1J1-3

匿名操作不可能シャプレイ値: 開環境での協力ゲームにおける効率的に表記／求解可能な解概念

Anonymity-proof Shapley Value:
Extending Shapley Value for Coalitional Games in Open Environments

大田 直樹[*1] 佐藤 恭史[*1] 岩崎 敦[*1] 横尾 真[*1] Vincent Conitzer[*2]
Naoki Ohta Yasufumi Satoh Atsushi Iwasaki Makoto Yokoo Vincent Conitzer

[*1]九州大学大学院システム情報科学府
Kyushu University Graduate School of ISEE

[*2]Duke University, Department of Computer Science
Duke University, Department of Computer Science

Coalition formation is an important capability for automated negotiation among self-interested agents. In order for coalitions to be stable, a key question that must be answered is how the gains from cooperation are to be distributed. Coalitional game theory provides many solution concepts for this. However, these traditional solution concepts are vulnerable to various manipulations in open anonymous environments such as the Internet. To address this, we developed several solution concepts that are robust against such manipulations. However, the required computational and representational costs of these anonymity-proof solution concepts are huge.

In this paper, we develop a new solution concept which we call the anonymity-proof Shapley value. We show that the anonymity-proof Shapley value is characterized by certain simple axiomatic conditions, always exists, and is uniquely determined. The computational and representational costs of the anonymity-proof Shapley value are drastically smaller than those of existing anonymity-proof solution concepts.

1. はじめに

他のエージェントと交渉し提携を組むことは自律的なエージェントの備えるべき重要な性質である．エージェントが安定した提携を組むためには，協力によって得られた利得をエージェント間でいかに分配するかという問題を解決する必要がある．この問題を解決する方法の一つとして，協力ゲーム理論を利用することが考えられる．協力ゲーム理論は，複数のエージェントが協力した際，協力の結果得た利得の分配方法に関する研究分野であり，解概念と呼ばれる利得の分配方法が提案されている．これらの解概念はマルチエージェントシステムの研究において利用されてきている [Agotnes 07, Conitzer 03, Shehory 98]．

さらに近年，インターネットの普及により，複数の企業，組織が動的，迅速に提携を構成することが可能／必要となったことから，協力ゲーム理論の適用分野は今後さらに拡大していくことが予想される．例えば，複数の小さなベンチャー企業が提携し，仮想的な組織 (バーチャルカンパニー) を動的に構成することで，より大規模かつ多様な顧客の要望に応えて利益を拡大することが可能である．このような場合，複数の小さな企業間で，提携によって得られる利益の分配方法が課題であり，協力ゲーム理論の研究成果の適用が期待される．

しかしながら，著者らは，シャプレイ値やコアといった伝統的な解概念は，インターネットのような匿名の開環境下において，エージェントが行うことが可能な不正操作に対し，耐性を持たず，そのためこれらの不正行為の影響を受けない匿名操作不可能な解概念（anonymity-proof outcome function）が必要であることを指摘しており，匿名操作不可能な解概念として匿名操作不可能コア等の解概念を提案している [Yokoo 05]．しかしながら，匿名操作不可能な解概念の表記量は，ゲームに参加するエージェントの数に対し指数的に増加する．また配分を決定するためには線形計画問題を解く必要があり，この線形計画法の変数の数も，やはりゲームに参加するエージェントの数に対し指数的に増加する．

そこで本論文では，新しい解概念である，匿名操作不可能シャプレイ値（anonymity-proof Shapley value, AP Shapley value）を提案する．匿名操作不可能シャプレイ値は協力ゲームにおいて，最も重要な解概念の一つであるシャプレイ値の公理系を，匿名操作不可能性を満足するために必要最小限に緩和した新しい公理系を満す唯一の解概念であり，計算量／表記量が匿名操作不可能コアに比べ極めて小さい．

連絡先: 大田 直樹，九州大学大学院 システム情報科学府 知能システム学専攻，福岡県福岡市西区元岡 744 九州大学ウエスト 2 号館 824 号室，電話;092-802-2999（内線 7932），Fax;092-802-3576，e-mail;ohta@agent.is.kyushu-u.ac.jp

2. モデル

従来の協力ゲームでは，特性関数をもとに利得の配分を決定してきた．特性関数 w とは，任意の提携（ゲームに参加しうるエージェントの集合）X を引数とする関数であり，$w(X)$ は提携 X に属するエージェントが協力して得る利得を示す．

