

TITLE

Linguistic Group Marking as an Emergent Phenomenon

AUTHORS

Cormac Moran¹, Judit Casas i Riu^{1,2}, Cedric Boeckx^{1,3,4,5}

AFFILIATIONS

¹University of Barcelona; Barcelona, Spain

²Radboud University; Nijmegen, Netherlands

³Institute of Neurosciences of the University of Barcelona; Barcelona, Spain

⁴Universitat de Barcelona Institute of Complex Systems (UBICS); Barcelona, Spain

⁵Catalan Institute for Advanced Studies and Research (ICREA); Barcelona, Spain

CORRESPONDENCE:

cedric.boeckx@ub.edu

ABSTRACT

This paper examines how linguistic group marking can emerge from the interaction between social ecology, inferential cognition, and cultural transmission. An agent-based model (ABM) simulates interactions within a structured population where agents use domain-general inference and social learning to interpret and adopt linguistic variants. The model implements a mechanism in which uncertainty about strangers' social affiliations increases reliance on social inference and investigates how variation in social memory and outgroup contact affects the resulting patterns of linguistic variation. Results show that under conditions of greater perceived social risk, the model produces greater use of inference and, subsequently, greater linguistic differentiation between groups. However, differentiation remains concentrated in particular linguistic features rather than producing a complete divergence of group repertoires. These findings suggest that linguistic group marking, characteristic of human language, can be understood as an emergent property of social conditions rather than a

specialised adaptation. The model therefore provides a possible sociocognitive mechanism contributing to the broader process of linguistic polyphilia, while more generally aligning with accounts of complex linguistic structures as arising from domain-general cognition interacting with cultural evolution.

Keywords: linguistic variation, group marking, polyphilia, agent-based modelling, cultural evolution

1. INTRODUCTION

The human linguistic condition is characterised by variation. It is the first thing that strikes linguists and non-linguists alike: there is so much variation. Humans exploit this variation for social purposes. Decades of work in sociolinguistics have documented the many ways variation correlates with ethnicity, locality, social class, stance or interactional style.¹⁻⁴ However, the factors contributing to the emergence of this social use of linguistic differences remain under-explored.

Language is perhaps the most salient example of what Cerrito et al. call polyphilia—a tendency humans possess to use differences as markers of group identity.⁵ According to Cerrito et al., this tendency is especially pronounced in *Homo sapiens*. They hypothesise that it likely emerged at a time of population expansion and increased intergroup contact. Properties like the arbitrariness of linguistic signs (extensive use of symbols) and ergodicity (memoryless transition probabilities between states of a linguistic system) make language particularly well suited to be used in this way, setting it apart from more utilitarian instances of cultural evolution such as tool use.⁶

This paper probes the forces that could have led to the emergence of linguistic polyphilia. It does so by resorting to agent-based modelling, a method that has yielded fruitful results in the study of language evolution, where the information that can be derived from the fossil record is sparse and very indirect.⁷⁻⁹

Seminal work by Kirby and colleagues has shown how design features of human language like compositionality can emerge as a trade-off between two opposing pressures: the need to be expressive (conveying as many distinct messages as possible) and the need to memorise what is being conveyed.^{10,11} Re-combining smaller meaning-bearing units in novel ways is an excellent way to meet this challenge.

But as Goldberg nicely expressed, human language design is more than this. Speakers optimise messages in a way that conforms to the conventions of their speech communities.¹² Convention sits next to composition at the heart of human language. Here we explore the possibility that (group) conventionality is also an emergent property: the outcome of a social ecology characteristic of *Homo sapiens*; an expanding group of cooperating individuals, leading to a steady flux of new in-group strangers, and a deep-seated tendency not to engage with foreign groups we encounter (out-group strangers). Our model explores the intuition that a variable social ecology and uncertainty about strangers' social affiliations can influence patterns of linguistic variation. Our results show that polyphilia can manifest itself in

language use through the interaction of risk-sensitive social inference and cultural transmission, amplifying initially weak correlations between language and social group. Crucially, no domain-specific, ‘strong’ bias is required for polyphilia to emerge. In line with Thompson et al., our results align with accounts of complex linguistic structures as arising from weak domain-general cognition interacting with cultural evolution in structured populations.¹³

2. MODEL DESIGN

The model formalises a scenario in which population expansion and increased intergroup contact create social ecologies in which individuals frequently encounter both in-group and out-group strangers (i.e. the group becomes more ‘anonymous’, in the sense of Moffett¹⁴). Assuming that in-group members make for more reliable cooperation partners than out-group members,^{5,15–18} this type of social ecology means personal familiarity can no longer be relied upon in selecting these preferred partners, and agents perceive the potential benefits of interactions with strangers to be more uncertain. Agents therefore require an ability to identify preferred in-group stranger partners, while distinguishing them from out-group strangers.

Agents possess a variable linguistic representation and generic cognitive abilities such as inference and social learning. They adapt to their environment through learning, tracking how often they encounter different classes of partners and the types of linguistic variants they use, and through imitating fitter agents.

The purpose of this model is to test the notion that linguistic group marking could be an emergent property, triggered by this type of social ecology. We hypothesise that these social conditions causing individuals to encounter unfamiliar partners would lead them to search for other indicators of these partners’ identities. This scenario, whereby agents need to discriminate preferred cooperating partners, could create selective pressure for increased recognisability. This in turn would gradually accentuate cultural differentiation between groups. The model presented here operationalises this causal chain.

2.1 Agents and Population Structure

The population consists of N agents divided equally between two groups (one in-group and one out-group for each agent). Each agent possesses:

1. a linguistic repertoire
2. estimates of how often they encounter strangers and how likely strangers are to belong to their own group
3. beliefs relating to the association of linguistic variants with social groups
4. a fitness level (w) based on agents' cooperative behaviour

Agents interact in a network whose topology is defined by two variables: social memory size (m), the number of the in-group that agents are personally familiar with, and out-group contact rate (p_{out}), the probability that an agent interacts with a member of the out-group.

The ratio of social memory size to in-group size is a proxy for the network density within groups, while the out-group contact rate represents the density of connections between groups (e.g. at 50% both groups are completely interconnected).

2.2 Linguistic representation

Language is represented in this model as a vector of three discrete linguistic features: $L = [f_1, f_2, f_3]$, where each feature f_i can take one of three categorical variants 1,2,3. For example, given this state space of 3 features $\times$ 3 variants, with a total of 27 possible combinations, a language might be represented as [3, 1,2].

Linguistic repertoires are initially random and have a small likelihood of being updated at each timestep. Agents update their linguistic repertoires via success-biased social learning, noisily copying a feature variant from a randomly fitter member of their ‘known’ group. There is also a small likelihood of agents randomly copying variants during interactions, accounting for cultural diffusion between groups and giving a secondary pathway for language change. This dynamic allows for the possibility of linguistic convergence and divergence over time, while also ensuring that the system does not become static.

In this model, language is treated as a set of non-semantic, arbitrary cultural features which may or may not be used to *index* group identity (in the sense of Eckert or Silverstein^{4,19}). The features can be seen as positions in the linguistic system where variation can occur (for example, the pronunciation of a certain phoneme, or the use of a given lexical item or grammatical structure). The variants, then, would be equivalent to sociolinguistic variables in the traditional sense.

