

New Directions in Behavioral Game Theory: Introduction to the Special Issue

Russell Golman  
Carnegie Mellon University

October 23, 2020

Russell Golman: Department of Social and Decision Sciences, Carnegie Mellon University, Email:  
[rgolman@andrew.cmu.edu](mailto:rgolman@andrew.cmu.edu).

Behavioral game theory accounts for how people actually make strategic decisions by incorporating social utility, limited iterated reasoning, and learning (Camerer, 2003). The papers in this special issue span this space of behavioral game theory research.

Seier (2020) in this special issue explores whether fairness in strategic games tends to be driven by intuitive or deliberative responses. Many people are willing to incur selfish costs to uphold norms of fairness, from promoting efficiency or equity to punishing others who violate these norms (Camerer and Fehr, 2004). Fair behavior could follow from deliberation, with self-control being used to do the right thing despite an intuitive inclination to be selfish, or it could be an intuitive response that is adaptive in naturalistic contexts, but that can be overcome deliberately in the lab in artificial contexts in which selfishness does not have reputational costs. Seier (2020) finds that people who give more intuitive answers on the cognitive reflection task tend to make more fair choices in strategic games: they give away more money in the dictator game, demand more money as receivers in the ultimatum game, and engage in more costly third-party punishment of norm violators in a multiplayer game. For many people, social utility is a fundamental element of their preferences.

Zhao (2020) in this special issue studies how the extent of iterated reasoning performed in a strategic decision depends on constraints on the other player's ability (as well as one's own ability) to engage in iterated reasoning. Using two-player guessing games in which strategy choices map cleanly onto levels of reasoning in a level- $k$  model (Nagel, 1995; Costa-Gomes and Crawford, 2006; Georganas et al., 2015), Zhao (2020) finds that players engage in more steps of reasoning when their opponents have been placed under a condition of lighter (rather than heavier) cognitive load, and this effect is stronger when players themselves are under lighter cognitive load, and thus able to engage in more steps of reasoning in the first place. That is, players are capable of recognizing that cognitive load may inhibit the reasoning ability of their opponents, and they respond appropriately. The observed pattern of behavior reflects an adaptive response that transcends the level- $k$  reasoning model. Other models, including logit quantal response equilibrium (McKelvey and Palfrey, 1995; Capra et al., 1999), noisy introspection (Goeree and Holt, 2001), and the dual accumulator model (Golman et al., 2020), can also account for limited iterated reasoning in guessing games, and manipulating the precision of logit responses in these models can also affect the depth of reasoning that an individual exhibits. A behavioral insight affirmed here, and consistent with all of these models, is that while people are boundedly rational, in that they are not capable of unlimited iterated reasoning, they do respond sensibly to changes in their opponent's incentives or constraints.

Guilfoos and Pape (2020) in this special issue study how strategic behavior changes as players play a game repeatedly (with new opponents) and get feedback. They econometrically estimate case-based learning (Gilboa and Schmeidler, 1995), reinforcement learning (Erev and Roth, 1998), and self-tuning experience weighted attraction (Ho et al., 2007), applied to Selten and Chmura's (2008) dataset of 864 subjects repeatedly playing one of twelve 2x2 games. Case-based learning fits the observed behavior best, and also best predicts out-of-sample choices for a held-out slice of the data. Comparing the models based on out-of-sample prediction ensures that the empirical support for case-based learning is not an artifact of model flexibility and overfitting.

Guisasola and Saari (2020) in this special issue introduce a coordinate system for the full space of 2x2 games that distinguishes changes in payoffs that exclusively affect: i) the selfish costs and benefits of one's own strategies averaged uniformly across the other player's strategies (the *individual preference*

*component*); ii) the dependence of these selfish costs and benefits on the choice of the other player's strategy (the *coordinative pressure component*); iii) the externality imposed on the other player by the choice of one's own strategy (the *pure externality component*); and iv) a constant level shift of all payoffs (the *kernel component*). The coordinate system is useful for a number of applications. This paper focuses on applying it to 2x2 potential games, including coordination games and anti-coordination games. Predictions based on individual selfish costs and benefits, including Nash equilibrium, risk-dominance (equivalently, the global maximum of the potential function), level- $k$  reasoning, quantal response equilibrium, noisy introspection, and the dual accumulator model, are invariant to changes in the pure externality component of a game. However, changes in the pure externality component of the game do affect social welfare. Thus, it is straightforward to design games that pose a tension between the strategy predicted by any model based on individual selfish costs and benefits and the strategy that maximizes social welfare. The empirical fact that people care about social welfare as well as other aspects of the interaction between the externality component of a game and the individual preference component (Charness and Rabin, 2002) indicates that any model of the individual reasoning process needs to be augmented with a model of social preferences to more fully capture behavior. The decomposition of 2x2 games in the coordinate system presented in this paper could be useful for experimental research by making it easier to independently test models of individual reasoning and models of social preferences.

