Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution: Comment

Heller, Yuval

Oxford

29 August 2013

Language, Meaning, and Games: A Model of

Communication, Coordination, and Evolution: Comment

Yuval Heller*(January 12, 2014)

Address: Nuffield College and Department of Economics, New Road, Oxford, OX1 1NF, United Kingdom. E-mail: 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 show that this result is caused by the discontinuity of preferences rather than by small lying costs per se. Finally, I discuss why discontinuity may not be an appealing assumption in evolutionary models. KEYWORDS: Lexicographic preferences, evolutionary stability, cheap talk. JEL: C73.

The game of Figure 1 is an example of a symmetric two-player coordination game in which each pure action is the best reply to itself. 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

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

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

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 before playing the game.

Several papers use the refinement of evolutionary stability to study such games. A symmetric Nash equilibrium is an evolutionarily stable strategy if it is a strictly better reply to all other best replies than those replies are to themselves (Maynard Smith & Price, 1973). 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. The existing literature (Section I) shows that non-efficient equilibria are also evolutionarily stable. In an influential paper, Demichelis & Weibull (2008) (henceforth, DW) show that enriching the model with lexicographic lying costs yields a sharp prediction: only the efficient outcome is evolutionarily stable (Section II).

The main contribution of this note (Section III) is the observation that DW's prediction does not hold for continuous lying costs: the set of stable outcomes remains qualitatively the same as in the cheap-talk model. I conclude in Section IV by showing that various kinds of discontinuous preferences (unrelated to lying costs) induce efficiency in coordination games, and that such discontinuities may not be an appealing assumption in evolutionary models.

I. Coordination Games with Cheap Talk

Definitions. Strategy σ is evolutionarily stable if for each strategy $\tau \neq \sigma$, we have: (1) $\pi(\sigma, \sigma) \geq \pi(\tau, \sigma)$ (i.e., σ is a symmetric Nash equilibrium), and (2) $\pi(\sigma, \sigma) = \pi(\tau, \sigma)$ implies that $\pi(\sigma, \tau) > \pi(\tau, \tau)$. It is neutrally stable (Maynard Smith, 1982) if the last inequality is replaced with a weak inequality.

Taylor & Jonker (1978) show that evolutionary stability implies *asymptotic stability* in the replicator dynamics (see extensions to other payoff-monotonic dynamics in Cressman, 1997; Sandholm, 2010): any sufficiently small change in the composition of the population results in a movement back toward the evolutionarily stable strategy. 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 yields payoff β and is a strictly better reply to any other best reply outside the set.[1]

Results. Schlag (1993) and Banerjee & Weibull (2000) characterize stable outcomes of coordination games with cheap talk. They show that if each player has two actions and the alphabet (the finite set of possible messages) is large, then only approximately efficient outcomes are evolutionarily stable. However, if there are more than two actions, then such games also admit inefficient evolutionarily stable strategies.[2] 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 Figure 1: players send each message with equal probability, and play c if they have sent the same message, and b otherwise.[3]

---

[1] 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.

[2] Demichelis (2013) has shown that when players play *long repeated* coordination games, then only efficient outcomes are evolutionarily stable (even when there are more than two actions).

[3] Similarly 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 qualitatively the same (but the presentation of the results is more cumbersome) in the alternative full normal form (à la DW) in which a strategy also specifies actions in counter factual cases. The translation of our results to the setup of full normal form is sketched as followed. Given a reduced-form pure strategy s, let $\tilde{S}(s)$ be the set of full-form pure strategies that are

II. Lexicographic Lying Costs

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. Specifically, DW assume that the alphabet contains a non-informative message ϕ that is always "truthful", for each action (say, a) there exists a message that "promises" to play this action (denoted by $\hat{a}$), and a player suffers a positive lying cost if he sends message $\hat{a}$ and plays an action other than a. The lying 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 cost. Finally, DW use the following solution concept: an outcome is stable if it is (1) an outcome of a Nash equilibrium component of the lexicographic game, and (2) each equilibrium in this component is neutrally stable with respect to the material payoffs.[4]

