Language, Meaning, and Games: Comment

Heller, Yuval

Oxford

29 August 2013

Language, Meaning, and Games: Comment

Yuval Heller *(August 29, 2013)*

Address: Nuffield College and Department of Economics, New Road, Oxford, OX1 1NF, United Kingdom. Email: yuval26@gmail.com or yuval.heller@economics.ox.ac.uk.

Abstract

Demichelis & Weibull (AER 2008) show that adding lexicographic lying costs to coordination games with cheap talk yields a sharp prediction: only the efficient outcome is evolutionarily stable. I demonstrate that this result is caused by the discontinuity of preferences, rather than by small lying costs per se.

KEYWORDS: Lexicographic preferences, evolutionary stability, cheap talk. JEL: C73.

Demichelis & Weibull (2008) (henceforth, DW) study symmetric two-player $n \times n$ *coordination games* in which each pure action is the best reply to itself, as demonstrated in the game of Table 1. This game admits three pure strict equilibria: the Pareto-dominant (a, a), and the Pareto-dominated (b, b) and (c, c). Intuition suggests that equilibrium (a, a) is more plausible, especially if players can communicate before playing the game. However, all the above-mentioned equilibria satisfy all the standard (non-evolutionary) refinements, even when the game is extended to include a pre-play *cheap-talk* stage in which each player simultaneously sends a message from a finite alphabet before playing the game.

Several papers use the refinement of evolutionary stability to study such games.[1] A symmetric Nash equilibrium is an *evolutionarily (neutrally) stable strategy* if it is a strictly (weakly) better reply to all other best replies than these are to themselves (Maynard-Smith & Price, 1973; Maynard Smith, 1982).[2] The motivation for this notion is that a stable strategy, if adopted by a population of players, cannot be invaded by any alternative strategy that is initially rare.[3]

[1] Other papers use different evolutionary approaches to study coordination games, such as stochastic evolutionary stability (e.g., Young, 1993) and direct study of the dynamic process (e.g., Kim & Sobel, 1995).

[2] I.e., strategy σ is evolutionarily stable if (1) $\pi(\sigma, \sigma) \geq \pi(\tau, \sigma)$ for each strategy τ, and (2) $\pi(\sigma, \sigma) = \pi(\tau, \sigma)$ implies that $\pi(\sigma, \tau) > \pi(\tau, \tau)$. It is neutrally stable if the last inequality is replaced with a weak inequality.

[3] Any NSS is *Lyapunov stable* in the replicator dynamic (Bomze & Weibull, 1995): no small change in the composition of the population can take the population away from the state in which everyone follows the NSS,

Tab. 1: Payoff Matrix (π) of a Symmetric Two-Player Coordination Game

	a	b	c
a	3	0	0
b	0	2	0
c	0	0	1

Thomas (1985) extends the notion of stability from strategies to outcomes. A payoff profile (*outcome*) β is evolutionarily stable if there exists a set of neutrally stable strategies such that each strategy in the set (1) yields payoff β, and (2) is a strictly better reply to any other best reply outside the set.[4]

Schlag (1993) and Banerjee & Weibull (2000) characterize stable outcomes of coordination games with cheap talk. They show that if each player has only two actions and the alphabet is large, then only approximately efficient outcomes are evolutionarily stable. However, if there are more than two actions, then such games admit also inefficient evolutionarily stable strategies. Each such strategy must use all the messages in the alphabet with positive probability (otherwise, the behavior against the unused messages "drifts" until it becomes favorable for mutants to use these messages and play the Pareto-dominant action profile.) In particular, the following strategy, σ^*, is evolutionarily stable in the game of Table 1: players send each message with equal probability, and they play c if they have sent the same message and play b otherwise.[5]

DW's Result. DW adapt the cheap-talk model by assuming that some messages have a pre-specified meaning regarding the action that the sender intends to play, and that there are small lying costs. These costs are modeled lexicographically: only when comparing two outcomes with the same material payoffs (π) do players care about the lying costs, and in this case, they prefer the outcome with the smaller lying costs. DW (Prop. 1) show that this adaptation yields a sharp prediction: the Pareto-dominant payoff is the unique evolutionarily stable outcome of the lexicographic game (even when there are more than two actions). The intuition of this result is as follows. The lexicographic preferences imply that players cannot be indifferent between two pure strategies that induce different lying probabilities. DW assume that the alphabet

and any ESS is *asymptotically stable* (Taylor & Jonker; see extensions to other dynamics in Cressman (1997); Sandholm (2010)): any sufficiently small change results in a movement back toward the ESS.

[4] Formally, outcome β is evolutionarily stable if there exists a closed and non-empty set of strategies X, such that each strategy $\sigma \in X$ satisfies: (1) $\pi(\sigma,\sigma) = \beta$, (2) neutral stability, and (3) for each strategy τ: $\pi(\sigma,\sigma) = \pi(\tau,\sigma)$ and $\pi(\sigma,\tau) = \pi(\tau,\tau)$ imply that $\tau \in X$. Thomas (1985) (see also Weibull, 1995) shows that such sets are asymptotically stable.

