

Research



Cite this article: Macanovic A, Tsvetkova M, Przepiorka W, Buskens V. 2024 Signals of belonging: emergence of signalling norms as facilitators of trust and parochial cooperation. *Phil. Trans. R. Soc. B* **379**: 20230029. <https://doi.org/10.1098/rstb.2023.0029>

Received: 31 May 2023

Accepted: 2 September 2023

One contribution of 15 to a theme issue ‘Social norm change: drivers and consequences’.

Subject Areas:

behaviour, cognition, evolution

Keywords:

social norms, trust, cooperation, signalling, intergroup conflict, group identity

Authors for correspondence:

Ana Macanovic

e-mail: a.macanovic@uu.nl

Wojtek Przepiorka

e-mail: w.przepiorka@uu.nl

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6980681>.

Signals of belonging: emergence of signalling norms as facilitators of trust and parochial cooperation

Ana Macanovic^{1,2}, Milena Tsvetkova³, Wojtek Przepiorka¹ and Vincent Buskens^{1,2}

¹Department of Sociology/ICS and ²Centre for Complex Systems Studies, Utrecht University, Utrecht, The Netherlands

³Department of Methodology, London School of Economics and Political Science, London, WC2A 2AE, UK

AM, 0000-0003-0800-5271; MT, 0000-0002-3552-108X; WP, 0000-0001-9432-8696; VB, 0000-0002-4483-7238

Mechanisms of social control reinforce norms that appear harmful or wasteful, such as mutilation practises or extensive body tattoos. We suggest such norms arise to serve as signals that distinguish between ingroup ‘friends’ and outgroup ‘foes’, facilitating parochial cooperation. Combining insights from research on signalling and parochial cooperation, we incorporate a trust game with signalling in an agent-based model to study the dynamics of signalling norm emergence in groups with conflicting interests. Our results show that costly signalling norms emerge from random acts of signalling in minority groups that benefit most from parochial cooperation. Majority groups are less likely to develop costly signalling norms. Yet, norms that prescribe sending costless group identity signals can easily emerge in groups of all sizes—albeit, at times, at the expense of minority group members. Further, the dynamics of signalling norm emergence differ across signal costs, relative group sizes, and levels of ingroup assortment. Our findings provide theoretical insights into norm evolution in contexts where groups develop identity markers in response to environmental challenges that put their interests at odds with the interests of other groups. Such contexts arise in zones of ethnic conflict or during contestations of existing power relations.

This article is part of the theme issue ‘Social norm change: drivers and consequences’.

1. Introduction

At times, mechanisms of social control reinforce norms that appear individually, or even collectively, costly or ‘wasteful’ [1]—such as mutilation practises [2] or extensive body tattoos [3]. We test the conjecture that such norms emerge as outcomes of signalling games [4–7] in contexts where groups could benefit from parochial cooperation—i.e. cooperation with members of one’s own group. The resulting signalling norms prescribe behaviours that mark individuals’ belonging to a certain group, thereby helping distinguish between ingroup ‘friends’ with aligned interests and outgroup ‘foes’ with opposing interests [5,6]. We build on insights from literature on costly signalling and parochial cooperation to shed light on the emergence of social norms prescribing the displaying of signs of group belonging.

Research on signalling has shown that costly behaviours which signal individuals’ cooperative intent can be part of an (evolutionarily stable) Nash equilibrium and enable observers of these signals to distinguish between cooperators and defectors [4,7–12]. These signals are reliable if only cooperators can afford to send them, either because cooperators incur lower signalling costs or gain higher cooperation benefits compared to untrustworthy defectors [13]. While some have suggested that norms prescribing costly signalling can

emerge from arbitrary behaviours introduced by a single individual [5], others have indicated that more substantial 'shocks' are needed to shift a population from a non-signalling to a signalling equilibrium [4,9]. At the same time, the literature on parochial cooperation has shown that cooperation conditional on group belonging emerges and persists in the presence of intergroup competition or conflict [14,15] within groups of relatively small sizes [16]. This strand of literature has mostly assumed that cooperation can be conditioned on apparent innate features (i.e. conspicuous physical characteristics) that double as hard-to-fake markers of group identity [16–20].

We bring together similar—albeit often differently conceptualized—concepts from these two literatures to understand the conditions that favour the emergence of signalling norms as facilitators of parochial cooperation [6,21]. Following the literature on parochial cooperation, we put the emphasis on intergroup conflict and the resulting need to identify ingroup members and cooperate conditional on group identity. In line with the literature on signalling, we operationalize indicators of group identity as symbolic—and often costly—markers that can, but do not have to, be displayed by individuals [22,23]. In particular, we build on the trust game with signalling previously introduced to derive hypotheses about the emergence of signalling norms [6] and incorporate it in an agent-based model. Our agent-based model allows us to study the dynamics of norm emergence and parochial cooperation that intergroup conflict may bring about. Our research question is: under what conditions do signalling norms emerge to facilitate parochial cooperation?

Imagine two large hunter–gatherer bands that are in conflict over limited resources in their shared habitat [24]. A group that has spent a very long time hunting away from their band is coming home and approaching one of the settlements. Others in this band might want to welcome the hunters, hoping for a share in the kill. Yet, as the group has been away for so long, the band cannot as easily recognize which band this particular group belongs to. If they are from the same band, the incoming hunters will be cooperative; if they are from another band, the group might have come to raid the settlement instead. To prevent being mistaken for members of an enemy group and win the trust of their band, the approaching group of hunters can display a reliable sign of band belonging, such as a difficult to fake local dialect or a tattoo with group-specific features [25]. Once the band members recognize that this signal corresponds to their own group, they can rest assured and welcome the arrivals; otherwise, they can play safe and stay away from interacting with the group of strangers that is approaching their settlement.

In particular, we test the hypothesis that, in contexts where an outgroup is frequently encountered—such as in zones of ethnic conflict [21,26,27], prisons with rival gangs [28,29] or during contestations of existing power relations [30,31]—group members will be more willing to invest considerable resources in costly displays of group belonging [4,6,21,32,33]. We call behaviours that change or reinforce observers' beliefs about someone's belonging to a social group signalling norms [6]. We consider signals that are hard to fake (for instance, because they require inaccessible group-specific knowledge) and cannot be discriminated against by the outgroup (for example, because they are unknown to them [34,35]). We first provide analytical results

on the conditions under which signal display and recognition can be preferred strategies of group members. We then employ agent-based simulations to understand how likely random changes in individual behaviours are to push populations from a non-signalling state into a state where the display of signals and their recognition are widely adopted—and the signalling norm is accepted [5]. Finally, we relax several model assumptions to understand an even broader range of conditions that favour the emergence of signalling norms.

Our study contributes to the literature on norm emergence and change in multiple ways. First, our agent-based model enhances our understanding of how signalling norms can solve trust dilemmas often arising in real-world exchange situations [13]. Second, we show that norms prescribing costly signalling of group identity can emerge and facilitate parochial cooperation in minority groups from initial conditions without signalling and signal recognition. Third, we show how costless signals of group identity support parochial cooperation in groups of all sizes. We thereby reveal a crucial difference in the dynamics of signalling trait emergence between norms of sending costly and costless signals. Fourth, we discuss how signalling norm emergence becomes more challenging for minority groups if signals can be recognized by the outgroup or there is significant noise in the perception of an individual's group identity. Our findings shed light on various real-life situations, such as how dialects [36] or specific types of scars [23] can be used to distinguish between ethnic groups in contexts of ethnic conflict or resource competition. Further work in this domain can help understand how groups coordinate on specific signalling norms in similar contexts, and how established signalling norms can hamper cooperation between groups once intergroup conflict is resolved.

