L

Learning and Evolution of Social Norms*

Rodrigo Harrison[1], Mauricio G. Villena[2]
[1]Instituto de Economía, Pontificia Universidad Católica de Chile, Macul, Santiago, Chile
[2]Escuela de Negocios, Universidad Adolfo Ibáñez, Peñalolén, Santiago, Chile

Synonyms

Conventions; Customs; Learning (and evolution) of social norms; Social rules

Definition

Social norms can be understood as standards of behavior that are based on widely shared beliefs of how individual group members ought to behave in a given situation (Horne 2001) (see Voss 2001). *The group can be a family, an organization, or a society. Members may follow the norm voluntarily if their individual preferences are consistent with the normative behavior, or they might be enforced by punishment if the differences between individual preferences and normative behavior result in a violation of the norm.*

While social norms can be modeled using alternative theoretical learning models (see for instance, Young 1998), in this brief review we focus on the basic elements of evolutionary game theory (EGT), which has been widely used to formally study the conditions under which social norms may emerge and be established in society (Weibull 1996; Vega-Redondo 1996).

Theoretical Background

One of the key research questions regarding social norms is how they can emerge in different social environments. While norms are typically taken as given in much of the economic and sociological literature, EGT tools allow us to formally model social norms dynamics. Indeed, when EGT concepts, which have thus far been mainly applied in biology to analyze animal behavior, are applied to the socioeconomic context they are mostly used to study the development of social norms in society. As Mailath (1998: 1348) explains: *Since evolutionary game theory studies populations playing games, it is also useful for studying social norms and conventions. Indeed, many of the motivating ideas are the same.*"

EGT does not assume optimizing behavior per se, though it does retain the idea that individuals adjust their behavior in response to persistent differentials in material incentives. In other words, while agents do pursue individual material payoffs, which in these models represent evolutionary success, i.e., fitness, they are not always in a position to obtain straightaway the payoffs an optimizing agent would obtain. This may be due to *social norms of behavior* restricting the course of action of individuals, in such a way as to prevent them from adjusting their behavior toward the optimal strategy immediately (it takes time to change a social norm), or it may be just because individuals do not realize what is the best strategy at once. However, if this situation persists in time, some individuals will start adopting the more efficient strategy and therefore receiving a higher payoff than the rest of the population. In the long run, the rest of the population will start imitating this more profitable course of action. Thus, the incumbent norm will be replaced by this new, more successful, strategy, which in time will be adopted as the new norm of behavior in the population. *In this sense, evolutionary models can be interpreted as models of learning*, where individuals learn about the game on a trial-and-error basis, and where more efficient behavior, in evolutionary terms, tends to be imitated.

The evolutionary approach to social norms has proved to be complementary to the extensive economic and sociological literature on norms. In particular, the concepts of Evolutionary Stable Strategy (ESS) and Replicator Dynamics (RD) are the more basic tools used in the analysis of social norm dynamics. A typical framework in which these concepts are applied is one where individuals are *repeatedly drawn at random from a large population* to play a symmetric two-person game. An ESS is a strategy, which, if adopted by a population of agents, cannot be invaded by any alternative strategy that is initially rare. An

---

* This review is based on Villene and Villena (2004).

Norbert Seel (ed.), *Encyclopedia of the Sciences of Learning*, DOI 10.1007/978-1-4419-1428-6,

ESS is an equilibrium refinement of the Nash equilibrium (NE). Hence, an ESS is an NE which is "evolutionarily" stable, meaning that once it is fixed in a population, natural selection alone is sufficient to prevent alternative (mutant) strategies from successfully invading.

The criterion of evolutionary stability emphasizes the role of mutations in an evolutionary process – a mutation mechanism. However, a selection mechanism is also required that favors some varieties over others. This is precisely the role of *the RD*, which does not embrace any mutation mechanism at all. Robustness against mutations is indirectly taken care of by dynamic stability criteria. The replicator permits the analysis of a genuinely diverse range of behavior (i.e., a polymorphic profile of strategies) as opposed to the concept of ESS, which makes good theoretical sense only when it represents a monomorphic situation.

In order to better exemplify the modeling of social norms using EGT, let us now formalize the concept of replicator dynamics. Let us consider a game with n pure strategies. If an agent playing strategy i meets an agent adopting strategy j, the payoff to i is π_{ij}. Assuming that $p = (p_1, \ldots, p_n)$ is the probability of meeting each type in the population, the expected payoff to an i-player is then $\pi_i(p) = \sum_{j=1}^{n} p_j \pi_{ij}$. Hence, the average payoff in the game becomes $\overline{\pi}(p) = \sum_{i=1}^{n} p_j \pi_i(p)$. Consequently, in this setting the RD in a polymorphic population is given by

$$\frac{dp_i}{dt} = p_i(\pi_i(p) - \overline{\pi}(p)) \quad (\text{all } i), \tag{1}$$

where $\overline{\pi}(p)$ denotes the average fitness of the population. Equation 1 is called the replicator equation.