DW (Proposition 1) show that this adaptation yields a sharp prediction: the Pareto-dominant payoff is the unique stable outcome of the lexicographic game. The intuition of this result is as follows. Recall that each inefficient

---

equivalent to s (different from s only in counter factual cases): each full-form strategy $\tilde{s} \in \tilde{S}(s)$ sends the same message m as strategy s, and chooses the same action as strategy s after any history of messages in which the player's own message is m. Given a reduced-form mixed strategy σ, let $\tilde{\Sigma}(\sigma)$ be the set of full-form mixed strategies that are equivalent to σ: a full-form mixed strategy $\tilde{\sigma}$ is in $\tilde{\Sigma}(\sigma)$ iff for each pure reduced-form strategy s, $\tilde{\sigma}\left(\tilde{S}(s)\right) = \sum_{\tilde{s} \in \tilde{S}(s)} \tilde{\sigma}(\tilde{s}) = \sigma(s)$. Minor adaptations to the proof of the main observation in Section III show that it holds also in the setup of full normal form; i.e., $\tilde{S}(\sigma_\epsilon)$ is an evolutionarily stable set (Thomas, 1985) in G_ϵ.

[4] The first requirement (being a Nash equilibrium of the lexicographic game) implicitly assumes (à la Volij, 2002) that the influence of mutants is smaller than the influence of the lying costs. This is demonstrated in the symmetric game of Figure 2 (adapted from Samuelson & Swinkels, 2003). Strategy a is not a Nash equilibrium of the lexicographic game (because strategy b yields the same material payoffs and higher secondary payoffs against strategy a). Observe that incumbents who follow strategy a strictly outperform mutants who follow strategy b in any post-entry population (i.e., $\pi(a, (1-\epsilon) \cdot a + \epsilon \cdot b) > \pi(b, (1-\epsilon) \cdot a + \epsilon \cdot b)$ for each $\epsilon > 0$). However, DW's notion of stability does not allow this advantage against mutants to compensate for the disadvantage in the secondary payoffs. Finally, note that under the opposite assumption (i.e., mutants are more influential than lying costs, à la Binmore & Samuelson, 1992) one can show that the set of evolutionarily stable outcomes in the lexicographic game is the same as in the cheap-talk game with no lying costs.
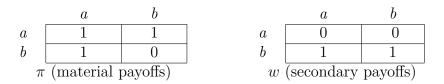
	a	b
a	1	1
b	1	0

π (material payoffs)

	a	b
a	0	0
b	1	1

w (secondary payoffs)

Fig. 2: Strategy a is not "DW stable" in this symmetric lexicographic game.

stable strategy must use all messages with positive probability. Assume to the contrary that there exists a Nash equilibrium of the lexicographic game in which all the messages are used with positive probability. By definition, message ϕ does not induce any lying cost. The lexicographic preferences imply that players cannot be indifferent between two pure strategies with different lying costs. Thus, all messages must induce zero lying costs. In particular, a player who sends message $\hat{a}$ ($\hat{b}$) always plays action a (b). This leads to a contradiction: if a player sends $\hat{b}$ and his opponent sends $\hat{a}$, then the player achieves a strictly higher payoff by deviating to play action a instead of b.

III. Continuous Lying Costs

In this section, I show that DW's prediction does not hold for continuous lying costs. For each $\epsilon > 0$ define G_ϵ as the coordination game with communication in which the (continuous) payoff function, π_ϵ, is equal to π (the material payoffs of Table 1) minus ϵ times the lying costs (see Chen *et al.*, 2008 for a similar model of lying costs). In what follows, I show that for any sufficiently small ϵ the game G_ϵ admits an evolutionarily stable strategy σ_ϵ close to σ^*. For concreteness, Table 1 describes σ_ϵ in an example in which the alphabet includes four messages $\left\{ \phi, \hat{a}, \hat{b}, \hat{c} \right\}$ and there is a fixed lying cost equal to 1.