2. Theory

(a) Modelling the trust game with two groups

We incorporate the trust game with signalling [6] in an agent-based model to map the boundary conditions for the emergence of signalling norms and their impact on individual and group outcomes. The trust game (electronic supplementary material, figure S1) is a sequential game where one individual (the truster) chooses whether or not to trust the other individual (the trustee), who then decides whether to honour or abuse that trust [37]. In the trust game with signalling, the trustee can signal their identity to the truster before the truster makes a decision on whether to trust the trustee.¹ Using a trust game to study the emergence of signalling has two main advantages over using other games such as the Prisoner's Dilemma. First, the trust game models sequential decisions which are characteristic of many real-world interactions (e.g. market exchanges or hiring decisions). Second, in the trust game, player roles are not interchangeable and pay-offs are, therefore, asymmetric. Both the sequential nature and asymmetry of the trust game allow us to attach behaviour to either trust or cooperation. This is unlike simultaneous-move and symmetric dilemmas (such as the Prisoner's Dilemma), where it is unclear whether it is fear or greed that drives behaviour [40].

We adopt the trust game with signalling [6] to model exchange relations between members of two groups with

conflicting interests (electronic supplementary material, figure S2). We set the pay-offs such that a trustee always has an incentive to honour the trust of an ingroup truster and abuse the trust of an outgroup truster. Hence, trusting an ingroup trustee results in a pay-off of R_I for both parties [26], whereas trusting an outgroup trustee results in a pay-off of S for the truster and T for the trustee. If trust is not given, there is no exchange and both parties receive pay-off P . The pay-offs are ordered as follows: $R_I > T > R_O > P > S$. R_O denotes the pay-off of both parties if a trustee honours the trust of an outgroup truster. By setting $R_I > T > R_O$, we hard-code the conflicting group interests into our model, so that trustees do not have an incentive to honour outgroup trusters' trust. This allows us to examine the role of signalling in contexts where, for instance, environmental constraints make capturing resources from neighbouring groups attractive [24,29]. In our model, trustees can pay a cost c to signal their group identity and trusters can condition their trust on these signals.

Our model makes three important assumptions: (i) signals are hard or impossible to fake [32]; (ii) recognizing signals requires group-specific knowledge, so that members of the outgroup do not recognize them [34]; and (iii) trusters reveal their group identity when placing trust, which allows trustees to (not) honour trust conditional on a truster's group belonging.² With these assumptions our model captures situations in which trusters, but not trustees, lack information about the other party's group belonging (and, thereby, intentions). In §4, we discuss the consequences of relaxing these three assumptions.

(b) Conditions for signalling norm existence

In this section, we outline the conditions under which trusters consider group identity signals and trustees bear the costs of sending them. Individuals are first assigned a role of a truster or a trustee at random and are then randomly matched with another individual of the opposite role. An individual will meet a member of their group with a probability α that is equal to the group's share in the total population (p_i where $i \in \{1, 2\}$ refers to either of two groups). We build on the suggestion that, the lower the probability of encountering ingroup members, the more likely a group is to develop a signalling norm [6]. More precisely, with $\alpha \leq \alpha^*$ (in our set-up: minority groups) trusters encounter their ingroup so rarely that they are best off distrusting any trustee in the absence of signalling. With $\alpha > \alpha^*$ (majority groups), trusters encounter their ingroup frequently enough to prefer trusting any trustee in the absence of signalling (electronic supplementary material, equations S1 and S2). In theorem 2.1, we establish the conditions necessary for the equilibrium where, within a group, all group members signal and conditionally trust (i.e. give trust conditional on having observed the ingroup's signal). In theorem 2.2, we establish the existence of an equilibrium without signalling and signal recognition. We provide proofs and the full analysis in the electronic supplementary material, S1. Finally, we define signalling norm emergence as the process during which a group in a non-signalling equilibrium moves to a signalling equilibrium.

Theorem 2.1. *There exists a Nash equilibrium where trusters condition their trusting on signals of the ingroup and trustees display group identity signals (signalling equilibrium) if and only if the*

signalling costs are offset by the benefits of parochial cooperation established with the help of signalling ($c \leq \alpha[R_I - P]$, where $R_I - P$ captures the benefits of cooperation).

Theorem 2.2. *In a group with $\alpha \leq \alpha^*$, there exists a Nash equilibrium where trusters unconditionally distrust all trustees and trustees do not display group identity signals (non-signalling equilibrium). In a group with $\alpha > \alpha^*$, there exists a non-signalling equilibrium where trusters unconditionally trust all trustees and trustees do not display group identity signals.*

To understand how signalling norms emerge, we evaluate how likely a group is to move from a non-signalling into a signalling equilibrium from random changes in strategies of trustees or trusters (e.g. a trustee suddenly starts sending some signal of group identity). Note that we only consider a signalling norm to have emerged if trustees signal their identity *and* trusters recognize and condition their trust on these signals. We provide the full analysis in the electronic supplementary material, S1.2. Our analysis shows that trusters in a group where $\alpha \leq \alpha^*$ are willing to recognize signals, rather than unconditionally distrust all trustees, as long as the share β of signalling ingroup trustees satisfies the following condition:

$$\beta \geq \beta^* = 0. \quad (2.1)$$

In the group where $\alpha > \alpha^*$, trusters recognize signals, rather than unconditionally giving their trust to all trustees, if the share β of signalling ingroup trustees satisfies the following condition:

$$\beta \geq \beta^* = 1 - \frac{(1 - \alpha)(P - S)}{\alpha(R_I - P)}. \quad (2.2)$$

In both groups, trustees will be willing to bear the cost c of sending a signal of their group identity if the share γ of ingroup trusters who recognize ingroup signals satisfies the following condition:

$$\gamma \geq \gamma^* = \frac{c}{\alpha(R_I - P)}. \quad (2.3)$$

These conditions show that the threshold β^* is always lower in groups with $\alpha \leq \alpha^*$. That is, trusters in minority groups are always willing to trust conditional on having observed the signal, which further ensures that the threshold γ^* needed for trustees to signal is more easily reached. Trusters in a group with $\alpha > \alpha^*$ only start conditioning their trusting behaviour on group identity signals if a sufficiently high number of trustees already signals their identity. We therefore expect that random changes in strategies of trustees and trusters will be more likely to shift the minority, rather than majority, groups into the signalling equilibrium.

(c) Modelling signalling norm emergence in populations

Our formal analysis allows us to formulate expectations regarding the likelihood of groups to shift to the signalling equilibrium under certain conditions. However, to understand the dynamics of signalling norm emergence from random variations in individual agent behaviours across different conditions, we use an agent-based model with social learning dynamics.

