Parochial cooperation and the emergence of signalling norms

Wojtek Przepiorka[1] and Andreas Diekmann[2,3]

[1]Department of Sociology/ICS, Utrecht University, The Netherlands
[2]Department of Sociology, University of Leipzig, Germany
[3]Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland

WP, 0000-0001-9432-8696

Why do people adorn themselves with elaborate body piercings or tattoos, wear obstructing garbs, engage in life-threatening competitions and other wasteful and harmful but socially stipulated practices? Norms of cooperation and coordination, which promote the efficient attainment of collective benefits, can be explained by theories of collective action. However, social norms prescribing wasteful and harmful behaviours have eluded such explanations. We argue that signalling theory constitutes the basis for the understanding of the emergence of such norms, which we call signalling norms. Signalling norms emerge as a result of the uncertainty about who is a friend and who is a foe. The need to overcome this uncertainty arises when different groups compete for scarce resources and individuals must be able to identify, trust and cooperate with their fellow group members. After reviewing the mechanisms that explain the emergence of cooperation and coordination norms, we introduce the notion of signalling norms as markers of group distinction. We argue that adherence to signalling norms constitutes a commitment promoting parochial cooperation rather than a quality-revealing signal facilitating partner choice. We formalize our argument in a game-theoretic model that allows us to specify the boundary conditions for the emergence of signalling norms. Our paper concludes with a discussion of potential applications of our model and a comparison of signalling norms with related concepts.

This article is part of the theme issue 'The language of cooperation: reputation and honest signalling'.

1. Introduction

Why do people incur costs in terms of time, effort and other resources to create benefits for non-kin others, even in the absence of formal institutions that incentivize such cooperation? Next to psychological mechanisms that explain cooperation by means of, for example, other-regarding preferences [1–3], there is a number of social mechanisms that explain cooperation while maintaining the assumption of only self-regarding individuals. It has been argued and shown that people's embeddedness in ongoing social relations can promote cooperation because of expectations of direct and indirect reciprocity [4–6]. That is, cooperation emerges when people can expect their deeds to be rewarded and their misdeeds to be punished by either the same interaction partners in repeated encounters [7–9] or by new interaction partners that learn these people's reputations through third-parties [10–14]. However, these mechanisms are also in place to enforce collectively inefficient (i.e. wasteful) and even harmful behaviours [15–23].

Footbinding, female genital mutilation, extensive body tattoos, lethal initiation rituals, excessive feasts, obstructive dressing codes and other behaviours that do not seem to serve collective nor individual interests have been enforced through selective incentives. It has been argued that such behaviours exist owing to cultural inertia and disappear only with a time-lag in response to changing environmental and social conditions [24,25]. This argument suggests,

however, that these behaviours were, at some point, functional in serving the collective. Another argument suggests that wasteful and harmful behaviours are part of a repertoire that is invoked by individuals to set group boundaries and facilitate cooperation within groups (i.e. parochial cooperation) [26–30]. It has been shown that arbitrary markers can evolve to group individuals with similar but unobservable traits that facilitate cooperation between individuals within these groups [31–33]. Yet, it remains a puzzle why the distinguishing markers of many social groups can only be acquired through wasteful and harmful behaviours, which bind workforce, mobility, time and other resources that could be used to create collective benefits.

We argue that signalling theory [34–38] constitutes the basis for the understanding of the emergence of wasteful and harmful behaviours as markers of group distinction, which we call *signalling norms*. We first review how the emergence of cooperation and coordination norms has been construed in the social and behavioural sciences. We then introduce the notion of signalling norms and argue that a demand for signalling norms arises because of information asymmetries pertaining to individuals' commitments to a social group [39]. That is, while cooperation and coordination norms emerge out of a demand for reducing negative externalities of individuals' behaviours, signalling norms emerge out of a demand for reducing uncertainty about who is a friend and who is a foe.

We devise a game-theoretic model which allows identifying the conditions for the emergence and evolution of signalling norms. In particular, we point out that it is the commitment rather than the signalling aspect of the adherence to signalling norms that reduces uncertainty about the cooperative intent of members of a social group. In other words, rather than an unobserved individual trait that can be inferred from norm-abiding behaviour, norm-abiding behaviour impedes cooperative relations with members of other groups and thus increases trust and cooperation within social groups.

We conclude our paper with a discussion of potential applications of our model and a comparison of signalling norms with two related concepts—collective effervescence and credibility-enhancing displays (CREDs).

2. Cooperation and coordination norms

Social norms can be defined as rules guiding individual behaviours in social interactions that are sustained by shared expectations and sanctions [40–42]. Social norms help to overcome social dilemmas that arise when people, by purposefully following their beliefs and preferences, in sum produce outcomes that make them worse off than intended. Norms, such as quiet zones in train compartments, dedicated smoking areas at train stations and airports, signs indicating appropriate walking behaviour on escalators, laws proscribing the use of doping in sports, drinking of alcohol below a certain age or the production of excessive noise late at night [40,41,43], all serve to reduce the negative externalities that individual actions can impose on others. Such externalities are noise, cigarette smoke, traffic jams, excessive competition, increasing healthcare costs and so on.[1] Social dilemmas are thus situations in which a gap between individual and collective rationality exists [44–47].

We distinguish between cooperation and coordination dilemmas [48,49]. Cooperation dilemmas refer to situations in which individuals have conflicting interests and, through the best pursuit of their own interests, reach a worse state than they could have reached by following an alternative plan of action. Coordination dilemmas refer to situations in which individuals' interests are (partly) aligned, but the multiple ways of best pursuing these interests make it difficult for these individuals to tacitly coordinate on reaching a better state. Accordingly, a distinction is made between cooperation norms, which help overcome cooperation dilemmas, and coordination norms, which help to overcome coordination dilemmas [48]. In the literature, cooperation norms are also called injunctive norms and coordination norms are also called conventions [50–53]. However, the term convention implies a lack of normative force guiding individual behaviours, which is often unjustified in the light of existing sanctioning mechanisms that help to overcome coordination dilemmas [54,55]. Coordination norms can thus also be injunctive norms. Moreover, the emergence of both cooperation and coordination norms can be explained by means of the same theoretical framework (figure 1a).

According to Coleman [40], social norms are collective-level (i.e. macro-level) constructs that emerge when individual behaviours create negative externalities (at the micro-level) to an extent that produces a demand for regulating these behaviours. A particular norm can emerge through bargaining, communication, collective decision making or other mechanisms that we will not discuss further here [41,56]. Once in place, these norms are maintained by positive and negative sanctions (i.e. selective incentives).

At this point, it is important to make explicit that social norms are directed at focal actions performed by target individuals and benefit the norm beneficiaries [40]. For example, a non-smoking norm is directed at the focal action smoking performed by smokers (the targets of the norm) and benefit the non-smokers (the norm beneficiaries). Adherence to the norm benefits the norm beneficiaries directly by reducing the negative externalities created by the targets. However, the targets and beneficiaries of a social norm do not have to be mutually exclusive groups. Even in the previous example, one could argue that smokers benefit from non-smoking norms as well. However, the emergence of other types of norms such as dress codes, table manners, jargon and tastes, which are often considered as norms of group distinction [57–59], are more difficult to explain with Coleman's argument [55].

