The Flouted Naming Game: Contentions and Conventions in Culture

Harold. P. de Vladar
Konrad Lorenz Institute
Klosterneuburg, Austria
MTA Ecological Research Centre, Hungarian Academy of Sciences
Tihany, Hungary
Harold.deVladar@kli.ac.at

ABSTRACT

Naming Games are AI platforms that can account for how conventions in language and culture are achieved. This approach, however, does not account for other cultural features such as contentions. By contentions it is meant that agents learn other agents' traits but decide not to transmit these. This work introduces a version termed Flouted Naming Game which allows convergence to stable states of cultural contentions as an alternative outcome of the Naming Game. Which regime is achieved depends on a basic asymmetry on the cognitive reward of two opposing cultural forms. Moreover, it is found that there is a sharp phase transition between the two behavioural strategies. The transition point is sensitive to population size: larger populations can maintain contentions on a wider range of parameters than smaller populations.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; Machine learning; • **Applied computing** → **Sociology**; *Life and medical sciences*;

KEYWORDS

Language evolution, cultural change, phase transition, cognition

ACM Reference Format:
Harold. P. de Vladar. 2018. The Flouted Naming Game: Contentions and Conventions in Culture. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3205651.3208265

1 CULTURAL CONVENTIONS AND CONTENTIONS

The *Naming Game* is an AI framework to study language change [12]. It consists of a population of agents that interact in pairs to communicate features of the environment through utterances. A central purpose of naming games is to understand how language is conventionalised [4, 5]. By conventionalisation it is meant that agents converge to the use of a single construction to refer to a specific concept. However, many features in language and culture do not conventionalise and a diversity of constructions are maintained in the population. More generally, this also applies to several cultural features: most contemporary cultures do not fully conventionalise regarding ideals, fashion, musical preferences, religion, etc. Instead, contentions play a fundamental role in maintaining behavioural diversity. This is a central yet unadressed aspect of human societies, at least from the perspective of language games and other cognitive theories of language and culture [3].

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3208265

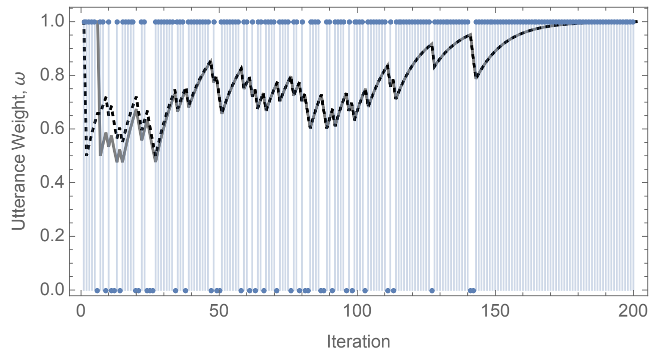

Figure 1: A minimal (2-player) naming game indicating the dynamics of the utterances (blue dots) and of the association weights of each player (black-dashed and gray-solid lines). This example results in conventionalisation, as any naming game with canonical update will. Both utterances have the same chance (50%) of being fixed. $\lambda = 0.09$

This work addresses some minimal sufficient conditions for cultural contentions based on the Naming Game. In some versions, many agents interact without embodiment, i.e. in a digital framework [12]. Embodiment is necessary in order to account for variability arising from perceptive differences (e.g. subjectivity) [13]. However, digital versions, although often lacking this perceptual variability, allow understanding other features, such as those pertaining to the size of the population of agents, their spatial arrangement and other kinds of population dynamics [11]. In its minimal version only two agents interact, a suitable set up to compare with embodied systems where substantial complexity of language can evolve [2]. Here, the digital approach is taken as a first step.