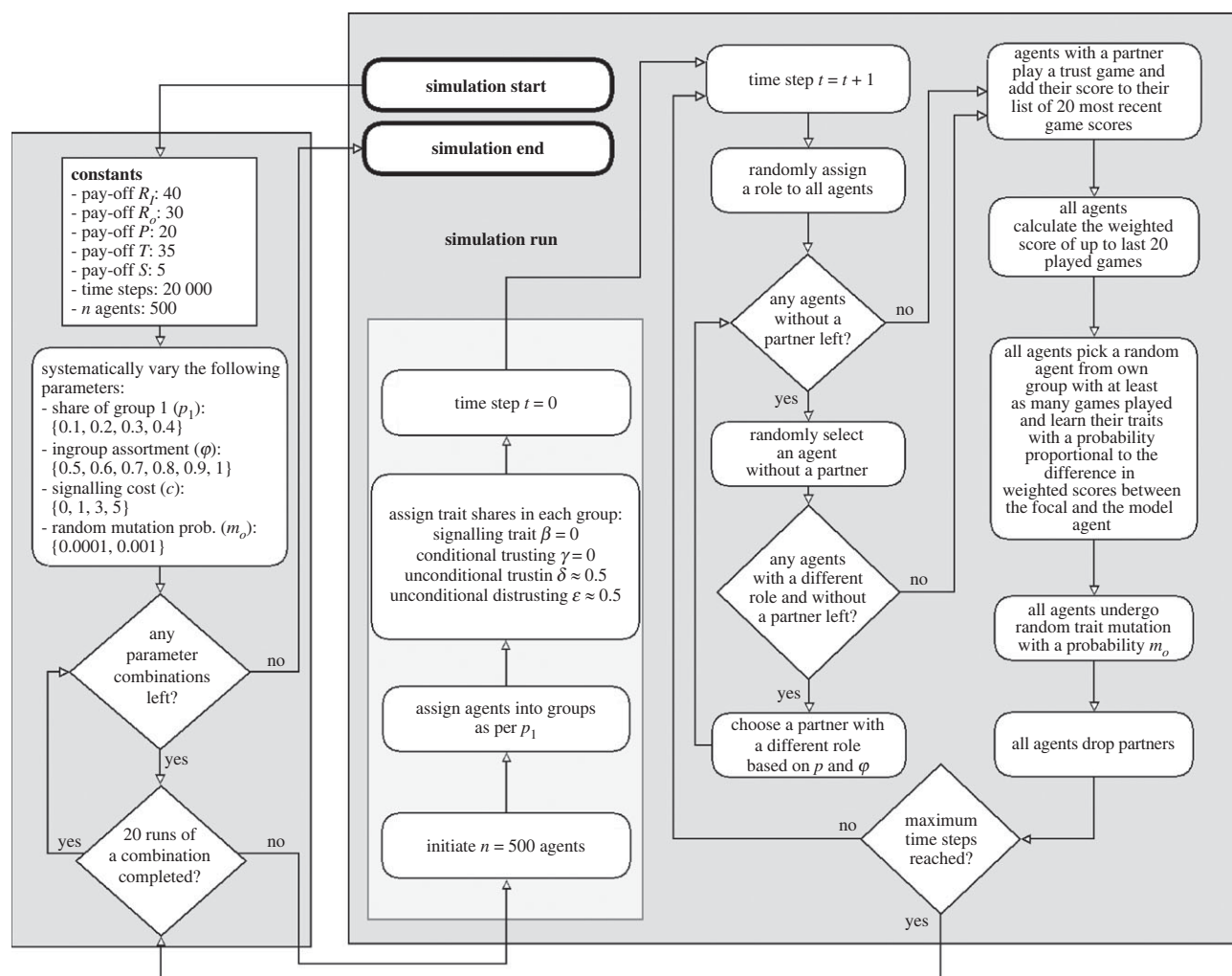


Figure 1. Flowchart of the simulation process. For each combination of parameters, we run 20 independent randomly initiated simulations. After all available agents have played the trust game, all agents undergo social learning and random mutation before the simulation moves onto the next time step—for a maximum of 20 000 steps.

Agents in our model have two main traits that determine their strategy in each of the roles they can assume in the game: the signalling and the trusting trait. The signalling trait determines whether an agent signals their group identity at a predefined cost c when acting as a trustee. We vary signalling costs to capture a range of different behaviours that can convey information about group membership. The trusting trait defines whether an agent will, in the role of a truster, trust unconditionally, distrust unconditionally, or trust conditionally (upon having recognized a signal of their ingroup). There are six possible strategies resulting from different combinations of these two traits.

We simulate a population of 500 agents randomly assigned to group 1 with probability p_1 and group 2 with probability $p_2 = (1 - p_1)$. Group sizes and memberships are held fixed. Agents assume the role of a truster or a trustee with equal probability in each round and are then matched with another agent of the opposite role with whom they play the trust game once (if without a match, an agent sits out the round). As in our theoretical model, the probability of meeting one's ingroup α is equal to the share of ingroup p in the population. Additionally, we introduce parameter ϕ capturing assortative matching with regard to group membership: the probability of meeting the ingroup, holding p fixed, increases with ϕ . In our agent-based model α , thus, depends on both p and ϕ .³ Agents cannot select whom they

are matched with and have no memory of past encounters. Such a set-up helps us model scenarios where group members frequently interact with strangers, which is a rather adverse environment for the evolution of parochial cooperation [41].

We set the trust game pay-offs such that $\alpha^* = 0.43$ (see figure 1 for pay-offs and the electronic supplementary material for α^* calculation). Given this, we choose values of p to test scenarios with α both below and above this threshold. We do this so that the group with $\alpha \leq \alpha^*$ always constitutes a minority in the population and faces a group with $\alpha > \alpha^*$ that constitutes the majority in the population. As our focus is on the emergence of the two traits that constitute the signalling norm (signal sending and recognition), we initialize the model without any signalling ($\beta_1 = \beta_2 = 0$) or conditional trusting (i.e. signal recognition; $\gamma_1 = \gamma_2 = 0$) in either of the groups and randomly assign agents to unconditional trusting and unconditional distrusting in both groups ($\delta_1 \approx \delta_2 \approx 0.5$ and $\epsilon_1 \approx \epsilon_2 \approx 0.5$). We introduce random mutations of agents' traits with probability m_o , and vary this value to capture different propensities of agents to introduce new behaviours with regard to identity signalling [42].

After playing a trust game, agents can update their strategies by observing and copying both of the traits of their more successful ingroup members in a mechanism resembling social learning [43,44]⁴ and, thereafter, can also