One could argue that because the focal actions targeted by norms of group distinction are conventional (e.g. veiling as a sign of a woman's religiosity and belonging to the Muslim community), these norms must be construed as coordination norms. In this case, however, it would remain difficult to pinpoint what the negative externalities are that such norms of group distinction serve to reduce. What kind of coordination dilemmas (or cooperation dilemmas) do these norms help to solve? Rather than improving collective outcomes, many of these norms promote collectively wasteful and individually harmful behaviours (e.g. religious garbs that impede mobility and vision).

Although we are not the first to stumble over this puzzle [15–23], the previous scholarship provides plausible explanations for the maintenance of inefficient norms but is less explicit about their emergence. For example, it has been
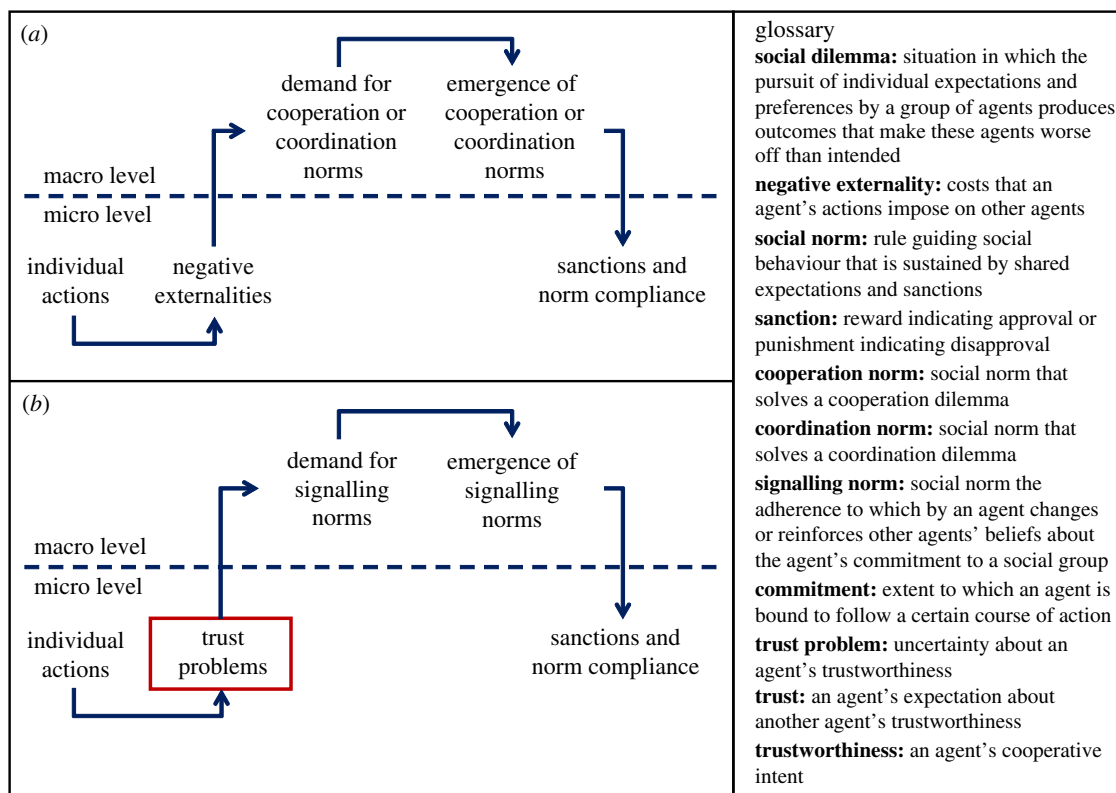
Figure 1. The emergence of social norms. (Online version in colour.)

argued that inefficient norms are adhered to because adherence is positively sanctioned by others, and since people value social relationships and approval, they adhere to these norms and enforce them [19,42]. At the same time, norms of group distinction, by definition, establish group boundaries that prevent the formation of social relations between members of different groups. For example, the demeanour of someone that wears a religious dress covering their entire body (including their head) to indicate their group affiliation is more likely to keep people at bay than establish new relations with members of other groups.

In what follows we propose that introducing a new type of social norms, signalling norms, can help to explain the emergence of wasteful and harmful markers of group distinction and bridge seemingly incompatible strands of literature on social norms [40,56].

3. Signalling norms

We define signalling norms as social norms, the adherence to which by person A changes or reinforces person B's beliefs about A's commitment to a social group. The members of this social group are the beneficiaries of the signalling norm. However, unlike for cooperation and coordination norms, the actions that signalling norms are directed at do not necessarily benefit the beneficiaries directly and may even harm other parties or the target of the norm (e.g. signalling norms prescribing illegal behaviour) [60]. A signalling norm benefits its beneficiaries indirectly by the information adherence to the signalling norm produces. Adherence to a signalling norm produces information about A's commitment to the group of beneficiaries. In other words, signalling norms are social norms the adherence to which is a signal of trustworthiness (i.e. cooperative intent) observed

by the beneficiaries of the signalling norm (see also [61–63]). The emergence of signalling norms can thus be explained by the prevalence of uncertainty about and the importance of knowing who is a friend and who is a foe.

Posner [39,64] was the first to suggest that people's adherence to social norms can be conceived as signals of these people's trustworthiness. Posner's main idea is that social norms are the equilibrium outcomes of signalling games [65,66]. Such equilibria arise mainly owing to a demand for the distinction between long-term and short-term types. These types are defined in terms of their discount factors, i.e. their propensity to engage in repeated interactions. Long-term types are more likely to be available for future interactions than short-term types. Because in repeated interactions cooperation can become a self-regarding best response [4,7–9], long-term types are more trustworthy [67] and thus more desirable interaction partners than short-term types. However, because individuals' discount factors are not directly observable, there is uncertainty about a potential partner's type. In this case, signalling behaviour (e.g. adherence to a social norm) can help to overcome the information gap allowing for repeated, cooperative interactions between long-term types. In order for types to be distinguishable, it must hold that only the long-term types can afford to engage in signalling behaviour.

Posner [39] applies his theoretical ideas to explain different, seemingly irrational, social behaviours like gift giving, voting or discrimination. However, Posner conceives of any norm-abiding behaviour as a potential signal that helps to distinguish long-term and short-term types [68]. By introducing the notion of signalling norms, we aim to specify the conditions under which the emergence and maintenance of social norms can be explained by signalling theory [34–38].

Like for cooperation and coordination norms, a necessary condition for the emergence of signalling norms is a demand

for such norms (figure 1*b*). However, the demand for signalling norms does not arise from the need to proscribe actions that produce negative externalities. The demand for signalling norms arises from trust problems and people's need to belong to a group of similar and like-minded others [69].

