Design and Secure Evaluation of Side-Choosing Games

RUIYANG XU, Northeastern University
KARL LIEBERHERR, Northeastern University

We present an important, general class of new games, called side-choosing games (SCGs), for "gamifying" problem solving in formal sciences using plausibility checking. Applications of SCGs include (1) peer-grading in teaching to (2) studying the evolution of knowledge in formal science communities to (3) organizing algorithm competitions. We view SCGs as a new and general model for formulating formal problems that need to be solved using human computation and our interest in this paper is on how to evaluate a set of SCGs about the same problem fairly and effectively. We observe that a specific kind of collusion, where players lie about their strength and sacrifice themselves, could bias the evaluation of SCGs dramatically. Following the idea of Social Choice Theory in the sense of Arrow, we take an axiomatic approach to guarantee that a specific kind of collusion is impossible. We prove the Collusion-Resistance Theorem and related results as a general principle for designing collusion-resistant evaluations for SCGs. The Collusion-Resistance Theorem is surprising: it tells us to be *indifferent* to wins but to count certain kinds of losses for scoring players and ranking them. If collusion is not an issue, we offer a family of useful ranking functions which are not collusion-resistant.

TO BE DELETED: Limit: 18 pages. July 24-28, '16 The Netherlands. 11/30/15, Submission deadline: February 23, 2016.EC'16 will be co-located with the 5th World Congress of the Game Theory Society (GAMES 2016), May 30, 2016: Camera-ready version of accepted papers due.

Our focus areas: (1) Theory and Foundations and (2) Applied Game Theory.

Our topics from CFP: Social Choice and Logic, Collective Intelligence, Game-theoretic models of competitions.

Version Feb. 5

1. INTRODUCTION

A side-choosing game (SCG) $H = \langle G, GS, Q, p_x, p_y \rangle$ is based on an extensive form two-player game G between players p_x and p_y with perfect information and without ties, i.e., there is always a winner and a loser[1]. G is a game between two players, *white* and *black*. GS is a game state of G (i.e., a node of the game tree of G). Q is a proposition on the game state GS of the form: *white* has a winning strategy (*white* moves first).

The players p_x, p_y of H have their preferred, static side (*white* = *Proponent* or *black* = *Opponent*), depending on whether they believe Q or !Q to be true. The players are free to choose their static side before the game. But during the game the players must have opposite dynamic sides which we implement by making (per game) at most one of them the devil's advocate (or forced). Fig. 1 illustrates the difference between static and dynamic assignments. The dynamic assignment is made by a trusted third party fairly. The details of this assignment are not important to this paper as we focus on evaluating the game results abstracting from this assignment.

[1] Extensive form games are widely used to model multi-agent sequential decision making. They are represented by game trees.

Author's addresses: R. Xu and K. Lieberherr, College of Computer and Information Science, Northeastern University; email: lieber@ccs.neu.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

EC'16, July 24–28, 2016, Maastricht, The Netherlands. ACM 978-1-4503-3936-0/16/07.
http://dx.doi.org/10.1145/XXXXXXX.XXXXXXX

	static		dynamic	
player	white	black	white	black
p_x	proponent		free	
p_y		opponent		free
p_x	proponent		free	
p_y	proponent			forced

Fig. 1. SCG explain

The side-choosing game $\langle G, GS, Q, p_x, p_y \rangle$ produces a game result row consisting of (1) the **winner** (p_x or p_y) (2) the **loser** (p_x or p_y) and (3) at most one **forced** player (p_x or p_y or 0, if none was forced). A set of game results produced by multiple binary SCGs is called an SCG-Table which we often interpret as a labeled multi-graph (see Table I and the corresponding graph in Fig. 3 and for more graphs Fig. 6 and 7).

1.1. Examples

Consider the chess position GS in Fig. 2 as a side-choosing game[2]. The game G is chess, modified so that winning for white means to mate the black king in 2 moves. The proposition Q says that white starts in GS and wins. We have two players, Alice and Bob, who study G and GS and make their side choices. Alice believes she can win as white (she "sees" the mate in two) and therefore her side choice is *white*. Bob does not see the mate in two and therefore he wants to be *black*. The game is played and Alice wins (how is left as an exercise to the reader; there is only one optimal move for white.). The game result is given by the row= (**winner**="Alice", **loser**="Bob", **forced**=0). If *white* plays b3 and *black* f3 then *white* mates with Qb2. *White* wins but only because *black* made a mistake. The correct move for *black* is c4 (not f3) and *white* cannot mate in the next move. This example gives the wrong impression that playing one perfect game reveals the winning strategy (solution). But this is not the case most of the time.

Next we consider a large family of examples of side-choosing games: the family consists of semantic games [Kulas and Hintikka 1983] with side choice added. The game G is defined by an interpreted logical sentence between white (proponent, existential quantifier) and black (opponent, universal quantifier). The outermost quantifier of the sentence determines who moves first. For white, the game is about making the sentence true by assigning values to variables. Black tries to prevent this. Side-choosing games exist for many different logics such as first-order predicate logic, higher-order logics and independence-friendly logic. A first important subfamily are claims in formal sciences with side choice added. A second important subfamily are claims related to algorithm specifications for algorithm competitions (a la TopCoder, see topcoder.com or Kaggle, see kaggle.com).

1.2. Motivation

Why is the concept of SCG useful? A game state GS of an extensive form game G is a model of a claim. Claims and problem solving are ubiquitous in human and robotic reasoning. What is surprising to some readers is that any formal claim can be brought to the extensive game form. The idea behind the translation is to express the claim as

[2]Taken from the collection: SEVENTY-FIVE CHESS PROBLEMS by John Thursby, Trinity Coll., Cambridge, 1883.

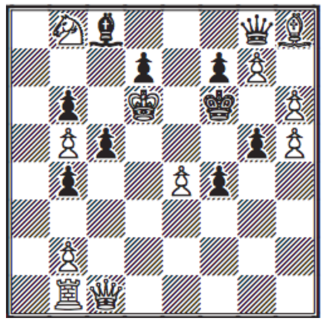

Fig. 2. An SCG example: white mates in two

an interpreted predicate logic sentence as mentioned above. An SCG (the playing of the game) can be viewed as a plausibility check of a claim. When I argue that a claim is true but then I cannot defend it using a plausibility check, there must be something wrong with my argument. We show how to aggregate multiple such plausibility checks in a robust manner so that players cannot collude against other players [Abdelmeged 2014].