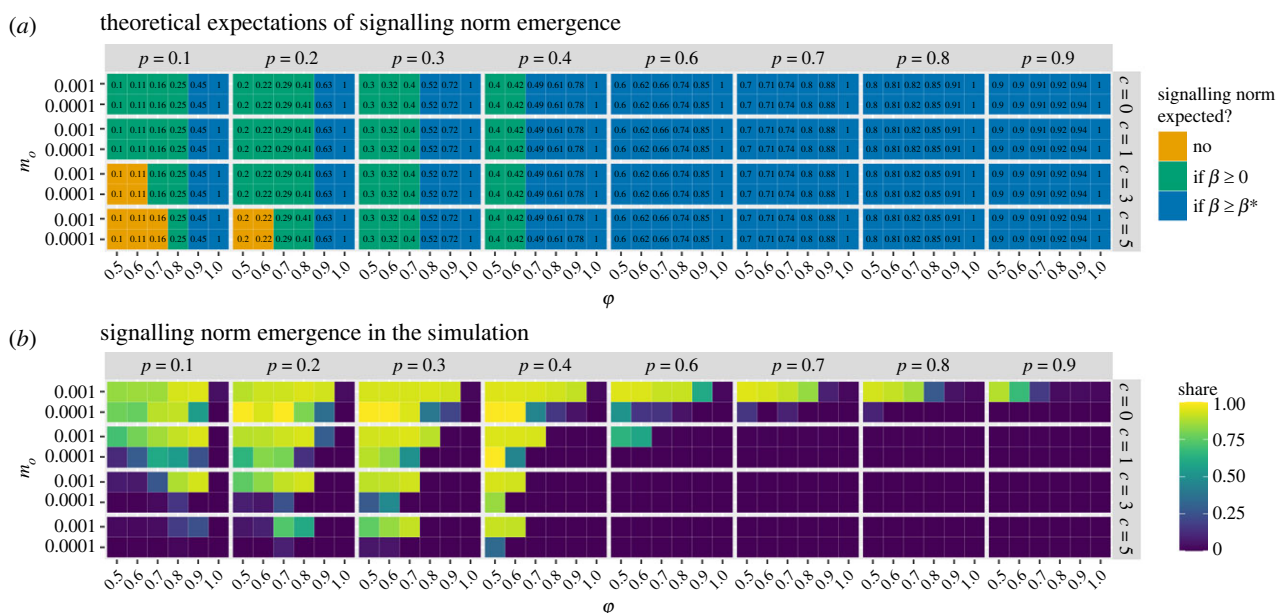


Figure 2. (a) Theoretical expectations regarding the signalling norm emergence given our analytical results. Tile colours correspond to different emergence conditions (orange corresponds to the condition in theorem 2.1, green in equation (2.1) and blue in equation (2.2)) and tile labels correspond to the α values resulting from different combinations of p and ϕ . (b) The results of our agent-based simulations. Tile colouring reflects the share of agents with a strategy including both signalling and trusting conditional on signals averaged across the last 100 rounds given the share of ingroup (p) and the assortment parameter (ϕ) under varying signal costs c and random mutation parameters. We average the results over 20 independent runs for each parameter combination. See the electronic supplementary material, figures S4 and S5 for an overview of shares of other strategies.

undergo a random mutation to one of their traits. We simulate the evolution of agent traits for 20 000 time steps, evaluating each combination of parameters (see the electronic supplementary material, table S3) in 20 independently initialized simulation runs. Figure 1 gives an overview of the full simulation procedure. Finally, we also run reference simulations for 2500 time steps following the same procedure, but excluding a possibility of signalling norm emergence. These simulations serve as a baseline for comparing the outcomes of the signalling equilibrium to those of a non-signalling one (electronic supplementary material, figure S3).

3. Simulation results

(a) Signalling and parochial cooperation

In figure 2a, we show the expectations regarding signalling norm emergence according to our formal analysis (electronic supplementary material, equations S1 and S2, and S1.2). In figure 2b, we show the results of our agent-based simulations.⁵ Figure 2b shows the prevalence of the signalling norm (i.e. agent strategy including both a signalling and a conditionally trusting trait) within each group averaged across the last 100 rounds of simulations. Henceforth, we refer to groups with $p < 0.5$ as minority groups, and groups with $p > 0.5$ as majority groups. For ease of interpretation, we focus on cases when $\phi = 0.5$ (therefore, $\alpha_i = p_i$) unless stated otherwise.

As we predict based on our analytical results, minority groups develop signalling norms (as $\beta \geq 0$) when signal costs are low enough.⁶ For example, when $c = 1$ and random mutations are high ($m_0 = 0.001$) almost all minority groups fully develop a signalling norm (second row in figure 2b, upper tick on the y-axis). In minority groups, trusters are indifferent between distrusting unconditionally and trusting upon observing the signal (equation (2.1)). Thus, if a few trustees

start signalling (following a random mutation), they are likely to encounter trusters who have started—and continued—conditionally trusting frequently enough to reap the benefits of parochial cooperation. These benefits offset the costs of signalling, allowing social learning to spread both signalling and conditional trusting in the group. However, when $c > 1$, groups with $\alpha \leq 0.2$ cannot afford to send signals.⁷ Furthermore, if mutation rates are low ($m_0 = 0.0001$), the signalling norm spreads more slowly and does not become as widely adopted (also see next subsection).

In most cases, signalling norms do not emerge in groups with $\alpha > \alpha^*$ if signals are costly. Trusters in these majority groups are unconditionally trusting and only willing to condition their trusting behaviour on a signal if the share of signalling trustees is sufficiently high (equation (2.2)). Therefore, even if some trustees do start signalling, social learning is likely to bring any trusters who start conditionally trusting back to unconditional trust instead; in turn, as long as the share of conditionally trusting trusters remains too low (equation (2.3)), the number of trustees willing to bear the costs of sending signals will remain low as well.⁸ Yet, when signals are costless ($c = 0$), trustees who start signalling will have no incentive to stop doing so and signalling can spread through social learning. This, in turn, makes conditional trusting more attractive for trusters and establishes the signalling norm in some majority groups as long as their α is sufficiently low (otherwise, a very large share of signalling is needed for conditional trust to proliferate) and the random mutations are sufficiently high (topmost row in figure 2b).

Figure 2 also shows that smaller minority groups adopt signalling more, and larger minority groups less, at higher levels of assortment (i.e. as their α increases, see also figure 2a and electronic supplementary material, table S2 for easier interpretation). We find that, overall, groups with similar α values