In this sense, signalling norms solve a particular cooperation dilemma, namely the trust dilemma [70], and not by the focal actions they prescribe, but by the information these focal actions produce. Like cooperation and coordination norms, once in place, signalling norms are enforced by sanctions. Because individuals adhering to signalling norms are perceived as more trustworthy by the norm beneficiaries (i.e. members of a particular group), these individuals are positively sanctioned by the greater trust that is placed in them. Correspondingly, individuals that do not adhere to signalling norms are negatively sanctioned by not being trusted. The following example will further illustrate the idea.

Muslim women veil as a sign of their religiosity and belonging to the Muslim community. To pass as a Muslim woman requires much more than wearing the right garb, but the sum of signalling norms Muslim women adhere to makes compliance with these norms a very strong signal of being a true member of the Muslim community. Against the often corroborated predictions of modernization and secularization theory, recent evidence suggests that highly religious Muslim women veil more in response to modernizing forces such as access to education, urban living and contacts with non-Muslims [71]. One explanation that Aksoy and Gambetta put forward for their finding is that these women veil more to signal their religiosity and commitment to the Muslim community because the surrounding temptations jeopardize their reputations for modesty (see also [72]).

This example illustrates the signalling and commitment aspects of the adherence to signalling norms (see also [39 p. 29], [73]). The signalling aspect captures the possibility that by adhering to a signalling norm, people convey information about their otherwise unobservable traits (e.g. beliefs, values, preferences and interdependencies) [34,35,62,63]. The commitment aspect captures the possibility that by adhering to a signalling norm, people constrain the number of alternative interaction partners or make themselves vulnerable to punishment by third-parties [74–76]. However, conceptualizing markers of group distinction as signalling norms makes it difficult to explain the emergence of these norms as signals of people's unobservable traits. The main objection is that sanctions enforcing the adherence to signalling norms could incentivise people to adhere to these norms irrespective of their underlying traits [77]. This is all the more true given the behaviours targeted by signalling norms (e.g. veiling) are mostly conventional and therefore unlikely to be causally related to the unobservable trait of interest (e.g. religiosity, trustworthiness). It is thus the commitment rather than the signalling aspect of the adherence to signalling norms that reduces uncertainty about the cooperative intent of (candidate) members of a social group.[2] Yet, the cost-benefit differential of adhering to a signalling norm must be large enough for the beneficiaries of the norm to decide whether to establish and maintain or deny (new) relationships with the targets of the norm.

In line with the commitment aspect, religious practices such as veiling foment discrimination from outgroup members, which makes ingroup members less competitive outside of their group. In other words, by subjecting themselves to discrimination, ingroup members reduce opportunities for cooperation outside of their group and thereby increase the costs of leaving their group [82,83]. Relatedly, Posner [39] emphasizes the importance of group membership for the evolution of cooperation. He argues that agents signal their loyalty (i.e. long-term cooperative intent) with their ingroup by discriminating against members of the outgroup (see also [84]). Again, such signals of group loyalty are reliable as by discriminating against members of the outgroup, agents forgo potential benefits from cooperation with them.[3]

One important implication of adherence to signalling norms being about commitment rather than signalling is that the type-separating behaviour *per se* does not have to be costly (which does not preclude it from being collectively wasteful and harmful for others) [86]. By adhering to signalling norms as markers of group distinction, individuals 'burn bridges', that is, they make themselves unavailable for cooperative interactions with members of an outgroup. This lack of outside opportunities, in turn, makes individuals adhering to signalling norms valid and trustworthy interaction partners for members of the ingroup (i.e. the beneficiaries of the signalling norm). What makes adherence to signalling norms type separating is the individual wants to earn benefits from repeated interactions with members of a particular (in)group.[4] The next section devises a game-theoretic model that captures the main tenets of signalling norms.

4. Game-theoretic model

Posner's [39,64] idea that social norms are separating equilibrium outcomes of signalling games that help to distinguish long-term types from short-term types has been formalized in a game-theoretic model by Przepiorka & Diekmann [87] (also see [88–90]). The following is an edited and extended version of the main part of the model section in Przepiorka & Diekmann [87]. A more detailed game-theoretic analysis is provided in the electronic supplementary material published along with the paper by Przepiorka & Diekmann [87].

The binary trust game (figure 2) represents a cooperation dilemma that cannot be overcome by rational and self-regarding players. While it is rational for the truster not to place trust, both the truster and the trustee would be better off if the trust was placed and honoured. In this sense the trust game is similar to the Prisoner's Dilemma game, although the two games differ in three respects: (i) the trust game is sequential (player 1 moves first and player 2 observes player 1's action) whereas the Prisoner's Dilemma game is simultaneous (both players decide without knowing what the other player does); (ii) by defecting (i.e. not placing trust) in the trust game, player 1 precludes player 2 from making a decision; from (i) and (ii) it follows that (iii) in the trust game only player 2 can exploit a cooperative player 1 whereas in the Prisoner's Dilemma game player 1 can also exploit a cooperative player 2. In both games, the Nash equilibrium is reached through mutual defection.

However, the assumption that trustees always abuse trust is neither realistic nor useful. If trusters were certain about the trustworthiness of trustees, the notion of trust would be superfluous. The trust problem arises from trusters' uncertainty about trustees' preferences and constraints, which determine these trustees' decisions in an interaction. In the model presented here, we assume two types of trustees that differ in their discount factors and can be characterized as
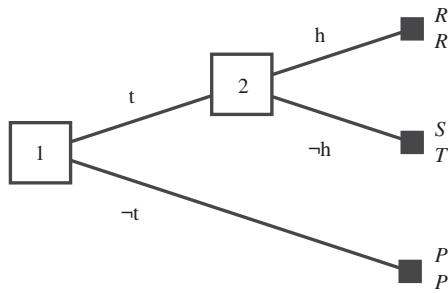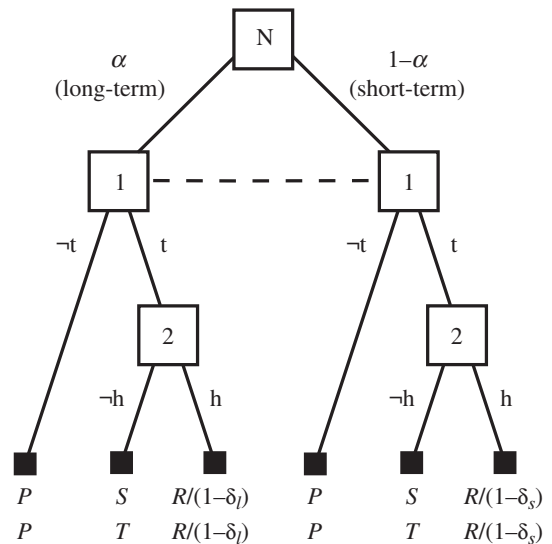
Figure 2. In the binary trust game, the truster (player 1) first decides whether to place trust (t) or not to place trust (¬t). If the truster decides not to place trust, the interaction is terminated and both parties receive payoff P. If, instead, the truster decides to trust, it is the trustee's turn (player 2) to choose whether or not to honour that trust (h or ¬h). If the trustee honours the trust, both players receive payoff R. If the trustee does not honour the trust, the trustee receives payoff T while the truster receives S. The payoffs are ordered so that the trustee abuses trust if the truster places it (i.e. $T > R > P$) and the truster prefers not to place trust rather than find his or her trust abused (i.e. $R > P > S$). The letters T, R, P and S stand for temptation, reward, punishment and sucker's payoff, respectively, and are commonly used to denote payoffs in the Prisoner's Dilemma game [91].