We study the ranking of players, given a table of game results. We want to find the most meritorious players in a robust fashion so that collusion between players cannot prevent the most meritorious players to be recognized. For example, we want to prevent that a player lies about its strength and "sacrifices" himself by intentionally losing. This might rank undeserving players highly and will destroy the purpose of SCGs. For an explicit example of collusion, see section 5.9.

Finding the most meritorious player is an interesting problem because there can be a lot of noise in a game result table. For example, a player may correctly choose the white side but then not have the necessary skills to defend the correct side. Or a game state GS may be considered a winning state for white until a new player joins who knows how to win as black. To bring order into this complex situation, we propose an axiomatic approach.

2. CONTRIBUTIONS

The contributions of this paper are (1) a new concept called SCG for gamifying problem solving using plausibility checking of claims, (2) a set of high-level axioms defining desirable properties of ranking functions and their interpretation in the context of a 4 argument scoring function (3) a new proof technique, called monotonicity constraint reduction, for reducing predicate logic sentences (representing axioms) to monotonicity constraints where the reasoning is simpler, (4) The Collusion-Resistance Theorem and related properties as an application of the monotonicity constraint reduction. The Collusion-Resistance Theorem suggests an efficient static implementation of the axioms (a ranking function based on the monotonicity constraints) avoiding any kind of dynamic checking. (5) We make a contribution to the secure design of techno-social systems where a kind of security (collusion resistance) is built-in by design.

We have been working on SCGs for a few years which resulted in a dissertation [Abdelmeged 2014]. We have gradually simplified our model to its core[3]. Initially we had strengthening of claims as a primitive and we had a claim language with all kinds of bells and whistles. We reduced our claim language first to interpreted predicate logic

[3]A history of the simplifications is here
http://www.ccs.neu.edu/home/lieber/evergreen/specker/publications/scg-publs.html

sentences and then to game positions in extensive form games. A four page summary of some of Abdelmeged's dissertation's results, but with simpler concepts and simplified proofs, are published [Abdelmeged et al. 2016] as a AAAI 2016 workshop paper. But there is no journal or conference paper published on the topic. Our work has been steadily improved based on feedback from conference submissions, presentations (Harvard, ETH Zurich, Technical University Darmstadt). Our most recent discovery, now included in this paper, was the observation that our three axioms were not independent in the context of our scoring function. Now our starting point are only two axioms: NNEW and CR.

The rest of the paper is organized as follows. Section 3 describes pertinent related work. Section 4 introduces the main concepts, including an interpretation in terms of graphs. It introduces the axioms NNEW and CR. Section 5 provides a correspondence between our axioms and monotonicity constraints which provides the basis for many of our proofs. It introduces our main formal result, the Collusion-Resistance Theorem and several ramifications of it. It also introduces an equilibrium concept for SCGs. Section 6 discusses the broad applications of SCGs and section 7 mentions future work.

3. RELATED WORK

Our concept of a side-choosing game is very broad but has not been formally studied before. We were influenced by semantic games which have a long history in logic [Kulas and Hintikka 1983]. Falsifiability as promoted by Karl Popper and many others was another strong influence. A claim is falsifiable if there is an argument which proves the claim to be false. We use a weaker form of falsifiability which we call plausibility checking. A claim C for which player p_x is a proponent is plausibility checkable by a player p_y if there is an argument involving p_x and p_y that brings p_x into a contradiction with respect to C. The argument is an interactive "debate" (p_y winning a game against p_x) but it does not prove that the claim is false, in general.

In [Rubinstein 1980], Rubinstein provides an axiomatic treatment of tournament ranking functions that bears some resemblance to ours. Rubinstein's treatment was developed in a primitive framework where "beating functions" are restricted to complete, asymmetric relations. Rubinstein showed that the points system, in which only the winner is rewarded with a single point is *completely* characterized by the following three *natural* axioms:

— anonymity which means that the ranks are independent of the names of players,
— positive responsiveness to the winning relation which means that changing the results of a player p from a loss to a win, guarantees that p would have a better rank than all other players that used to have the same rank as p, and
— Independence of Irrelevant Matches (IIM) which means that the relative ranking of two players is independent of those matches in which neither is involved.

Our important CR axiom is at least as strong as Rubinstein's IIM.

Our work also falls into the field of axiomatic approaches to measures on graphs, an area active since the 1980s and inspired by earlier Social Choice theory [Gupte and Eliassi-Rad 2012; Rubinstein 1980]. Graphs are mapped to rankings of nodes or edges while axioms constrain the space of those mappings. The measures which satisfy all axioms are of interest. [Rubinstein 1980] is one of the first papers in this space. We compress edge information of the graph into a total ordering (ranking) of the graph nodes under the control of axioms that give us desirable properties. Starting from a list of axioms, which such a ranking must satisfy, we characterize functions that satisfy all the axioms (see section 5.6). We then show that there is a range of ranking functions that satisfy this characterization (see section 5.7).

Table I. $Legal(P)$ for a two players set

winner	loser	forced
1	2	1
1	2	2
1	2	0
2	1	1
2	1	2
2	1	0

[Simpson 2014] provides a comprehensive overview of techniques to "Combined Decision Making with Multiple Agents".Our work differs by working with multiple arguing or debating agents who have to defend their decisions. The concept of collusion is not mentioned in [Simpson 2014] while it is central to our analysis of SCGs.

While information and quality elicitation has been studied thoroughly for crowdsourcing mechanism designs, ranging from analyzing incentive design strategies for different platforms like Kaggle and MTurk [Easley and Ghosh 2015] to reasonable ranking of user-generated content (UCG) on Amazon or Quora [Ghosh 2012], there is less literature that addresses crowdsourcing designs which are based on the SCG mechanism. We think this new game offers for many use-cases useful benefits compared to "traditional" online crowdsourcing mechanisms, which justifies further studies and analysis.