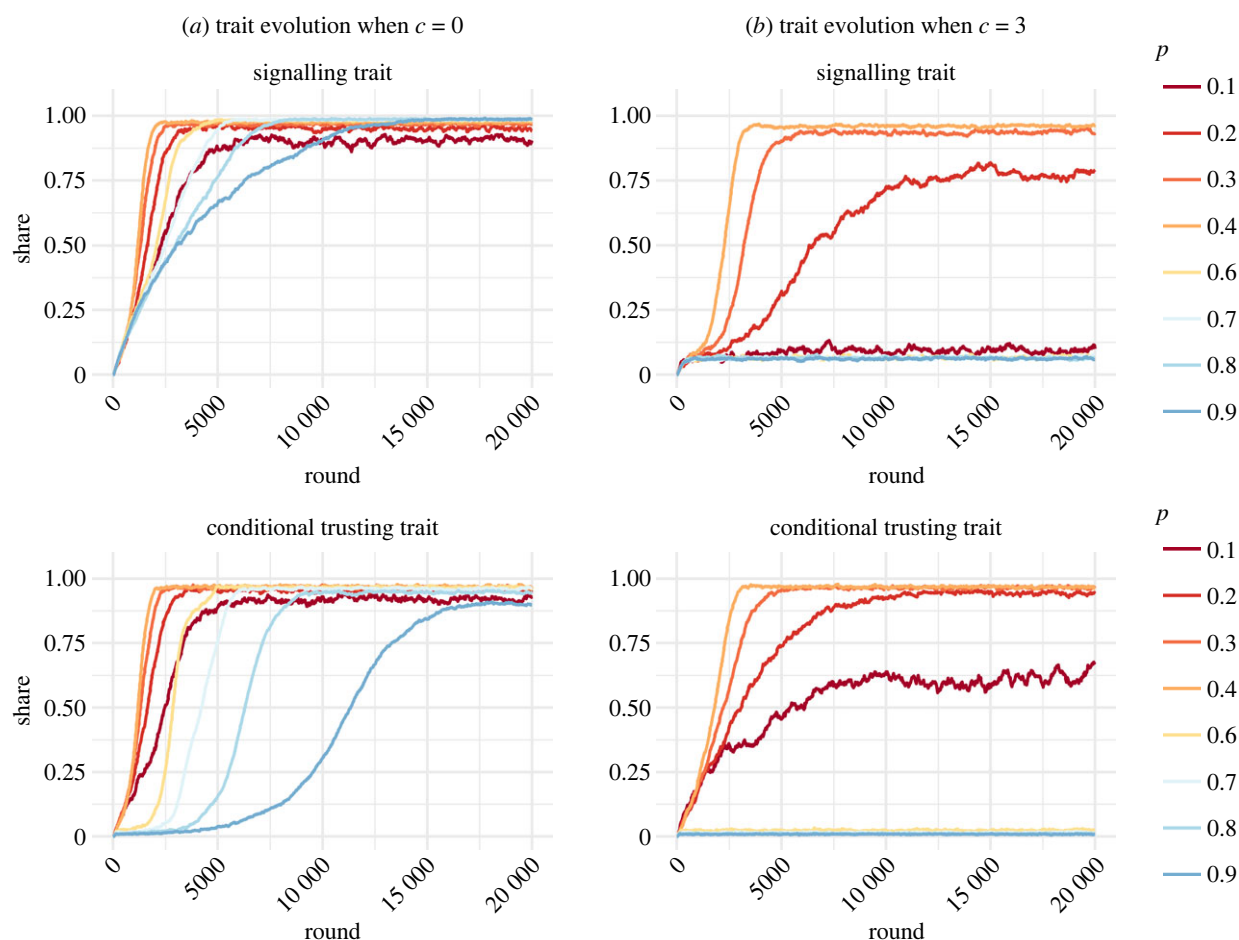


Figure 3. Evolution of signalling and conditional trusting traits over 20 000 rounds if signals are costless ($c = 0$) (a), or costly ($c = 3$) (b); there is no assortment ($\varphi = 0.5$) and random mutations are more common ($m_0 = 0.001$). Shares on the y-axis represent share of different traits within each group. We average the results over 20 independent runs for each parameter combination.

stemming from different combinations of p and φ develop comparable shares of the signalling and conditional trusting strategy (electronic supplementary material, figure S6). Yet, as we show in the next subsection, dynamics of signalling norm emergence can differ depending on the group size.

Overall, groups that develop a (costly) signalling norm are better off than they were in the absence of signalling (see the electronic supplementary material, figure S7 panel A for the average pay-offs obtained by agents within each group compared to a non-signalling baseline scenario). Yet, while a group's signalling and conditionally trusting behaviours are not dependent on the outgroup's strategies in our model, changes in outgroup trait distributions still affect group pay-offs. For instance, despite obtaining benefits from establishing parochial cooperation with the help of signals, minority groups net losses when signals are costless ($c = 0$). This is owing to the fact that the emergence of signalling norms (and parochial cooperation) in majority groups closes the door for the minority groups' exploitation of the majority. When signals are costly, despite bearing the costs of sending signals, minority groups that adopt signalling and conditional trusting see an increase in average pay-offs per agent that mainly stems from truster benefits (see the electronic supplementary material, figure S7 panels B and C). However, if signalling behaviour appears in the group, but is not followed by sufficient conditional trusting (e.g. the electronic supplementary material, figure S7 red tiles when $p = 0.6$ and $c > 0$ in panel A), groups net losses because

trustees bear the costs of signalling, but do not reap the benefits of parochial cooperation.

(b) Dynamics of signalling norm emergence

To understand the dynamics of norm emergence we show trait change dynamics in figure 3, comparing a scenario with costless signals (a) and a scenario with signals of intermediate cost (b). The figure shows the coevolution of the signalling trait (upper half of panel) and the conditional trusting trait (lower half of panel). For ease of interpretation, we only consider scenarios without any group assortment ($\varphi = 0.5$, $\alpha = p$); electronic supplementary material, figures S9–S12 show detailed trait evolution plots with 95% confidence intervals. In figure 3a, we show that, if signals are costless ($c = 0$), moderate minority groups ($p \geq 0.2$) adopt the signalling trait the fastest, followed by moderate majority groups ($0.6 \leq p \leq 0.7$). Costless signalling fixates quickly, followed by conditional trusting; yet, the latter spreads with a smaller delay in minority than in majority groups. These results are in line with the results from our analytical model (equation (2.2)): in groups with $\alpha > \alpha^*$, conditional trusting only pays off once a sufficiently high share of ingroup trustees signal their identity.

Figure 3b shows that, under intermediate signal costs ($c = 3$), most minority groups ($0.2 \leq p \leq 0.4$) develop signalling and conditional trusting traits. Majority groups do not adopt conditional trusting as the share of signalling trustees never reaches the threshold needed for conditional trusting to

be a viable strategy (equation (2.2)). Note that, unlike in the costless signalling scenario, the proliferation of signalling traits depends on (and lags behind) the presence of a sufficiently large share of the conditional trusting trait.

In figure 2, we showed that signalling norms do not fixate when mutation rates are low ($m_0 = 0.0001$). Inspecting the dynamics of these simulations suggests that signalling norms develop in some simulation runs, but not in others. Signalling takes off only once a sufficiently high share of trusters adopts signal recognition; but, depending on the dynamics of individual runs, this might not happen in every run (five out of twenty runs when $p = 0.3$, $\varphi = 0.5$ and $c = 3$, as shown in the electronic supplementary material, figure S14).⁹ The lower propensity of agents to randomly adopt signalling and conditional trusting makes it more difficult for social learning to boost the coevolution of the two traits in a manner that supports the wide adoption of the signalling norm.

Inspecting different combinations of p and φ that result in the same α value, we observe similar rates, but different dynamics of signalling norm emergence. Smaller minority groups with a higher ingroup assortment (e.g. $p = 0.3$ and $\varphi = 0.7$) take longer to develop the signalling norm and show more variability across runs compared to larger groups with lower assortment, and same α (e.g. $p = 0.4$ and $\varphi = 0.5$, $\alpha = 0.4$ for both; see the electronic supplementary material, figures S17 and S18).

4. Relaxing model assumptions

Here, we briefly discuss how relaxing some of the model assumptions affects the conditions under which signalling norms emerge. Full results are discussed at length in the electronic supplementary material, S2. We relax each of the assumptions individually, and not simultaneously. First, we relax the assumption that trustees can only send ingroup signals, allowing them to send outgroup signals in order to exploit the conditionally trusting outgroup. However, sending outgroup signals is not an equilibrium strategy for most groups, minority and majority alike. Only when a rather small minority group—in which benefits from parochial cooperation are low—faces a rather large majority group in which a signalling norm emerged will there be an equilibrium with outgroup signalling. In these cases, minority trustees send outgroup signals (while minority trusters unconditionally distrust) and majority trusters conditionally trust (thereby reaping the benefits of parochially cooperating with the signalling ingroup and, at times, bearing the losses of being exploited by the outgroup).

Second, we relax the assumption that trusters do not recognize outgroup signals, allowing trusters to distrust only those trustees that display outgroup signals. Signalling and signal recognition is still an equilibrium in several cases: (i) if both groups are in a signalling equilibrium; (ii) if one group is in a signalling equilibrium, whereas trustees in the other do not signal and trusters discriminate against outgroup signals. In this set-up, trustees are willing to bear lower costs of signalling compared to our main analysis. This is because signalling now includes a possibility of being distrusted by outgroup trusters—who could be abused without consequences in the main model. Our simulation results show that signalling norms still emerge in this set-up, but in a somewhat narrower set of minority groups (electronic supplementary material, figures S20 and S21).

Finally, we relax the assumption that trustees always perceive the trusters' group identity correctly upon being given trust. We explore the threshold of error (i.e. noise) in trustees' perception of truster identity below which signalling and signal recognition remain equilibrium strategies. Our analytical results suggest that very high levels of noise can hamper signalling norm emergence as relative group size decreases or signalling cost increases (electronic supplementary material, figure S22). The level of noise groups of different sizes can tolerate in a signalling equilibrium depends on the pay-off structure.

