが 1 という情報が決定している場合、そのカードは何色であるかに関わらず、プレイ可能なカードとなる。また、もし緑の花火が 3 まで完成していれば、G4 のカードはプレイ可能カードである。

3.2.5: 破棄可能カードの定義

もし、あるカードかこれ以上プレイ可能とならないという確定的な情報を持つ場合、そのカードを破棄可能カードと定義する。例えば、場に赤 4 の花火が完成している場合、R1,R2,R3 のカードはいずれも破棄可能カードである。また、黄色の花火が 5 まで完成している場合、全ての黄色のカードは、数字の情報があるなしに関わらず、破棄可能カードとなる。

3.2.6: 重複カード

あるカードと同じものが存在し、それが公開されていない場合、このカードは重複カードと定義される。例えば、R2 というカードが 1 枚存在し、それ以外のカードが公開されていなければ、R2 は重複カードと定義される。

4. 戦略

本章ではエージェントが使用した代表的な 5 つの戦略を説明する。それぞれの戦略におけるエージェントの視点を図 1 の通り表す。

4.1. 完全戦略

Hanabi の各プレイヤーが最高得点を取る場合は、お互いに情報を得ている場合である。この場合には、両方のプレイヤーは常に最適な手を打つことができる。Hanabi のルール上、両プレイヤーから見えない山札などの情報にアクセス出来ない限り、この戦略に勝てる戦略は存在しない。得点の比較のため、最高得点を取るような戦略を持つプログラム作成した。この戦略は、以下の様な手順で推移する。

1. もしプレイヤーがプレイ可能カードを持っていれば、そのカードをプレイする。
2. もし相手がプレイ可能カードを持っていれば、プレイヤーは情報を教えて相手に番を回す（両者が情報を保持している場合、情報を教える意味は無いため、実質的にターンスキップと同じである）
3. もしプレイ可能なカードが両者に無い場合は
   - A) もしプレイヤが破棄可能カードを持っていたら、それを捨てる
   - B) もしプレイヤが破棄可能カードを持っておらず、重複カードを持っていた場合、それを捨てる
   - C) もしプレイヤが破棄可能カードも重複カードも持っていない場合には、手の中から一番大きい数字のカードを捨てる。