This paper is based on Ahmed Abdelmeged's dissertation [Abdelmeged 2014]. The dissertation is based on semantic games to which side-choice was added, and does not explicitly define side-choosing games. However, the proof of the Collusion-Resistance Theorem does not rely on the details of semantic games. Therefore, we introduced side-choosing games in this paper to have an appropriate context for formulating and proving the Collusion-Resistance Theorem. The proofs in this paper have been simplified through the systematic use of monotonicity constraints.

4. MAIN THEORY

We discuss ranking the players based on an SCG-table T under the axiom of collusion-resistance. When collusion-resistance does not hold, there are SCG-tables T for which a meritorious player is not top-ranked. This will frustrate meritorious players and therefore we enforce the axiom of collusion-resistance which is based on the concept of control. A player is not in control in a game if she does not participate or she loses while forced. Note that if a player is forced we cannot blame her when she loses. Collusion-resistance is formalized by expressing that adding a row where player p_x is not in control, will keep the ranking of p_x with respect to other players p_y invariant. It turns out that collusion-resistance is linked to the concept of fault: a player makes a fault if she loses while not forced.

Our ranking approach prevents sybil attacks. In an online competition, several sybils might enter and help others to win thereby preventing the strong players to win. In the presence of collusion-resistance sybils have no effect on overpowering perfect players.

Let P be the set of all players. $Legal(P)$ is the set of legal game results for P. For example, if $P = \{1, 2\}$, then Table I gives the table of all possible game results for 2 players. $Legal(P)$ contains $n \cdot (n-1) \cdot 3$ rows where n is the number of players in P. Fig. 3 gives the same information as a labeled multi-graph. 0 is used to indicate that no one was forced. In the graph notation, if no one is forced, no * is used on the edge.

$R(P)$ is the multiset (or bag) of possible game results for P allowing for repetition in the game results: $R(P)$ is a pair $(Legal(P), m)$, where m is the multiplicity function $m : Legal(P) \to \mathbb{N}_{\geq 1}$.

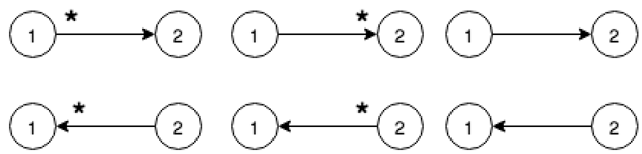

Fig. 3. Graph interpretation of Table I
(* = forced)

For our theory we define a few basic predicates: $\forall p_x \in P, \forall r \in R(P)$

$$participant(p_x, r) \Leftrightarrow p_x \text{ is a participant in the game } r \tag{1}$$
$$win(p_x, r) \Leftrightarrow p_x \text{ won the game } r \tag{2}$$
$$loss(p_x, r) \Leftrightarrow p_x \text{ lost the game } r \tag{3}$$
$$forced(p_x, r) \Leftrightarrow p_x \text{ is forced to choose a side in the game } r \tag{4}$$
$$control(p_x, r) \Leftrightarrow participant(p_x, r) \wedge \neg forced(p_x, r) \tag{5}$$
$$fault(p_x, r) \Leftrightarrow loss(p_x, r) \wedge \neg forced(p_x, r) \tag{6}$$

4.1. Graph Interpretation

Our theory can be visualized in terms of directed, labeled multi-graphs. The nodes are players and the labeled edges are game results. An edge points from winner to loser. The labels on the edges indicate who is forced. We call those graphs *SCG-graphs*. Our axiomatic approach to evaluation is based on counting different edge kinds incident with a node. It is based on local counts and ignores the structure of the SCG-graph. In Fig. 3 we give the graph of game results in Table I. In the future we plan to study such graphs using more holistic techniques that take paths in the graphs into account.

4.2. High-Level Ranking Axioms

We define a pre-order $\preceq^T$, called the weakly better relation, $\forall T \subseteq R(P)$. We want the ranking relation to satisfy the following axioms defined in terms of table extensions. We formulate the axioms dynamically in terms of game events. In principle, the axioms could be enforced dynamically each time a game result is produced but this would be very expensive (involving all pairs of players). We will propose an efficient technique to enforce the axioms through clever design of the ranking mechanism, requiring no dynamic checking at all.

— **NNEW**: Winning cannot lower your rank:

$$\forall p_x, p_y \in P, \forall T \subseteq R(P), \quad \forall r \in \{r \,|\, r \in R(P) \wedge win(p_x, r)\} \quad [p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y] \tag{7}$$

— **NPEL**: Losing cannot increase your rank:

$$\forall p_x, p_y \in P, \forall T \subseteq R(P), \quad \forall r \in \{r \,|\, r \in R(P) \wedge loss(p_y, r)\} \quad [p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y] \tag{8}$$

— **CR**: Games you don't control don't lower your rank.

$$\forall p_x, p_y \in P, \forall T \subseteq R(P),$$
$$\forall r \in \{r \,|\, r \in R(P) \wedge \neg control(p_x, r)\}$$
$$[\, p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y \,] \quad (9)$$

This axiom allows that games you don't control might improve your rank. Indeed, this flexibility is needed for interesting ranking functions. If we don't allow this flexibility, there are no interesting ranking functions in the class we study. For an explicit example of collusion and a ranking function which is not CR, see section 5.9.

It is beneficial to split the CR axiom into two more basic properties both for understanding CR and for proving implications of the ranking axioms. The Harmless Non-Participation property protects the players against the negative effects of non-participation. The Harmless Devil's Advocate property protects the forced players against disadvantages.

— **Harmless Non-Participation**: Games you don't participate in don't lower your rank.

$$\forall p_x, p_y \in P, \forall T \subseteq R(P),$$
$$\forall r \in \{r \,|\, r \in R(P) \wedge \neg participate(p_x, r)\}$$
$$[\, p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y \,] \quad (10)$$

— **Harmless Devil's Advocate**: Games where you are forced don't lower your rank.

$$\forall p_x, p_y \in P, \forall T \subseteq R(P),$$
$$\forall r \in \{r \,|\, r \in R(P) \wedge forced(p_x, r)\}$$
$$[\, p_x \preceq^T p_y \Rightarrow p_x \preceq^{T \cup \{r\}} p_y \,] \quad (11)$$

Fig. 5 summarizes the monotonicity constraints for the two new axioms.