Minimal Naming Games. In the standard version of the naming game [2], which here is referred to as "Canonical Naming Game" (CNG), the agent that takes the speaker role (S) chooses an object from an environment and identifies a property (or set of properties)

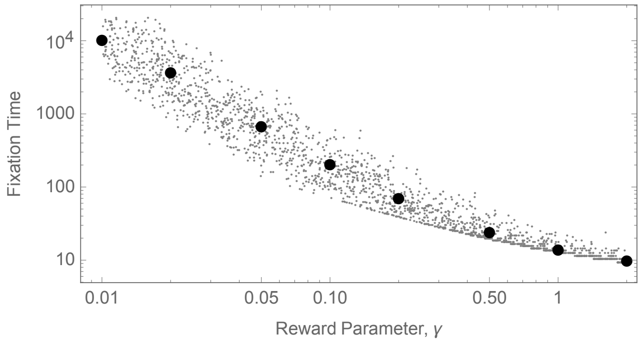

Figure 2: In minimal (2-player) naming games with canonical update (Eq. 1) the time to conventionalisation decreases with the learning rate λ.

of it. Then, the speaker renders an utterance to refer to that property (either chooses a word from a list associated to the properties or, if no word exists, it invents one) and transmits it to the agent that has the hearer role (H). In embodied systems transmission is acoustic, but in digital agents the information is directly copied. Then, H checks in its list whether it knows the word and if so it refers to it (e.g. by pointing) and both agents transmit an agreement signal. If H does not know the word it transmits a no-agreement signal. Then S points at the object that has the property in question and H learns what properties of the referential object are associated with the new utterance.

Each agent may have several words for any given property (either invented by the agent or learnt from another agent). For each word, any given agent has a weight that associates it to the property [8, 14]. On each successful round the weight of the utterance that was used is updated (reinforced) in a Hebbian way. For example, for two utterances, the canonical update in each agent α is $\omega \rightarrow \omega + \gamma$ when the focal utterance u_1 occurs and $\omega \rightarrow \omega - \gamma$ when the alternative utterance u_2 occurs. Normalising this is equivalent to

$$\Delta\omega_{(U)} = \begin{cases} \lambda(1-\omega) & \text{if } U = u_1 \\ -\lambda\omega & \text{if } U = u_2 \end{cases} \tag{1}$$

where $\lambda = \gamma/(1+\gamma)$ is the learning rate. Note that some instances of the language game do not normalise the weights.

In general, each agent α has an $\omega_{\alpha,i}$ for each utterance u_i, but for the sake of the explanation the agent subscript is dropped, keeping in mind that each agent has its own weight system. The weight ω has two functions. First, it is the association between an utterance and a concept. (Here, however, the concept is left implicit.) Second, it is the probability that the focal utterance is chosen to be used when the agent takes the S role.

The minimal naming game is a non-embodied simplification where there are only two players and the referential is tacitly assumed. The minimal CNG focuses only on the dynamics of the utterances. Figure 1 depicts a realisation of a minimal CNG. Unless otherwise stated, in this work the CNG refers to its minimal version.

The CNM has only one parameter, namely, the learning rate λ [8]. Invariably, the dynamics always result in conventionalisation with equal probability of fixation (absorption) for each utterance.