一方で，開環境での協力ゲームでは，エージェントがもつ能力をより詳細に記述する必要があるため，エージェントの持つスキルという概念を導入する [Yokoo 05]．これにより，開環境においてエージェントが可能な操作を明確に定義できる[*1]．

定義 1（スキルの特性関数） スキルの特性関数 $v : 2^T \rightarrow \Re$（T はスキルの全体集合）は，任意のスキルの集合 S に対し，S を持つエージェントが協力した際に得る利得 $v(S)$ を与える．

このスキルの特性関数 v は，エージェントの特性関数 w よりも詳細な情報を含んでいる．例えば，エージェント i が持つスキルの集合を S_i とする．この時，任意の提携 X について，$S_X = \bigcup_{i \in X} S_i$ とする時，$w(X) = v(S_X)$ が成立し，v から w を求めることができるが，逆は不可能である．また，特性関数 v は 0-正規化され，かつ単調性を満たしている．すなわち，単独のスキルに関する特性関数の値は 0 であり，あるスキル

[*1] ここでは，“スキル” という用語を，エージェントの持つ技能という意味だけでなく，エージェントの所有する資源等も含めた広い意味で用いている．

の集合に対し，マイナスの貢献をするスキルは存在しないと仮定する．

各エージェントへの利得の配分は以下のようになされる; まずメカニズムデザイナと呼ばれる特別なエージェントの存在を仮定する．このメカニズムデザイナは参加する可能性のある全てのスキルの集合 T および T に関して定義された特性関数 v を知っている*2．その上でエージェント i は提携に参加する意思がある場合に，自らの持つスキルをメカニズムデザイナに申告し，メカニズムデザイナは申告された情報をもとに参加者間での利得の分配方法を決定する．

匿名の開環境下において，任意のエージェントは，以下に示すスキルの隠蔽，架空名義の利用，共謀の操作及びこれらを組み合わせた操作を利用できると仮定する [Yokoo 05].

スキルの隠蔽： スキルの集合 S_i を所持するエージェント i は，一部のスキルを隠蔽し，$S'_i \subseteq S_i$ しか持たないように振舞うことができる．

架空名義： スキルの集合 S_i を所持するエージェント i は，複数の名義を用いて複数のエージェントのように振舞うことができる．この時，各エージェントの持つスキルは S_i の互いに素な部分集合となる．

共謀： 任意のエージェントの集団は，一人のエージェントとして振舞うことができる．この時，この一人のエージェントの持つスキルの集合は，共謀するエージェントが持つスキルの和集合となる．

これらの操作のうち，架空名義と共謀は，以下の利得の分配方法を用いることで防止できる：まずメカニズムデザイナは，各エージェントの申告したスキルの和集合を S とした場合，スキル s が受け取る非負の利得 $\pi(s, S)$（以下 π を利得関数と呼ぶ）を決定する．この際に，どのエージェントが s を所有しているかという情報は利用しない．エージェントが受け取る利得は，利得関数をもとに決定する．具体的にはスキルの集合 S_i を持つエージェント i には $\sum_{s \in S_i} \pi(s, S)$ だけ利得を与える．

上記の方法を用いた場合，スキルに配分される利得は，そのスキルをどのエージェントが保有しているかには依存しないため，架空名義／共謀は無意味となる．よって，利得関数が以下に示すようにスキルの隠蔽の影響を受けなければ，この利得関数によって表現される解概念は匿名操作不可能性を満足する．

定義 2 (匿名操作不可能な利得関数) 以下の条件を満たす利得関数 π は匿名操作不可能である：

$$\forall S_i, \forall S'_i, \forall S,\ S'_i \subseteq S_i, S_i \cap S = \emptyset,$$

$$\sum_{s \in S'_i} \pi(s, S'_i \cup S) \leq \sum_{s \in S_i} \pi(s, S_i \cup S).$$

この定義は，あるエージェント i がスキル S_i を持ち，その他のエージェントがスキル S を持つとき，エージェント i がスキルを S'_i しか持たないと申告しても利得が増加しないことを意味している．この匿名操作不可能な利得関数の条件を満たす解概念（以下，匿名操作不可能な解概念）として，匿名操作不可能コア等が提案されている [Yokoo 05].