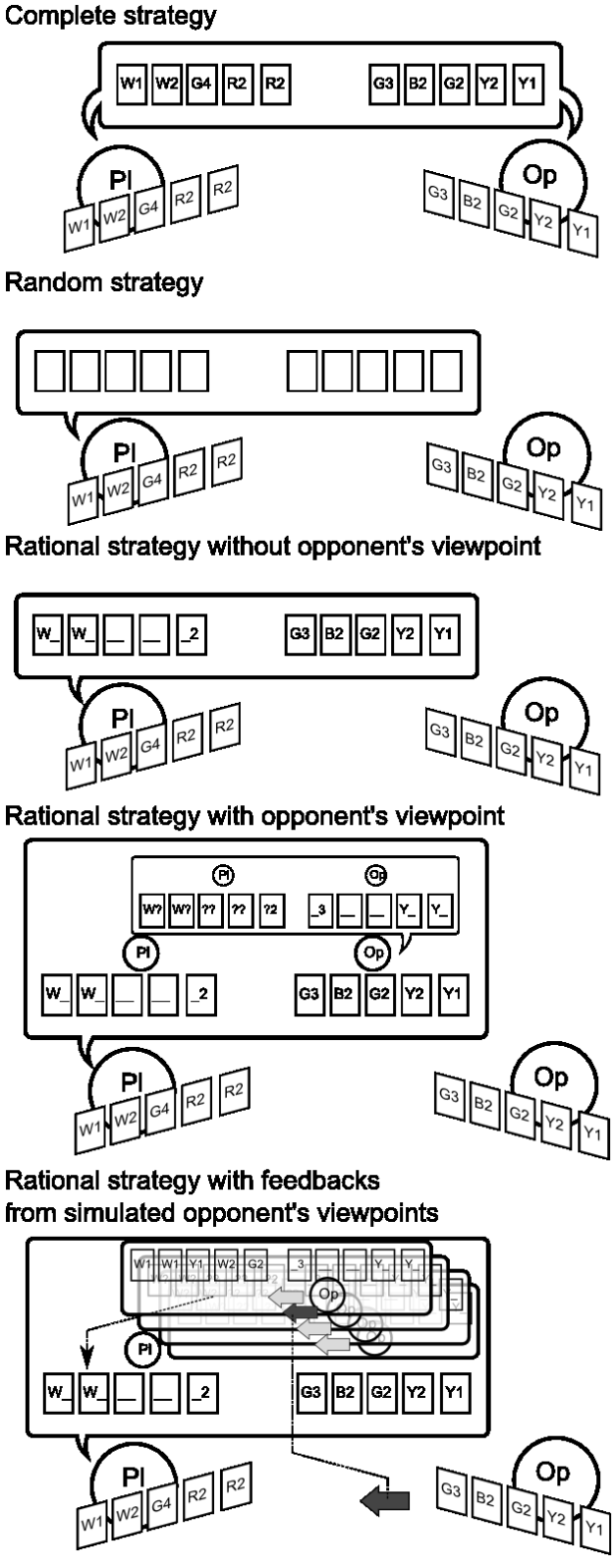

図 1　代表的な戦略 5 種

4.2. ランダム戦略

ランダム戦略では、プレイヤーはカードに対するあらゆる情報を持たない。本ランダム戦略では、30%の確率で情報提示、40%の確率でランダムなカードの破棄、30%の確率でランダムなカードのプレイを行った。

4.3. 他者視点のない合理的戦略

他者視点のない合理的戦略では、各プレイヤの手は図 1 のように表される。この条件下では、プレイヤは自身の手の記憶を持つが、相手が何を覚えているか、という記憶は持たない。

1. もしプレイヤーがプレイ可能カードを持っていれば、そのカードをプレイする。
2. もしプレイヤが破棄可能カードを持っていたら、それを捨てる
3. もし相手がプレイ可能カードを持っていれば、プレイヤーはそのカードの色か数字情報をどれか一つ教える
4. もし相手がプレイ可能カードを持っていない場合、プレイヤは相手のカードを 1 枚ランダムに選び、そのカードの色か数字の情報を教える
5. もし相手がプレイ可能なカードを持っておらず、プレイヤに情報トークンが存在しない場合、プレイヤは自分のカードを 1 枚ランダムに選び、それを捨てる。

4.4. 他者視点のある合理的戦略

本戦略では、他者視点のない合理的戦略と同じ手順でプレイを行うが、プレイヤは相手の視点を持つ。このため、一度教えた情報を教えることがない。

4.5. 他者視点のシミュレートをフィードバックする合理的な戦略

本戦略では、他者視点のある合理的戦略と同じようにプレイを行うが、4.3 のステップ 4 で、相手がプレイ可能なカードをもっていない場合、相手の出した手から、相手の視点をシミュレートし、その結果として自分のカードを予想する、という行動を行う。これは、相手が自分と同じように考える、という前提を元にした推論である。この予想の手順は以下のとおりとなる。

1. プレイヤは自分が持つ手の可能な組み合わせ集合 H を全て考える（例えば、H={{R1, R1, G2, G2, W1}, {R1, R1, G2, G2, W2}…}）
2. プレイヤは、一手前のゲームの状態 D_{pre}, P_{pre}, T_{pre} を再現する。そして、H の一つ一つの要素を P_{pre} に当てはめる。
3. 当てはめた P_{pre} の各要素に対して、シミュレートした F_{op} の結果を計算する（$A_{hyp}=F_{op}(D_{pre}, P_{pre}, T_{pre}, W_{hyp\_op})$）

4. もしシミュレートした結果 A_{hyp} が、相手が言って前に行った現実の手と一致しない場合、その手を集合 H から取り除く。
5. H の要素がなくなるまで、step2 に戻る