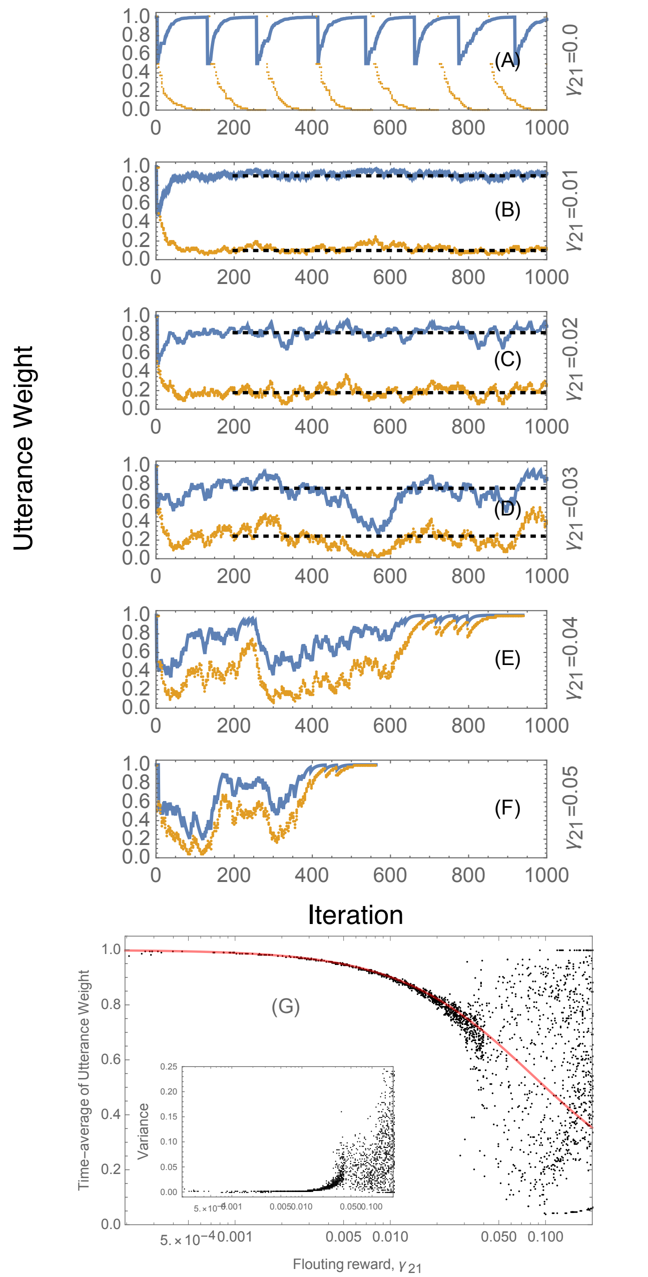

Figure 3: Association weights in Flouted Naming Games. Individual based simulations are of only two agents. (A) $\gamma_a = 0$; (B) $\gamma_a = 0.01$ (C) $\gamma_a = 0.02$; (D) $\gamma_a = 0.03$; (E) $\gamma_a = 0.04$; (F) $\gamma_a = 0.05$. Yellow: weight for the focal utterance in its cognate type. Blue: weight for the focal utterance in the opposite individual type. Black dashed lines: theoretical expectation. (G) Mean (and variance, inset) of the stationary associative weight ω in an ensemble of simulations with different (randomly chosen) values of γ_a. Red line: theoretical expectation; black dots: individual based simulations. In all simulations (A-G) $\gamma_0 = 0.1$. There are only two types of agents and two initial utterances.