4.3. Scoring Functions

Next we define natural scoring functions which we will use for ranking.

$$wf_{p_x}(T) = \text{the win count of } p_x \text{ in } T \text{ in a forced position}$$
$$wu_{p_x}(T) = \text{the win count of } p_x \text{ in } T \text{ in an unforced position}$$
$$lf_{p_x}(T) = \text{the loss count of } p_x \text{ in } T \text{ in a forced position}$$
$$lu_{p_x}(T) = \text{the loss count of } p_x \text{ in } T \text{ in an unforced position}$$

It's obvious that given table $T' = T \cup \{r\}$ and $X \in \{wf, wu, lf, lu\}$ the following transitional relations hold:

$$X_{p_x}(T') = \begin{cases} X_{p_x}(T) + 1 & \text{if } X \text{ happens in } \{r\} \\ X_{p_x}(T) & \text{otherwise} \end{cases}$$

4.4. Ranking Axioms With Scoring Functions

Now we define the weakly better relation using a scoring function $U : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$. Lower U means better player. We define a pre-order $\preceq^T_U$ below in (12). For convenience, we drop the subscript and refer to it simply as $\preceq^T$. We want to assign to each player a score solely based on the players' demonstration of ability. We use the above four counting functions, based on wins and losses and whether a player was forced, to calculate a player's score. We formally define the ranking relation as,

$$\forall p_x, p_y \in P, \forall T \subseteq R(P)[\, p_x \preceq^T p_y \Leftrightarrow U(wf_{p_x}(T), wu_{p_x}(T), wf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), wu_{p_y}(T), lu_{p_y}(T))\,] \quad (12)$$

A Venn diagram of the modified axioms (using U) is in Fig. 4. From now on, when we refer to an axiom, we mean the instantiation of the axiom for the above 4 argument scoring function U. For simplicity, we refer to the instantiated axiom Y still as Y although the axioms are different.

4.5. Discussion of Axioms

The axioms are formulated for general ranking functions however we focus on the special case where a "natural" four argument scoring function U is used (see 12). Using this restriction has surprising implications on the axioms. It even changes the implications relationships between the axioms. Consider the axiom CR which is formulated in terms of participation. Clearly, CR does not imply NPEL because NPEL does not even refer to participation. However, if we assume that ranking is done with the natural four-argument scoring function, CR now implies NPEL (see Fig. 4). There are many other scoring functions that could be used that are refinements of the U above. We found that they all give the same theory but there are more monotonicity constraints.

4.6. Universal Domain

From equation 12, it is clear that for every logically possible game result table T, we have a valid preorder. This implies that our ranking relation satisfies the Universal Domain property.

4.7. Anonymity

From equation 12 it is clear that the scoring function ignores the identity of the player in calculating the score. Hence, the ranking relation $\preceq^T$ is unaffected by changing labels and therefore anonymous.

4.8. Monotonicity of U

We score a player solely based on the player's wins and losses and whether forced or unforced. One interesting property of the parameters of U for a particular player is that when we add a new game to the existing game result table T, at most one parameter increments. This allows us to define the following notations:

$U \uparrow_x$: U is monotonically non-decreasing on parameter x

$U \downarrow_x$: U is monotonically non-increasing on parameter x

$U \uparrow\downarrow_x$: U is indifferent on the parameter x

The axioms NNEW and CR imply that U must be either argument-wise monotonically non-decreasing or non-increasing. If U would fluctuate on one of the arguments one of the axioms NNEW or CR would not hold (see Fig. 4). The monotonicity constraints are a tool to implement the axioms efficiently. Our plan is to combine the monotonicity constraints of the axioms and to find ranking functions which satisfy all of them.

5. TRANSLATING AXIOMS INTO MONOTONICITY CONSTRAINTS

We use the following reduction technique, called monotonicity constraint reduction, to prove properties of ranking relations: We map the predicate logic statements corresponding to the axioms into the space of monotonicity constraints of the scoring function U. We combine the monotonicity constraints and map the result back to predicate logic statements about the ranking relations. We show that this reduction is correct. Recognizing that functional monotonicities are hidden behind the properties has simplified our proofs. Compare with [Abdelmeged 2014]. In this section, we reformulate the axioms as equivalent monotonicity constraints.

5.1. Collusion-Resistance (CR)

Given $T' = T \cup \{r\}$, we reformulate CR as follows:

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T)) \Rightarrow U(wf_{p_x}(T'), wu_{p_x}(T'), lf_{p_x}(T'), lu_{p_x}(T')) \leq U(wf_{p_y}(T'), wu_{p_y}(T'), lf_{P_y}(T'), lu_{p_y}(T')) \quad (13)$$

Considering the definition of "not in control", we split CR into two separate cases: Harmless Non-participation and Harmless Devil's Advocate.

5.2. Harmless Non-Participation

Game results where p_x did not participate. Then p_y may have won or lost in a forced or unforced position against some third player p_z.

Let us consider the row $\{r\}$ where p_y wins over p_z in a forced position, given $T' = T \cup \{r\}$ and since:

$$U(wf_{p_x}(T'), wu_{p_x}(T'), lf_{p_x}(T'), lu_{p_x}(T')) = U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T))$$

$$U(wf_{p_y}(T'), wu_{p_y}(T'), lf_{P_y}(T'), lu_{p_y}(T')) = U(wf_{p_y}(T) + 1, wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T))$$

from 10 we have:

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T) + 1, wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T)) \quad (14)$$

From equations 12 and 14, we get the monotonicity constraint,

$$U \uparrow_{wf} \quad (15)$$

Similarly, let us consider the case $\{r\}$ where p_y wins over p_z in an unforced position, given $T' = T \cup \{r\}$ we have:

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T) + 1, lf_{p_y}(T), lu_{p_y}(T)) \quad (16)$$