2.3 Cooperative decision: cognitive architecture

The basic structure of an interaction is as follows: agents are paired according to the interaction structure, they decide whether or not to cooperate, they receive fitness effects depending on the accuracy of their decision, and they learn from the interaction.

The decision to cooperate is taken based on agents' estimation of their partners identity. Known in-group agents are always trusted as cooperation partners. Cooperating with an unfamiliar partner carries structural risk: coordinated in-group cooperation yields payoffs, but out-group cooperation incurs penalties. Hence, when interacting with strangers, agents must evaluate their partner's group identity before deciding to cooperate. To do so, they dynamically select between a cognitively cheap, prior-based heuristic strategy, and a more expensive Bayesian inference strategy based interpreting on partners' linguistic signals. This strategy selection is modelled as a rational choice driven by the agent's perception of interaction risk and the calculated expected utility of information.

2.3.1 Risk calculation

Agents assess the structural risk of their social environment in selecting a cooperation strategy. This assessment relies on two priors, shaped continuously by the agent's interaction history:

1. $\hat{p}_{stranger}$: the estimated frequency of encountering strangers
2. $\hat{p}_{in}$: The estimated probability that a stranger belongs to the agent's own group

These priors combine to generate a single metric of perceived interaction risk (r), calculated as:

$$r = \hat{p}_{stranger} \times (1 - 2|\hat{p}_{in} - 0.5|) \quad (1)$$

This formula captures an important cognitive trade-off. The second term, $(1 - 2|\hat{p}_{in} - 0.5|)$, represents identity ambiguity, which peaks at 1 when the likelihood of a stranger being in-group or out-group is equal, dropping to 0 when strangers homogeneously belong to one or the other group.

By weighting this by $\hat{p}_{stranger}$, r captures the *salience* of that ambiguity. Where agents meet few strangers, these interactions have less of an impact on agents' fitness and their perception of risk is lower; where strangers make up a larger proportion of agents' interactions, they become more relevant and risk perception increases accordingly.

2.3.2 Strategy choice

To determine whether to deploy cheap intuition (the heuristic strategy) or the more cognitively costly Bayesian inference, agents calculate and compare the expected utility of both cognitive strategies based on the model's underlying payoff structure.

The heuristic strategy represents the default strategy, given the greater cost of inference. The agent completely ignores the partner's linguistic signal, choosing instead to make a blind guess using only their interaction structure prior ($\hat{p}_{in}$). The agent estimates what their payoff from the interaction would be if they always blindly cooperated versus if they always blindly defected, according to the likelihood that their partner is a preferred in-group member:

$$EU_{cooperate} = (\hat{p}_{in} \times payoff) - ((1 - \hat{p}_{in}) \times penalty) \quad (2)$$

$$EU_{defect} = ((1 - \hat{p}_{in}) \times payoff) - (\hat{p}_{in} \times penalty) \quad (3)$$

The baseline utility of the heuristic strategy is simply the higher of these two paths:

$$EU_{heuristic} = \max(EU_{coop}, EU_{defect}) \quad (4)$$

The higher cost of Bayesian inference is due to its providing more accurate information about a partner's identity through decoding their linguistic signal. The expected utility of inference reflects a priori assumptions based on environmental risk (as represented by r), and is calculated using the formula

$$EU_{inference} = ((r \times payoff) + ((1 - r) \times EU_{heuristic})) - C_{inf} \quad (5)$$

where C_{inf} is a small, fixed inference cost.

Once the expected utility of both strategies is evaluated, the agent calculates the net utility gain of shifting to Bayesian inference:

$$\Delta EU_{strat} = EU_{inference} - EU_{heuristic} \quad (6)$$

The final choice between strategies is formalised using a logit choice rule, yielding a smooth probability:

$$P(\text{inference}) = \frac{1}{1 + e^{-s\Delta EU_{strat}}} \quad (7)$$

This equation embodies the value of information (VoI). When the environment is highly unpredictable ($r \approx 1$), the blind heuristic is expected to fail or deliver poor returns, making inference a more valuable strategy. Conversely, where the interaction structure is perceived as safer (lower r), the cognitive cost to inference makes the heuristic a preferred strategy.

2.3.3 Bayesian Inference

When an agent decides to use the inference strategy, they use a naïve Bayesian framework to estimate their partner's social group based on the linguistic variants they observe. This relies on their knowledge from previous interactions, counts of how frequently they have observed members of different groups use specific linguistic variants.

Following Bayes' rule, given an observed signal $s = [v_1, v_2, v_3]$, the posterior likelihood that this signal was produced by a member of group g is proportional to the prior probability that the speaker belongs to that group, weighted by the likelihood of observing the signal given that group:

$$P(g|s) \propto P(g)P(s|g) \quad (8)$$

Here, the prior probability $P(g)$ is derived from the agent in-group stranger estimate ($\hat{p}_{in}$). Assuming conditional independence between linguistic features (i.e. a naïve independence assumption), the likelihood factorises as

$$P(s|g) = \prod_{f=1}^F P(v_f|g) \quad (9)$$

with $P(v_f|g)$ calculated from the relative frequency the agent has observed a member of group g produce the variant v of the feature f . This means the agent treats each linguistic feature as an independent piece of evidence. The posterior probability is obtained by normalisation over groups:

$$P(g|s) = \frac{P(g) \prod_f P(v_f|g)}{\sum_{g'} P(g') \prod_f P(v_f|g')} \quad (10)$$

From a cognitive perspective, this process can be interpreted as the agent accumulating evidence for competing group identities. As more linguistic features associated with a particular group are observed, the posterior probability assigned to that group increases, providing the agent with a stronger basis for deciding whether the interaction partner should be trusted.

2.3.4 Cooperation decision

The final decision (cooperation or defection) depends on the strategy selected. No matter what strategy is used, the final decision to cooperate or defect is made probabilistically using

a logit choice rule based on the expected utility differential of each action. When using the heuristic strategy, the previously expected utility values calculated using formulae shown in (2–3) are used. For the inferential strategy, expected utility of cooperating is calculated using the same expected utility formula, replacing $\hat{p}_{in}$ with the posterior probability that their partner belongs to their in-group $P(in|s)$, based on observed signal s .

2.4 Interaction outcomes

2.4.1 Fitness and payoffs

An agent’s fitness is driven by its behaviour while cooperating — specifically, how well it can coordinate its cooperating behaviour with its partner’s identity. The payoff matrix is shown below in Table 1.

Table 1: Payoffs from cooperation action

Action	In-group partner	Out-group partner
Cooperate	+1	−1
Defect	−1	+1

Cooperation with in-group members gives a positive payoff, while cooperating with out-group members is penalised. Conversely, defecting against out-group members is rewarded, while doing so against in-group members is penalised. Agents also receive secondary payoffs based on whether their own linguistic signal successfully elicits cooperation from in-group partners, working the same way as in Table 1. This is to represent the fact that the signalling system serves a dual purpose: it is both a tool for categorising others and being recognised oneself.