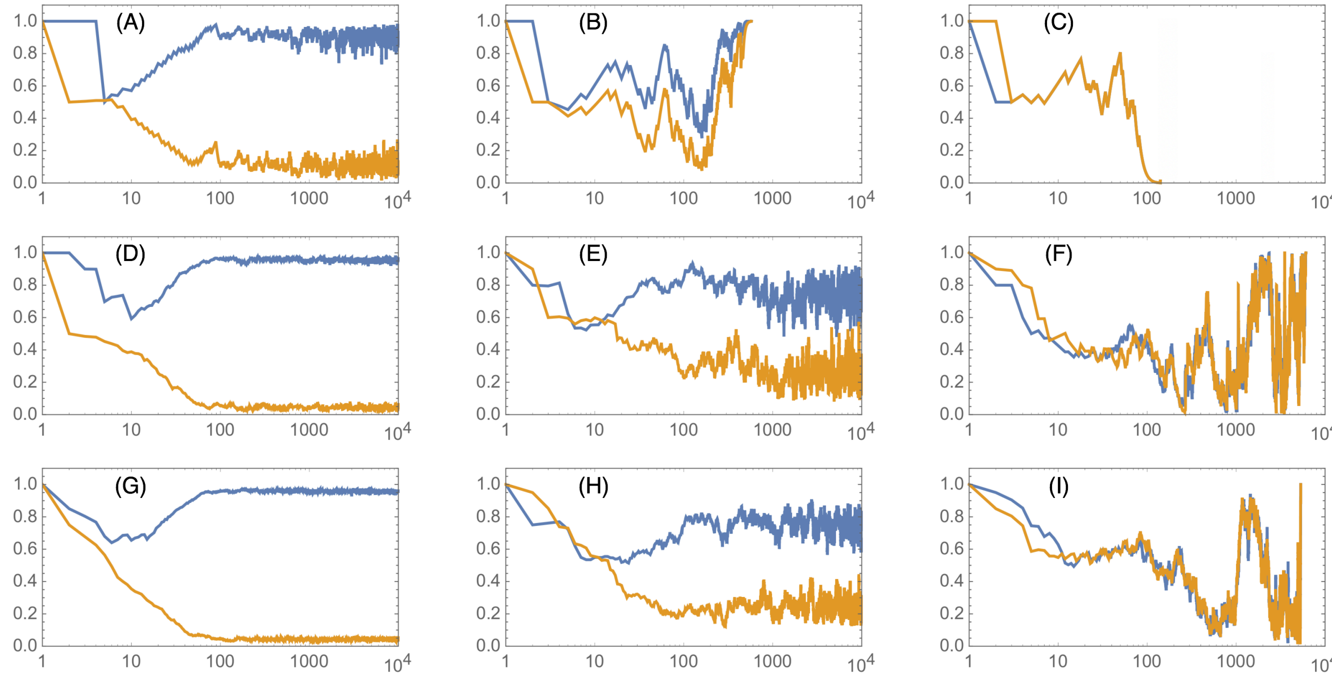

Figure 4: Association weights in many-player Flouted Naming Games. Yellow: population average of the weight for the focal utterance in its cognate type. Blue: population average of the weight for the focal utterance in the opposite individual type. (A) $\gamma_a = 0.01, N = 2$; **(B)** $\gamma_a = 0.01, N = 10$; **(C)** $\gamma_a = 0.01, N = 20$; **(D)** $\gamma_a = 0.05, N = 2$; **(E)** $\gamma_a = 0.05, N = 10$; **(F)** $\gamma_a = 0.05, N = 20$; **(G)** $\gamma_a = 0.1, N = 2$; **(H)** $\gamma_a = 0.1, N = 10$; **(I)** $\gamma_a = 0.1, N = 20$. **Otherwise as in Fig. 3.**

This result is also true for many utterances, provided they have equal starting probability of being uttered.

2 FLOUTED NAMING GAME

As explained in the introduction, the CNG cannot explain contentions in language. For example, why do the British use "football" and the Americans "soccer" to refer to the same sport, whilst nevertheless knowing the meaning of the alternative word? One explanation is demographic isolation: Britons use the word "football" and even when people from the UK know what "soccer" is, the former term is reinforced much more often than the latter. The contrary happens for US Americans.

A second (non-exclusive) alternative is introduced in this work, namely that rewarding preferences are asymmetric. In the previous example, a Briton may reinforce but less strongly (e.g. with lower learning rate) the American term, as compared with the reinforcement strength it uses for the British term. A necessary addition in this framework is that agents must have an identity that correlates them to a (preferred) utterance. Mathematically this is modelled with an update function $\omega \longrightarrow \omega + \gamma_0$ when the cognate utterance u_0 is used and $\omega \longrightarrow \omega - \gamma_a$ when the alternative utterance u_a is used. γ_a is named "flouting reward". Normalising:

$$\Delta\omega^0_{(U)} = \begin{cases} \lambda_0(1-\omega^0) & \text{if } U = u_0 \\ -\lambda_a\omega^0 & \text{if } U = u_a \end{cases} \tag{2}$$

where, as above, $\lambda_i = \gamma_i/(1+\gamma_i)$ and it is assumed that $\lambda_a < \lambda_0$. Note that each agent has an independend reward system. Which is the cognate utterance for each agent, depends on their type.

The NG with update (2) is dubbed *Flouted Naming Game* (FNG).

Through the rest of the paper it will be shown that this rewarding asymmetry is enough to maintain contentions. Moreover, it will also be shown that to reach conventionalisation it is not necessary that $\gamma_a = \gamma_0$ but, instead there is a threshold that sharply separates two qualitative regimes: contention and conventions.