上記のプロセスにより、プレイヤは自分の手の可能集合 $H_{estimate}$ を求めることができる。

その後、我々はヒューリスティクスを用いて持っているカードを推測する。プロセスは以下のとおりである。集合のうち、最も登場したカードの総数を x、次に登場したカードの総数を y とする。x が y の a 倍より大きい場合に、そのカードを持っているとかんがえる。

具体例を考えてみる。例えば、花火がもしひとつも完成していないときに、相手のプレイヤが「あなたの一番右のカードは緑」と教えるとする。このカードに対し情報が与えられる可能性は、プレイ可能カードであるときに大きくなる。従って計算により、このカードを緑、と教えるのは、このカードが G1 であるとき、という可能性が多くなる、以上の手続きより、このカードを G1 と推測し、1 が出せるという合理的規則に基づいて、プレイヤはこのカードをプレイする。

5. シミュレーション結果

5 つの戦略について、それぞれコンピュータで 100 回シミュレートした結果と、人間のプレイヤで 20 回シミュレートした結果を比較した。条件はいずれも同条件である。計算では、持てるカードが 2 枚の場合、5 枚の場合のそれぞれの計算を行った。資源の制限のため、戦略 5 の再帰的なシミュレーション推定は直前の一回のみ行った。事前のシミュレーションにより、a=2.5 という値が戦略 5 において最も高得点であったため、この値を採用した。

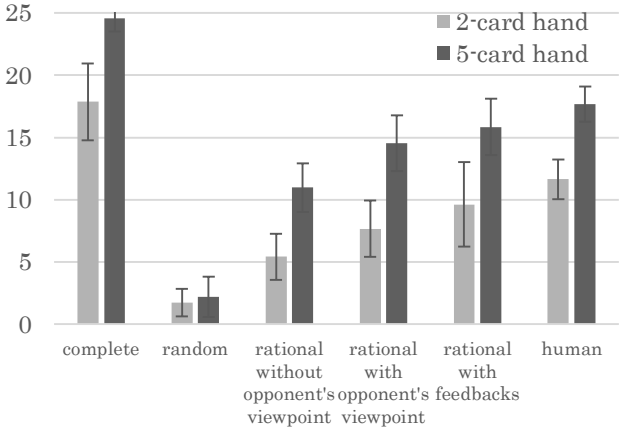

図 2　戦略の変化

結果を図 2 に示す。持てるカードが 2 枚の場合、完全戦略の際の得点は 17.86 (SD 3.07) 、ランダムの際は 1.74 (SD 1.10) 、他者視点のない合理的戦略では 5.41 (SD 1.83)、他者視点のある合理的戦略では 7.66 (SD 2.28) 、他者視点のシミュレートをフィードバックする合理的戦略では 9.61 (SD 3.40) 、人間同士の試合では 11.65 (SD 1.60) となった。持てるカードが 5 枚の場合、完全戦略の際の得点は 24.6 (SD 1.10) 、ランダムの際は 2.20 (SD 1.60) 、他者視点のない合理的戦略では 10.97 (SD 1.94)、他者視点のある合理的戦略では 14.53 (SD 2.24) 、他者視点のシミュレートをフィードバックする合理的戦略では 15.85 (SD 2.26) 、人間同士の試合では 17.7 (SD 1.42) となった。ANOVA 検定を行った結果、全ての対において、$p<.05$ となり、全群の有意差が示された。本結果は、他者視点のシミュレートをフィードバックする合理的戦略が、人間同士の試合の次に良い結果であることを示唆する。

6. 考察と結論

本結果は、他者の視点から自己の視点を推測するシミュレーションが、ゲームの得点上昇に対して有意に働いしていることを示唆する。たとえば、D={W:5, R:5, B:0, Y:3, G:2} 、W_{pre_pl}={C_{pre_pl}=(__,_1,__,Y_,__)、C_{pre_op}=(Y1, R4, Y4, B2, Y4)}、かつ W_{pre_op}={Cpre_pl=(Y3, B1, G1, Y2, B1),、C_{pre_op}=(__, __, _4, __, _4)}というとき、プレイヤが相手に対して 1 のカードを教えた場合があった (C_{op}=(_1,__,_4,__,_4))。このとき、教えられたカードは Y1 である可能性がもっとも高いとこのプログラムは判断し、Y1 をプレイしている。同様の状況でただの合理的戦略では手を決定できず、このような決断ができなかった。