2.4.2 Learning

After every interaction, agents update their priors to reflect what they just experienced. As described, this occurs on two distinct levels: interaction structure tracking and linguistic tracking.

Agents update their beliefs about the frequency of interaction with strangers ($\hat{p}_{stranger}$) after every interaction, given that this information is always available to them. On the other hand, learning related to true group identity of a stranger partner only takes place probabilistically. This represents imperfect observation; we can imagine someone deduces a

stranger’s identity via other means, e.g. being told by someone else, but they only get this definitive proof occasionally. When the stranger’s true group identity is revealed, the agent updates their language beliefs and the in-group stranger prior ($\hat{p}_{in}$). They register the specific linguistic variants their partner used and attribute them to that partner’s actual group. Over generations of interactions, this allows agents to map out the linguistic repertoires of different social groups.

2.5 Output metrics

The model produces both behavioural and linguistic metrics. Together, these metrics allow us to evaluate whether linguistic variation becomes socially informative and the circumstances under which this occurs. A summary of metrics recorded and the theoretical concepts they are intended to represent is shown in Table 2.

Table 2: Metrics recorded and their conceptual equivalents

Metric	Concept
Perceived risk	Agents’ perception of interaction risk
Proportion of inference used	Reliance on social signalling
Jensen–Shannon divergence	Emergence of distinct group linguistic varieties

2.5.1 Perceived risk and inference use

Agents’ response to the sociodemographic parameters is recorded via perceived risk. The proportion of inference used records the frequency with which agents choose this strategy when interacting with strangers. These metrics allow us to test whether linguistic differentiation emerges specifically under conditions where the interaction risk makes social inference beneficial.

2.5.2 Linguistic differentiation

The primary measure of group marking is the normalised Jensen–Shannon (JS) divergence of the groups’ linguistic distributions. JS divergence measures the degree of overlap between groups in the feature–variant space, ranging from 0 (identical distributions) to 1 (completely distinct distributions). Higher values therefore indicate stronger linguistic differentiation, likely indicating use of language for group marking.

To examine the patterns of differentiation across linguistic features, the JS divergence per feature (minimum, median and maximum features) was also recorded. If a social signalling system emerges, this metric allows us to examine how it is implemented in agents' languages.

3. RESULTS

We tested the model with a population of 200 agents to investigate the effect of our inputs (the experimental sociodemographic parameters p_{out} and m) on the outputs (the metrics). Each simulation was run for 3000 timesteps. A preliminary inspection indicated that this duration was sufficient for the system to approach a stable state under most parameter combinations. To account for stochastic variation, each parameter combination was replicated 30 times.

The results reported here represent the mean value of each metric during the final 500 timesteps of a simulation, thereby characterising the long-run behaviour of the system rather than transient early dynamics. All simulations were implemented in NetLogo and executed using BehaviorSpace (see Appendix).

3.1 Cognitive and linguistic metrics are sensitive to sociodemographic factors

We first examined how system-level outcomes vary under different sociodemographic parameters. Risk perception showed a clear gradient, increasing with out-group contact and decreasing with social memory size (Figure 1A). Group-level Jensen–Shannon divergence showed a less gradual pattern, instead separating broadly into regions of higher and lower divergence (Figure 1B). Higher values for both variables occurred under conditions of limited social memory and high out-group contact.

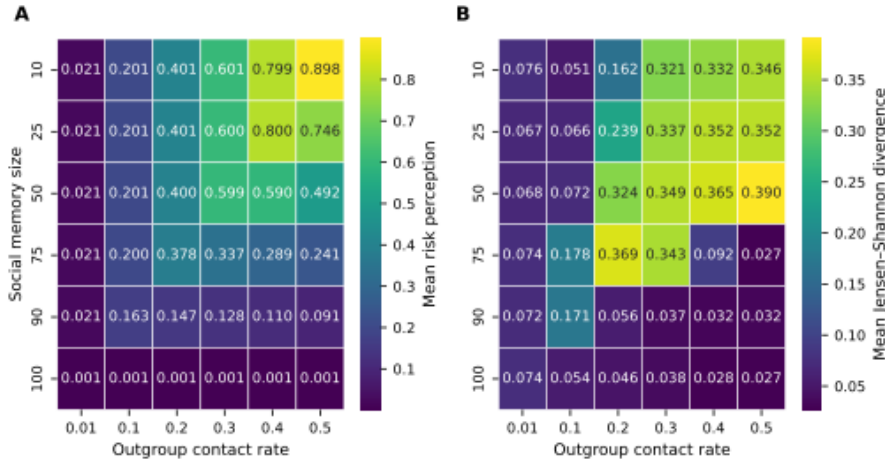


Figure 1: Risk perception (A) and on global Jensen–Shannon divergence (B) results under different sociodemographic parameters

A multiple linear regression assessed the direct effects of social memory size, out-group contact rate and risk perception on JS divergence. The model explained a large proportion of the variance ($R^2 = 0.812$). Risk perception was the strongest predictor, showing a clear positive association with divergence ($\beta = 0.61$). However, out-group contact also had a significant direct negative effect ($\beta = -0.23$), despite its positive indirect relationship. This may reflect within-interaction copying, whereby greater cross-group contact facilitates the diffusion of linguistic material between groups. Social memory size had only a weak, non-robust direct association with divergence, suggesting that its primary contribution is indirect through risk perception.

Despite the positive relationship between risk perception and JS divergence, the groups' language distributions never completely diverged: maximum JS divergence remained below 0.4 (Figure 1B). Thus, group marking emerged under some conditions, but the resulting linguistic systems retained substantial similarity.

3.2 Risk perception drives inference use

We next examined the relationship between perceived risk and the use of inference. As expected from our model design, the relationship is best captured by a logistic function, which provided a significantly better fit than a linear alternative according to AIC comparisons ($\Delta AIC = 27.94$). This indicates that inference adoption remains low at lower levels of perceived risk but increases sharply once a critical threshold is reached, levelling off once again at high-risk values. Figure 2 maps inference use onto risk perception values.

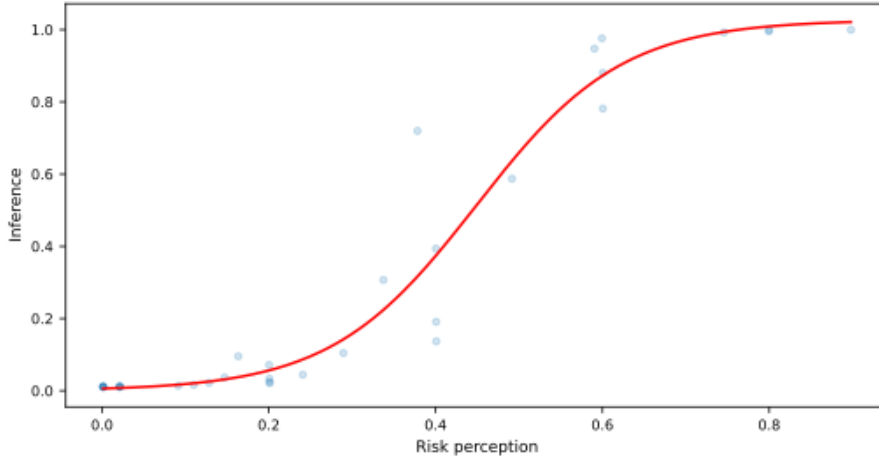


Figure 2: Proportion of inference use at different levels of risk perception

3.3 The effect of inference

Next, we assessed how inference use relates to linguistic divergence. Aggregating group-level Jensen–Shannon divergence across binned levels of inference revealed a clear positive relationship, as shown in Figure 3A. This pattern is supported statistically, indicating a strong association between inference use and divergence (Spearman's $\rho = 0.821$). The trend is not fully monotonic, however, with global JS divergence reaching a peak around 0.9 inference use before dropping slightly.

