Author for correspondence:
Junhui Wu
e-mail: wujunhui@psych.ac.cn

THE ROYAL SOCIETY
PUBLISHING

Honesty and dishonesty in gossip strategies: a fitness interdependence analysis

Junhui Wu[1,2,†], Szabolcs Számadó[3,4,†], Pat Barclay[5], Bianca Beersma[6], Terence D. Dores Cruz[6], Sergio Lo Iacono[7], Annika S. Nieper[6], Kim Peters[8], Wojtek Przepiorka[7], Leo Tiokhin[9] and Paul A. M. Van Lange[10]

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[3]Department of Sociology and Communication, Budapest University of Technology and Economics, H-1111 Budapest, Hungary
[4]CSS-RECENS, Centre for Social Sciences, H-1097 Budapest, Hungary
[5]Department of Psychology, University of Guelph, Guelph, Ontario, N1G 2W1, Canada
[6]Department of Organization Sciences, Vrije Universiteit Amsterdam, 1081HV Amsterdam, The Netherlands
[7]Department of Sociology/ICS, Utrecht University, 3584CS Utrecht, The Netherlands
[8]University of Exeter Business School, Exeter EX4 4PU, UK
[9]Human Technology Interaction Group, Eindhoven University of Technology, 5600MB Eindhoven, The Netherlands
[10]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, 1081BT Amsterdam, The Netherlands

JW, 0000-0003-3224-3595; PB, 0000-0002-7905-9069; TDDC, 0000-0002-4792-8469;
WP, 0000-0001-9432-8696; PAMVL, 0000-0001-7774-6984

Gossip, or sharing information about absent others, has been identified as an effective solution to free rider problems in situations with conflicting interests. Yet, the information transmitted via gossip can be biased, because gossipers may send dishonest information about others for personal gains. Such dishonest gossip makes reputation-based cooperation more difficult to evolve. But when are people likely to share honest or dishonest gossip? We build formal models to provide the theoretical foundation for individuals' gossip strategies, taking into account the gossiper's fitness interdependence with the receiver and the target. Our models across four different games suggest a very simple rule: when there is a perfect match (mismatch) between fitness interdependence and the effect of honest gossip, the gossiper should always be honest (dishonest); however, in the case of a partial match, the gossiper should make a choice based on their fitness interdependence with the receiver and the target and the marginal cost/benefit in terms of pay-off differences caused by possible choices of the receiver and the target in the game. Moreover, gossipers can use this simple rule to make optimal decisions even under noise. We discuss empirical examples that support the predictions of our model and potential extensions.

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

1. Introduction

Humans recurrently encounter situations characterized by conflicting interests with ingroup members, outgroup members or strangers. In such mixed-motive situations, the decision to cooperate or compete has important consequences for the involved parties [1,2]. Promoting cooperation in such situations is especially important, because of the benefits of cooperation for collective welfare. Yet, encouraging every individual to cooperate by paying personal cost to benefit others is often challenging [3]. This paradox has stimulated researchers from various disciplines to investigate pathways that facilitate cooperation. One factor that has captured

researchers' attention is gossip, which refers to the process of sharing information that is positive, negative or neutral about absent others with one or more receivers [4–9]. Researchers increasingly realize the important social functions of gossip [9–14]. Specifically, to the extent that gossip provides information about others' trustworthiness, it allows receivers of gossip to detect potential cheaters and selectively cooperate with deserving others [5,15–17]. Moreover, when people are unable to directly punish free riders, gossip can be used as a low-cost form of punishment that can impose reputational costs on free riders [15,18]. Of course, the fact that gossip involves low cost does not mean that it is risk-free: when the gossip target finds out the identity of the gossiper who shares the negative information, they may punish or ostracize the gossiper [19]. In sum, there is a growing body of research suggesting that gossip is a key mechanism that promotes and sustains cooperation [14,20–27].

However, there is at least one important barrier to gossip serving its functions in everyday life: people may be motivated to share dishonest information about others for personal benefit. The possibility that gossip can be biased or dishonest [28,29] may make it difficult for reputation-based cooperation to evolve [9,30,31]. For example, when people manipulate gossip in ways that result in free riders having a good reputation, free riders may use this good reputation to mislead and exploit cooperators. Similarly, if cooperators are falsely assigned bad reputations, an otherwise potentially cooperative interaction between two cooperators may break down because they do not trust each other, leading to mutual defection and confirming the (initial false) bad reputations [32]. Indeed, research shows that gossip does not stimulate cooperation if it is false or inaccurate at sufficiently high levels [11,33,34], notwithstanding the evidence that in certain situations, inaccurate gossip can still support cooperation [35,36]. Because the positive impact of gossip on cooperation thus seems limited to situations in which gossip reflects true information about others, it is important to shed light on the question of when gossip is likely to be honest or dishonest.

