Chapter 18

Frontiers: Behavioral, Experimental, and Modern Directions

The final chapter surveys where game theory is being tested, revised, and extended. Each section states the core departure from the classical theory, the leading models, and pointers into the literature.

18.1 Experimental game theory: the evidence

Fifty years of laboratory and field experiments yield robust regularities:

- **Ultimatum/dictator games**: proposers offer 40–50%; low offers are rejected (Example 6.4); even *dictators* (no rejection possible) give around 20%—evidence for fairness preferences.
- **Public goods**: initial contributions of 40–60% decay with repetition; costly *punishment* opportunities restore high cooperation (Example 7.2); anti-social punishment varies across societies.
- **Beauty contests**: winning guesses far above the rationalizable 0 (Example 3.7)—people iterate reasoning only 1–3 steps.
- **Coordination**: subjects find focal points reliably; risk-dominance predicts behavior in Stag Hunts better than payoff-dominance; conventions once established are sticky.
- **Auctions**: overbidding relative to risk-neutral theory in first-price auctions; the winner's curse afflicts inexperienced common-value bidders exactly as Example 10.5 suggests.
- **Mixed strategies**: professionals (soccer, tennis) approximate equilibrium mixing; novices alternate excessively.

The pattern: theory performs best with experienced players, high stakes, competitive market-like settings, and strategic simplicity; it needs behavioral amendment where fairness, cognition limits, or novelty dominate.

18.2 Behavioral game theory: amended models

Social preferences. Fehr–Schmidt inequity aversion: $u_i = x_i - \alpha_i \max(x_j - x_i, 0) - \beta_i \max(x_i - x_j, 0)$; Bolton–Ockenfels ERC; Rabin's intentions-based reciprocity (kindness begets kindness) built on *psychological game theory*, where payoffs depend on beliefs themselves.

Level-k and cognitive hierarchy. Level-0 acts nonstrategically (random or focal); level-k best-responds to level-$(k-1)$; estimated levels concentrate on 1–2, organizing beauty-contest and auction data (Nagel; Camerer–Ho–Chong).

Quantal response equilibrium (McKelvey–Palfrey 1995). Players best-respond *noisily* (logit choice), in equilibrium with each other's noise. QRE rationalizes systematic deviations (e.g. own-payoff effects in Matching Pennies) while nesting Nash as the zero-noise limit.

Limited foresight and complexity. Finite automata playing repeated games, sampling equilibria, and models of costly reasoning connect to the computational themes of Chapter 17.

18.3 Stochastic games and dynamic equilibria

Shapley's *stochastic games* (1953)—state variables evolving with play—underlie modern dynamic oligopoly (Markov perfect equilibrium, Maskin–Tirole), resource economics, and multi-agent reinforcement learning. Deep results: existence of stationary equilibria, the *uniform value* of zero-sum stochastic games (Mertens–Neyman 1981), and the algorithmic theory now central to multi-agent AI.

18.4 Games with a continuum of players and mean-field games

Large anonymous populations justify *nonatomic games* (Schmeidler) and *mean-field games* (Lasry–Lions 2007; Huang–Caines–Malhamé): each agent optimizes against the population *distribution*, whose evolution is generated by the agents' own optimal behavior—a fixed point in distribution space, described by coupled Hamilton–Jacobi–Bellman and Kolmogorov equations. Applications: crowd dynamics, energy markets, systemic risk, and large-scale multi-agent learning.

18.5 Information design and Bayesian persuasion

Mechanism design fixes the game and designs incentives; *information design* fixes incentives and designs *what players learn*.

Example 18.1: Bayesian persuasion (Kamenica–Gentzkow 2011)

A prosecutor (sender) commits to an investigation policy (a signal about guilt) to persuade a judge (receiver) who convicts iff the posterior probability of guilt is at least $\frac{1}{2}$; the prior is 0.3. Full disclosure convicts 30% of defendants. Optimal persuasion: a signal that says "guilty" always when guilty and with probability $\frac{3}{7}$ when innocent—making the posterior upon "guilty" exactly $\frac{1}{2}$. Conviction rate: $0.3 + 0.7 \cdot \frac{3}{7} = 0.6$, double full disclosure. The general solution method—*concavification* of the sender's value over posterior beliefs—has made information design a flourishing field (ratings, stress tests, recommendation systems, traffic apps).

18.6 Game theory and AI

- **Superhuman play**: checkers solved (2007); heads-up limit poker essentially solved (2015); Go, chess, shogi via self-play reinforcement learning (AlphaZero 2017); no-limit poker

(Libratus, Pluribus); Diplomacy with natural-language negotiation (Cicero 2022)—each a milestone in computing (approximate) equilibria of ever richer games.

- **Multi-agent learning**: convergence of independent learners, the emergence of cooperation and communication among agents, and the game dynamics of GAN training (a zero-sum game whose last-iterate behavior can cycle—Chapter 12's theory of averages vs. iterates, live in ML).
- **Mechanism design by/for AI**: automated mechanism design; auction design via deep learning; incentive alignment questions for LLM agents negotiating, bidding, and delegating on behalf of humans—game theory's next large application domain.
- **AI safety as game design**: principal–agent framings of alignment, debate and verification games, and commitment/credibility problems between humans and capable systems.

18.7 A reading roadmap

Consolidate (advanced undergraduate). Osborne, *An Introduction to Game Theory*; Gibbons, *Game Theory for Applied Economists*; Dixit–Nalebuff, *The Art of Strategy* (popular but sound).

Graduate core. Fudenberg–Tirole, *Game Theory*; Osborne–Rubinstein, *A Course in Game Theory* (free online); Myerson, *Game Theory: Analysis of Conflict*; Maschler–Solan–Zamir, *Game Theory* (most comprehensive and mathematical).

Fields. Krishna, *Auction Theory*; Börgers, *An Introduction to Mechanism Design*; Roth–Sotomayor, *Two-Sided Matching*; Mailath–Samuelson, *Repeated Games and Reputations*; Weibull, *Evolutionary Game Theory*; Sandholm, *Population Games and Evolutionary Dynamics*; Nisan–Roughgarden–Tardos–Vazirani (eds.), *Algorithmic Game Theory* (free online); Roughgarden, *Twenty Lectures on Algorithmic Game Theory*; Camerer, *Behavioral Game Theory*; Perea, *Epistemic Game Theory*.

Primary sources worth reading in the original. Nash (1950); Schelling, *The Strategy of Conflict* (1960); Aumann (1974, 1976); Myerson (1981); Rubinstein (1982); Kreps–Wilson (1982); Milgrom–Weber (1982); Kamenica–Gentzkow (2011).

Key Idea: The unity of the subject

Look back at the arc of this book: a handful of ideas—best response, equilibrium, credibility, incentive compatibility, conditioning on others' information, the shadow of the future—generate the entire edifice, from penalty kicks to spectrum auctions to blockchain protocols. Mastering game theory is not memorizing models; it is learning to see every interactive situation through these few lenses. That skill, once acquired, is permanent.