[5] Similar to Schlag (1993), I use the reduced normal form in which a strategy does not specify what action a player would take in the counter-factual case when he sends another message than the message specified by his strategy. The results are similar in the alternative form (à la DW) in which a strategy also specifies actions in counter-factual cases. Specifically, the set of strategies that are equivalent to σ^* (and only differ in counter-factual cases) is an evolutionarily stable set (Thomas, 1985).

contains a message ϕ that is always "truthful" and messages that "promise" to play specific actions. Observe that in any Nash equilibrium at least one of the "promising messages" induces a positive probability of lying. Thus, there must be unused messages in any Nash equilibrium of the lexicographic game (DW, Lemma 3) and inefficient strategies, such as σ^*, cannot be evolutionarily stable.[6]

Discontinuity and Stability. Samuelson & Swinkels (2003) argue that "lexicographic constructions are interesting to the extent that they provide convenient approximations for cases in which preferences are continuous, but some considerations are very much more important than others." Motivated by this view, I study whether DW's lexicographic construction approximates continuous preferences, and I show that the answer is negative: the prediction depends crucially on the discontinuity of preferences. For each $\epsilon > 0$ define G_ϵ as the coordination game with communication in which the (continuous) payoff function is equal to π (the material payoffs of Table 1) minus ϵ times the lying costs. Observe that for any sufficiently small ϵ the game G_ϵ admits the following evolutionarily stable strategy σ_ϵ: (1) at the second stage, players play as in σ^*: c if they sent the same message and b otherwise, and (2) at the first stage, each message's probability of being sent is adjusted relative to the equal probabilities in σ^*: messages that yield higher expected lying costs are sent with slightly lower probabilities, such that all messages yield the same expected payoff in G_ϵ given the play at the second stage.[7] Thus, any continuous game with small lying costs admits an evolutionarily stable strategy close to σ^*. Similarly, one can show that every evolutionarily stable strategy of the cheap-talk game is a limit of evolutionarily stable strategies in every converging sequence of games with continuous small lying costs.[8]

Finally, observe that various kinds of discontinuities (unrelated to lying costs) yield the same sharp prediction as DW. For example, assume that there are two messages in the alphabet, m and m', such that m' is slightly harder to pronounce than m and bears an additional lexicographic "pronunciation cost." This discontinuity implies that these two messages cannot both be used in any equilibrium, and hence only efficient outcomes are stable. Thus, *discontinuous preferences induce efficiency in coordination games with communication, whereas continuous small lying*

[6] DW's notion of evolutionary stability implicitly assumes (à la Volij, 2002) that the influence of mutants is smaller than the influence of the lying costs. That is, if $\pi(\tau,\sigma) = \pi(\sigma,\sigma)$ and τ yields a higher lying cost, then mutants who play τ cannot invade a population of incumbents who play σ even if the mutants obtain higher material payoffs in any post-entry population (i.e., $\pi(\tau,(1-\epsilon)\cdot\sigma+\epsilon\cdot\tau) > \pi(\sigma,(1-\epsilon)\cdot\sigma+\epsilon\cdot\tau)$ for each $\epsilon > 0$). Note that under the opposite assumption (i.e., mutants are more influential than lying costs, à la Binmore & Samuelson, 1992) the set of evolutionarily stable outcomes in the lexicographic game is the same as in a cheap-talk game with no lying costs.

[7] Due to continuity: (1) there always exist probabilities that balance the payoffs, and (2) strategy σ_ϵ converges to σ^* as $\epsilon \to 0$. Minor adaptations to the proofs of Schlag (1993) and Banerjee & Weibull (2000) show that σ_ϵ is an evolutionarily stable strategy in G_ϵ for a sufficiently small ϵ.

[8] Moreover, this result holds regardless of the relative magnitude of the lying costs and the mutation barriers (as defined in Samuelson & Swinkels, 2003).

costs do not influence the set of stable outcomes.

References

Banerjee, Abhijit, & Weibull, Jörgen W. 2000. Neutrally stable outcomes in cheap-talk coordination games. *Games and Economic Behavior*, **32**(1), 1–24.

Binmore, K.G., & Samuelson, L. 1992. Evolutionary stability in repeated games played by finite automata. *Journal of economic theory*, **57**(2), 278–305.

Bomze, Immanuel M, & Weibull, Jörgen W. 1995. Does neutral stability imply Lyapunov stability? *Games and Economic Behavior*, **11**(2), 173–192.

Cressman, Ross. 1997. Local stability of smooth selection dynamics for normal form games. *Mathematical Social Sciences*, **34**(1), 1–19.

Demichelis, Stefano, & Weibull, Jörgen W. 2008. Language, meaning, and games: A model of communication, coordination, and evolution. *The American Economic Review*, 1292–1311.

Kim, Yong-Gwan, & Sobel, Joel. 1995. An evolutionary approach to pre-play communication. *Econometrica*, 1181–1193.

Maynard Smith, J. 1982. *Evolution and the theory of games*. Cambridge University Press.

Maynard-Smith, J., & Price, G.R. 1973. The Logic of Animal Conflict. *Nature*, **246**, 15.

Samuelson, Larry, & Swinkels, Jeroen M. 2003. Evolutionary stability and lexicographic preferences. *Games and Economic Behavior*, **44**(2), 332–342.

Sandholm, William H. 2010. Local stability under evolutionary game dynamics. *Theoretical Economics*, **5**(1), 27–50.

Schlag, Karl H. 1993. *Cheap talk and evolutionary dynamics*. Univ. of Bonn Disc. Paper B-242.

Taylor, P.D., & Jonker, L.B. 1978. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, **40**(1), 145–156.

Thomas, Bernhard. 1985. On evolutionarily stable sets. *J. of Math. Biology*, **22**(1), 105–115.

Volij, Oscar. 2002. In defense of DEFECT. *Games and Economic Behavior*, **39**(2), 309–321.

Weibull, Jörgen W. 1995. *Evolutionary game theory*. The MIT press.

Young, H Peyton. 1993. The evolution of conventions. *Econometrica*, 57–84.