We then compared how global Jensen–Shannon divergence is distributed across the feature–variant space. Figure 3B shows substantial variation between features, with the gap between the most- and least-divergent features increasing significantly with inference use, as indicated by a Spearman rank-order correlation between inference use and the maximum–minimum divergence gap ($\rho = 0.86$). This pattern suggests that increasing inference use selectively amplifies divergence in more diagnostic features rather than producing a uniform split in the feature–variant space. At high inference use, the maximally divergent feature reaches a peak at ≈ 1.0 , while divergence in the remaining features decreases. This may indicate that once a highly diagnostic feature provides sufficient information for classification, other features contribute less to the resulting differentiation.

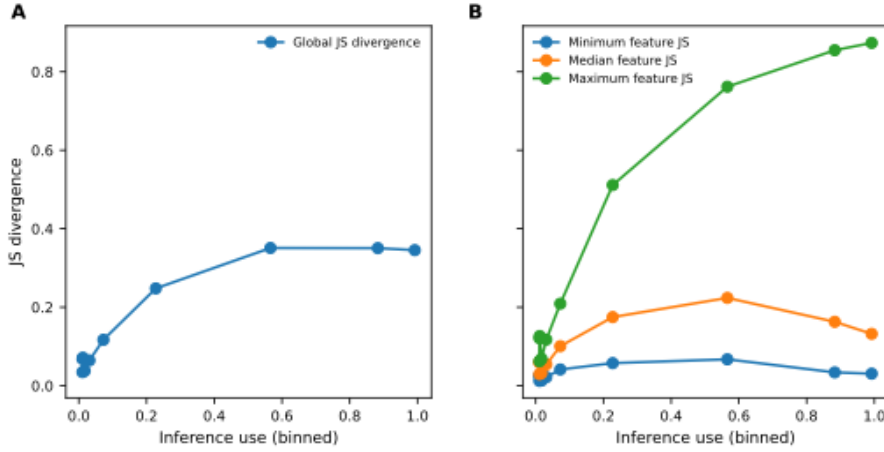


Figure 3: Global Jensen–Shannon divergence across levels of inference use (A) and variation in feature-level divergence (B)

3.3 Causal intervention

To directly test the causal role of inference, we compared counterfactual model variants in which inference is either always enabled or completely disabled, holding all other parameters constant. The results of these are shown in Figure 4A–B, using the same colour gradient as in Figure 1B. Jensen–Shannon divergence was consistently higher when inference was always enabled, with the difference supported by a Wilcoxon signed-rank test ($W = 22$, $p < 0.001$). Across conditions, the always-inference variant produced a mean increase in JS divergence of approximately 0.3. This counterfactual comparison indicates that inference does not merely correlate with divergence but contributes directly increased linguistic differentiation.

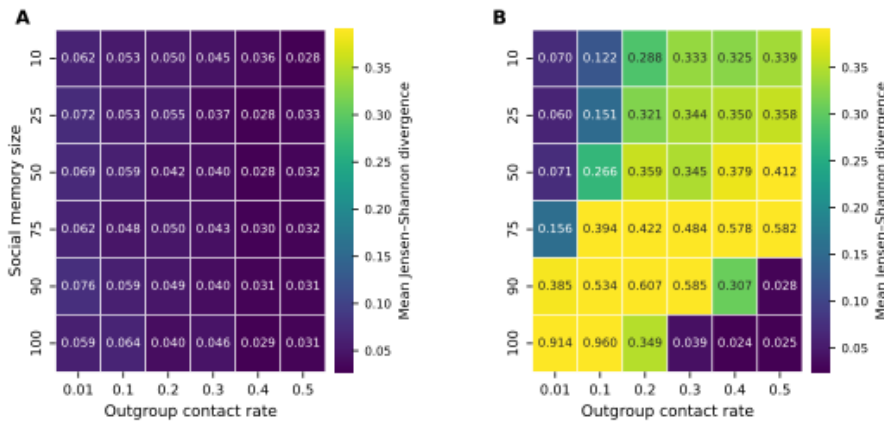


Figure 4: Global JS divergence with inference always enabled (A) and completely disabled (B)

3.4 Temporal validation

Finally, we examined temporal dynamics to assess whether the observed relationships were maintained over time. Time-series data from high- and low-risk regimes were analysed to

track changes in inference use and global Jensen–Shannon divergence. Figures 5A–B show the mean values of these variables over the initial 500 time steps.

The temporal patterns are broadly consistent with the relationships identified above. In the high-risk scenario, inference use increases sharply, followed by a corresponding increase in JS divergence after a lag of approximately 200 time steps. In the low-risk regime, both inference use and JS divergence remain quite low, although the latter variable increases to a limited extent over time. This suggests that some amount of group linguistic divergence can occur through cultural transmission within groups. The stronger temporal association between inference use and subsequent divergence in the high-risk regime is consistent with inference contributing to the amplification of linguistic differentiation.

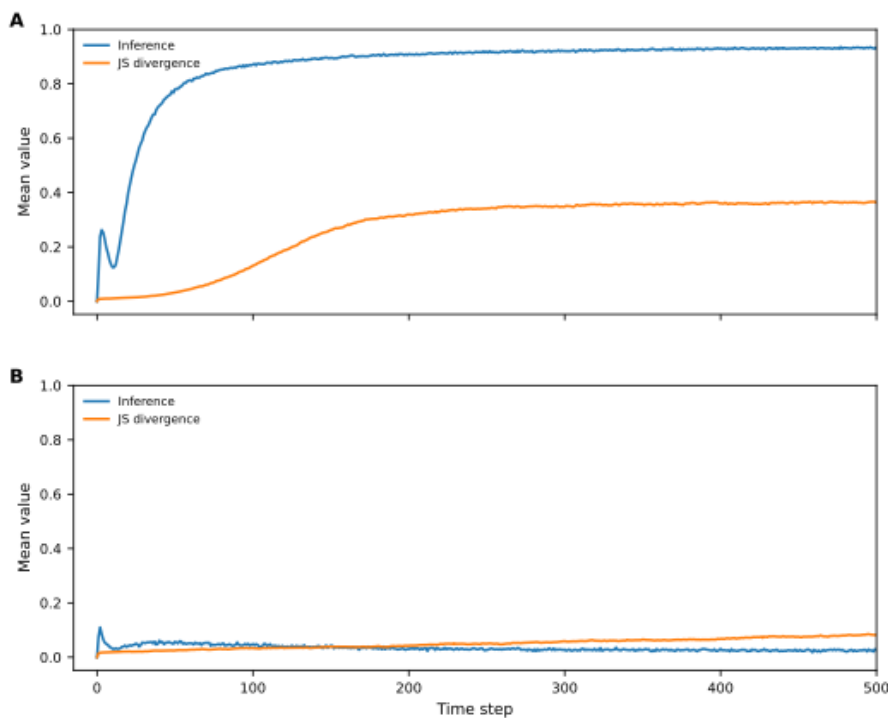


Figure 5: Evolution of cognitive and behavioural variables over time in a high-risk scenario (A) and a low-risk one (B)