From equations 12 and 16, we get the monotonicity constraint,

$$U \uparrow_{wu} \quad (17)$$

Using a similar argument, for the case where p_y loses over p_z in a forced position, we have

$$U \uparrow_{lf} \quad (18)$$

Also, for the case where p_y loses over p_z in an unforced position, we have

$$U \uparrow_{lu} \tag{19}$$

5.3. Harmless Devil's Advocate

We consider game results where p_x is forced. First, we consider game results $\{r\}$ where p_x was forced and lost against some third player p_z, given $T' = T \cup \{r\}$ and since:

$$U(wf_{p_x}(T'), wu_{p_x}(T'), lf_{p_x}(T'), lu_{p_x}(T')) = U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T) + 1, lu_{p_x}(T))$$

$$U(wf_{p_y}(T'), wu_{p_y}(T'), lf_{P_y}(T'), lu_{p_y}(T')) = U(wf_{p_y}(T), wu_{p_y}(T), lf_{P_y}(T), lu_{p_y}(T))$$

from 11 we have:

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T) + 1, lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), lf_{p_y}(T), lu_{p_y}(T)) \tag{20}$$

From equations 12 and 20, we get the monotonicity constraint,

$$U \downarrow_{lf} \tag{21}$$

Then, we consider the case where p_x was forced and won against some third player p_z. And similar to the analysis above, we shall have:

$$U \downarrow_{wf} \tag{22}$$

An observation of completeness shows that Harmless Devil's Advocate still holds when p_x plays exactly against p_y. This is because of:

$$U \uparrow\downarrow_{lf} \wedge U \uparrow_{wu} \Rightarrow (wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T) + 1, lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T) + 1, lf_{p_y}(T), lu_{p_y}(T)) \tag{23}$$

and

$$U \uparrow\downarrow_{wf} \wedge U \uparrow_{lu} \Rightarrow U(wf_{p_x}(T) + 1, wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), lf_{p_y}(T), lu_{p_y}(T) + 1) \tag{24}$$

Now, CR can be summarized in terms of monotonicity constraints as,

$$U \uparrow\downarrow_{wf} \wedge U \uparrow_{wu} \wedge U \uparrow\downarrow_{lf} \wedge U \uparrow_{lu} \tag{25}$$

5.4. Non Negative Effect of Winning (NNEW)

Let us consider a game result $\{r\}$ where p_x won against a third player p_z. p_x could have won either in a forced or unforced position.

First, considering the case where p_x wins over p_z in a forced position, we have,

$$U(wf_{p_x}(T) + 1, wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), lf_{p_y}(T), lu_{p_y}(T)) \tag{26}$$

From equations 12 and 26, we get the monotonicity constraint,

$$U \downarrow_{wf} \tag{27}$$

Similarly, for the case where p_x wins over p_z in an unforced position, we have

$$U \downarrow_{wu} \tag{28}$$

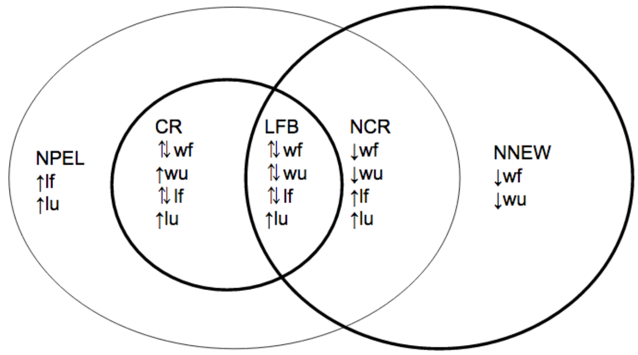

Fig. 4. Relations among NNEW, NPEL, CR. Their monotonicity constraints cover entire oval.

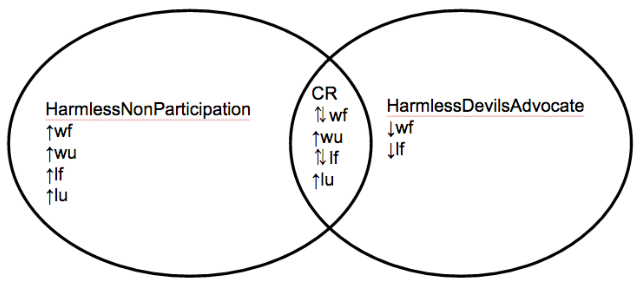

Fig. 5. CR split into two parts

Summarizing the monotonicity constraints, we have,

$$U \downarrow_{wf} \wedge U \downarrow_{wu} \tag{29}$$

5.5. Non Positive Effect of Losing (NPEL)

Let us consider a game result $\{r\}$ where p_y lost against a third player p_z.

First, considering the case where p_y loses over p_z in a forced position, we have,

$$U(wf_{p_x}(T), wu_{p_x}(T), lf_{p_x}(T), lu_{p_x}(T)) \leq U(wf_{p_y}(T), wu_{p_y}(T), lf_{p_y}(T) + 1, lu_{p_y}(T)) \tag{30}$$

From equations 12 and 30, we get the monotonicity constraint, $U \uparrow_{lf}$. Similarly, for the case where p_y loses over p_z in an unforced position, we have $U \uparrow_{lu}$. Summarizing the monotonicity constraints, we have,

$$U \uparrow_{lf} \wedge U \uparrow_{lu} \tag{31}$$

5.6. Local Fault Based (LFB)

As we want the ranking relation to satisfy all the three properties NNEW, NPEL and CR, from equations 25, 29 and 31, we get the monotonicity constraints,

$$U \uparrow\downarrow_{wf} \wedge U \uparrow\downarrow_{wu} \wedge U \uparrow\downarrow_{lf} \wedge U \uparrow_{lu} \tag{32}$$

This tells us that the scoring function should be monotonically non-decreasing on faults and indifferent on other parameters. We call the ranking relation that uses a scoring function that satisfies equation 32 as Local Fault Based (LFB). The monotonicity constraints in equation 32 can be easily reformulated in predicate logic.