5. Discussion and conclusion

We combine insights from literature on signalling and parochial cooperation to understand how signalling norms evolve to facilitate parochial cooperation in groups with conflicting interests. Using a trust game allows us to model encounters where giving trust to an untrustworthy other exposes one to significant risks, while not giving trust can prevent one from successfully cooperating with the ingroup. Our formal analysis suggests that, while following a signalling norm can be an equilibrium for different groups, groups who face a lower share of potential ingroup cooperators are more likely to shift from a non-signalling to a signalling equilibrium. We use agent-based simulations to examine the emergence of traits related to signalling norms under different signalling costs, mutation rates, relative group shares, and extent of ingroup assortment.

Our agent-based simulations confirm that groups which constitute a minority in the population—and are, thus, less likely to encounter cooperative ingroup members—more easily develop (costly) signalling norms. Signalling norms facilitate cooperation with the ingroup while avoiding exploitation by the outgroup. However, when ingroup assortment is higher—for instance, because members of small groups live close to each other—signalling becomes less beneficial for promoting parochial cooperation [45]. We show that the sufficiently wide recognition of signals (i.e. conditional trusting) opens the door for costly signalling.

Broadly in line with the literature on tag-based cooperation [46], we find that costless signals which cannot be recognized by the outgroup emerge to facilitate parochial cooperation rather easily across the groups. Costless signals improve majority group benefits as they allow the majority to avoid exploitation by the minority. This is, however, at the expense of minority groups who lose the opportunity to exploit otherwise trusting outgroup members. These results showcase how, in set-ups where groups have conflicting interests, group benefits sensitively depend on the actions of the outgroup (also see [47]).

These findings remain robust even when relaxing several of our model assumptions. We assume that sending signals of group identity requires cultural knowledge not easily accessible to the outgroup, which makes these signals hard to 'fake' [22]. Yet, even if we allow groups to send outgroup signals, this strategy can only benefit minority groups that are rather small and are facing a majority outgroup that has already developed a signalling norm. Thereby, we specify the conditions under which outsiders could benefit from learning the (costly) 'secret handshakes' of a group and exploit parochial cooperators by defecting instead of cooperating

[48,49]. In our set-up, sufficiently large groups can sustain parochial cooperation in spite of such exploitation. Extensions of our model could further explore other scenarios, such as those where groups can adopt new signals and establish cooperation before the signal is 'hijacked' again, resulting in cycles of dominance of different reliable identity markers [19,20,22].

Our model considers signals that are so group specific that they remain undetectable to the outgroup [34,35]. If signals are unknown to the outgroup, group members can both reap the benefits of parochial cooperation with the ingroup and exploit the unsuspecting outgroup. Relaxing this assumption, we find that, if signals are easily recognized and the outgroup can discriminate against them, smaller minority groups become less likely to develop signalling norms as they cannot bear the losses from discrimination. Small minority groups will probably benefit from developing signals that are not as easily recognized by the outgroup. In addition, our simulations show that the emergence of signalling norms in most groups remain undisturbed even as we introduce significant amounts of noise to the perception of group identities during encounters. It is the smallest minority groups that are the most sensitive to increases in uncertainty about the counterpart's identity.

Existing work has suggested that incidental individual behaviours can develop into social norms prescribing signals of some qualities in the population [5]. Our results show that this can indeed be the case in groups that constitute a minority in the population as long as these incidental behaviours are frequent enough. Further, majority groups in the population cannot move from the non-signalling equilibrium. This is because they fail to reach the high threshold at which sending costly signals becomes beneficial. Developing a signalling norm in majority groups might, thus, require a stronger stochastic shock to introduce sufficiently wide signal recognition, for instance, via a centralized or persuasive intervention [9,50–52].

Research has suggested that selectively interacting with one's group allows parochial cooperation to emerge [45,53–56]. We, however, allow agents to condition cooperation on the partner's signal (or the absence thereof) in a randomly mixed population. This allows us to model how signals can emerge to support parochial cooperation even in adverse contexts where individuals have to face unknown others—as is often the case in large, complex societies where one cannot rely on knowledge from past interactions [41]. We further explore the interaction between relative group size and the likelihood of interacting with the ingroup. Our results highlight that isolated minority communities who infrequently encounter the outgroup benefit less from signals of group identity. At the same time, minority groups situated at group boundaries who encounter the outgroup more frequently obtain more benefit from being able to tell friends and foes apart. Even holding the probability of encountering the ingroup constant, smaller minority groups take longer to establish signalling norms compared to larger minority groups.

We analyse the conditions under which a norm prescribing the signalling of group identity can emerge to facilitate parochial cooperation. Combining our model with existing work that evaluates how groups coordinate on a specific signal among the multitude of potential candidates could help understand signalling norm emergence more generally

[5,57–61]. To gain a more realistic understanding of how large numbers of group members come to recognize a specific signal as a reliable marker of group identity, future work could allow agents to choose their partners based on displayed signals or consider the possibility that certain individuals can send signals at a lower cost, kick-starting their recognition in the population [33,57]. Moreover, in our model, individuals that do not follow the signalling norm of their group are (indirectly) sanctioned through the withdrawal of trust. Future research could investigate the effect of direct sanctioning and the variation in the strength of norm enforcement on the emergence of signalling norms [62].

Further, our model assumes that coherent groups with conflicting interests exist before signalling norms can emerge. This is the case when, for instance, geographical boundaries or kinship ties determine group formation and constrain resource sharing with other groups. Generalizing our insights even further, future work could consider modelling the coevolution of groups and signalling norms [63]. In this regard, studying shifts in the role of signalling norms due to changes in relative group sizes (i.e. when majorities become minorities and vice versa) seems a promising avenue for future research [64]. Finally, our model captures contexts where individuals successfully cooperate within groups, but face environmental restrictions that put their interests at odds with those of other groups [24,31,65]. Extending our model to situations where periods of intergroup conflict are interrupted by peaceful coexistence can help understand how, by supporting parochial cooperation, signalling norms that emerge in times of conflict could hamper intergroup cooperation during peaceful times [24,66].

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The code of the agent-based simulation is available from the OSF repository: <https://osf.io/nxafv/> [67].

Supplementary material is available online [68].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. A.M.: conceptualization, formal analysis, investigation, methodology, project administration, software, visualization, writing—original draft, writing—review and editing; M.T.: conceptualization, methodology, supervision, writing—review and editing; W.P.: conceptualization, formal analysis, project administration, supervision, writing—review and editing; V.B.: conceptualization, formal analysis, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. We received no funding for this study.

Acknowledgements. A.M. gratefully acknowledges financial support by the Swaantje Mondt travel fund.

Endnotes

¹Trust game with signalling resembles the hostage trust game [38–40].

²This could be because trustees acquire additional information about trusters after their signalling decision is made (e.g. the group from our example in the Introduction recognizes, by its appearance, the village they are approaching only after having decided to display a signal of their group identity) or because the truster reveals their group membership through the act of giving trust (e.g. by using a specific dialect when addressing the trustee).

³In the electronic supplementary material, table S2, we show α values resulting from different combinations of p and φ parameters tested in our models.

⁴Agents compare their weighted pay-offs from the last 20 rounds to the equivalent pay-offs of a randomly chosen ingroup agent who has played a comparable number of games; the more successful the model agent, the more likely the focal agent is to adopt their traits (see the electronic supplementary material, S3.1). We exclude learning from the outgroup since the two groups' interests are in conflict.

⁵In the electronic supplementary material, S4.9, we provide evidence of the robustness of our simulations to changes in several parameters.