*2 メカニズムデザイナは，実際にエージェントが所有しているスキルの集合を正しく知っている必要はなく，エージェントが所有するスキルがスキルの全体集合 T に含まれていることのみが必要である．すなわち，メカニズムデザイナはエージェントが保有する可能性のあるスキルの上界を知っていることを仮定する．

注意すべきこととして，利得関数 π の第二引数の値域は T の任意の部分集合であり，$n = \mid T \mid$, すなわち参加しうるスキルの数を n とした時，T の部分集合の個数は 2^n となることがある．メカニズムデザイナは，実際にはどのスキルが参加しているかを正確に知ることはできないため，生じうるすべての場合に関して，あらかじめ利得関数を準備しておく必要がある．このため，従来提案されている匿名操作不可能な解概念には，表記量／計算量が膨大であるという問題点があった．

本論文では，この問題点を解決する新しい解概念である匿名操作不可能シャプレイ値を提案する．匿名操作不可能シャプレイ値の特徴は，T に関するシャプレイ値から，$S \subset T$ に対する利得関数の値を on-demand で計算可能である点であり，この特徴により，表記量／計算量を大幅に削減可能となっている．

3. シャプレイ値

本章では匿名操作不可能シャプレイ値のもととなる解概念であるシャプレイ値の定義とその性質について述べる．

定義 3 (シャプレイ値) スキルの全体集合 T に対してある順列 o を与えたとき，$S(o, s)$ を順列 o において，s に先行するスキルの集合を示すものとする．特性関数が v で与えられるとき，シャプレイ値は以下で定義される．

$$sh_v(s, T) = \frac{1}{|T|!}(\sum_o v(S(o, s) \cup \{s\}) - v(S(o, s))).$$

あるスキルの集合に，スキル s が加わることによる特性関数の増加値をスキルの限界効用と呼ぶ．スキル s のシャプレイ値は，すべての順列に関して，スキル s の限界効用の平均値となっている．実際にシャプレイ値を算出する例を示す．

例 1 参加しうるスキルの集合 $T = \{a, b, c\}$，以下のように定義されるスキルの特性関数 v をもつゲームを考える．

- $v(\{a, b, c\}) = v(\{a, b\}) = v(\{a, c\}) = 300$,
- その他の部分集合 S について，$v(S) = 0$.

このゲームのシャプレイ値 $sh_v(s, T)$ は以下のように定まる．

- スキル a は順列 $(b, c, a), (c, b, a), (b, a, c), (c, a, b)$ についてそれぞれ限界効用 300 を得るので，シャプレイ値は $sh_v(a, T) = \frac{300*4}{3!} = 200$ となる．
- スキル b, c のシャプレイ値は同じように計算することで $sh_v(b, T) = sh_v(c, T) = \frac{300}{3!} = 50$ となる．

シャプレイ値はパレート効率性，ヌルスキル公理，対称性，加法性を同時に満たす唯一の解概念である [Shapley. 53]．しかしながら，シャプレイ値は匿名操作不可能な解概念ではない [Yokoo 05]．以下にシャプレイ値が匿名の開環境における不正操作のうち，スキルの隠蔽に脆弱である例を示す．

例 2 例 *1* で示したゲームについて考える．この時スキルの集合 $\{b, c\}$ を所持する参加者がスキル c を隠蔽したと仮定する．するとこのゲームは，参加しうるスキルの集合 $T' = \{a, b\}$，$v'(\{a, b\}) = 300$，その他の部分集合 S について，$v'(S) = 0$ と定義される特性関数 v' をもつゲームへと変化する．この協力ゲームのシャプレイ値 $sh_{v'}(s, T')$ は $sh_{v'}(a, T') = sh_{v'}(b, T') = 150$ と決定される．よってスキルの集合 $\{b, c\}$ を所持する参加者が存在した場合，その参加者はスキル c を隠蔽することで利得を 100 から 150 へと不正に増加できる．

このようにシャプレイ値は匿名の開環境に適用できない．次の章ではシャプレイ値をベースに定義した，匿名操作不可能な解概念である匿名操作不可能シャプレイ値を定義する．

4. 匿名操作不可能シャプレイ値

本章では本論文で我々が提案する解概念である匿名操作不可能シャプレイ値の詳細を示す．

定義 4 (匿名操作不可能シャプレイ値) $sh_v(s,T)$ を，特性関数 v とスキルの全体集合 T で構成されるゲームにおけるスキル s のシャプレイ値とする．匿名操作不可能シャプレイ値は $s \in S \subseteq T$ を満たす任意の s, S について以下の条件を満たす利得関数 π である．

$$\pi(s,S) = \begin{cases} v(S) \cdot \frac{sh_v(s,T)}{\sum_{t \in S} sh_v(t,T)} & \text{if } v(S) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