Figure 3. In the trust game with incomplete information, nature (N) moves first and determines the trustee's type to be long-term or short-term with probability α or $(1 - \alpha)$, respectively. The probability α (as the entire game) is common knowledge and the fact that the truster does not know whether the trustee is long-term or short-term is denoted by the dashed line. If the truster places trust, the long-term trustee honours trust while a short-term trustee does not. In the first case, the truster receives payoff $R/(1 - \delta_l)$; in the second case, the truster's payoff is S. The truster prefers placing trust if the trustee is a long-term type over not placing trust at all and is most reluctant to trust a short-term trustee (i.e. $R/(1 - \delta_l) > P > S$).

either long-term or short-term [39,64]. A discount factor stands for a trustee's probability of engaging in another interaction with the truster, but it can also be interpreted as the trustee's time preference. Here we adhere to the former interpretation as it better conveys the commitment aspect of the adherence to signalling norms (see the previous section).

Given a long-term trustee's discount factor (δ_l), he or she strictly prefers to engage in repeated interactions with a truster over a one-time abuse of trust (i.e. $R/(1 - \delta_l) > T > P$).[5] Given a short-term trustee's discount factor (δ_s), the expected payoff from repeated interactions with a truster is strictly smaller than his or her payoff from a one-time abuse of trust (i.e. $T > R/(1 - \delta_s) > P$). Hence, the long-term trustee and the short-term trustee will differ in what they do in the trust game, as long as their discount factors differ such that

$$\delta_l > \frac{T - R}{T} > \delta_s. \tag{4.1}$$

Note that an implicit assumption of our model is that an interaction between a truster and a trustee ends if the truster does not place trust or the trustee abuses placed trust. This assumption implies that only the long-term trustee will be deterred from abusing a truster's trust as he or she would otherwise forgo the higher future benefits from repeated cooperative interactions with the truster. The short-term trustee's potential future benefits from repeated interactions with the truster are too small for him or her to resist the temptation of abusing the truster's trust right away. Of course, failing to be trusted by a truster (a second time) does not prevent a short-term trustee to try to gain and abuse the trust of another truster elsewhere. However, from the perspective of the trusters that are members of a particular group, a short-term trustee ceases to be part of the equation after one encounter. This is plausible, for example, if one considers that trustees that were not trusted or that abused trust are banned from any further interactions with the trusters that are members of a particular group ([84,93,94]; also see endnote 6 on this assumption). However, these trusters remain

uncertain about the trustworthiness of every new trustee that attempts to interact with one of them.

A truster's uncertainty about a trustee's type can be accounted for in the trust game with incomplete information (figure 3) [70,73,95–97]. Given the probability α to meet a long-term trustee and the payoff structure, a truster only trusts if the expected payoff from trusting is higher than the payoff from not doing so. That is, if

$$\alpha R \frac{1}{1 - \delta_l} + (1 - \alpha)S > P. \tag{4.2}$$

After solving equation (4.2) for α, it can be shown that a truster will abstain from placing trust if α is less than the threshold value α^*, where:

$$\alpha^* = \frac{(P - S)(1 - \delta_l)}{R - S(1 - \delta_l)}. \tag{4.3}$$

Under these conditions (i.e. $\alpha < \alpha^*$), the truster and the long-term trustee could attain a more beneficial outcome if the trustee were able to communicate his or her type credibly.

The model can be extended so that the trustee can initially choose whether or not to send a signal at cost c (figure 4). In order for the truster to interpret the trustee's type, the signal must be type separating. Then, a separating equilibrium can emerge in which the long-term trustee sends a signal, a short-term trustee does not (i.e. type-separating behaviour), and the truster places trust only if a signal has been sent by the trustee. The signal is type separating if the long-term trustee can afford to send it while the short-term trustee cannot.[6] That is, if

$$R \frac{1}{1 - \delta_l} - P > c > T - P. \tag{4.4}$$

In our case, sending a signal means adhering to a signalling norm. Note that equation (4.4) implies that the costs of adhering to a signalling norm do not have to differ across trustee types for adherence to be type separating; what matters is the cost-benefit differential [34,98]. That is, what makes adherence to a signalling norm type separating is not the signalling costs alone but the long-term trustee's benefit from interacting with the truster repeatedly, which the short-term trustee does not obtain [86]. In other words, adherence to signalling norms is rewarded (i.e. positively sanctioned) through cooperation by members of the ingroup. However, under what conditions are signalling norms likely to emerge?

If $\alpha < \alpha^*$, the separating equilibrium is collectively more beneficial than the equilibrium without a signalling opportunity. That is, if signalling is not possible, trusters will abstain from placing trust and all will receive payoff P. In the case with signalling opportunity, however, trusters' expected payoff is $\alpha R/(1-\delta_l) + (1-\alpha)P$ and trustees receive $\alpha[R/(1-\delta_l) - c] + (1-\alpha)P$. If $\alpha > \alpha^*$, a separating equilibrium can also emerge but does not improve collective gains (i.e. the sum of trusters' and trustees' expected payoffs). Without a signalling opportunity, trusters will always place trust and their expected payoff is $\alpha R/(1-\delta_l) + (1-\alpha)S$, with trustees receiving $\alpha[R/(1-\delta_l)] + (1-\alpha)T$. It can be shown that these collective gains are always larger than the collective gains in the separating equilibrium, if $c > T - P$. In other words, if $\alpha > \alpha^*$, a pooling equilibrium, where both trustee types do not send signals (i.e. type-pooling behaviour) and trusters always place trust, is collectively more beneficial.

Our model thus implies that signalling norms are more likely to emerge if $\alpha < \alpha^*$, than if $\alpha > \alpha^*$. Put differently, if the group of norm beneficiaries (i.e. the trusters) expects the proportion of trustworthy individuals in the target population (i.e. the trustees) to be relatively small, they will be more likely to introduce wasteful markers of group distinction and demand their fellow group members to adopt these markers. However, note that the members of the target population do not have to be untrustworthy in general for a demand for signalling norms to emerge among the group of norm beneficiaries; it is the group of norm beneficiaries that defines the trust problem by determining the benefits that accrue through repeated interactions of their members among each other and the costs that accrue through these members' interactions with outgroup members. How norm beneficiaries define the trust problem that shapes their demand for signalling norms will further depend on their competition with other groups seeking access to scarce resources. Finally, both access to and competition for scarce resources will depend on competing groups' relative power, external shocks such as natural and human-made disasters and technological innovations (see also [99]). Signalling norms must, therefore, be seen as a side-product of group formation processes that take place in a wider social context that shapes cooperation and conflict between groups ([30,84,100,101 pp. 277–288]).