4. DISCUSSION

Our results suggest that linguistic group marking can be produced without requiring a specific bias, relying instead on the interaction of an uncertain social ecology (typical of *Homo sapiens*) with generic inferential cognition and social learning (plausibly available in *Homo*). This result aligns more with work focusing on group marking as an emergent cultural behaviour in contexts where individuals need to coordinate with individuals outside their

immediate social network.¹⁷ Regarding the evolved capacities required to produce this dynamic, the model suggests that more generic cognitive abilities such as inference²⁰ and social learning^{21,22} are sufficient to explain the emergent phenomena observed. Indeed, recent empirical work has demonstrated that young children actively exploit linguistic variation to infer speaker group membership when group–marker mappings are consistent.²³ In general, these results align with work focusing on the creation of complex cultural systems from the interactions of weak biases and domain-general abilities in different ecological conditions, amplified by cultural evolution.^{10,11,24,25}

In particular, Thompson et al. show that weak constraints can have equivalent effects to strong ones where adaptive cultural evolution amplifies their effects.¹³ In the model presented here, the use of inference to make judgements about group identity could be seen as a weak bias, given that it is an ability all agents have access to; however, the use of this ability is not a given and only takes place where the social ecology makes it advantageous to do so. This weak bias is amplified by the cultural evolution of linguistic features, creating a feedback loop that consistently produces the same outcome of linguistic group marking under the right conditions.

These findings are also relevant to the broader account of polyphilia proposed by Cerrito et al.⁵ Our model provides evidence that group marking could emerge in the specific sociodemographic context they suggest (an expanding population leading to greater contact with both in- and out-group strangers), as well as suggesting concrete sociocognitive processes by which group linguistic differentiation could emerge.

Our results also align with work emphasising the emergence of features of language under interactions between individual-level communicative needs and group-level constraints. Our results showing the transition from a discrimination system mostly based on personal familiarity to group-level conventions is related to work discussing the emergence of conventionality and systematicity in larger groups.^{26–29} An expanding population accommodating in-group strangers accelerates regularization and the wide-spread adoption of conventions, unlike smaller speech communities, which can maintain more complex languages (as evidenced in village sign languages³⁰). In our model, it is the frequency of contact with the out-group that increase the importance of linguistic generalization and group-wide convention, placing linguistic variation as an index not of individual identity but group identity.

The model-variant and time-series analyses provided further support for the proposed relationship between inference use and linguistic differentiation. The mediating variables, perceived risk and inference use, are an individual-level cognitive bridge (the control parameter) between the sociodemographic inputs and the observable patterns of linguistic variation. We propose that this could constitute a basic mechanism creating pressure for social signalling systems to emerge: ambiguity regarding strangers' social identities, leading individuals to rely on observable cultural features as probabilistic cues. Where certain cultural features gain in diagnostic power, they are selected for and diffuse within groups.

The Jensen–Shannon results, used here as a proxy for group differentiation, do not indicate gradual divergence of the groups' language distributions. Instead, we observe a more abrupt shift in which one of the two strategies becomes dominant in each group. This pattern is partly consistent with the logistic formulation used to model strategy choice, although the sparsity of data in the transitional zone makes it difficult to fully characterise the transition. One possible explanation is that group-marking strategies may generate reinforcing feedback, but this remains speculative within the present model.

The results show the emergence of social meaning from arbitrary linguistic variation, a result consistent with the transition from first- to second-order indexicality in the sense of Silverstein.¹⁹ The agents use correlations between linguistic variants and social categories as probabilistic cues when making judgements about individuals. The reverse pathway, whereby increasing linguistic differentiation could itself increase use of inference, is also a plausible driver of this behaviour, although not tested here. It is also interesting to note that the results do not show full group divergence of group repertoires; instead, specific features which are more strongly indexical are widely diffused among groups, while the rest of the state space remains fairly equally distributed amongst groups. The emergence of highly indexically salient features ('shibboleths') is an empirically described phenomenon in sociolinguistics.³¹ Other results have shown how under conditions of contact, group stratification is not uniform in all areas of language: some domains, e.g. lexicon, diverge strongly, while borrowing causes other to become more similar.^{32,33}

Limitations and future expansions

The model is deliberately minimal and should therefore be interpreted as a test of plausibility of the proposed mechanism rather than a realistic representation of human linguistic behaviour. Several extensions would allow the robustness and the scope of our findings to be assessed further.

Firstly, the model represents language as a small vector of independent, discrete features. This abstraction allows the relationship between social inference and linguistic variation to be isolated while omitting many of the properties of real-world linguistic systems. In particular, real sociolinguistic variables often exhibit complex correlations, both because of linguistic structure and because speakers may package multiple features into broader styles or registers.^{4,34} Additionally, linguistic features in this model function only as carriers of social meaning, without any referential meaning. Future work could incorporate richer linguistic and communicative representations to test the plausibility of our results in more complex situations.

In addition, agents belong to only two groups and interact according to simple, fixed sociodemographic variables. Real populations are typically characterised by multiple overlapping identities and more complex patterns of interaction. Similarly, social memory size is used as the sole determiner of in-group network structure, a somewhat inelegant proxy for network density. Although this captures one aspect of network structure, it does not represent other important properties such as clustering or hierarchical social organisation. Extending the model to include multiple groups, nested identities, or dynamic network formation would allow investigation of whether the same mechanisms operate under more realistic social conditions. Of particular interest would be the introduction of new groups or group splitting after stable signalling systems have already emerged, allowing the adaptability of linguistic marking systems to be examined.

Thirdly, the model is intentionally agnostic with respect to the exact processes through which successful linguistic behaviour spreads. As with agent-based models more broadly, representing these processes requires reducing complex cognitive mechanisms to a limited set of computational rules. While theoretical models assume a semi-intentional nature in the construction of linguistic personae^{4,19,34,35}, our model represents language as being relatively faithfully transmitted, without the possibility of strategic language use. This limits the extent to which the model can address how speakers actively manipulate linguistic behaviour to signal or negotiate social identities.

The results also raise questions about the directionality of the relationship between linguistic differentiation and inference. The intervention and temporal analyses provide evidence that greater use of inference contributes to subsequent increases in linguistic differentiation, consistent with our proposed pathway from social risk through inference to linguistic divergence. However, the present model does not directly test the reverse pathway:

whether increasingly diagnostic linguistic features subsequently increase individuals' reliance on inference. Such a feedback loop is theoretically plausible, particularly if cultural transmission amplifies features that become useful for social classification, but it remains untested here. Future work could therefore manipulate the diagnosticity of linguistic features independently of risk perception and examine whether this alters inference use. More generally, extensions of the model could investigate whether other factors beyond risk perception, including different network structures, influence patterns of linguistic differentiation.