⁶See the electronic supplementary material, S4.6 for results from simulations initiated with full signalling and signal recognition in

both groups. These results are in alignment with our theoretical expectations in figure 2a.

⁷Recall from theorem 2.1 that costs need to satisfy the condition $c \leq \alpha(R_I - P)$ to be affordable for trustees.

⁸For instance, for a group with $p = 0.8$, the threshold value as per equation (2.2) would be $\beta^* \geq 0.81$ (electronic supplementary material, table S4).

⁹Electronic supplementary material, figure S16 shows how, in a similar scenario with $c = 5$, a signalling norm emerges in only one out of 20 runs by time step 20 000.

References

- Gavrilets S, Tverskoi D, Sánchez A. 2024 Modelling social norms: an integration of the norm-utility approach with beliefs dynamics. *Phil. Trans. R. Soc. B* **379**, 20230027. (doi:10.1098/rstb.2023.0027)
- Mackie G. 1996 Ending footbinding and infibulation: a convention account. *Am. Sociol. Rev.* **61**, 999. (doi:10.2307/2096305)
- Gambetta D. 2009 *Codes of the underworld: how criminals communicate*. Princeton, NJ: Princeton University Press.
- Henrich J. 2009 The evolution of costly displays, cooperation and religion. *Evol. Hum. Behav.* **30**, 244–260. (doi:10.1016/j.evolhumbehav.2009.03.005)
- Posner EA. 1998 Symbols, signals, and social norms in politics and the law. *J. Leg. Stud.* **27**, 765–797. (doi:10.1086/468042)
- Przepiorka W, Diekmann A. 2021 Parochial cooperation and the emergence of signalling norms. *Phil. Trans. R. Soc. B* **376**, 20200294. (doi:10.1098/rstb.2020.0294)
- Smith EA, Bliege Bird R. 2005 Costly signaling and cooperative behavior. In *Moral sentiments and material interests: the foundations of cooperation in economic life* (eds H Gintis, S Bowles, R Boyd, E Fehr), pp. 115–148. Cambridge, MA: MIT Press.
- Barclay P, Bliege Bird R, Roberts G, Számadó S. 2021 Cooperating to show that you care: costly helping as an honest signal of fitness interdependence. *Phil. Trans. R. Soc. B* **376**, 20200292. (doi:10.1098/rstb.2020.0292)
- Gintis H, Smith EA, Bowles S. 2001 Costly signaling and cooperation. *J. Theor. Biol.* **213**, 103–119. (doi:10.1006/jtbi.2001.2406)
- Harms W, Skyrms B. 2009 Evolution of moral norms. In *The Oxford handbook of philosophy of biology* (ed. M Ruse), pp. 434–450. Oxford, UK: Oxford University Press.
- Smith EA, Bliege Bird RL. 2000 Turtle hunting and tombstone opening. *Evol. Hum. Behav.* **21**, 245–261. (doi:10.1016/S1090-5138(00)00031-3)
- Fehrler S, Przepiorka W. 2013 Charitable giving as a signal of trustworthiness: disentangling the signaling benefits of altruistic acts. *Evol. Hum. Behav.* **34**, 139–145. (doi:10.1016/j.evolhumbehav.2012.11.005)
- Przepiorka W, Berger J. 2017 Signaling theory evolving: signals and signs of trustworthiness in social exchange. In *Social dilemmas, institutions, and the evolution of cooperation* (eds B Jann, W Przepiorka). Berlin, Germany: De Gruyter.
- Choi J-K, Bowles S. 2007 The coevolution of parochial altruism and war. *Science* **318**, 636–640. (doi:10.1126/science.1144237)
- De Dreu CKW, Balliet D, Halevy N. 2014 Parochial cooperation in humans: forms and functions of self-sacrifice in intergroup conflict. In *Advances in motivation science*, pp. 1–47. Amsterdam, The Netherlands: Elsevier.
- Antal T, Ohtsuki H, Wakeley J, Taylor PD, Nowak MA. 2009 Evolution of cooperation by phenotypic similarity. *Proc. Natl Acad. Sci. USA* **106**, 8597–8600. (doi:10.1073/pnas.0902528106)
- Hammond RA, Axelrod R. 2006 The evolution of ethnocentrism. *J. Confl. Resolut.* **50**, 926–936. (doi:10.1177/0022002706293470)
- Hartshorn M, Kaznatcheev A, Shultz T. 2013 The evolutionary dominance of ethnocentric cooperation. *J. Artif. Soc. Soc. Simul.* **16**, 7. (doi:10.18564/jasss.2176)
- Jansen VAA, van Baalen M. 2006 Altruism through beard chromodynamics. *Nature* **440**, 663–666. (doi:10.1038/nature04387)
- Riolo RL, Cohen MD, Axelrod R. 2001 Evolution of cooperation without reciprocity. *Nature* **414**, 441–443. (doi:10.1038/35106555)
- Blanton RE. 2015 Theories of ethnicity and the dynamics of ethnic change in multiethnic societies. *Proc. Natl Acad. Sci. USA* **112**, 9176–9181. (doi:10.1073/pnas.1421406112)
- Cohen E. 2012 The evolution of tag-based cooperation in humans: the case for accent. *Curr. Anthropol.* **53**, 588–616. (doi:10.1086/667654)
- Sosis R, Kress H, Boster J. 2007 Scars for war: evaluating alternative signaling explanations for cross-cultural variance in ritual costs. *Evol. Hum. Behav.* **28**, 234–247. (doi:10.1016/j.evolhumbehav.2007.02.007)
- De Dreu CKW, Gross J, Fariña A, Ma Y. 2020 Group cooperation, carrying-capacity stress, and intergroup conflict. *Trends Cogn. Sci.* **24**, 760–776. (doi:10.1016/j.tics.2020.06.005)
- Moffett MW. 2013 Human identity and the evolution of societies. *Hum. Nat.* **24**, 219–267. (doi:10.1007/s12110-013-9170-3)
- Castro L, Toro MA. 2007 Mutual benefit cooperation and ethnic cultural diversity. *Theor. Popul. Biol.* **71**, 392–399. (doi:10.1016/j.tpb.2006.10.003)
- McElreath R, Boyd R, Richerson PJ. 2003 Shared norms and the evolution of ethnic markers. *Curr. Anthropol.* **44**, 122–130. (doi:10.1086/345689)
- Demello M. 1993 The convict body: tattooing among male American prisoners. *Anthropol. Today* **9**, 10. (doi:10.2307/2783218)
- Doğan G, Glowacki L, Rusch H. 2022 Are strangers just enemies you have not yet met? Group homogeneity, not intergroup relations, shapes ingroup bias in three natural groups. *Phil. Trans. R. Soc. B* **377**, 20210419. (doi:10.1098/rstb.2021.0419)
- Mann M. 1993 *The sources of social power*, 1st edn. Cambridge, UK: Cambridge University Press.
- Kroneberg C, Wimmer A. 2012 Struggling over the boundaries of belonging: a formal model of nation building, ethnic closure, and populism. *Am. J. Sociol.* **118**, 176–230. (doi:10.1086/666671)
- Cohen E, Haun D. 2013 The development of tag-based cooperation via a socially acquired trait. *Evol. Hum. Behav.* **34**, 230–235. (doi:10.1016/j.evolhumbehav.2013.02.001)
- Dumas M, Barker JL, Power EA. 2021 When does reputation lie? Dynamic feedbacks between costly signals, social capital and social prominence. *Phil. Trans. R. Soc. B* **376**, 20200298. (doi:10.1098/rstb.2020.0298)
- Smaldino PE, Flamsom TJ, McElreath R. 2018 The evolution of covert signaling. *Sci. Rep.* **8**, 4905. (doi:10.1038/s41598-018-22926-1)
- van der Does T, Galesic M, Dunivin ZO, Smaldino PE. 2022 Strategic identity signaling in heterogeneous networks. *Proc. Natl Acad. Sci. USA* **119**, e2117898119. (doi:10.1073/pnas.2117898119)
- Bugarski R. 2012 Language, identity and borders in the former Serbo-Croatian area. *J. Multiling. Multicult. Dev.* **33**, 219–235. (doi:10.1080/01434632.2012.663376)
- Dasgupta P. 1988 Trust as a commodity. In *Trust: making and breaking cooperative relations* (ed. D Gambetta), pp. 49–72. Oxford, UK: Blackwell Publishing.
- Raub W. 2004 Hostage posting as a mechanism of trust: binding, compensation, and signaling. *Ration. Soc.* **16**, 319–365. (doi:10.1177/1043463104044682)
- Raub W, Keren G. 1993 Hostages as a commitment device. A game-theoretic model and an empirical