5. Discussion

In her seminal book, Ullmann-Margalit [48] distinguishes between three types of social norms. These are norms of cooperation, norms of coordination and norms of partiality. While norms of partiality concern the distribution of resources, cooperation and coordination norms emerge spontaneously or

by design and may sustain cooperation in social dilemma situations [40]. However, there are also social norms that do not fit into these three categories [56]. Examples are norms of etiquette, dress codes, table behaviour and other kinds of seemingly wasteful and even self-harming behaviours. We call these norms 'signalling norms' to convey the idea that adherence to these norms signals a person's commitment to a particular group. The emergence of signalling norms can be explained by models stemming from signalling theory developed in biology and economics [36–38].

By combining insights from signalling theory with rational choice theories of the emergence of social norms from sociology, we highlight how wasteful but socially stipulated behaviour can promote the evolution of parochial cooperation in humans. Signalling norms emerge when a group of trusters is uncertain about the trustworthiness of others they want to interact with in their common interest. In accordance with Posner's idea [39,64], we define trustworthy types by their long-term interests in repeated interactions while untrustworthy types have short-term interests only. Thus, long-term types are expected to reciprocate trustful actions while short-term types defect right from the beginning of an interaction. Note that the short-term and long-term types are not defined through their other-regarding preferences. The model that we outlined here can explain trusters placing trust and trustees behaving trustworthily even if both parties only pursue their own material interests. Put differently, signalling norms facilitate trust and cooperation even among Homines oeconomici. Of course, the model can be extended by including other-regarding preferences of honesty [86]. Long-term types may develop an intrinsic value of honest behaviour making cooperation more likely even with members of an outgroup.

To be sure, signalling theory and our model cannot predict the specific content of a signalling norm that might evolve in a situation requiring trust. Explanations that, for example, tattoo norms had evolved in delinquent communities for identifying their members are ex-post. Any type of behaviour could have evolved for this purpose and a variety of factors contribute to the emergence of particular norms. However, the theory can predict in which situations, determined by the parameters of our model, signalling norms are more likely to emerge. Consider once more the case that the proportion of long-term interaction partners is below the threshold denoted by α^* (equation (4.3)). Then, the theory predicts that the probability of signalling norms arising and the strictness of compliance with these norms are larger than in a situation where the proportion of long-term types is above the threshold. Moreover, the threshold may vary; it depends on the game payoff parameters R, P, S and the 'intensity' of the long-term interest δ_l (the discount factor). These parameters can be systematically varied to derive hypotheses that can be tested in controlled experiments [87]. Observations from field or survey studies are also valuable for inspecting the theory. For example, Aksoy & Gambetta's [71] explanation for why Muslim women veil more in a more secular environment is in accordance with our model predictions. Relatedly, Patel [102] observes that the proliferation of veiling instigates pious groups to invest more in signals to make a distinction between more and less religious groups. The latter is an example of a signalling norm that is adjusted in response to a reoccurring need for distinguishing true from would-be believers.

Although signalling norms are more likely to emerge in situations requiring trust, the trustees must be willing to
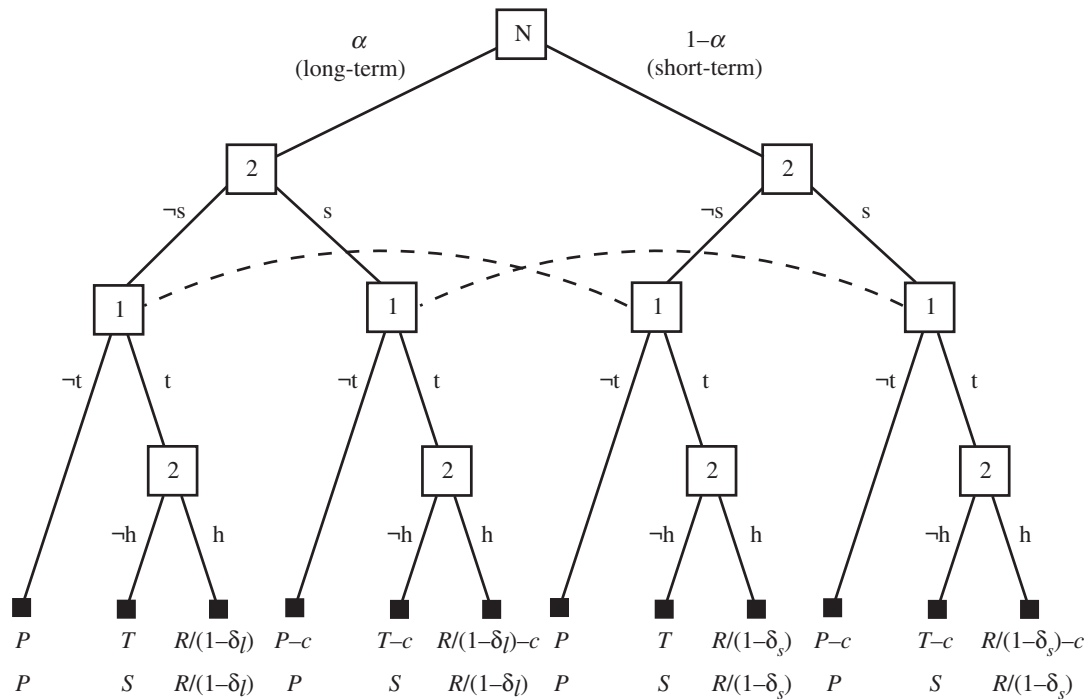
Figure 4. In the signalling trust game too, first nature (N) determines the trustee's type, but before the truster decides whether or not to place trust, the trustee decides whether or not to send a signal. The signal is associated with a cost c for the trustee and this cost is incurred first and irrespective of what the truster and the trustee do thereafter. Before the truster decides whether or not to place trust, they observe the trustee's signalling decision.

pay the price for the signal. The price of truthful information has to be large enough to exclude cheaters and free-riders. Consequently, signalling norms are not always the most efficient solution to trust problems. There are other means to reduce uncertainty about trustees' trustworthiness in social exchange. Contract law and legal sanctions deter cheating but come with transaction and enforcement costs. Legal solutions to trust problems are also not viable in extra-legal contexts. Signalling norms are thus more likely to emerge if the costs of contracts and legal enforcement are too high or formal trust-building mechanism absent altogether. We suppose that this is one reason why signalling norms are common in delinquent communities. For example, extensive body tattoos used to denote membership in a criminal organization such as the Japanese yakuza or the Russian vory, and these signs of membership were protected through the punishment of non-members that adopted them [34,103]. In various sub-cultures and in delinquent groups signalling norms promote social order in the absence of state-sanctioned enforcement mechanisms [104].

