主観的メンタル・モデルと他者

合理的思考の技術 Lecture 14

小林憲正

Department of Value and Decision Science (VALDES)
Tokyo Institute of Technology

July 23, 2012

Outline

1 思いやり Benevolence

2 ハイパーゲーム Hypergames

黄金律 The Golden Rule -- 真のアガペの愛とは?

Do unto others as you would have them do unto you.

Jesus (c. 5 BC - AD 32) in the Gospels,
Matthew 7:12, Luke 6:31, Luke 10:27

None of you truly believes until he loves for his brother what he loves for himself.

Muhammad (c. AD 571 - 632) in a Hadith.

(quotes from Wikipedia English)

他者の心の考慮と共有知識 Common Knowledge

黄金律で複雑なのは、他者もあなたと同様に思いやりを持っている可能性を考える必要があることだ。

Example

Ann ``does unto Bob as Ann would have Bob do unto Ann.''
But,
Bob ``does unto Ann as Bob would have Ann do unto Bob.''

この例は、フィードバック的なネスト構造を有する。
Q. Ann と Bob はどのようにこのネスト構造を解くことができるだろう？
When a stable solution exists to this infinite chain of reasoning and is known to all the members, the solution is common knowledge (a jargon of logic!).
i.e.) Ann knows that Bob knows that Ann knows ...

思いやりのモデル化 -- 相互依存効用 Interdependent Utilities

Definition (相互依存効用 Interdependent utilities)

社会的結果(資源配分など) $s \in S$ から得られる相互依存効用は一般に以下の関数形で表現可能

$$U_A(s) = F_A(s, U_B(s))$$

$$U_B(s) = F_B(s, U_A(s))$$

思いやり Benevolence F_i が U_j の増加関数

意地悪 Spite F_i が U_j の減少関数

相互依存と利己的効用

$i \in N$ が $s \in S$ より得る（他者の効用を全く考慮しない）利己的効用を $u_i(s)$ とすると、思いやりを加味した効用全体は最も簡単には以下のように重み付き和（線形結合）で表現可能。

相互依存効用の線形モデル [2]

$$U_A(s) = F_A(s, U_B(s)) = u_A(s) + \lambda_A U_B(s)$$

$$U_B(s) = F_B(s, U_A(s)) = u_B(s) + \lambda_B U_A(s)$$

λ_i は思いやりパラメータ：

思いやり Benevolence $\lambda_i > 0$

意地悪 Spite $\lambda_i < 0$

利己的 Selfish $\lambda_i = 0$

相互依存効用の線形モデルの解

$\lambda_A \lambda_B < 1$ とすると、線形モデルは発散せず、相互依存効用を $u_A(s), u_B(s)$ で以下のように表現可能。

$$U_A(s) = G_A(u_A(s), u_B(s)) = \frac{1}{1 - \lambda_A \lambda_B}(u_A(s) + \lambda_A u_B(s))$$

$$U_B(s) = G_B(u_A(s), u_B(s)) = \frac{1}{1 - \lambda_A \lambda_B}(u_B(s) + \lambda_B u_A(s))$$

$\lambda_A \lambda_B \geq 1$ は、思いやりにポジティブ・フィードバックがかかって、発散している状況。

応用 -- パイの配分

Ann と Bob に対する配分をそれぞれ $100s$%, $100(1-s)$% とする。最適な配分は？

Example

Assume $u_A(s) \equiv 0$ and $u_B(s) \geq 0$. Then, the optimal allocation for both Ann and Bob allocates 0 to Ann and 1 to Bob.
However, this result does not imply that Bob is selfish. He takes all the cake only because he knows that makes Ann happiest as well.

Outline

1 思いやり Benevolence

2 ハイパーゲーム Hypergames

単純ハイパーゲーム [1]

Definition (単純ハイパーゲーム Simple Hypergame)

単純ハイパーゲームは、以下の主観ゲームの組 $(G^i)_{i \in N}$:

- $N = \{1, ..., n\}$ はプレーヤーの集合,
- $G^i = \langle N^i, A^i, u^i \rangle$ プレーヤー$i \in N$ が認識する主観的ゲーム.

- ハイパーゲームにおける主観的ゲームは、プレーヤーの主観的なメンタル・モデルに基づき、物理的世界とは一般には独立。
- 各プレーヤーはハイパーゲーム状況それ自体の具体的な中身（例えば、他のプレーヤーの主観的ゲーム）を一般には知ることはできない。

他のプレーヤーの利得のみの誤認識 [4]

Definition (ベース・ゲーム Base Game)

$G = \langle N, A, u \rangle$ はハイパーゲーム $(G^i)_{i \in N}$ より導出されるベース・ゲームである iff:

- $A_i = A_i^i$
- $u_i = u_i^i$

Definition (利得のみ誤認識を含むハイパーゲーム)

ベース・ゲームについて利得のみ誤認識しているハイパーゲームは以下を満たす。$\forall i, j \in N$:

- $N^i = N$
- $A_j^i = A_i$

経験と整合的な主観的ゲームの均衡 [3]

Definition (主観的ゲームの均衡の経験との整合性)

s^{i*} を利得のみの誤認識を含むハイパーゲームにおける主観的ゲーム G^i の均衡とする.
主観的ゲームの均衡の列 $(s^{i*})_{i \in N}$ が経験と整合的 iff
$\forall i \in N \forall j \neq i, s_j^{i*} = s_j^{j*}$.

他のプレーヤーの利得のみの誤認識がある状況では、以下の定理の証明は容易。

Theorem

シンプル・ハイパーゲーム $(G^i)_{i \in N}$ の主観的ゲームの均衡の列 $(s^{i*})_{i \in N}$ が経験と整合的ならば、$(s_{i*})_{i \in N}$ は導出されたベース・ゲームの均衡である。

[1] Peter G. Bennett.
Hypergames: Developing a model of conflict.
Futures, 12(6):489--507, 1980.

[2] Theodore C. Bergstrom.
Systems of benevolent utility functions.
Journal of Public Economic Theory, 1(1):71--100, 1999.

[3] Yasuo Sasaki.
Preservation of misperceptions--stability analysis of hypergames.
In Proceedings of ISSS, 2009.

[4] Yasuo Sasaki, Norimasa Kobayashi, and Kyoichi Kijima.
Hypergames with mixed strategies -- equilibria and comparative statics.
In GDN Montreal, 2007.