- test of some scenarios. *J. Econ. Behav. Organ.* **21**, 43–67. (doi:10.1016/0167-2681(93)90039-R)
40. Raub W, Weesie J. 2000 Cooperation via hostages. *Anal. Krit.* **22**, 19–43. (doi:10.1515/auk-2000-0102)
 41. Smaldino PE. 2019 Social identity and cooperation in cultural evolution. *Behav. Processes* **161**, 108–116. (doi:10.1016/j.beproc.2017.11.015)
 42. Bliege BR, Smith EA. 2005 Signaling theory, strategic interaction, and symbolic capital. *Curr. Anthropol.* **46**, 221–248. (doi:10.1086/427115)
 43. Masuda N, Ohtsuki H. 2007 Tag-based indirect reciprocity by incomplete social information. *Proc. R. Soc. B* **274**, 689–695. (doi:10.1098/rspb.2006.3759)
 44. Rendell L *et al.* 2010 Why copy others? Insights from the social learning strategies tournament. *Science* **328**, 208–213. (doi:10.1126/science.1184719)
 45. García J, van den Bergh CJM. 2011 Evolution of parochial altruism by multilevel selection. *Evol. Hum. Behav.* **32**, 277–287. (doi:10.1016/j.evolhumbehav.2010.07.007)
 46. García J, van Veelen M, Traulsen A. 2014 Evil green beards: tag recognition can also be used to withhold cooperation in structured populations. *J. Theor. Biol.* **360**, 181–186. (doi:10.1016/j.jtbi.2014.07.002)
 47. Helbing D, Johansson A. 2010 Cooperation, norms, and revolutions: a unified game-theoretical approach. *PLoS ONE* **5**, e12530. (doi:10.1371/journal.pone.0012530)
 48. Miller JH, Butts CT, Rode D. 2002 Communication and cooperation. *J. Econ. Behav. Organ.* **47**, 179–195. (doi:10.1016/S0167-2681(01)00159-7)
 49. Nettle D. 1997 Social markers and the evolution of reciprocal exchange. *Curr. Anthropol.* **38**, 93–99. (doi:10.1086/204588)
 50. Gavrillets S. 2020 The dynamics of injunctive social norms. *Evol. Hum. Sci.* **2**, e60. (doi:10.1017/ehs.2020.58)
 51. Gavrillets S, Richerson PJ. 2022 Authority matters: propaganda and the coevolution of behaviour and attitudes. *Evol. Hum. Sci.* **4**, e51. (doi:10.1017/ehs.2022.48)
 52. Efferson C, Ehret S, von Flüe L, Vogt S. 2024 When norm change hurts. *Phil. Trans. R. Soc. B* **379**, 20230039. (doi:10.1098/rstb.2023.0039)
 53. Fletcher JA, Doebeli M. 2009 A simple and general explanation for the evolution of altruism. *Proc. R. Soc. B* **276**, 13–19. (doi:10.1098/rspb.2008.0829)
 54. Kim J-W, Hanneman RA. 2014 Coevolutionary dynamics of cultural markers, parochial cooperation, and networks. *J. Confl. Resolut.* **58**, 226–253. (doi:10.1177/0022002712468691)
 55. Riolo RL. 1997 The effects of tag-mediated selection of partners in evolving populations playing the iterated Prisoner's Dilemma. SFI Working Paper no. 97-02-016. Santa Fe NM: Santa Fe Institute.
 56. Takesue H. 2020 From defection to ingroup favoritism to cooperation: simulation analysis of the social dilemma in dynamic networks. *J. Comput. Soc. Sci.* **3**, 189–207. (doi:10.1007/s42001-019-00058-4)
 57. Barker JL, Power EA, Heap S, Puurtinen M, Sosis R. 2019 Content, cost, and context: a framework for understanding human signaling systems. *Evol. Anthropol. Issues News Rev.* **28**, 86–99. (doi:10.1002/evan.21768)
 58. Bell AV. 2020 A measure of social coordination and group signaling in the wild. *Evol. Hum. Sci.* **2**, e34. (doi:10.1017/ehs.2020.24)
 59. Centola D, Baronchelli A. 2015 The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl Acad. Sci. USA* **112**, 1989–1994. (doi:10.1073/pnas.1418838112)
 60. Przepiorka W, Szekely A, Andrighetto G, Diekmann A, Tummlini L. 2022 How norms emerge from conventions (and change). *Socius Sociol. Res. Dyn. World* **8**, 237802312211245. (doi:10.1177/23780231221124556)
 61. Schelling TC. 1980 *The strategy of conflict*. Cambridge, MA: Harvard University Press.
 62. Molho C, De Petrillo F, Garfield Z, Siewe S. 2024 Cross-societal variation in norm enforcement systems: a review. *Phil. Trans. R. Soc. B* **379**, 20230034. (doi:10.1098/rstb.2023.0034)
 63. Efferson C, Lalive R, Fehr E. 2008 The coevolution of cultural groups and ingroup favoritism. *Science* **321**, 1844–1849. (doi:10.1126/science.1155805)
 64. Alvarez-Benjumea A, Winter F, Zhang N. 2024 Norms of prejudice: political identity and polarization. *Phil. Trans. R. Soc. B* **379**, 20230030. (doi:10.1098/rstb.2023.0030)
 65. Rodrigues AMM, Barker JL, Robinson EJH. 2022 From inter-group conflict to inter-group cooperation: insights from social insects. *Phil. Trans. R. Soc. B* **377**, 20210466. (doi:10.1098/rstb.2021.0466)
 66. He Q-Q, Yu J, Tang S-H, Wang M-Y, Wu J, Chen Y, Tao Y, Ji T, Mace R. 2024 Jeans and language: kin networks and reproductive success are associated with the adoption of outgroup norms. *Phil. Trans. R. Soc. B* **379**, 20230031. (doi:10.1098/rstb.2023.0031)
 67. Macanovic A, Tsvetkova M, Przepiorka W, Buskens V. 2023 Code from: Signals of belonging: emergence of signalling norms as facilitators of trust and parochial cooperation. OSF repository. (<https://osf.io/nxafxv/>)
 68. Macanovic A, Tsvetkova M, Przepiorka W, Buskens V. 2024 Signals of belonging: emergence of signalling norms as facilitators of trust and parochial cooperation. Figshare. (doi:10.6084/m9.figshare.c.6980681)