In the same vein, it has been argued that in pre-state societies, religions contributed to solving cooperation problems arising in increasingly large human groups ([101, pp. 462–464]). Human groups, by defining markers of group distinction related to religious beliefs and practices, established group identities through which members of different groups could be discerned. Because being a member of a religious group often came with considerable benefits, either provided directly by the collective or indirectly through group reputations [105], religious groups had to protect their identifying markers against free-riders, who would adopt these markers to earn the benefits without contributing to the collective good. Already Durkheim [106] suggested that religious groups, to overcome the free-rider problem, established various kinds of collective activities (e.g. communal praying) that would allow them to monitor the commitment of their members to the group. Durkheim called these religious activities *collective effervescence*. However, it has been objected that collective effervescence, although effective in promoting in-group solidarity [107], do not signal religious identities beyond the group in which they are practised. Henrich [108] argued that members of religious communities could signal their commitment more broadly by engaging in so-called CREDs (see also [109–111]). The more costly it is to perform CREDs, the stronger is the signal that the person performing them is committed to the beliefs and values of the group. CREDs can range from cheap (e.g. wearing a certain garb) to costly (e.g. praying five times a day) to very costly (e.g. self-mutilation) and are performed by individuals in public but not necessarily as part of a collective ritual.

Both collective effervescence and CREDs are related to signalling norms in that they signal individuals' commitments to a group. However, collective effervescence and CREDs are limited to religious identities and the performance of CREDs is mostly voluntary. Signalling norms thus offer explanations for the widespread use of markers of group distinction by individuals also in other than religious contexts (e.g. social classes, criminal organizations).

Endnotes

[1]Positive externalities can also create a demand for social norms that prescribe certain behaviours [40].

[2]The distinction between the signalling and commitment aspect of signalling norms is, in principle, equivalent to the distinction between signalling and screening mechanisms in games with

asymmetric and incomplete information [78–81]. We thus argue that signalling norms are screening devices that emerge out of a group's demand to identify trustworthy group members. We prefer the term 'signalling norms' over 'screening norms' because the screening mechanism can be subsumed under the general and more widely known signalling theory framework. We prefer the term 'commitment' over 'screening' because the former reflects better the bonding aspect of group memberships.

[3]Group identity (i.e. social identity) can thus be conceived as a system of signalling norms the adherence to which separates the members of the group from the non-members. This is an important point to make because it links the literature that conceives of social norms as regulating social interactions [42] with the literature that equates social norms with social identities [85].

[4]Markers of group distinction can also be imposed on members of a group by members of another group to facilitate discrimination and oppression. So-called badges of shame must be kept distinct from signalling norms as the members of the oppressed group do not collectively agree to and sanction wearing them.

[5]The present value of the reward from repeated cooperation is $R/(1 - \delta) = R + \delta R + \delta^2 R + \delta^3 R + \ldots$ (see, e.g. [92]).

[6]In our model, we assume that only mutual cooperation can go on forever but not mutual defection. We make this assumption because it makes our model more comprehensible. We obtain similar results under the assumption that in an infinitely repeated trust game the truster and the trustee use a so-called trigger strategy, which starts defecting forever once the other party abstains from cooperation (e.g. [92]). Then, equation (4.1) becomes $\delta_l > (T - R)/(T - P) > \delta_s$, equation (4.3) becomes $\alpha^* = (P_l - S - \delta_s P_s)/(R_l - S - \delta_s P_s)$ and equation (4.4) becomes $R_l - P_l > c > T - P$, where R_l, P_l and P_s are, respectively, $R/(1 - \delta_l)$, $P/(1 - \delta_l)$ and $P/(1 - \delta_s)$ (see the electronic supplementary material, data S1 in [87]).

References

1. Balliet D, Parks C, Joireman J. 2009 Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Process. Intergroup Relat.* **12**, 533–547. (doi:10.1177/1368430209105040)

2. Fehr E, Schmidt KM. 2006 The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In *Handbook of the economics of giving, altruism and reciprocity, volume 1* (eds S-C Kolm, JM Ythier), pp. 615–691. Amsterdam, The Netherlands: North Holland.

3. van Lange PAM. 1999 The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *J. Pers. Soc. Psychol.* **77**, 337–349. (doi:10.1037/0022-3514.77.2.337)

4. Buskens V, Raub W. 2013 Rational choice research on social dilemmas: embeddedness effects on trust. In *The handbook of rational choice social research* (eds R Wittek, T Snijders, V Nee), pp. 113–150. Stanford, CA: Stanford University Press.

5. Nowak MA, Sigmund K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131)

6. Trivers RL. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57. (doi:10.1086/406755)

7. Axelrod R. 1984 *The evolution of cooperation.* New York, NY: Basic Books.

8. Dal Bó P. 2005 Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *Am. Econ. Rev.* **95**, 1591–1604.

9. Fudenberg D, Maskin E. 1986 The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* **54**, 533–554. (doi:10.1257/000282805775014434)

10. Giardini F, Wittek R (eds) 2019 *The Oxford handbook of gossip and reputation.* Oxford, UK: Oxford University Press.

11. Przepiorka W, Norbutas L, Corten R. 2017 Order without law: reputation promotes cooperation in a cryptomarket for illegal drugs. *Eur. Sociol. Rev.* **33**, 752–764. (doi:10.1093/esr/jcx072)

12. Roberts G, Raihani N, Bshary R, Manrique HM, Farina A, Samu F, Barclay P. 2021 The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. *Phil. Trans. R. Soc. B* **376**, 20200290. (doi:10.1098/rstb.2020.0290)

13. Sylwester K, Roberts G. 2013 Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evol. Hum. Behav.* **34**, 201–206. (doi:10.1016/j.evolhumbehav.2012.11.009)

14. Wu J, Balliet D, van Lange PAM. 2016 Gossip versus punishment: the efficiency of reputation to promote and maintain cooperation. *Sci. Rep.* **6**, 23919. (doi:10.1038/srep23919)

15. Abbink K, Gangadharan L, Handfield T, Thrasher J. 2017 Peer punishment promotes enforcement of bad social norms. *Nat. Commun.* **8**, 609. (doi:10.1038/s41467-017-00731-0)

16. Bliege Bird R, Smith EA. 2005 Signaling theory, strategic interaction, and symbolic capital. *Curr. Anthropol.* **46**, 221–248. (doi:10.1086/427115)

17. Efferson C, Vogt S, Elhadi A, Ahmed HEF, Fehr E. 2015 Female genital cutting is not a social coordination norm. *Science* **349**, 1446–1447. (doi:10.1126/science.aaa7978)

18. Gambetta D, Origgi G. 2013 The LL game: the curious preference for low quality and its norms. *Politics Phil. Econ.* **12**, 3–23. (doi:10.1177/1470594X11433740)

19. Horne C. 2009 *The rewards of punishment: a relational theory of norm enforcement.* Stanford, CA: Stanford University Press.