本研究は、相手のシミュレーションによる自己推定による協力課題を解いている、と捉えることができる。この得点を生存要因ととらえると、認知科学における心の理論が人間にもたらす効用を推測できる[15][16]。他者の意図を推定することが、生存に結びつくような課題が存在する、ということになる。本研究の成果を元に、知能を発達させる様々な課題について、引き続き検討を行っていく所存である。

謝辞

本研究は JSPS 科研費 26118006A の助成を受けたものです。

参考文献：

[1] R. W. Byrne and A. Whiten, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, USA, 1989.

[2] Z. Chen, Z. Su, S. Ma, X. Wu, and Z. Luo, “Biomimetic modeling and three-dimension reconstruction of the artificial bone.,” *Comput.*

Methods Programs Biomed., vol. 88, no. 2, pp. 123–30, Nov. 2007.

[3] J. Luft and H. Ingham, "The Johari Window: a graphic model of awareness in interpersonal relations," *Hum. relations Train. news*, vol. 5, no. 9, pp. 6–7, 1961.

[4] S. des Jahres, "Spiel des Jahres Award," *Spiel des Jahres*, 2013. [Online]. Available: http://www.spieldesjahres.de/cms/front_content.php?idcatart=1121&id=828. [Accessed: 31-Aug-2014].

[5] B. Abramson, "Control strategies for two-player games," *ACM Comput. Surv.*, vol. 21, no. 2, pp. 137–161, Jun. 1989.

[6] K. Krawiec and M. G. Szubert, "Learning n-tuple networks for othello by coevolutionary gradient search," in *Proceedings of the 13th annual conference on Genetic and evolutionary computation - GECCO '11*, 2011, p. 355.

[7] S. Gelly, L. Kocsis, M. Schoenauer, M. Sebag, D. Silver, C. Szepesvári, and O. Teytaud, "The grand challenge of computer Go," *Commun. ACM*, vol. 55, no. 3, p. 106, Mar. 2012.

[8] S. Ganzfried and T. Sandholm, "Game theory-based opponent modeling in large imperfect-information games," in *International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 533–540.

[9] D. Billings, D. Papp, J. Schaeffer, and D. Szafron, "Opponent Modeling in Poker," in *AAAI Conference on Artificial Intelligence*, 1998, pp. 493–499.

[10] D. Billings, N. Burch, A. Davidson, R. Holte, J. Schaeffer, T. Schauenberg, and D. Szafron, "Approximating Game-Theoretic Optimal Strategies for Full-scale Poker," in *International Joint Conference on Artificial Intelligence*, 2003, pp. 661–668.

[11] M. L. Ginsberg, "GIB: Imperfect Information in a Computationally Challenging Game."

[12] D. Whitehouse, E. J. Powley, and P. I. Cowling, "Determinization and information set Monte Carlo Tree Search for the card game Dou Di Zhu," in *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)*, 2011, pp. 87–94.

[13] G. Zlotkin and J. S. Rosenschein, "Incomplete information and deception in multi-agent negotiation," in *Proceedings of the 12th international joint conference on Artificial intelligence*, 1991, pp. 225–231.

[14] K. Wärneryd, "Evolutionary stability in unanimity games with cheap talk," *Econ. Lett.*, vol. 36, no. 4, pp. 375–378, Aug. 1991.

[15] L. M. Hiatt and J. G. Trafton, "A Cognitive Model of Theory of Mind," in *International Conference on Cognitive Modeling*, 2010, pp. 91–96.

[16] M. C. Frank and N. D. Goodman, "Predicting Pragmatic Reasoning in Language Games," *Science (80-.).*, vol. 336, no. 6084, p. 998, 2012.

[17] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Developing a model of robot behavior to identify and appropriately respond to implicit attention-shifting," *Proc. 4th ACMIEEE Int. Conf. Hum. Robot Interact. HRI 09*, pp. 133–140, 2009.