3 SIMULATIONS

A simulation approach is taken and consists of the following steps:

(1) **Initialisation.** A population of N (≥ 2, even) agents is created where each agent is initialised with three features: Type t (a static label), a starting list of ν utterances $\mathcal{U}_\nu$, their associated weights (a vector $\{\omega_1, \ldots, \omega_\nu\}$), and the reward parameters (a vector $\{\gamma_1, \ldots, \gamma_\nu\}$). In this work only one type (canonical game) or two types (flouted game) are defined. More types can be considered but in this study only two utterances and two types are considered. Each agent is initialised with one specific type and utterance.

(2) **Interaction.** The N players are randomly paired. In each pair, one agent is given the role of speaker, S, and one the role of hearer H; then S chooses to utter an element $U \in \mathcal{U}^{\mathsf{S}}$ with probability $\omega_{\mathsf{S},i}$, $i = 1, 2, \ldots, \nu$.

(3) **Agreement.** If H recognises the utterance, i.e. if $U \in \mathcal{U}^{\mathsf{H}}$, it signals agreement (T); otherwise it signals no-agreement (F).
(4) **Update.** If there is agreement each agent updates its weights according to Eq. 1 or 2 –depending on whether the game is canonical or flouted. If there is no-agreement then there is no update but H adds the new utterance u_a to its list and initialises its weight with probability $\omega_{\mathsf{H},a} = 1/\nu$, where ν is the number of utterances in its list.
(5) **Forgetting.** If $\exists i\ \omega_{\alpha,i} < \epsilon$ then the weight is set to zero: $\omega_{\alpha,i} = 0$ and its corresponding utterance u_i is deleted from the list.
(6) **Halting.** Steps 2-6 are iterated R times or until all individuals have only one and the same utterance, u_f, in their lists. When the latter criterion is fulfilled, u_f is said to be fixed.

For the purposes of this work, the weights of each utterance in each agent and the utterance used by S are tracked in each round.

4 RESULTS

A first aspect to note in the CNG is how fast an utterance fixes, i.e. the time some utterance to reach $\omega_{\alpha,i} = 1\forall\alpha$. This is termed as "fixation time". According to intuition, stronger learning rates results in shorter fixation times, at least in average. Figure 2 shows results for an ensemble of realisations of CNG. The observed variability is due to the stochastic nature in which the agents interacts. Otherwise, the CNG has a time scale defined by the learning rate λ and there is no further dynamical or behavioural richness.

Figure 3A shows that in the FNG diversity can be maintained even when agents learn the alternative utterances. First, if there is no reward for the alternative utterance ($\gamma_a = 0$) the agents inevitably forget it. Eventually they re-learn it and, again, forget it. This results on semi-regular cycles and no utterance ever fixes.

If the flouting reward is non-zero but small, agents show preference for their cognate utterance. As the flouting reward increases, the preferences for the cognate utterance diminishes and the fluctuations increase (Fig. 3B-D). Yet, the contention behaviour can be maintained for a range of values. The preference weights fluctuate around a well-defined expected value but no utterance can fix in this regime. Therefore there is no expected fixation time.

In the experiments in Fig. 3, somewhere between $\gamma_a = 0.03$ and 0.04 there is a point after which the contention is not stable anymore and one of the utterances fixes. As $\gamma_a \to \gamma_0$ the time to fixation decreases, converging to that of the CNG (data not shown). However, to achieve conventionalisation it is no necessary that there is full symmetry in the rewarding system (as is the case for the CNG, where $\gamma_0 = \gamma_a$). In other words, the FNG does not need to fully reduce to the CNG in order to allow for conventions. Instead, there is a range of cognitive values that allow for conventions to be reached even under asymmetric preferences.

In Appendix A it is calculated that the expected value of the preferences in stationary state is $\langle\omega\rangle = \lambda_0/(\lambda_0 + \lambda_a)$. This expectation holds only in the flouting regime; in the conventionalisation regime it has no correspondence with the outcomes (Fig. 3G).

Summarising, allowing for different learning rates for different utterances, as proposed in this work, allows for different kinds of behavioural outcomes including contentions and conventions.