The sketch of the proof of this result is as follows. Strategy σ^* is an isolated quasi-strong equilibrium (i.e., it assigns a positive probability to every pure strategy that is a best reply to σ^*), and this implies (Jansen, 1981) that σ^* is an essential equilibrium (i.e., all nearby continuous games admit an equilibrium close to σ^*). Thus, game G_ϵ admits a nearby Nash equilibrium σ_ϵ. Moreover, the quasi-strongness of σ^* implies that, for a sufficiently small ϵ, strategy σ_ϵ has the same support as σ^*. This implies that σ_ϵ has the following properties:

m	$\Pr_{\sigma_\epsilon}(m)$	$\pi(\sigma_\epsilon, \sigma_\epsilon\mid \text{ sending } m)$	$\Pr_{\sigma_\epsilon}(\text{lying }\mid m)$	$\pi_\epsilon(\sigma_\epsilon, \sigma_\epsilon\mid m)$
ϕ	$\frac{1}{4}+\frac{\epsilon}{2}$	$\frac{7}{4}-\frac{\epsilon}{2}$	0	$\frac{7}{4}-\frac{\epsilon}{2}$
$\hat{a}$	$\frac{1}{4}-\frac{\epsilon}{2}$	$\frac{7}{4}+\frac{\epsilon}{2}$	1	$\frac{7}{4}-\frac{\epsilon}{2}$
$\hat{b}$	$\frac{1}{4}+\frac{\epsilon}{4}$	$\frac{7}{4}-\frac{\epsilon}{4}$	$\frac{1}{4}+\frac{\epsilon}{4}$	$\frac{7}{4}-\frac{\epsilon}{2}$
$\hat{c}$	$\frac{1}{4}-\frac{\epsilon}{4}$	$\frac{7}{4}+\frac{\epsilon}{4}$	$\frac{3}{4}+\frac{\epsilon}{4}$	$\frac{7}{4}-\frac{\epsilon}{2}$

Tab. 1: The evolutionarily stable strategy σ_ϵ for an alphabet $\left\{\phi, \hat{a}, \hat{b}, \hat{c}\right\}$ and a fixed lying cost equal to 1 (which is paid in case a player "promises" to play one action and instead plays another action; this cost is multiplied by ϵ in the game G_ϵ). To simplify the presentation, the table neglects terms that have a magnitude of $o(\epsilon^2)$. The exact probabilities of the messages in σ_ϵ are: $Pr(\hat{c}) = \frac{1}{4} - \frac{\epsilon\cdot(2+\epsilon)}{8-4\epsilon^2}$, $Pr(\hat{b}) = \frac{1}{4} + \frac{\epsilon\cdot\left(2+\epsilon-\epsilon^2\right)}{(8-4\cdot\epsilon^2)\cdot(1+\epsilon)}$, $Pr(\hat{a}) = \frac{1}{4} - \frac{\epsilon}{2} + \frac{\epsilon^2}{8-4\epsilon^2}$.

(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.

Finally, minor adaptations to the proofs of Schlag (1993) and Banerjee & Weibull (2000) (which show why σ^* is evolutionarily stable in G_0) imply that σ_ϵ is evolutionarily stable in G_ϵ. The intuition is as follows: (1) each strategy that induces different behavior at the second stage is strictly outperformed (because the second-stage behavior of σ_ϵ is the unique strict best reply against σ_ϵ), and (2) each strategy $\tau \neq \sigma_\epsilon$ that differs in the distribution of messages at the first stage is strictly outperformed in any nearby post-entry population. The reason for the latter observation is that if a message m is sent with higher (lower) probability in τ than in σ_ϵ, then it is played more (less) often by the opponent in the post-entry population relative to the pre-entry frequency,

and, as a result, conditional on sending message m, there is a higher (lower) probability of facing an opponent who sends the same message and playing the worse action profile (c, c). As a result, the expected payoff conditional on sending message m is lower (higher) in the post-entry population than in the pre-entry payoff. Thus, the "mutant" strategy τ assigns higher (lower) probabilities to messages that induce lower (higher) expected payoffs, and, as a result, it is outperformed by the incumbent strategy σ_ϵ.

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.[5]

IV. Discussion of Discontinuity

Observe that various kinds of discontinuities (unrelated to lying costs) yield the same sharp prediction as DW's lexicographic lying costs. 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.