Jamison (2020) in this special issue explores the role of pre-play cheap talk among players with common knowledge of rationality. Whereas cheap talk is often dismissed as not credible because it is easily imitated, it may actually be informative when players have partially aligned incentives (Crawford and Sobel, 1982; Farrell and Rabin, 1996) or social preferences (Golman, 2016) such that conditional on a statement being interpreted correctly, an individual wants to make the statement in the first place. In the absence of pre-play communication, common knowledge of rationality implies that players will choose rationalizable strategies, but not necessarily successfully coordinate on a Nash equilibrium. Jamison (2020) shows that cheap talk allows rational players to reach (only) efficient Nash equilibria. Understanding pre-play cheap talk among rational players give us a benchmark for studying pre-play cheap talk in laboratory games and in the real world, a context in which players are boundedly rational, may have incomplete information, and may have uncertainty or biases about each other's social preferences (Di Tella et al., 2015). These behavioral elements allow communication to be informative in new and interesting ways (Charness, 2000; Chaudhry and Loewenstein, 2019; Crawford, 1998; Ellingsen & Östling, 2010; Gneezy, 2005; Gurney and Loewenstein, 2020).

## References

- Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.
- Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, 97, 55-95.
- Capra, C. M., Goeree, J. K., Gomez, R., & Holt, C. A. (1999). Anomalous behavior in a traveler's dilemma?. *American Economic Review*, 89(3), 678-690.

- Charness, G. (2000). Self-serving cheap talk: A test of Aumann's conjecture. *Games and Economic Behavior*, 33(2), 177-194.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3), 817-869.
- Chaudhry, S. J., & Loewenstein, G. (2019). Thanking, apologizing, bragging, and blaming: Responsibility exchange theory and the currency of communication. *Psychological Review*, 126(3), 313-344.
- Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5), 1737-1768.
- Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, 78(2), 286-298.
- Crawford, V., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50, 1431-1451.
- Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *American Economic Review*, 105(11), 3416-42.
- Ellingsen, T., & Östling, R. (2010). When does communication improve coordination?. *American Economic Review*, 100(4), 1695-1724.
- Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 848-881.
- Farrell, J., & Rabin, M. (1996). Cheap talk. *The Journal of Economic Perspectives*, 10(3), 103-118.
- Georganas, S., Healy, P. J., & Weber, R. A. (2015). On the persistence of strategic sophistication. *Journal of Economic Theory*, 159, 369-400.
- Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*, 110(3), 605-639.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1), 384-394.
- Goeree, J. K., & Holt, C. A. (2001). Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, 91(5), 1402-1422.
- Golman, R. (2016). Good manners: signaling social preferences. *Theory and Decision*, 81(1), 73-88.
- Golman, R., Bhatia, S., & Kane, P. B. (2020). The dual accumulator model of strategic deliberation and decision making. *Psychological Review*, 127(4), 477-504.
- Guilfoos, T., & Pape, A. D. (2020). Estimating Case-Based Learning. *Games*, 11(3), 38.  
doi:10.3390/g11030038
- Guisasola, S., & Saari, D. (2020). With Potential Games, Which Outcome Is Better? *Games*, 11(3), 33.  
doi:10.3390/g11030033
- Gurney, N., & Loewenstein, G. (2020). Filling in the Blanks: What Restaurant Patrons Assume About Missing Sanitation Inspection Grades. *Journal of Public Policy & Marketing*, 39(3), 266-283.

Ho, T. H., Camerer, C. F., & Chong, J. K. (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, 133(1), 177-198.

Jamison, J. (2020). Valuable Cheap Talk and Equilibrium Selection. *Games*, 11(3), 34.  
doi:10.3390/g11030034

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1), 6-38.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, 85(5), 1313-1326.

Seier, M. (2020). The Intuition of Punishment: A Study of Fairness Preferences and Cognitive Ability. *Games*, 11(2), 21. doi:10.3390/g11020021

Selten, R., & Chmura, T. (2008). Stationary concepts for experimental 2x2-games. *American Economic Review*, 98(3), 938-66.

Zhao, W. (2020). Cost of Reasoning and Strategic Sophistication. *Games*, 11(3), 40.  
doi:10.3390/g11030040