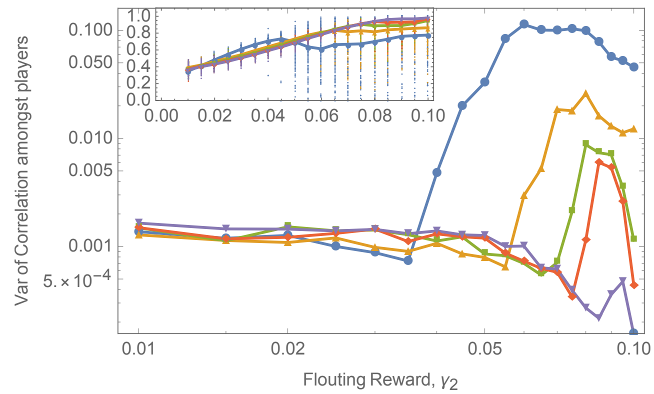

Figure 5: Correlation amongst the weights of the players (variance) and mean (inset). Blue: $N = 2$; Yellow: $N = 5$; Green: $N = 10$; Red: $N = 20$; Purple: $N = 50$. Otherwise as in Figs. 3-4.

4.1 Many-player games

A question of ultimate interest is whether the flouting behaviour, as well as the sharp transition to conventionalisation, hold also for many agents. For now, the study is limited to randomly "well-mixed" interactions where the population is constituted by equal proportions of two types of agents. Other structured populations are also of interest, but left out for now.

Results show that larger populations stabilise the contentions. For example, Fig. 4 shows the average preference for each type under different setups of N and of γ_a. Each column shows runs with the same parameters γ_a but different population sizes N, whereas in each row runs are with equal population size but varying the flouting reward γ_a.

Increasing the population size decreases the strength of fluctuations, although not their expected value. Furthermore, as it is appreciable in the middle column, larger populations stabilise the contentions, even under the conditions where 2-player games reach conventionalisation. This pattern is consistently observed across simulation repetitions (data not shown). Once the value of flouting reward converges to that of the cognate reward, conventionalisation is achieved, irrespective of the sizes of the population.

4.2 Phase transition

A conspicuous result is that as $\gamma_a \to \gamma_0$ the preferences amongst types –not only amongst players– become more strongly correlated (i.e. they follow more similar trajectories). The population average $\overline{\omega}^{(t)}$ for each type t can be seen as a vector, and thus the correlation $\rho = \operatorname{corr}\left(\overline{\omega}^{(1)}, \overline{\omega}^{(2)}\right)$ can be calculated. In any given realisation, this correlation ρ is an estimator, i.e. a random variable on an ensemble or replicate experiments. The ensemble of runs shows certain "statistical mechanical" properties, such as mean and variance of ρ, in relation to a "control parameter", in this case, γ_a (Fig. 5). The correlation $\langle\rho\rangle$ becomes stronger in a smooth way as γ_a increases but the variance, $\langle\Delta\rho^2\rangle$, shows an interesting pattern, decreasing until reaching a minimum and then off-shooting again.

The analogy with statistical mechanics is sufficiently precise as to argue about the existence of a genuine phase transition[1]. Following the analogy, the existence of a critical point $\gamma_a = \hat{\gamma}$ where $d\langle\Delta\rho^2\rangle/d\gamma_a = 0$, is a defining criterion of a phase transition. An important fact, beyond this formal notion, is that the value $\hat{\gamma}$ is congruent with the separation between the flouting and conventionalisation states which can be properly interpreted as two alternative phases of the game dynamics.

As populations become larger, the point $\hat{\gamma}$ increases. This critical point is an indication of the range of values for which a behaviour of contention is observed and increases with the size of the population. This property is consistent with the observation stated above that larger populations stabilise the flouting behaviour. However, as $N \to \infty$, $\hat{\gamma} \to \gamma_0$ which means that in very large populations any small asymmetry amongst preferences rewards create contentions.