From Eq. 1 it transpires that according to the replicator equation, the strategies that grow are those that perform better than average, and that generally the best performing strategies grow the fastest. In this framework, an NE is a stationary point of the dynamic system. On the other hand, each stable stationary point is an NE and an asymptotically stable fixed point is a perfect equilibrium. Moreover, evolutionary stability becomes a sufficient (but not necessary) condition for asymptotic stability if only pure strategies can be inherited.

In what follows we present a simple application of the concept of RD in the modeling of social norms.

Cooperative Versus Noncooperative Social Norms

Let us consider a doubly symmetric two-player game with two pure strategies and payoff matrix:

$$A = \begin{array}{cc} & \begin{array}{cc} C & NC \end{array} \\ \begin{array}{c} C \\ NC \end{array} & \begin{pmatrix} 6 & 0 \\ 4 & 3 \end{pmatrix} \end{array} \tag{2}$$

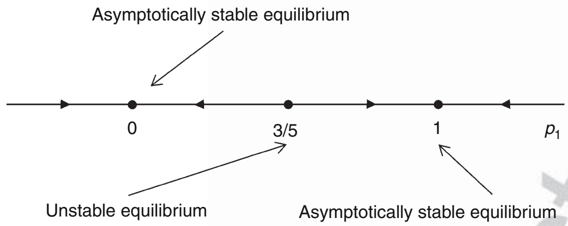
Since C-C > NC-C and NC-NC > C-NC, we have that this game is a coordination game. We can think of this game, for example, as a two-person common property resource game in which the common resource is an inshore fishery exploited by two fishermen, and that each agent can exploit the fishery choosing between two different levels of effort, e.g., fishing effort might be measured by the number of standardized vessels operating in a fishery during a particular day. In particular, here we consider a low fishing effort, C, which we call cooperative, and a high fishing effort, NC, which we call noncooperative. From the payoff matrix it can be inferred that if both players choose the cooperative fishing effort, they will be better off than if both players use the noncooperative fishing effort, i.e., a payoff of 6 against one of 3. This could be the case if both players adopt the large fishing effort, the stock could be harvested to a level where extraction gets more difficult and therefore not as profitable as in that case where both fishermen use the low fishing effort giving thus more time to the stock to recover. Playing in a cooperative manner is not without its risks, since if one plays cooperatively and the other noncooperatively the player can end up receiving nothing while his/her opponent gets a payoff of 4. In terms of our example this makes sense, since, as we have assumed here, cooperation means using a lower effort to exploit the resource, which, depending on the relation between efforts, can imply that the other individual using a larger effort can be able to harvest the stock down to a level where it is not more profitable for individual 1 to continue in business or even can harvest the entire stock and there will then be nothing left for individual 1. In any case the cooperative individual will lose revenue by using a lower effort than the other individual who uses a larger effort. Finally, if considering the risk of playing cooperative both players decide to use the noncooperative fishing effort then they get a return of 3, which is lower than that obtained if both players decide to play cooperative, getting a return of 6.

Consequently, according to the basic principles of traditional game theory, it is evident that here both players (strictly) prefer the strategy profile C-C, which gives payoff 6 to each player. Indeed, C-C is a strict NE. However, the pure strategy profile NC-NC is also a strict NE, resulting in payoff 3 to each player. If one player expects the other to play strategy NC with sufficiently high probability, then his or her unique optimal action is to play strategy NC as well. The game has a third Nash equilibrium, which is mixed. This corresponds to the symmetric pair (x, x) where $x = 3/5, 2/5$, the payoff to each player in this equilibrium being 18/5. All Nash equilibria are clearly perfect: Two are strict, and one is interior.

Now we suppose that within the population there is a proportion of players using the cooperative strategy C, and other of players adopting the noncooperative strategy NC which we denote p_1 and p_2 respectively. We also have the identity $p_1 + p_2 = 1$. Thus, we get the following replicator equation:

$$\dot{p}_1 = p_1(1 - p_1)(5p_1 - 3). \qquad (3)$$

In order to see how solutions of (3) change over time, let us draw the associated phase portrait.