I conclude by arguing why discontinuity of preferences may not be an appealing assumption in an evolutionary setup. The standard approach in the evolutionary literature (see, e.g., Maynard Smith, 1982) is that preferences are fitnesses (reproductive success). As argued by Samuelson & Swinkels (2003), the fitness function should allow continuous trade-offs over various objectives, and lexicographic constructions are interesting only if they provide convenient approximations for cases in which preferences are continuous, but some considerations are much more important than others. This note shows that

---

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

lexicographic lying costs do not provide a convenient approximation in the current setup.

Another approach is that only material preferences are fitnesses, while lying aversion has no direct effect on fitness, and hence it might be lexicographic. However, I demonstrate that even in this case it is problematic to perform a static stability analysis with lexicographic preferences. The problem arises because a dynamically stable state does not have to be a Nash equilibrium of the lexicographic game (unlike in continuous games, in which all the dynamically stable state are Nash equilibria; see Nachbar, 1990).

Consider the following simple example: agents in a large population are randomly matched and play at each round a Hawk-Dove game with lexicographic preferences for playing *hawk* (as described in Figure 3). The type of each agent determines his action - either a *hawk* or a *dove*. The frequency of types evolves according to the replicator dynamics (Taylor & Jonker, 1978), where the material payoffs are fitnesses. Every so often, a randomly chosen agent switches his type if he strictly prefers the other type (given the aggregate behavior of the population). One can see that the population converges from any initial state towards $(0.5, 0.5)$, and then remains forever in a small neighborhood of this state, even though $(0.5, 0.5)$ is not a Nash equilibrium of the lexicographic game (which admits no symmetric Nash equilibria).[6]

Similarly, one can extend this dynamic analysis to coordination games with pre-play communication and show that strategy σ^* (or a small neighborhood around it) is stable in various plausible evolutionary dynamics regardless of the exact way in which the small lying aversion is modeled; in particular, σ^* is dynamically stable (though not a Nash equilibrium) with a lexicographic lying aversion.

---

[6] The argument would remain the same even if there were symmetric Nash equilibria (e.g., $(0.5, 0.5)$ also remains stable if one adds a third strategy, which is a pure symmetric Nash equilibrium).

	dove	*hawk*
dove	2	1
hawk	3	0

π (material payoffs)

	dove	*hawk*
dove	0	0
hawk	1	1

w (secondary payoffs)

Fig. 3: Hawk-Dove game with lexicographic preferences for playing *hawk*. The game admits no Nash equilibria. The strategy $(0.5, 0.5)$ is stable in plausible evolutionary dynamics.

References

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

Binmore, Kenneth., & Samuelson, Larry. 1992. Evolutionary stability in repeated games played by finite automata. *Journal of Economic Theory*, **57**(2), 278–305.

Chen, Ying, Kartik, Navin, & Sobel, Joel. 2008. Selecting cheap-talk equilibria. *Econometrica*, **76**(1), 117–136.

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

Demichelis, Stefano. 2013. Efficient coordination in repeated games: Behavioral maxims. Unpublished, http://ssrn.com/abstract=2322757.

Demichelis, Stefano, & Weibull, Jörgen W. 2008. Language, meaning, and games: A model of communication, coordination, and evolution. *The American Economic Review*, **98 (4)**, 1292–1311.

Jansen, M.J.M. 1981. Regularity and stability of equilibrium points of bimatrix games. *Mathematics of Operations Research*, **6**(4), 530–550.

Maynard Smith, John. 1982. *Evolution and theTtheory of Games*. Cambridge University Press.

Maynard Smith, John, & Price, George R. 1973. The logic of animal conflict. *Nature*, **246**, 15.

Nachbar, John H. 1990. Evolutionary selection dynamics in games: Convergence and limit properties. *International Journal of Game Theory*, **19**(1), 59–89.

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. University of Bonn Discussion Paper B–242.

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

Thomas, Bernhard. 1985. On evolutionarily stable sets. *Journal of Mathematical 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 GameTheory*. The MIT Press.