Cristina Bicchieri and
Alexander Funcke

Norm Change: Trendsetters and Social Structure: Appendix

THE SOCIAL NORM GAME

The following game $G(k_i)$ represents a situation in which a social norm exists, and a player has to decide whether to follow it (F) or to transgress (T).

		Follow (*F*)	Transgress (*T*)
$G(k_i)$:	Follow (*F*)	$a_{11}(k)$	$a_{12}(k)$
	Transgress (*T*)	$a_{21}(k)$	$a_{22}(k)$

Let $\pi : k_i \times \sigma_1 \times \sigma_2 \rightarrow U$ be the payoff function for the game, where k_i is the subjective norm sensitivity of individual i, and σ_1, σ_2 are the row and column player's strategy, respectively. The utility a player receives from successful coordination on the strategy prescribed by the norm increases with her sensitivity $\frac{da_{11}(k_i)}{dk_i} > 0$; for all other strategy profiles, the utility typically decreases (but may remain constant) with an increase in sensitivity.

With π being the payoff function for $G(k_i)$, let

$$\pi(k_i, \sigma_1 \sigma_2) = h(\pi(k_i, \sigma_1, \sigma_2), a_i),$$

where a_i is the individual's risk-sensitivity and perception combined, and h is an absolute risk function (e.g. exponential utility) from the payoff and a_i to a risk-adjusted utility.

Given a belief (empirical expectation) about the ratio of F-choosers, x, in the relevant (sub) population, the expected utility of the risk adjusted game is

$$E_i[U(F) \mid x] = x\pi(k_i, F, F) + (1 - x)\pi(k_i, F, T)$$
$$E_i[U(T) \mid x] = (1 - x)\,\pi(k_i, T, T) + x\pi\,(k_i, T, F).$$

Now, players interact on a network and have to form beliefs about the ratio of followers in their relevant population. An important element of expectation formation is the importance a player assigns to her neighbors' actions versus the actions of the general population. The relative weight assigned to a local network versus the entire population for a specific norm is captured by the parameter β. If β is equal to zero, this implies that all influence on our expectation stems from the network at large. Conversely, if β equals one, then all influence stems from the individual's local network (her direct neighbors). In what follows, for simplicity, we are assuming a single β for the whole population, even if in reality the may vary between individuals.

Let's see how β matters as we are forming empirical expectations. Let x_i^0 be the weighted ratio of followers (F) in the neighborhood of i and in the population at-large at the current time, then i's expectation about x_i^0 is

$$E_i\,[x_i^t] = \beta(\sum_{j \in N_i} \frac{I^t{}_j\,(F)}{|(N_i|}) + (1 - \beta)\,(\sum_{j \in P-_i} \frac{I^t{}_j\,(F)}{|P-1|}\,,$$

where β is the relative weight assigned to the neighbors rather than the population, P is the whole population, N_i is the neighborhood of individual i, and $I^t{}_j(\sigma)$ is an indicator function for whether individual j chose the action σ at a previous time step.

In deciding what action to choose, players are not myopic but instead consider expected behavior over time. When discussing trendsetters, we highlighted a particular characteristic they share, i.e. self-efficacy. A trendsetter is an early adopter who thinks she may be able to sway the rest of population to adopt her behavior. In the model, the utility of the choice to follow the norm is calculated with an expectation that the population will play the "follow" strategy indefinitely. The utility of the choice to transgress, however, depends on how the player thinks her choice will influence others' choices. Self-efficacy is a key aspect of expectation formation in our model, as a person with a high level of self-efficacy will believe that her choice will have a significant effect on others' choices to transgress. Conversely, people with low self-efficacy will believe that their choices are inconsequential.

Adoption of new behavior in a population typically increases along an S-curve (Rogers 1962); thus it is not far-fetched to assume that an individual's mental model of the dynamics of adoption is an S-curve. The difference in steepness of an individual's S-curve represents her perceived self-efficacy, i.e., her belief about how fast she will be followed by the rest of the population as she transgresses the norm. Let

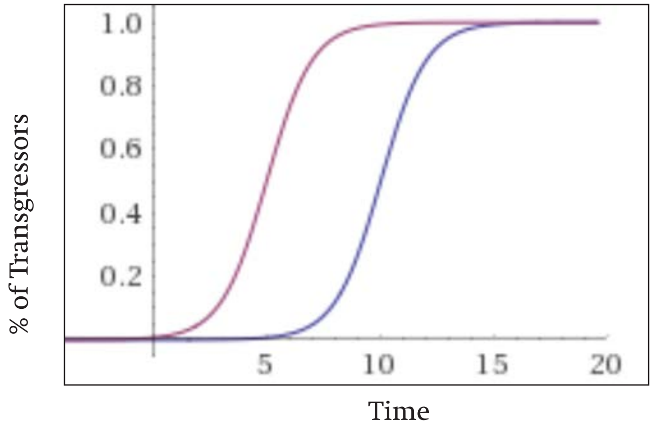
$$S_f(t) = \frac{1}{1 + e^{(f-t)}}$$

be the S-curve function that maps, for an individual with self-efficacy $f \in [1, \infty]$, time t to her expected number of T-followers at t. The functional form implies that the mental model's tipping point occurs f time steps after the first move. In Figure 1 we can see how the first S-curve has a tipping point, f, at 5, and the second at 10.

The trendsetter who chooses to transgress the standing norm will have the following expectation about the ratio of T-choosers in the relevant population further at time $t+1$:

$$E_i[1 - x_i^{t+1}] = s_{f_i}\left(s_{f_i}^{-1}(1 - E_i[x_i^t])\right).$$

Figure 1.



Note that we use the inverse S-curve to determine how far along the mental model's S-curve a player currently considers herself given the observed ratio of T-choosers in the previous period.

However, if the would-be trendsetter chooses not to transgress, we assume that her expectation of the ratio of T-choosers in the population is zero:

$$E_i[1 - x_i^t] = 0.$$

At the time of each choice a player will always consider the expected dynamics, and its expected utility, till the end of time, τ,

$$E_i\ [V(\sigma_i, \tau)] = \Sigma_{t=0}^{\tau}\ \gamma^{\tau}\ (E_i[x_i^t]\pi\ (\sigma_i,\ F) + (1 - E_i[x_i^t]\pi(\sigma_i,\ T)),$$

for $\sigma_i \in$ {T, F} and where $\gamma \in [0,1]$ is the time discounting factor.

An individual i will thus choose to transgress (T) rather than follow the norm (F) if and only if

$$E_i[V(T,\ \tau)] \geq E_i[V(F,\ \tau)].$$

We call the first person(s) in the population that has/have a preference for transgression "trendsetter(s)."