REFERENCES

1. Labov, W. The Social Motivation of a Sound Change. *WORD* **19**, 273–309 (1963).
2. Labov, W. *The Social Stratification of English in New York City*. (Cambridge University Press, 2006). doi:10.1017/cbo9780511618208.
3. Milroy, L. *Language and Social Networks*. (Basil Blackwell, 1987).
4. Eckert, P. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annu. Rev. Anthropol.* **41**, 87–100 (2012).
5. Cerrito, P. *et al.* Pleistocene origins of cultural and linguistic diversification: how Homo sapiens and Neanderthals differed. Preprint at <https://doi.org/10.32942/x2c35z> (2025).
6. Bickel, B., Giraud, A.-L., Zuberbühler, K. & van Schaik, C. P. Language follows a distinct mode of extra-genomic evolution. *Phys. Life Rev.* **50**, 211–225 (2024).
7. Lekvam, T., Gambäck, B. & Bungum, L. Agent-based modeling of language evolution. *Proc. 5th Workshop Cogn. Asp. Comput. Lang. Learn. CogACLL* 49–54 (2014) doi:10.3115/v1/w14-0510.
8. Romanowska, I. *Agent-Based Modeling for Archaeology*. (SFI Press, 2021). doi:10.37911/9781947864382.
9. Wilensky, U. & Rand, W. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. (The MIT Press, 2015).

10. Kirby, S., Tamariz, M., Cornish, H. & Smith, K. Compression and communication in the cultural evolution of linguistic structure. *Cognition* **141**, 87–102 (2015).
11. Kirby, S., Griffiths, T. & Smith, K. Iterated learning and the evolution of language. *Curr. Opin. Neurobiol.* **28**, 108–114 (2014).
12. Goldberg, A. E. Usage-based constructionist approaches and large language models. *Constr. Fram.* **16**, 220–254 (2024).
13. Thompson, B., Kirby, S. & Smith, K. Culture shapes the evolution of cognition. *Proc. Natl. Acad. Sci.* **113**, 4530–4535 (2016).
14. Moffett, M. W. Human Identity and the Evolution of Societies. *Hum. Nat.* **24**, 219–267 (2013).
15. Axelrod, R. & Hamilton, W. D. The Evolution of Cooperation. *Science* **211**, 1390–1396 (1981).
16. Hare, B. Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annu. Rev. Psychol.* **68**, 155–186 (2017).
17. McElreath, R., Boyd, R. & Richerson, P. J. Shared Norms and the Evolution of Ethnic Markers. *Curr. Anthropol.* **44**, 122–130 (2003).
18. Pisor, A. C. & Ross, C. T. Parochial altruism: What it is and why it varies. *Evol. Hum. Behav.* **45**, 2–12 (2024).
19. Silverstein, M. Indexical order and the dialectics of sociolinguistic life. *Lang. Commun.* **23**, 193–229 (2003).
20. Bartolo, R. & Averbeck, B. B. Inference as a fundamental process in behavior. *Curr. Opin. Behav. Sci.* **38**, 8–13 (2021).
21. Baldini, R. Success-biased social learning: Cultural and evolutionary dynamics. *Theor. Popul. Biol.* **82**, 222–228 (2012).
22. Laland, K. N. Social learning strategies. *Learn. Behav.* **32**, 4–14 (2004).

23. Chen, Y., Tran, L. & Saffran, J. Children infer speaker groups from variation in morphosyntactic patterns. *Child Dev.* aacag137 (2026) doi:10.1093/chidev/aacag137.
24. Culbertson, J. & Kirby, S. Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Front. Psychol.* **6**, (2016).
25. Smith, K. Learning Bias, Cultural Evolution of Language, and the Biological Evolution of the Language Faculty. *Hum. Biol.* **83**, 261–278 (2011).
26. Hawkins, R. D. *et al.* From partners to populations: A hierarchical Bayesian account of coordination and convention. Preprint at <https://doi.org/10.48550/arXiv.2104.05857> (2021).
27. Raviv, L., Meyer, A. & Lev-Ari, S. Larger communities create more systematic languages. *Proc. R. Soc. B Biol. Sci.* **286**, 20191262 (2019).
28. Lupyan, G. & Dale, R. Language Structure Is Partly Determined by Social Structure. *PLoS ONE* **5**, e8559 (2010).
29. Wray, A. & Grace, G. W. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* **117**, 543–578 (2007).
30. Meir, I., Israel, A., Sandler, W., Padden, C. A. & Aronoff, M. The influence of community on language structure: Evidence from two young sign languages. *Linguist. Var.* **12**, 247–291 (2012).
31. Busch, B. & Spitzmüller, J. Indexical borders: the sociolinguistic scales of the shibboleth. *Int. J. Sociol. Lang.* **2021**, 127–152 (2021).
32. Epps, P. Amazonian linguistic diversity and its sociocultural correlates. in *Language Dispersal, Diversification, and Contact* 275–290 (Oxford University Press, 2020). doi:10.1093/oso/9780198723813.003.0016.
33. Graff, A. *et al.* An inverse correlation between linguistic and genetic diversity. Preprint at <https://doi.org/10.1101/2024.12.18.628602> (2024).

34. Agha, A. The social life of cultural value. *Lang. Commun.* **23**, 231–273 (2003).
35. Agha, A. Enregisterment. *Oxford Research Encyclopedia of Linguistics* (2024)
doi:10.1093/acrefore/9780199384655.013.1023.

FUNDING

C.B. was supported by Project PID2023-146627NB-100 (Spanish Ministry of Science, Innovation, and Universities CIENCIA/AEI), project 2021-SGR-313 (AGAUR/Generalitat de Catalunya) and a Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation.

AUTHOR CONTRIBUTIONS:

Conceptualization: C.M., C.B., J.C.R.; methodology: C.M., C.B., J.C.R.; data curation: C.M.; software: C.M.; formal analysis: C.M.; visualisation: C.M.; investigation: C.M., C.B., J.C.R.; writing – original draft: C.M., C.B.; writing – review and editing: C.M., C.B., J.C.R.; supervision: C.B.; funding acquisition: C.B.

DATA & CODE AVAILABILITY

The agent-based model, datasets and analysis code used in this study are available in the accompanying GitHub repository: <https://github.com/camroc10/emergent-group-marking>

APPENDIX: MODEL SPECIFICATION AND ANALYTICAL METHODS

A. Model overview

A.1 Purpose

This model investigates how different social ecologies and cultural transmission can affect patterns of linguistic variation. Specifically, it examines whether an uncertain, risky social ecology coupled with the generic cognitive abilities of social inference and social learning can cause linguistic group-marking to emerge. The model is intended as a minimal plausibility test for a causal chain where agents' perception of risk in interactions leads to increased reliance on linguistic cues to discern partners' identities and, consequently, social learning accentuates groups' linguistic distinctiveness.