すなわち，匿名操作不可能シャプレイ値は，T では通常のシャプレイ値と一致し，$S \subset T$ では，T でのシャプレイ値に基づいて，特性関数の値 $v(S)$ を比例配分したものとなっている．

実際に匿名操作不可能シャプレイ値を求めた例を示す

例 3 例 *1*で用いた協力ゲームについて考える．このゲームのシャプレイ値は，$sh_v(a,T) = 200,\ sh_v(b,T) = sh_v(c,T) = 50,$ であるため，このゲームの匿名操作不可能シャプレイ値 π は以下のように決定される．

- $\pi(a, \{a,b,c\}) = 200,$
- $\pi(b, \{a,b,c\}) = \pi(c, \{a,b,c\}) = 50,$
- $\pi(a, \{a,b\}) = \pi(a, \{a,c\}) = 240,$
- $\pi(b, \{a,b\}) = \pi(c, \{a,c\}) = 60.$
- $\{a,b,c\}, \{a,b\}, \{a,c\}$ 以外の部分集合 S と，$s \in S$ を満たすスキル s について，$\pi(s,S) = 0$

他のスキルが a の場合，スキル b, c を所持する参加者は，スキル b, c を所持していると申告した場合の利得は 100 であるのに対し，スキル c を隠蔽した場合の利得は 60 となり，利得の隠蔽の効果はない．実際，$S \subset T$ で各スキルに与えられる値は，$v(S)$ を T でのシャプレイ値で比例配分したものであるため，一般にスキルの隠蔽の効果がないことは明らかである．

匿名操作不可能シャプレイ値は T におけるシャプレイ値のみを算出／保持しておけば，任意の匿名操作不可能シャプレイ値を計算量 $O(n)$ で算出できる．よって，匿名操作不可能シャプレイ値の計算量／表記量は，通常のシャプレイ値と同等となり，既存の匿名操作不可能な解概念の計算量／表記量よりもはるかに小さい．

5. 匿名操作不可能シャプレイ値の性質

既存のシャプレイ値は，パレート効率性，ヌルスキル公理 (null property)，対称性 (symmetry)，加法性 (additivity) の全ての公理を満たす唯一の解概念である．しかし，シャプレイ値は匿名操作不可能を満足しない (スキルの隠蔽の効果がある) ため，これらの公理を必要最小限に緩和した新しい公理系を構築する．匿名操作不可能シャプレイ値は，この新しい公理系を満足する唯一の解概念であることを示す．

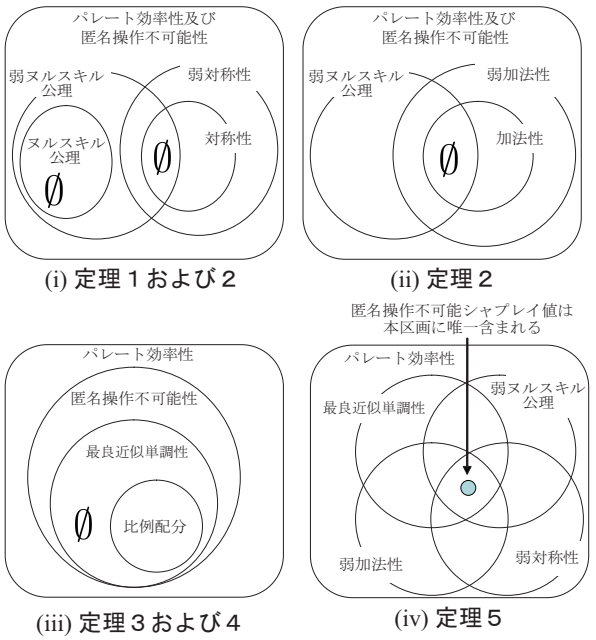

図 1: 匿名操作不可能シャプレイ値の公理系の構築

以下，シャプレイ値の公理系をなぜ／どのように弱めるかについて示す．匿名操作不可能性とパレート効率性に関しては，匿名の環境において解概念が当然満すべき性質であるためそのまま保持する．