— LFB: Games in which you don't make faults don't lower your rank.

$$\forall p_x, p_y \in P, \forall T \subseteq R(P), \quad \forall r \in \{r \,|\, r \in R(P) \wedge \neg fault(p_x, r)\} \quad [\, p_x \preceq_U^T p_y \Leftrightarrow p_x \preceq_U^{T \cup \{r\}} p_y \,] \tag{33}$$

— **Collusion-Resistance Theorem** We just proved for the instantiation of the axioms for the 4 argument scoring function U in 12:

$$(\text{NNEW} \wedge \text{CR}) = \text{LFB}$$

The following proper subset relationships can be shown using the same proof techniques (see Fig. 4):

$$\text{CR} \subset \text{NPEL}, \text{LFB} \subset (\text{NNEW} \cap \text{NPEL}), \text{LFB} \subset \text{CR}$$

The Collusion-Resistance Theorem tells us that collusion-resistant ranking functions have a simple form based on fault counting. There is an infinite family of such functions that can be used in the design of techno-social systems with guaranteed collusion resistance (see section 5.7).

5.7. A Family of Collusion-Resistant Rankings

When designing a techno-social system for solving precisely formulated problems there are many concerns to be addressed. Besides just using simple fault-counting, there are other weighted fault counting functions of interest. In the graph representation of an SCG-table T, we consider two kinds of edges: α edges going into p_x are edges where no one was forced. β edges going into p_x are edges where the winner was forced. In both cases p_x made a fault but we are counting the two kinds of edges differently. α (β) edges have weight $\alpha > 0$ ($\beta > 0$), respectively. The resulting scoring function U has the property that a high α (compared to β) encourages non-forced players to win. Other families of collusion-resistant rankings can be defined by considering finer-grained properties of game results.

5.8. A Simple Property of Fault Counting

We consider the ranking we get from the scoring function U which counts faults in a table T ($lu_{p_x}(T)$). A *quasi-perfect player* p_x is a player with zero fault counts ($lu_{p_x}(T) = 0$). (A perfect player is quasi-perfect but the converse does not necessarily hold because a quasi-perfect player may choose the wrong side of a claim and still successfully defend it because of weakness in the opponents.) A *top-ranked* player is a player for which there exists no stronger player in the ranking. We have the simple but desirable meritocracy property: for all SCG-Tables quasi-perfect implies top-ranked under fault-counting. This easily generalizes to: When a ranking is LFB then for all SCG-tables, quasi-perfect players are top ranked. Next we give an explicit counterexample for win counting.

5.9. Counterexamples for Win Counting

We assume that players can recognize each other and use that knowledge to alter their play. This assumption is satisfied in most applications even when the players are implemented in software.

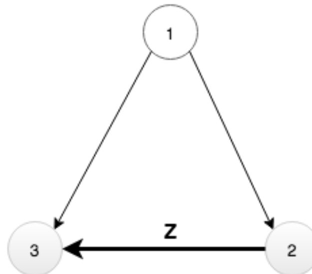

Fig. 6. Simplest Counterexample: Players 2 and 3 collude

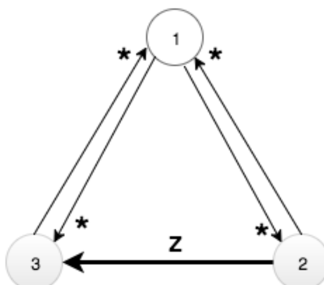

Table II. Game results for Fig. 7

player	win	loss	fault
1	2	2	0
2	z+1	1	0
3	1	z+1	z

Fig. 7. Player 3 lies about its strength

Under win counting, quasi-perfect players are not necessarily top-ranked. Win counting is defined by

$$U(p_x, T) = -(wf_{p_x}(T) + wu_{p_x}(T)) \tag{34}$$

The corresponding ranking function is NNEW and NPEL but ¬ CR. Fig. 6 gives the smallest counterexample (both in terms of number of players and number of game results).

We set $z = 3$. Player 2 is top ranked with the 3 wins and 1 fault. But, player 1, the quasi perfect player with no faults is not top ranked. The reason is that there is collusion: player 3 helped player 2 accumulate wins which helped to overrule the quasi-perfect player.

5.9.1. All Perfect with Liars. How many colluding players are needed to prevent a perfect player from winning under win counting? We show with an example that 2 out of n are enough provided the 2 players play enough games.

We have $n - 2$ perfect players and a total of n players. The tournament is basically a full round-robin tournament where the non-forced player always wins, except for the pair of colluding players. One player (the liar) helps the other player to accumulate wins. The helping games have multiplicity z. Fig. 7 shows the graph of game results for $n = 3$ with players numbered 1,2,3. Player 1 is perfect, 3 is the liar (lying about its strength) and 2 is being helped. Players 2 and 3 collude: player 3 (the liar) helps player 2 win points. Although player 3 is also perfect, it lies about its strength when it plays against player 2. Table II shows the game statistics: For $z \geq 2$ only fault-counting is collusion-resistant and player 1 is top-ranked.

5.10. Equilibria

Let Q be a subset of a set of players P. Q is in an equilibrium if games between players in Q are fault-free. This means that players in P agree on a construction for defending

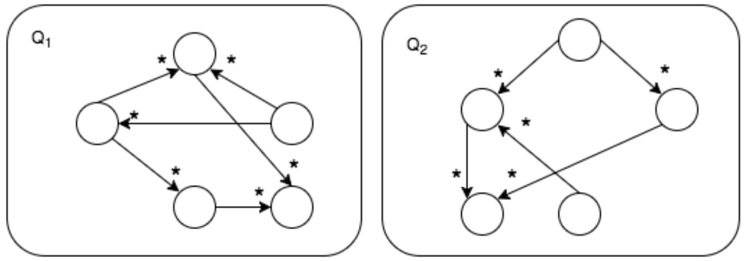

Fig. 8. Equilibrium of two disjoint subsets of P

the claim. They always win when they are not forced and they always lose when they are forced. There might be several islands of agreement represented by disjoint subsets of P. Let Q_1 and Q_2 be two disjoint subsets of P, each being in an equilibrium (see Fig. 8). What happens when a player in Q_1 plays against a player in Q_2? If the construction used in Q_2 is of "higher quality," the player in Q_2 will win when not forced. The quality of construction used by a player p_x might increase over time. p_x might have an insight which leads to a better solution shaking up an equilibrium. There will be again games leading to faults until the island of player p_x has learned about the new construction. This is a model of how knowledge in a formal science community might evolve.