A.2 Entities

The principal entities are individual agents and social groups. The population contains 200 agents divided equally between two fixed groups. Agents interact pairwise and maintain individual social, behavioural and knowledge states.

Each agent has a linguistic repertoire consisting of three discrete features, each of which can take one of three variants. Agents also maintain information about familiar interaction partners, their experience of stranger interactions, associations between linguistic variants and social groups, accumulated fitness, and estimates of the social interaction structure. These variables are described in Table A.

Table A: Agent-level properties

Variable	Description	Type
group-id	Social group membership	Fixed state
language	Linguistic repertoire	Mutable state
fitness	Accumulated fitness from interactions	Mutable state
knowns	Familiar in-group interaction partners	Mutable state
stranger-counts	Interaction experience counts	Learned state
ingroup-stranger-counts		
p-stranger-est	Interaction structure priors based on counts	Derived belief
p-ingroup-est		
risk-perception	Estimated interaction risk based on priors	Derived belief

language-counts	Variant-group association counts	Learned state
-----------------	----------------------------------	---------------

A.3 State variables and scales

The principal global parameters are population size (`pop-size`), number of groups (`num-groups`), number of linguistic features (`num-features`), number of variants per feature (`num-variants`), interaction rounds per timestep (`interaction-rounds`), social memory size (`social-memory-size`), out-group contact probability (`p-outgroup`), inference cost (`inference-cost`), and parameters governing learning and diffusion. Parameter values are summarised in the table in Section F.

The model has no explicit spatial environment. Social structure is instead represented through probabilistic partner selection and agents' sets of familiar in-group partners. Model time consists of abstract discrete interaction rounds.

A.4 Process overview and scheduling

Each timestep consists of a sequence of interaction, learning and population-level updating processes. First, agents engage in a fixed number of pairwise interactions. During each interaction, a partner is selected, the focal agent selects a cooperation strategy, and a cooperation or defection decision is made. Fitness effects are then assigned and relevant interaction and linguistic beliefs are updated. There is also a low probability of stochastic linguistic diffusion, where the focal agent copies one of their partner's feature variants. Following the interaction phase, social learning occurs noisily within known networks. Agent fitness is subsequently decayed, and model-level output measures are calculated.

The simulation schedule is therefore:

1. Partner selection and interaction
2. Strategy selection
3. Cooperation decision and payoff assignment
4. Interaction structure and linguistic belief updates
5. Linguistic diffusion
6. Social learning/noise
7. Fitness decay
8. Metric calculation

B. Design concepts

B.1 Basic principles

The model assumes that in-group members are, on average, more reliable cooperation partners than out-group members. As demographic expansion causes populations to become more ‘anonymous’, personal familiarity becomes insufficient for identifying appropriate cooperation partners. Agents must therefore use other information to distinguish in-group from out-group strangers.

The model proposes that linguistic variation can provide such information. However, using linguistic information requires cognitive investment, represented here by a costly Bayesian inference strategy. The model therefore links social ecology to linguistic differentiation through a sequence of mechanisms: social uncertainty increases perceived risk; increased risk increases the expected value of inference; inference creates selective value for linguistic cues to group membership; and social learning can subsequently amplify these differences.

B.2 Objectives

Agents seek to maximise their fitness from interactions. This requires distinguishing partners who are likely to cooperate successfully from those for whom cooperation negatively affects fitness.

B.3 Emergence

Group marking is not directly imposed on agents: they are not given an objective to maintain group-distinctive linguistic repertoires, nor are linguistic variants initially assigned fixed social meanings. Linguistic differentiation emerges from interactions between agents’ decisions, learning, and the cultural transmission. The model therefore treats group marking as an emergent population-level pattern arising from individual-level adaptive behaviours.

B.4 Sensing

Agents directly observe whether a partner is familiar or unfamiliar and observe the partner’s linguistic signal. The true group identity of a stranger is not available at the time of the cooperation decision. Agents may subsequently learn a stranger’s true group membership with probability p_{observe} .

B.5 Adaptation and learning

Agents adapt their behaviour through experience. They maintain and update estimates of stranger frequency and the probability that strangers belong to their in-group. When the true identity of a stranger is observed, agents also update associations between linguistic variants and social groups.

Linguistic repertoires are adapted through success-biased social learning within agents' known networks. Lower-fitness agents preferentially learn from higher-fitness members of their network, while stochastic noise introduces occasional variant changes. A separate low-probability diffusion process also allows linguistic material to spread between interacting individuals, regardless of their respective group identities.

B.6 Interaction

Interactions are pairwise and involve a focal agent and a selected partner. Partners are selected according to the out-group contact probability (p_{outgroup}), after which the partner's identity is sampled randomly from the relevant group. Interactions may therefore occur between familiar in-group partners, unfamiliar in-group partners, or out-group partners.

B.7 Stochasticity

Stochasticity is present throughout the model. Initial linguistic repertoires, interaction partner selection, strategy selection, cooperation decisions, stranger identity observation, social learning models and diffusion or copying noise are all determined stochastically.

B.8 Collectives

Agents belong to one of two social groups, which constitute exogenous fixed collectives. Group membership is assigned at initialisation and remain fixed throughout the simulation. The groups themselves do not have independent behavioural rules or state variables; rather, membership determines the social category relevant to cooperation, interaction structure and linguistic differentiation.

B.9 Observation

The principal model outcomes are perceived interaction risk, proportion of stranger interactions in which inference is used, and linguistic differentiation between groups. Linguistic differentiation is measured using global Jensen–Shannon divergence and feature-level Jensen–Shannon divergence.

The main parameter sweep records these measures for each combination of social memory size and out-group contact probability. Counterfactual model variants record global Jensen–Shannon divergence under fixed strategy conditions, while time-series experiments record inference use and linguistic differentiation at each timestep.

Model outputs are exported from BehaviorSpace as CSV files. Summary statistics are calculated over the final 500 timesteps of each simulation for the main analyses, while the time-series experiments retain observations at individual timesteps. Statistical analyses include multiple linear regression, comparison of linear and logistic fits, Spearman rank correlations and paired Wilcoxon tests. These data were visualised using heatmaps, scatter plots and line charts.

C. Details

C.1 Initialisation

At setup, 200 agents are created and divided equally between the two social groups. Each agent is assigned a linguistic repertoire by randomly selecting a variant for each of the three linguistic features. Interaction structure and language–group association counts are initialised using the specified pseudocounts. Each agent is then assigned a randomly selected in-group member to its set of familiar partners.

C.2 Input data

The model does not use any external data during simulation. All inputs are specified through model parameters and experimental conditions defined in the BehaviorSpace experiments.

C.3 Submodels

C.3.1 Partner selection

Interaction partners are chosen probabilistically. This is implemented in NetLogo by generating a random number: if lower than p_{outgroup} , a random agent from the out-group is chosen; else, a random in-group member is chosen, which may or may not be familiar to the agent.

C.3.2 Strategy selection

As indicated in the Model Design (Section 2.3.2), agents choose between heuristic and inference strategies probabilistically according to a logit choice rule, where an increase in a strategy’s expected utility made it more likely to be selected. Our model used a steep curve

($s = 15$, corresponding to the model parameter `decision-steepness`), yielding a sharp transition between strategies.