定義 5 (パレート効率性) 解概念は，その解概念を表す利得関数 π が以下の条件を満たすとき，かつその場合に限りパレート効率性を満たす：$\forall S,\ \sum_{s \in S} \pi(s,S) = v(S).$

まず定理 1 により，匿名操作不可能性とパレート効率性を満たす解概念はヌルスキル公理を満たさないことをしめす．この定理より，ヌルスキル公理を弱めることが不可避であるため，ヌルスキル公理を弱めた，弱ヌルスキル公理を提案する (図 1 (i) 左側)．

定理 1 ヌルスキル公理と，パレート効率性，匿名操作不可能性を同時に満たす解概念は存在しない．

証明は紙幅の都合上省略する (詳細は文献 [Ohta 08] を参照されたい)．

以下に弱ヌルスキル公理を定義する．まず最初にこの公理を定義するために必要な概念であるヌルスキルを定義する．

定義 6 (ヌルスキル) 以下の条件を満たすスキル s は，スキルの集合 S に対するヌルスキルである：$\forall S', S' \ni s, S' \subseteq S,\ v(S') = v(S' \setminus \{s\}).$

定義 7 (弱ヌルスキル公理) 解概念は，その解概念を表す利得関数 π が以下の条件を満たすとき，かつそのときに限り，弱ヌルスキル公理を満足する：もし s が T に関してヌルスキルならば，$\forall S \subseteq T,\ \pi(s,S) = 0.$

次に対称性と加法性をなぜ／どのように弱めるかについて述べる．定理 2 より匿名操作不可能性，パレート効率性，弱ヌルスキル公理を満足させる解概念は，対称性および加法性を満たさないことが導かれる．この定理より対称性と加法性を弱めることが不可避であるため，弱対称性および弱加法性という新しい公理を提案する (図 1 (i) 右側および (ii))．

定理 2　弱ヌルスキル公理，パレート効率性，および匿名操作不可能性を満たす解概念は対称性および加法性を満たさない．

証明は紙幅の都合上省略する (詳細は文献 [Ohta 08] を参照されたい)．

定理 2 より対称性と加法性を弱める必要がある．以下に対称性を弱めた公理である弱対称性と，加法性を弱めた公理である弱加法性を定義する．まず弱対称性を定義するために必要な概念である，あるスキル集合に関して対称なスキルという概念を定義する．

定義 8 (対称なスキル)　以下の条件を満たすスキル s, $s'(s \in S,\ s' \in S)$ は，スキルの集合 S に対して対称である：$\forall S' \subseteq S \setminus \{s, s'\}, v(S' \cup \{s\}) = v(S' \cup \{s'\})$.

定義 9 (弱対称性)　解概念は，その解概念を表す利得関数 π が以下の条件を満たすとき，かつそのときに限り弱対称性を満足する：任意のスキル s と s' が T に関して対称であるならば，$\forall S \subseteq T \setminus \{s, s'\},\ \pi(s, S \cup \{s\}) = \pi(s', S \cup \{s'\})$ および $\pi(s, S \cup \{s, s'\}) = \pi(s', S \cup \{s, s'\})$.

定義 10 (弱加法性)　解概念は，$v_1 + v_2 = v$ を満たす任意の特性関数 v_1, v_2, v について，その解概念を表す利得関数をそれぞれ π_1, π_2, π としたとき，これらの利得関数が以下の条件を満たすとき，かつそのときに限り，弱加法性を満足する：$\forall s \in T, \pi(s, T) = \pi_1(s, T) + \pi_2(s, T)$.

次に，最良近似単調性と呼ばれる新しい概念を導入する．定理 4 で示されるように，最良近似単調性は，匿名操作不可能性を含む，より強い概念となっている (図 1 (iii))．

定義 11 (最良近似単調性)　解概念は，その解概念を表す利得関数 π が以下の条件を満たすとき，かつそのときに限り最良近似単調性を満足する：任意のパレート効率性を満たす利得関数 π', $\forall s \in S',\ \pi(s, S') = \pi'(s, S')$ を満たす任意のスキルの集合 S' および $v(S) \neq 0$ $(S \subseteq S')$ を満たす S' の任意の部分集合 S について，以下の式が成立する

$$\begin{aligned} &\max_{s \in S} \left\{ \tfrac{\pi(s,S)}{\pi(s,S')} | \pi(s, S) + \pi(s, S') > 0 \right\} \\ &\leq \max_{s \in S} \left\{ \tfrac{\pi'(s,S)}{\pi'(s,S')} | \pi'(s, S) + \pi'(s, S') > 0 \right\} \end{aligned} \tag{2}$$