5.11. Independence of axioms

The two axioms NNEW and CR are independent as demonstrated by two examples of ranking functions. We choose any of the axioms and show that there is a ranking function that satisfies that axiom but not the other. A ranking function that is (1) NNEW and not CR is win counting; (2) CR and not NNEW is win counting for unforced players. The reader can confirm this by checking with Fig. 4.

6. SCG APPLICATIONS

After the introductory examples in section 1.1 and the theory, we outline now the breadth of applicability of SCGs. We motivate the importance of SCGs by describing applications, users and owners (also known as principals).

— **SCG Applications.**
  — **Formal Sciences** Formal sciences are disciplines concerned with formal systems, such as logic, mathematics, statistics, theoretical computer science, information theory, game theory, systems theory, and decision theory. A claim is defined using an *interpreted* predicate logic sentence. This is *not* an exercise in logic as the quantifiers are only used to define the tasks that the users must perform. The sentence is interpreted in a structure which might be defined by a complex program encoding the functionality best executed by computers.
  — **Formal Claims based on Simulation Environments** Robotics and biological sciences, etc. fall into this category. The structure in which the claim is interpreted is the simulation environment.

— **SCG Users.** Users are problem solvers or learners and they operate directly or indirectly. In **direct mode**, the users perform the moves themselves, maybe using software. In **indirect mode**, the users produce software that plays the SCG on their behalf. There is a simple SCG-interface that the software has to follow. Of course, indirect users must have software development skills.
The indirect mode is of central interest to us because it is a novel approach to develop software for computational problems using a group of people. The quality control of

the software is automated by running an online or offline tournament to determine the top-ranked software. The claim under consideration determines what quality means. Note that the SCG-interface implies that testing is an integral part of the solution.
Users of SCGs include:
  - **Students in high schools and universities.** They must understand the concept of a claim. Focus is on dissemination of knowledge through peer teaching and peer evaluation.
  - **Researchers.** Focus is on creation of new knowledge and its peer evaluation. Researchers propose claim variations.
  - **Citizen Scientists.** They might find innovative constructions that are imperfect. Experts might benefit from those ideas and correct them.
- **SCG Owners.**
Owners define claims. Some users also play the role of owners. Owners don't need expertise how to solve the problems.
Owners include: (1) Teachers and Professors. (2) Research Directors, Heads of Research Programs, Organizations like NSF, DARPA, ONR etc. (3) Program Chairs of conferences and Journal Editors. (4) Companies who need a specific computational problem solved for which no off-the-shelve solution is available. (5) Companies who are looking for employees with skills in a specific domain. E.g., Facebook organized a competition on kaggle.com and the winner got a Facebook job.

6.1. Applications of Side-Choosing Games to Existing Systems

Our study of side-choosing games is motivated by their potential to organize problem-solving competitions and by their successful use in CS education at Northeastern University. We believe SCGs are a foundation for platforms like TopCoder or Kaggle or for scientific human-computation tools like Foldit [Cooper et al. 2010].

- **Education in Formal Sciences.** Our favorite way of summarizing learning objectives for a formal science domain is to say that learners must demonstrate the skill of judging claims in the domain, choosing their side on the claim and then defending their side choice through game play against other students. The resulting peer-teaching and peer-grading is very attractive. A claim is representing a lab in which students learn and is chosen in such a way that solving the problem requires skills that students should have.
  - **Using piazza.com.** To post claims and to organize the playing of games related to those claims we used piazza.com. This worked very well, especially when we divided the Algorithms class into small groups of three students and kept the games in those small groups. The undergraduate students solved challenging problems like finding the worst-case input for the Gale-Shapley algorithm or optimally solving a product stress testing problem (see 6.1.1).
  - **Using our own software.** In software development classes we had the students develop "avatars" to play the game and we did a full-round-robin tournament evaluation of the avatars. The problem to be solved was a maximum constraint satisfaction problem (see 6.1.2).
- **Improving Evaluation in Problem-Solving Competitions for Computational Problems.** A significant advantage of our approach is that the evaluation of solutions is done by peers and not the competition organizer. This is relevant to systems like topcoder.com and various competitions like SAT-solver competitions. The competition organizer only acts in a role as referee. Instead of static benchmarks, dynamic benchmarks are developed through game play.
The quality of the solutions produced depends on the skills of the players who might not be motivated or not have the knowledge necessary to solve the problem.

To attract strong players either money or fame has to be given; a common theme in human computation.

6.1.1. Gale-Shapley Lab. We present an example from our Algorithms class. The students have studied the Gale-Shapley algorithm for producing a stable matching of n women with n men given their preferences. To get a better understanding of how the algorithm works (it is a loop), the students have to find for a given n a set of preferences which create the most number q of iterations of the algorithm. The claim $GSW = GaleShapleyWorstCaseClaim(n = 10, q = 30)$ says that for 10 women and men there is a set of preferences generating 30 iterations of the outer loop of the Gale-Shapley algorithm. And the claim is also that it is not possible to have more iterations with other preferences. The predicate logic representation of GSW automatically produces the following game between a P(roponent) and O(pponent): P produces an input $i(n)$ of preferences for n women and men. The algorithm is run on $i(n)$ and produces $q(n)$ iterations. If $q(n) < 30$, P has made a fault. If $q(n)$ is too small, O produces input $i_1(n)$ which is run and produces $q_1(n)$ iterations. If $q_1(n) > 30$, P has also made a fault. This is the essence of the semantic game behind the predicate logic formula specifying the problem.

6.1.2. Approximate MaxCSP Lab. We present a simple example of an algorithm development lab. We are interested in algorithms for approximately solving MaxCSP instances with guaranteed performance. We are considering Boolean constraint satisfaction problems of the following form: Each constraint is of the form $R(x_1, x_2, x_3)$ which is true when exactly one of the three Boolean variables is true. Given a CSP formula consisting of n variables we are interested in finding an assignment that satisfies the fraction τ_R of the constraints and we want to maximize τ_R. It turns out that $\tau_R = 4/9$. The SCG behind this problem has to deliver counterexamples (where the fraction t cannot be satisfied) if $t > \tau_R$ and to produce an assignment where the fraction t is satisfied, if $t \leq \tau_R$. Notice that in this context the algorithm designer needs not only to provide an algorithm which satisfies the required fraction of constraints but she also needs an algorithm that can produce ”hard” inputs.