5 DISCUSSION: CULTURE AND EVOLUTION

Between the lines, this work has made another modification from many-agent CNG. That is, instead of tracking individuals allowing for one interaction at a time, the population is mixed and interactions occur in parallel and in a synchronous way. This is analogous to the process of random mating in genetical evolution. Although seemingly a cosmetic difference, it slips in the "population point of view" that is a preclusion to understand evolution through selection. Selection amongst alternatives is implemented by S, when the agent chooses an utterance amongst its alternatives. Thus, at the population level selection is manifested as frequency-dependent, because at any given time it depends on the distribution of preferences.

The two different phases can be interpreted from the evolutionary biology perspective. In the flouting phase the population stabilises in a state that in genetics it is know as "polymorphic" (multiple forms are maintained; in this case, two). The type of selective mode that would favour this kind of state is know as disruptive selection.

If the flouting reward is larger than the critical point, then the population converges to one fixed state, albeit which, is entirely random. This is not unknown in some population genetics and evolutionary games; two different alleles or types can be equally fit but only one is able to fix.

Actually, it can be shown that the population-view of the FNG can be written as an evolutionary game in terms of payoff matrices. In this formulation, the entries depend on the relative value of the preferences. Formally deriving this analogy requires some detailed work and explantation which is beyond the scope (and length) of this paper. However, the discussion about the evolutionary nature of language and culture is an agitated subject [10]. Thus, it is pertinent to present this advance, stating that at least there is a formal analogy between the mathematical structure of evolution and that of naming games. (In this analogy, the language game corresponds to payoff matrix of a coordination game.)

Thus, it is legitimate to think of evolutionary dynamics of language, provided that a population perspective is taken.

[1] A phase transition is understood as the change from a "phase" to another and "phase" can be though as the form of the distribution of states, in this case, that of the correlation amongst types. The use of correlations to monitor phase transitions is common in statistical mechanics [7].

6 CONCLUSIONS AND FURTHER DIRECTIONS

Summarising, this work presents a new kind of Naming Game that can account for contentions of culture. The minimal requirement for contentions to occur are:

(1) Different agent identities
(2) Differential reward for cognate and non-cognate traits

and, in a preferred way, it might be added:

(3) Synchronous population update

so that it is possible to interpret the change in frequency of the utterances in an evolutionary way. (The focus in this paper, however has been on the cognitive mechanism behind it.)

Under these conditions, if the flouting reward is small enough, contentions are stable. Can this model be justified in terms of neurobiological processes? Weakening of some utterance weights can be achieved through lateral inhibition mechanisms [1], effectively decreasing the strength of an association between the concept being referred to and the alternative, non-cognate utterance [9].

One of the central results of this work is that in large populations, small asymmetries between preference rewards allow contentions. Although it could be argued on a statistical basis that larger populations have larger diversity, the results here presented indicate that there is a phenomenon of stabilisation of the flouting behaviour. Moreover, this is consistent with the large cultural and language diversities that exist in human populations.

On the other hand the result implies that, in large populations, conventionalisation is not fully stable: even small differences on the rewards (which is expected), most likely result on contentions.

The FNG extends the tenets of the CNM; still, this approach is clearly still an over-simplification to a complex cognitive system with a rich culture. Thus, the extent to which these over-simplifications are explanatory of actual cultural contentions is contingent the sufficiency of the requirements 1-3 above to more complex situations. Hence, the obvious direction to proceed is to implement the FNG in embodied populations in rich environments. This can account for interesting interactions amongst concepts and forms, reminiscent of a true culture [11].

From the neurobiological approach it seems a pressing issue to dig more into the possible ways in which preferences are constructed. It is known, although not so popularly, that neuronal rewards (e.g. dopamine system) employ prior information as a way to tune expectations. When expectations are fulfilled, rewards are more intense than when expectations are deceived [6]. This mechanism explains the subjective nature of preferences and might be of relevance for understanding behavioural bases for establishing social conventions and contentions. This implies that, in order to understand what creates cultural contentions, it is important to account for the dynamics of the rewarding system γ. (The assumption here is that different types have different expectations regarding the transmitted utterance, but these differences should be also an outcome of the dynamics.)

In a non-trivial way, and highly informed on neuronal mechanisms, it seems promising that Naming Game-like systems can explain for surrogates of cultural complexity, including the difference between contentions and conventions.