20. Opp K-D. 2018 Externalities, social networks, and the emergence of norms: a critical analysis and extension of James Coleman's theory. *Soc. Res.* **85**, 167–196.

21. Sosis R. 2005 Does religion promote trust? the role of signaling, reputation, and punishment. *Interdiscip. J. Res. Relig.* **1**, 1–30.

22. Willer R, Kuwabara K, Macy MW. 2009 The false enforcement of unpopular norms. *Am. J. Sociol.* **115**, 451–490.

23. Mackie G. 1996 Ending footbinding and infibulation: a convention account. *Am. Sociol. Rev.* **61**, 999–1017. (doi:10.2307/2096305)

24. Mesoudi A. 2011 *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences.* Chicago, IL: University of Chicago Press.

25. Ogburn WF. 1922 *Social change with respect to culture and original nature.* New York, NY: Huebsch.

26. Bernhard H, Fischbacher U, Fehr E. 2006 Parochial altruism in humans. *Nature* **442**, 912–915. (doi:10.1038/nature04981)

27. 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)

28. Habyarimana J, Humphreys M, Posner DN, Weinstein JM. 2009 *Coethnicity: diversity and the dilemmas of collective action.* New York, NY: Russell Sage Foundation.

29. Handley C, Mathew S. 2020 Human large-scale cooperation as a product of competition between cultural groups. *Nat. Commun.* **11**, 702. (doi:10.1038/s41467-020-14416-8)

30. Sherif M, Harvey OJ, White JB, Sherif CW. 1961 *Intergroup conflict and co-operation: the robber's cave experiment.* Norman, OK: University Book Exchange.

31. 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)

32. Efferson C, Lalive R, Fehr E. 2008 The coevolution of cultural groups and ingroup favoritism. *Science* **321**, 1844–1849. (doi:10.1126/science.1155805)

33. McNamara T. 2005 21st century Shibboleth: language tests, identity and intergroup conflict. *Lang. Policy* **4**, 351–370. (doi:10.1007/s10993-005-2886-0)

34. Gambetta D. 2009 *Codes of the underworld: how criminals communicate.* Princeton, NJ: Princeton University Press.

35. Gintis H, Smith EA, Bowles S. 2001 Costly signaling and cooperation. *J. Theor. Biol.* **213**, 103–119.

36. Spence MA. 1974 *Market signaling: informational transfer in hiring and related screening processes.* Cambridge, MA: Harvard University Press.

37. Zahavi A. 1975 Mate selection-A selection for a handicap. *J. Theor. Biol.* **53**, 205–214.

38. Zahavi A. 1977 The cost of honesty: further remarks on the handicap principle. *J. Theor. Biol.* **67**, 603–605.

39. Posner EA. 2000 *Law and social norms*. Cambridge, MA: Harvard University Press.

40. Coleman JS. 1990 *Foundations of social theory*. Cambridge, MA: The Belknap Press of Harvard University Press.

41. Hechter M, Opp K-D (eds) 2001 *Social norms*. New York, NY: Russell Sage Foundation.

42. Horne C, Mollborn S. 2020 Norms: an integrated framework. *Annu. Rev. Sociol.* **46**, 467–487. (doi:10.1146/annurev-soc-121919-054658)

43. Voss T. 2001 Game-theoretical perspectives on the emergence of social norms. In *Social norms* (eds M Hechter, K-D Opp), pp. 105–136. New York, NY: Russell Sage Foundation.

44. Dawes RM. 1980 Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193. (doi:10.1146/annurev.ps.31.020180.001125)

45. Kollock P. 1998 Social dilemmas: the anatomy of cooperation. *Annu. Rev. Sociol.* **24**, 183–214.

46. Rapoport A. 1974 Prisoner's Dilemma – recollections and observations. In *Game theory as a theory of conflict resolution* (ed. A Rapoport), pp. 17–34. Dordrecht, The Netherlands: Reidel.

47. van Lange PAM, Joireman J, Parks CD, van Dijk E. 2013 The psychology of social dilemmas: a review. *Organ. Behav. Hum. Decis. Process.* **120**, 125–141. (doi:10.1016/j.obhdp.2012.11.003)

48. Ullmann-Margalit E. 1977 *The emergence of norms*. Oxford, UK: Oxford University Press.

49. Przepiorka W. In press. Game-theoretic models. In *Research handbook on analytical sociology* (ed. G Manzo). Cheltenham, UK: Edward Elgar.

50. Bicchieri C. 2006 *The grammar of society: the nature and dynamics of social norms*. Cambridge, UK: Cambridge University Press.

51. Cialdini RB, Reno RR, Kallgren CA. 1990 A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* **58**, 1015–1026. (doi:10.1037/0022-3514.58.6.1015)

52. Hawkins RXD, Goodman ND, Goldstone RL. 2019 The emergence of social norms and conventions. *Trends Cogn. Sci.* **23**, 158–169. (doi:10.1016/j.tics.2018.11.003)

53. Lewis D. 1969 *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.

54. Przepiorka W *et al*. In preparation. How norms emerge from conventions (and change).

55. Guala F, Mittone L. 2010 How history and convention create norms: an experimental study. *J. Econ. Psychol.* **31**, 749–756. (doi:10.1016/j.joep.2010.05.009)

56. Elster J. 2007 *Explaining social behavior: more nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University Press.

57. Elias N. 1994 [1936] *The civilizing process. Sociogenetic and psychogenetic investigations*. Oxford, UK: Blackwell.

58. Bourdieu P. 1979 *La distinction: critique sociale du jugement*. Paris, France: Les Éditions du Minuit.

59. Veblen T. 1994 [1899] *The theory of the leisure class: an economic study of institutions*. New York, NY: Dover Publications.

60. Gambetta D. 2009 Signaling. In *The Oxford handbook of analytical sociology* (eds P Hedström, P Bearman), pp. 168–194. Oxford, UK: Oxford University Press.

61. Barclay P. 2004 Trustworthiness and competitive altruism can also solve the 'tragedy of the commons'. *Evol. Hum. Behav.* **25**, 209–220.

62. 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)

63. 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)

64. Posner EA. 1998 Symbols, signals, and social norms in politics and the law. *J. Leg. Stud.* **27**, 765–798.

65. Camerer CF. 2003 *Behavioral game theory*. Princeton, NJ: Princeton University Press.

66. Gintis H. 2009 *Game theory evolving: a problem-centered introduction to modeling strategic interaction*–2nd edn. Princeton, NJ: Princeton University Press.

67. Hardin R. 2002 *Trust and trustworthiness*. New York, NY: Russell Sage Foundation.

68. McAdams RH. 2001 Signaling discount rates: law, norms, and economic methodology. *Yale Law J.* **110**, 625–689.

69. Greene J. 2013 *Moral tribes: emotion, reason and the gap between us and them*. New York, NY: Penguin Books.

70. Dasgupta P. 1988 Trust as a commodity. In *Trust: making and breaking cooperative relations* (eds D Gambetta), pp. 49–72. Oxford, UK: Basil Blackwell.