6.2. Incentive and Trust

There are two kinds of incentives in SCG: the incentive (1) to be top-ranked which brings money or fame and (2) to get feedback during game play which builds skills and provides opportunity for learning. Incentive (2) suggests productive applications of SCGs in education.
Trust in the SCG approach is related to the belief that good work as a player (problem solver) will get rewarded and that it is not possible to be top-ranked without doing good work. There should be no sneaky ways to game the system. Trust can be broken in at least two ways: (1) by defining games but not checking that all game rules are perfectly followed and (2) by having tournaments and evaluations where you can succeed without hard work. This paper is addressing point (2). Point (1) is addressed by having reliable software to check the game rules related to the claims.

7. FUTURE WORK

The work in this paper abstracts away from who is proponent and who is opponent of a claim in a game. When the proponent/opponent information is considered we have a richer labeling structure on the edges of the SCG-graphs. Each edge gets a pair of static and dynamic labels where the dynamic labels are determined by the static labels plus the forcing information. Recall from the introduction that static labels provide the side-choices $\{P(roponent), O(pponent)\}$ and the dynamic labels provide the roles used when the game is played. We call those graphs *extended* SCG-graphs.

A player is called *consistent*, if it always uses the same static side-choice across all games. We plan to prove the following Collusion-Inconsistency Conjecture: For all ranking functions R (which are not LFB) and for all extended SCG graphs where there is a quasi-perfect player that is not top-ranked under R, there exists a player that is not consistent. This conjecture would prove that non-collusion-resistance implies inconsistency.

We call an SCG-graph *consistent* if it has a completion to an extended SCG-graph where all players are consistent. The SCG-graph in Fig. 6 is inconsistent because of the odd cycle and the fact that none of the players is forced. The SCG-graph in Fig. 7 is inconsistent too. We conjecture that the SCG-graph consistency problem is solvable in polynomial time. Note that a mapping from nodes to $\{P(roponent), O(pponent)\}$ that is compatible with the SCG-graph, serves as a witness for SCG-graph consistency. *Compatibility* of a node mapping is defined in terms of the forced labels: when the two nodes incident with an edge have the same value under the map then exactly one of the two nodes must be forced and if they have different values then none of the two nodes must be forced.

We want to study SCGs with imperfect information and with random moves. Independence-friendly logic and the corresponding semantic games are a good starting point.

An interesting question is what can be said about the truth value of a claim given an SCG-table of game results and information about the strength of the players.

Collusion is linked to trust in an SCG-Table T to find the best players. Collusion-resistance eliminates some collusion but there is still other collusion possible. To explore the link between trust and collusion is interesting future work. Trust can be improved by controlling the game scheduler to enforce Swiss-style scheduling, for example.

8. CONCLUSION

We propose the concept of side-choosing Game (SCG) as a generalization of extensive form games. SCGs are useful for organizing techno-social systems for problem solving in Formal Sciences. Considering that a specific kind of collusion might compromise the truth, we modeled the ranking of players functionally via two axioms or postulates: NNEW (Non-Negative Effect for Winning), and the crucial axiom CR (Collusion-resistance, which says that games where one is not in control cannot lower ones ranking, hence preventing gaming the game). We prove the Collusion-Resistance Theorem which states that ranking has to be based on fault counting.

What comes next? Our plan is to deploy SCG-based applications on the web and gather the benefits of collective intelligence. So far, we have already applied SCG-based ideas and tools in designing courses at Northeastern University from algorithm and software development courses to basic courses on spreadsheets and databases. And we were planning to build a tool that can be used in MOOCs or algorithm competitions. An implementation of a domain-specific language for human computation in formal sciences is a challenge that requires several algorithms to be developed. Why not develop those algorithms with SCG-based human computation effectively bootstrapping the system based on user feedback. We view SCG as the programming language for human computation to solve complex problems.

Another important area that needs further work is where players can propose new claims. A modular approach to solving claims is needed. For example, a complex claim C_1 might be reducible to a simpler claim C_2 so that a solution for C_2 implies a solution for C_1. We propose a formal study of claim relations which can themselves be captured as claims and approached with side-choosing games.

Acknowledgments We would like to thank Thomas Wahl and Ravi Sundaram for their influential feedback on earlier versions of the paper.

REFERENCES

Ahmed Abdelmeged. 2014. *Organizing Computational Problem Solving Communities via Collusion-Resistant Semantic Game Tournaments*. Ph.D. Dissertation. Northeastern University, Boston, MA, USA. Advisor-Karl Lieberherr.

Ahmed Abdelmeged, Ruiyang Xu, and Karl Lieberherr. 2016. Design and Secure Evaluation of Side-Choosing Games. In *Proceedings of AAAI 2016 Workshop on Incentive and Trust in E-Communities (WIT-EC'16)*.

Seth Cooper, Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite, Firas Khatib, Alex Cho Snyder, Michael Beenen, David Salesin, David Baker, and Zoran Popović. 2010. The challenge of designing scientific discovery games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games (FDG '10)*. ACM, New York, NY, USA, 40–47. DOI: http://dx.doi.org/10.1145/1822348.1822354

David Easley and Arpita Ghosh. 2015. Behavioral mechanism design: optimal crowdsourcing contracts and prospect theory. In *ACM SIGecom Exchanges (v.14 n.1)*. 89–94.

Arpita Ghosh. 2012. Social computing and user-generated content: a game-theoretic approach. In *ACM SIGecom Exchanges (v.11 n.2)*. 16–21.

Mangesh Gupte and Tina Eliassi-Rad. 2012. Measuring tie strength in implicit social networks. *Proceedings of the 4th Annual ACM Web Science Conference* (2012), pp. 109–118.

J. Kulas and J. Hintikka. 1983. *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications*. Springer. http://books.google.com/books?id=6GphI2_3u-sC

Ariel Rubinstein. 1980. Ranking the Participants in a Tournament. *SIAM J. Appl. Math.* 38, 1 (1980), pp. 108–111. http://www.jstor.org/stable/2100804

Edwin Simpson. 2014. *Combined Decision Making with Multiple Agents*. Ph.D. Dissertation. University of Oxford.