A Formal Separation Between Strategic and Nonstrategic Behavior

James R. Wright
Alberta Machine Intelligence Institute
University of Alberta

Joint work with Kevin Leyton-Brown (UBC)

Contributions:

1. A **formal definition** of strategic behavior that is not equivalent to perfect rationality
2. A **constructive characterization** that precisely distinguishes between **strategic** and **nonstrategic** behavior

Strategic vs. Nonstrategic Agents

- Behavioral game theorists often model **boundedly rational agents** (e.g., humans) using **iterative models** such as the **level-*k* model**:
  - *Level-0* agents: behave uniformly at random — Nonstrategic
  - *Level-1* agents: best respond to level-0 agents
  - *Level-2* agents: best respond to level-1 agents
  - ⋮

  (Level-1, Level-2, …: Strategic)

Question: What kinds of other behavior "count" as nonstrategic?

- Most work argues heuristically for certain rules
- (truthful reporting, largest number, etc.)

Uniform randomization	$a_i \in A_i$
Maxmin	$a_i \in \arg\max_{a_i'} \left[\min_{a_{-i}} u_i(a_i, a_{-i})\right]$
Maxmax	$a_i \in \arg\max_{a_i'} \left[\max_{a_{-i}} u_i(a_i, a_{-i})\right]$
Max total payoffs	$a_i \in \arg\max_{a_i'} \left[\max_{a_{-i}} \sum_j u_j(a_i, a_{-i})\right]$
Min unfairness*	$a_i \in \arg\min_{a_i'} \left[\min_{a_{-i}} u_{-i} \left\vert (a_i', a_{-i}) - u_i(a_i', a_{-i}) \right\vert \right]$
Nash equilibrium	$\{a_i \mid \exists a_{-i} : (a_i, a_{-i})$ is Nash equilibrium$\}$

Strategic Agents

Definition: A behavioral model is **strategic** if it is both **other responsive** and **dominance responsive**.

Definition:
A behavioral model is **dominance responsive** if, for every pair of games where some action is strictly dominant in one game and strictly dominated in the other, the model does not behave identically:

$$f_i(G) \neq f_i(G') \; \forall G, G' \text{ with } a_i^* \text{ dominant in } G \text{ and dominated in } G'$$

Definition:
A behavioral model is **other responsive** if there exists **any pair** of games that differ only in the payoffs of the **other agents** in which the model predicts different behavior:

$$\exists G, G' : f_i(G) \neq f_i(G') \wedge \forall a \in A : u_i(a) = u'_i(a)$$

Theorem: All of **QRE**, **Nash equilibrium**, **correlated equilibrium**, **cognitive hierarchy**, and **level-k**(*) are (profiles of) **strategic** behavioral models.

Elementary Behavioral Models

Definition:

A behavioral model f_i is **elementary** if it can be represented as $f_i(G) = h(\Phi(G))$, where:

- for all games $G = (N, A, u)$, for all $a \in A$, $\Phi(G)_a = \varphi(u(a))$,
- φ satisfies **no smuggling**, and
- h is an **arbitrary function** that maps $\mathbb{R}^A \to \Delta(A_i)$

Main Theorem: No **elementary** behavioral model f_i is **strategic**.