本論文では特性関数 v が単調であることを仮定しているため，$S \subseteq S' \subseteq T$ ならば $v(S) \leq v(S')$ が成立する．よって，任意の特性関数及び全ての S, S' に関して，$\pi(s, S) \leq \pi(s, S')$ が成立すると考えることは自然であるが，実際にはこの性質を満たす解概念は存在しない（詳しくは文献 [Ohta 08]）．最良近似単調性は，この性質を可能な限り近似的に満足することを意味する．

次に，最良近似単調性を満たす利得関数 π は，以下の定理で与えられる形式に限られること，また，匿名操作不可能性を満たすことを示す．

定理 3　$v(S) \neq 0$ が成立するスキルの集合 $S \subseteq T$ に対し，$\sum_{s \in S} \pi(s, T) \neq 0$ を満足させる利得関数 π は，全ての $S \subseteq T$, $s \in S$ に対して $\pi(s, S) = v(S) \cdot \pi(s, T) / \sum_{t \in S} \pi(t, T)$ （ただし $v(S) = 0$ の時，$\pi(s, S) = 0$ とする）を満たすとき，かつそのときに限り，最良近似単調性を満足する．

証明は紙幅の都合上省略する (詳細は文献 [Ohta 08] を参照されたい)．

定理 4　最良近似単調性を満たす利得関数 π は匿名操作不可能性を満足する．

この定理は，比例配分を行う場合にスキルの隠蔽の効果がないことから容易に導くことができる．

最後に匿名操作不可能シャプレイ値はパレート効率性，弱ヌルスキル公理，弱対称性，弱加法性そして最良近似単調性を唯一満たす解概念であることを示す．

定理 5　匿名操作不可能シャプレイ値は，パレート効率性，弱ヌルスキル公理，弱対称性，弱加法性，最良近似単調性を同時に満たし，かつ上記の公理系を満たす唯一の解概念である．

この定理は，パレート効率性，弱ヌルスキル公理，弱対称性，弱加法性より，スキルの全体集合 T に関して，利得関数はシャプレイ値と一致することが必要となりこと，また，最良近似単調性と定理 3 から容易に導くことができる．

6. おわりに

本論文では匿名操作不可能シャプレイ値という，表記量／計算量を劇的に減少させる新しい匿名操作不可能な解概念を提案した．この解概念は，スキルの全体集合での利得関数を算出／保持しておけば，任意のスキル集合に関する利得関数を必要に応じて簡単に算出できる．また，匿名操作不可能シャプレイ値は，弱ヌルスキル公理，弱対称性，弱加法性，最良近似単調性を満たす唯一の解概念であり，常に一意に定まることが保証される．今後の課題として，ジョブマッチングや腎移植ネットワークなどの現実の事例に対して，スキルの隠蔽の影響，提案した解概念の妥当性等を検討することが挙げられる．

参考文献

[Agotnes 07] Agotnes, T., Hoek, van der W., and Wooldridge, M.: Quantified Coalition Logic, in *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1181–1186 (2007)

[Conitzer 03] Conitzer, V. and Sandholm, T.: Complexity of determining nonemptiness of the core., in *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 613–618 (2003)

[Ohta 08] Ohta, N., Conitzer, V., Satoh, Y., Iwasaki, A., and Yokoo, M.: Anonymity-Proof Shapley Value: Extending Shapley Value for Coalitional Games in Open Environments., in *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (2008)

[Shapley. 53] Shapley., L. S.: A value for n-person games., in *Contributions to the Theory of Games*, pp. 307–317, Princeton University Press (1953)

[Shehory 98] Shehory, O. and Kraus, S.: Methods for task allocation via agent coalition formation., *Artificial Intelligence*, Vol. 101, No. 1-2, pp. 165–200 (1998)

[Yokoo 05] Yokoo, M., Conitzer, V., Sandholm, T., Ohta, N., and Iwasaki, A.: Coalitional games in open anonymous environments., in *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI)*, pp. 509–515 (2005)