A APPENDIX: EQUILIBRIUM WEIGHTS

The derivation is provided for only one of the agents and its cognate utterance. The calculation is equivalent for the alternative utterance. If ψ is the probability that the speaker (whoever it is) utters the u_0:

$$\Delta\omega = \begin{cases} \lambda_0(1-\omega) & \text{with prob. } \psi \\ -\lambda_a\omega & \text{with prob. } 1-\psi \end{cases} \tag{3}$$

Thus,

$$\langle\Delta\omega\rangle = \lambda_0(1-\omega^i)\psi - \lambda_a\omega^i(1-\psi) \tag{4}$$

Following Bayes' rule:

$$\psi \equiv \Pr[\mathcal{U} = u_0] = \Pr[\mathcal{U} = u_0 | t = t_0]\Pr[t = t_0] + \Pr[\mathcal{U} = u_0 | t = t_a]\Pr[t = t_a]$$

where t refers to the type. Since each individual has the same chance to be a speaker, $\Pr[t = t_0] = \Pr[t = t_a] = 1/2$. The conditional probabilities are precisely the weights that the individual of type t utters u_0, i.e.

$$\Pr[\mathcal{U} = u_0 | t = t_0] = \omega^0 \tag{5}$$

$$\Pr[\mathcal{U} = u_0 | t = t_1] = 1 - \omega^1 \tag{6}$$

Due to symmetry $\omega^0 = \omega^1 = \omega$, hence $\psi = 1/2$. In stationarity $\langle\Delta\omega\rangle_{\text{eq}} = 0$ thus

$$\lambda_0\left(1 - \langle\omega\rangle_{\text{eq}}\right)\frac{1}{2} - \lambda_a\langle\omega\rangle_{\text{eq}}\frac{1}{2} = 0$$

which results in

$$\langle\omega\rangle_{\text{eq}} = \frac{\lambda_0}{\lambda_0 + \lambda_a}\ , \tag{7}$$

as reported in Section 4.

ACKNOWLEDGMENTS

The author would like to acknowledge Luc Steels for feedback and comments and the KLI board for providing support for the attendance to and co-organisation of the workshop Evolution in Cognition. This work was partially supported by the Hungarian Academy of Sciences, project GINOP-2.3.2-15-2016-00057 "In the light of evolution".

REFERENCES

[1] Adam R Aron. 2007. The neural basis of inhibition in cognitive control. *The Neuroscientist* 13, 3 (2007), 214–228.
[2] Andrea Baronchelli, Maddalena Felici, Vittorio Loreto, Emanuele Caglioti, and Luc Steels. 2006. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment* 2006, 06 (2006), P06014.
[3] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81, 2 (2009), 591.
[4] Damon Centola and Andrea Baronchelli. 2015. The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences* 112, 7 (2015), 1989–1994.
[5] Bart De Vylder and Karl Tuyls. 2006. How to reach linguistic consensus: A proof of convergence for the naming game. *Journal of Theoretical Biology* 242, 4 (2006), 818–831.
[6] Wolfram Schultz. 2016. Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience* 17, 3 (2016), 183.
[7] H.Eugene Stanley. 1971. *Introduction to phase transitions and critical phenomena.* Oxford University Press.
[8] Luc Steels. 1995. A self-organizing spatial vocabulary. *Artificial Life* 2, 3 (1995), 319–332.
[9] Luc Steels. 1998. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence* 103, 1-2 (1998), 133–156.
[10] Luc Steels. 2011. Interactions between cultural, social and biological explanations. *Physics of Life Reviews* (Nov. 2011), 1–5.
[11] Luc Steels. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8, 4 (Dec. 2011), 339–356.
[12] Luc Steels. 2015. *The Talking Heads experiment.* Language Science Press.
[13] Luc Steels and Martin Loetzsch. 2012. The grounded naming game. *Experiments in cultural language evolution* 3 (2012), 41–59.
[14] Remi Van Trijp and Luc Steels. 2012. Multilevel alignment maintains language systematicity. *Advances in Complex Systems* 15, 03n04 (2012), 1250039.