71. Aksoy O, Gambetta D. 2016 Behind the veil: the strategic use of religious garb. *Eur. Sociol. Rev.* **6**, 792–806. (doi:10.1093/esr/jcw035)

72. Aksoy O, Gambetta D. 2021 The politics behind the veil. *Eur. Sociol. Rev.* **37**, 67–88. (doi:10.1093/esr/jcaa035)

73. Raub W. 2004 Hostage posting as a mechanism of trust: binding, compensation, and signaling. *Ration. Soc.* **16**, 319–365.

74. Gambetta D, Przepiorka W. 2019 Sharing compromising information as a cooperative strategy. *Sociol. Sci.* **6**, 352–379. (doi:10.15195/v6.a14)

75. Hardin R. 1996 Trustworthiness. *Ethics* **107**, 26–42.

76. Schelling TC. 1960 *The strategy of conflict*. Cambridge, MA: Harvard University Press.

77. Eriksson L. 2019 Social norms as signals. *Soc. Theory Pract.* **45**, 579–599. (doi:10.5840/soctheorpract2019123072)

78. Archetti M. 2011 Contract theory for the evolution of cooperation: the right incentives attract the right partners. *J. Theor. Biol.* **269**, 201–207. (doi:10.1016/j.jtbi.2010.10.025)

79. Archetti M, Scheuring I, Hoffman M, Frederickson ME, Pierce NE, Yu DW. 2011 Economic game theory for mutualism and cooperation. *Ecol. Lett.* **14**, 1300–1312. (doi:10.1111/j.1461-0248.2011.01697.x)

80. Kübler D, Müller W, Normann H-T. 2008 Job-market signaling and screening: an experimental comparison. *Games Econ. Behav.* **64**, 219–236.

81. Riley JG. 2001 Silver signals: twenty-five years of screening and signaling. *J. Econ. Lit.* **39**, 432–478.

82. Iannaccone LR. 1994 Why strict churches are strong. *Am. J. Sociol.* **99**, 1180–1211.

83. Sosis R. 2003 Why aren't we all Hutterites? Costly signaling theory and religious behavior. *Hum. Nat.* **14**, 91–127.

84. Takács K, Gross J, Testori M, Letina S, Kenny AR, Power EA, Wittek RPM. 2021 Networks of reliable reputations and cooperation: a review. *Phil. Trans. R. Soc. B* **376**, 20200297. (doi:10.1098/rstb.2020.0297)

85. Akerlof GA, Kranton RE. 2000 Economics and identity. *Q. J. Econ.* **115**, 715–753.

86. 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), pp. 373–392. Berlin, Germany: De Gruyter Oldenbourg.

87. Przepiorka W, Diekmann A. 2013 Temporal embeddedness and signals of trustworthiness: experimental tests of a game theoretic model in the United Kingdom, Russia, and Switzerland. *Eur. Sociol. Rev.* **29**, 1010–1023. (doi:10.1093/esr/jcs078)

88. Diekmann A. 2009 *Spieltheorie: Einführung, Beispiele, Experimente*. Reinbek, Germany: Rowohlt.

89. Diekmann A, Przepiorka W. 2010 Soziale Normen als Signale. Der Beitrag der Signaling-Theorie. In *Soziologische Theorie kontrovers. Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie* (eds G Albert, S Sigmund), pp. 220–237. Wiesbaden, Germany: VS Verlag.

90. Przepiorka W. 2009 Reputation and signals of trustworthiness in social interactions. Dissertation no. 18649, Zurich ETH, Zurich, Switzerland. (doi:10.3929/ethz-a-005994644)

91. Rapoport A, Chammah AM. 1965 *Prisoner's dilemma. a study of conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

92. Osborne MJ. 2009 *An introduction to game theory, international edition*. Oxford, UK: Oxford University Press.

93. Greif A. 1989 Reputation and coalitions in medieval trade: evidence on the Maghribi traders. *J. Econ. Hist.* **49**, 857–882. (doi:10.1017/S002205070000475)

94. Greif A. 1993 Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition. *Am. Econ. Rev.* **83**, 525–548.

95. Camerer C, Weigelt K. 1988 Experimental test of a sequential equilibrium reputation model. *Econometrica* **56**, 1–36.

96. Voss T. 1998 Vertrauen in modernen Gesellschaften. Eine spieltheoretische Analyse. In *Der Transformationsprozess: Analysen und Befunde aus dem Leipziger Institut für Soziologie* (eds R Metze, K Mühler, K-D Opp), pp. 91–129. Leipzig, Germany: Leipziger Universitätsverlag.

97. Bacharach M, Gambetta D. 2001 Trust in signs. In *Trust in society* (ed. KS Cook), pp. 148–184. New York, NY: Russell Sage Foundation.

98. Johnstone RA. 1997 The evolution of animal signals. In *Behavioral ecology: an evolutionary approach* (eds JR Krebs, NB Davies), pp. 155–178. Oxford, UK: Blackwell.

99. Romano A, Giardini F, Columbus S, de Kwaadsteniet EW, Kisfalusi D, Triki Z, Snijders C, Hagel K. 2021 Reputation and socio-ecology in humans. *Phil. Trans. R. Soc. B* **376**, 20200295. (doi:10.1098/rstb.2020.0295)

100. Hess NH, Hagen EH. 2021 Competitive gossip: the impact of domain, resource value, resource scarcity and coalitions. *Phil. Trans. R. Soc. B* **376**, 20200305. (doi:10.1098/rstb.2020.0305)

101. van Tubergen F. 2020 *Introduction to sociology*. London, UK: Routledge.

102. Patel DS. 2012 Concealing to reveal: the informational role of Islamic dress. *Ration. Soc.* **24**, 295–323. (doi:10.1177/1043463112440683)

103. Varese F. 2005 *The Russian mafia: private protection in a new market economy*. Oxford, UK: Oxford University Press.

104. Ellickson RC. 1991 *Order without law: how neighbors settle disputes*. Cambridge, MA: Harvard University Press.

105. Weber M. 2002 [1920]. *The protestant ethic and the spirit of capitalism*. Oxford, UK: Blackwell.

106. Durkheim E. 2021 [1912]. *The elementary forms of religious life*. Oxford, UK: Oxford University Press.

107. Power EA. 2018 Collective ritual and social support networks in rural South India. *Proc. R. Soc. B* **285**, 20180023. (doi:10.1098/rspb.2018.0023)

108. Henrich J. 2009 The evolution of costly displays, cooperation and religion: credibility enhancing displays and their implications for cultural evolution. *Evol. Hum. Behav.* **30**, 244–260. (doi:10.1016/j.evolhumbehav.2009.03.005)

109. Norenzayan A. 2013 *Big Gods: how religion transformed cooperation and conflict*. Princeton, NJ: Princeton University Press.

110. Power EA. 2017 Discerning devotion: testing the signaling theory of religion. *Evol. Hum. Behav.* **38**, 82–91. (doi:10.1016/j.evolhumbehav.2016.07.003)

111. 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)