C.3.3 Bayesian inference

For each group and linguistic feature, agents store counts representing variant–group associations. These counts are converted to conditional probabilities $P(v_f|g)$ when inference is performed. To ensure numerical stability, Bayesian inference is implemented in log space. The product over features is replaced by a sum of log-likelihoods:

$$\log P(g|s) \propto \log P(g) + \sum_f \log P(v_f|g)$$

where s denotes the observed linguistic signal and v_f the variant observed for feature f .

These log-scores are converted into probabilities via a softmax transformation over groups:

$$P(g|s) = \frac{e^{\text{score}(g)}}{\sum_{g'} e^{\text{score}(g')}}$$

This is equivalent mathematically to direct Bayesian normalisation but avoids numerical underflow from multiplying small probabilities.

C.3.4 Cooperation decision and payoff

The final cooperation action is chosen probabilistically, as described in Section 2.3.4. This logit choice rule uses the same `decision-steepness` values as before ($s = 15$).

Agents' `fitness` is then updated according to the adaptiveness of their action given their partners identity, as described in Section 2.4.1.

C.3.5 Diffusion

There is a low probability (`diffusion-rate`) per interaction that agents copy a random variant from their interaction partner.

C.3.6 Interaction learning

Agents store counts based on prior interaction experience, forming a beta distribution. Agents store counts based on two types of interaction partners: strangers (both groups), and specifically in-group strangers. Counts are stored in pairs: α (occurrence of given type) and β (not the given type).

Stranger counts are incremented by one at each timestep; in-group stranger counts are only updated occasionally when true group identity is revealed (at a rate of `p-observe`).

These counts are then converted into probabilities, the priors `p-stranger-est` and `p-ingroup-est`, using the mean formula:

$$P(\alpha) = \frac{\alpha}{\alpha + \beta}$$

An agent's risk perception is calculated as described in Section 2.3.1 using these priors.

Agents also store counts of variant-group associations. Given they rely on observing true group identity, they are also only updated at a rate of `p-observe`. Each feature-variant-group position observed in the agent's array is incremented by 1. This is the distribution used to calculate the probabilities for Bayesian inference.

C.3.7 Known partners update

If both agents belong to the same in-group and have space in their social memory (as limited by `social-memory-size`), they both add each to their `knowns agentset`.

C.3.8 Social learning

Social learning takes place within known networks. A random percentage in the range `[0, social-learning-rate]` of the worst performing agents by fitness is selected to learn at each timestep. A higher-fitness agent is chosen from their known network to serve as a model. The learners adopt the models' full language, with the of one feature possibly being set to a random variant (miscopied) at a likelihood of `noise-rate`.

C.3.9 Fitness decay

Fitness is decayed at each timestep, ensuring it does not spiral endlessly. After the social learning phase, all agents' fitness is set to 95% of its previous level.

C.3.10 Metrics calculation

Linguistic differentiation between groups was quantified using the Jensen-Shannon (JS) divergence between group-level linguistic distributions. For each group, frequencies were calculated separately for every feature-variant combination and pooled into a single probability distribution of length $(F \times V)$, where (F) is the number of linguistic features and (V) the number of variants per feature. Given two group distributions (Q) and (R) , the midpoint distribution was defined as

$$M = \frac{Q + R}{2}$$

JS divergence was then calculated as

$$JS(Q, R) = \frac{1}{2}D_{KL}(Q \parallel M) + \frac{1}{2}D_{KL}(R \parallel M)$$

where D_{KL} denotes the Kullback–Leibler divergence. Values were normalised by $\ln 2$, producing a measure bounded between 0 and 1, where 0 indicates identical linguistic distributions and 1 maximal divergence.

Feature-level JS divergence was calculated using the same procedure, except that distributions were constructed from the variants of a single feature rather than all feature–variant combinations. For a feature with (V) variants, the resulting distributions therefore had length (V). The minimum, median and maximum feature-level JS divergence across all features were recorded.

Other metrics recorded included mean risk-perception and the proportion of inference used in stranger interactions.

D. Experimental design and analysis

D.1 Parameter sweep

All parameter sweeps were executed using NetLogo’s BehaviorSpace tool. The experimental parameters were varied as follows:

- `p-outgroup`: [0.01, 0.1, 0.2, 0.3, 0.4, 0.5]
- `social-memory-size`: [10, 25, 50, 75, 90, 100]

Each parameter combination was simulated 30 times for 3000 timesteps, more than enough time for the model dynamics to stabilise. Simulation outputs were exported to CSV files. Statistical analyses were performed on parameter-level aggregates computed over the final 500 timesteps of each run.

D.2 Counterfactual model variants

The model variants used to isolate inference’s role in the causal chain, described in the Causal Intervention section, used identical parameter values as the main model, varying only the availability of cooperating strategies to agents. This bypassed the strategy choice, with agents always using either the heuristic or inferential strategies, depending on the variant. In this case, the only metric recorded was global Jensen-Shannon divergence. Once again, the analyses were performed on parameter-level aggregates from the final 500 timesteps of each run.

D.3 Time-series experiments

The time series data were obtained from two parameter regimes, corresponding to a high-risk regime (`p-outgroup` = 0.3, `social-memory-size` = 20) and a low-risk regime (`p-outgroup` =

0.05, social-memory-size = 40). These models were also simulated 30 times and ran for 3000 timesteps, while metrics were recorded at each tick. The analyses were performed on the average values per timestep across the 30 runs.

E. Software

The model was implemented in NetLogo version 7.0.3 using the array extension (<https://www.netlogo.org/>). Parameter sweeps were executed using NetLogo's BehaviorSpace tool. Output data were exported as CSV files and analysed in Jupyter Notebook using Python 3.12.7 with the packages pandas, NumPy, SciPy, statsmodels, Seaborn and Matplotlib.

F. Parameter table

Parameter	Description	Value
<i>Structural parameters:</i>		
pop-size	Population size	200
num-groups	Number of social groups	2
num-features	Number of linguistic features per agent	3
num-variants	Number of linguistic variants per feature	3
interaction-rounds	Number of interactions per timestep	3
social-memory-size	Maximum number of individuals an agent can be familiar with	sweep (see D.1)
<i>Interaction parameters</i>		
p-outgroup	Out-group contact rate	sweep (see D.1)
inference-cost	Cognitive cost of inference in expected utility calculation	0.3
decision-steepness	Sensitivity of logit choice rule	15
payoff	Reward for adaptive cooperation	1
penalty	Penalty for maladaptive cooperation	1
<i>Learning parameters</i>		
p-observe	Probability of observing true group identity	0.2
social-learning-rate	Maximum proportion of agents updated per timestep	0.05
noise-rate	Mutation probability during social learning	0.33
diffusion-rate	Probability of feature copying during interaction	0.001
Initial pseudocounts		1
Learning increment (counts)		1
<i>Experimental design:</i>		
Simulation length		3000 timesteps
Analysis window (final)		500 timesteps
Runs per condition		20