Hence, it is clear that the steady states $p_1 = 0$, and $p_1 = 1$ are asymptotically stable, while $p_1 = 3/5$ is unstable. In other words, if one starts to the left of 3/5, i.e., where the population playing C, cooperative, is a rather small proportion of the total population, the system tends to the steady state $p_1 = 0$, i.e., the cooperative population is wiped out. If one starts anywhere to the right of 3/5, the system tends to the steady state $p_1 = 1$, i.e., the population adopting the noncooperative strategy is wiped out. The unstable equilibrium at $p_1 = 3/5$ is the boundary, or separatrix, between the region of attraction of $p_1 = 0$ and that of $p_1 = 1$.

In this example we have used the concept of the RD to analyze the evolution of a population where there is a proportion of players using the cooperative strategy C, and other of players adopting the noncooperative strategy NC. We can interpret these two strategies as two different *social norms*, one cooperative and the other noncooperative. The result presented here clearly shows that in this particular example, the emergence of one social norm as the dominant one depends on the initial number of people who subscribe to each norm of behavior. In particular, if, initially, less than 60% of the total population adheres to the cooperative social norm, then the noncooperative one will become the dominant in the long run and people adopting the cooperative strategy will be wiped out. Otherwise, the cooperative social norm will become the dominant and the population adopting the noncooperative strategy will be wiped out. This clearly points to the importance of initial conditions, which somehow determine future developments, and to the relevance of studying the historical context when analyzing social norms in specific settings.

From this simple example it can also be inferred that there can be some conflicts between social norms and that some norms of behavior are not always positive in terms of society's welfare. Indeed, it can be noted that the RD does not reject the socially inefficient profile NC-NC, i.e., where players use the noncooperative fishing effort. In this sense a socially inefficient norm of behavior, e.g., always use strategy NC when meeting, may be evolutionarily (asymptotically) stable. Certainly, depending on the initial population adhering to the cooperative social norm, the noncooperative convention can become the dominant in the long run and people adopting the cooperative strategy will be wiped out.

Important Scientific Research and Open Questions

Finally, there are many interesting research projects related to learning and the evolution of social norms that could be highlighted: (a) the *"economic anthropology"* of Herbert Gintis and Samuel Bowles, which is based mainly on EGT tools, reviewing topics such as *the importance and origins of reciprocity, fairness and cooperation in primitive societies, and the measure of social norms and preferences using experimental games* (see Bowles 2004; Gintis 2000); (b) the work on the *"evolution of preferences"* as developed by Werner Güth (see Heifetz 2005); (c) the study of the *"evolution of social norms in specific economic settings"* – an excellent example here is provided by the work of Sethi and Somanathan (1996) which examines the problem of the exploitation of a common property resource within an evolutionary game theoretic framework– and (d) the *"evolution of rationality,"* where social norm-guided behavior, which is associated with a nonrational conduct, is contrasted with rational, optimizing, behavior (see, Banerjee and Weibull 1994) (see, Vega-Redondo 1996: 85).

Cross-References

► Learning and Evolutionary Game Theory

References

Bowles, S. (2004). *Microeconomics: Behavior, institutions, and evolution.* Princeton: Princeton University Press.

Gintis, H. (2000). *Game theory evolving*. Princeton: Princeton University Press.

Heifetz, A., Shannon C., Spiegel, Y. (2007). The dynamic evolution of preferences. *Economic Theory*, 32(2), 251–286.

Horne, C. (2001). Sociological perspectives on the emergence of social norms. In M. Hecheter & K. D. Opp (Eds.), *Social norms*. New York: Russell Sage.

Mailath, G. J. (1988). Do people play nash equilibrium? Lessons from evolutionary game theory. *Journal of Economic Literature, 36*(3), 1347–1374.

Sethi, R., & Somanathan, E. (1996). The evolution of social norms in common property resource use. *The American Economic Review, 86*(4), 766–788.

Vega-Redondo, F. (1996). *Evolution, games, and economic behavior.* New York: Oxford University Press.

Villena, M. G., & Villena, M. J. (2004). Evolutionary game theory and Thorstein Veblen's evolutionary economics: Is EGT Veblenian?. *Journal of Economic Issues*, *38*(3), 585–610.

Voss, T. (2001). Game theoretical perspectives on the emergence of social norms. In M. Hecheter & K. D. Opp, (Eds.) *Social norms*, Nueva York: Russell Sage Foundation.

Weibull, J. (1996). *Evolutionary game theory.* London: MIT Press.

Young, H. P. (1998). *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton: Princeton University Press.