Here, we use a novel approach to model and analyse honest and dishonest gossip as a strategic behaviour in line with models from biological signalling theory [37,38]. The key insight of signalling theory is that signals are adaptations shaped by marginal costs and marginal benefits of different behaviours, and that the ultimate function of the signaller's behaviour is to maximize their fitness [39,40]. The goal of 'honest signalling' models is to analyse the conditions under which this fitness optimization will result in an honest equilibrium [38,41–43], such that the receiver receives reliable information about the signaller from the signals, and this information in turn helps the receiver to achieve higher fitness. While seminal signalling models investigate pairwise interactions between signallers and receivers [37,38,41], the bare minimum for gossip is a triad of a signaller (i.e. gossiper), a receiver and a target. This implies that while our basic approach is the same, conclusions from seminal signalling models cannot be directly transferred to gossip.

In addition, our approach differs from the traditional approach of indirect reciprocity models that address the effect of noise on the evolution of cooperation [30,35,44]. Noise is usually modelled in terms of errors (e.g. an error of perception or error of judgement in assigning reputations) in these models. The conclusion of these models is that a reputation system that supports the evolution of cooperation is usually robust up to a (not too high) level of noise [30,35,44]. The key difference between the strategic approach and the 'dishonesty as noise' approach is

that the level of noise is due to external factors and it is usually assumed to be fixed [30,35,44]. By contrast, the frequency of a strategy is determined by internal factors, and a successful strategy will spread through the population. For instance, a dishonest gossip strategy can increase its frequency in the population if it is more successful than an honest one. In the long term, this can challenge the evolution of cooperation even if the frequency of dishonest gossip strategy is low at the start of a simulation [34].

All in all, we propose that it is important to analyse honest and dishonest gossip as a strategic behaviour shaped by marginal costs and marginal benefits of different behaviours. In order to describe these marginal costs and marginal benefits in a gossip triad, we draw on recent theoretical developments on fitness interdependence [45–48]. Specifically, we propose that gossipers will choose a gossip strategy (i.e. honest or dishonest gossip) that maximizes their fitness benefits and minimizes their fitness costs, which are in part determined by the levels of fitness interdependence between the gossiper and the other two parties (i.e. the target and the receiver) that together constitute the gossip triad. Different from a recent review on how interdependence among the actors in the gossip triad affects when people do not gossip [49], we analyse the conditions under which people share honest versus dishonest gossip—a topic that has hardly received any attention—and thus provide a novel and meaningful contribution to the theoretical developments on gossip honesty. In the following sections, we first outline the fitness interdependence perspective and illustrate the types of cues that people use to infer their fitness interdependence with others. We then specify how fitness interdependence between the gossiper and the other two parties in the gossip triad may relate to the gossiper's honest or dishonest gossip strategy using a modelling approach. We end by discussing empirical examples that support the predictions of our models, and potential extensions in future research.

2. Fitness interdependence and the situational cues for fitness interdependence inferences

Fitness interdependence refers to the extent to which one or several organisms influence each other's success in replicating their genes [45–48]. Fitness interdependence is slightly different from interdependence theory [50,51] and functional interdependence theory [52]: the latter two theories categorize social situations into several dimensions of how people affect one another (e.g. the degree to which each person can determine their own outcomes; covariation of interests), whereas fitness interdependence focuses solely on fitness interests. While (functional) interdependence theories primarily focus on the proximate mechanisms of behaviours, the concept of fitness interdependence extends these theories by focusing on the ultimate causes of behaviours [53]. Fitness interdependence can be reflected in a stake index (s), which depicts the extent to which changes in one's fitness relate to changes in another's fitness [48]. For instance, imagine two people who must work together for their livelihood—each one's well-being depends on the other doing well. More formally, imagine someone paying a personal cost c ($c > 0$) to provide a benefit b to a recipient, but at the same time also gaining a secondary benefit (sb) that is a function of the recipient's gain. When $sb > c$, a larger value of s suggests that the individual will be more likely to act altruistically, i.e. to provide a benefit to the recipient despite the personal cost this entails. Similarly, welfare tradeoff
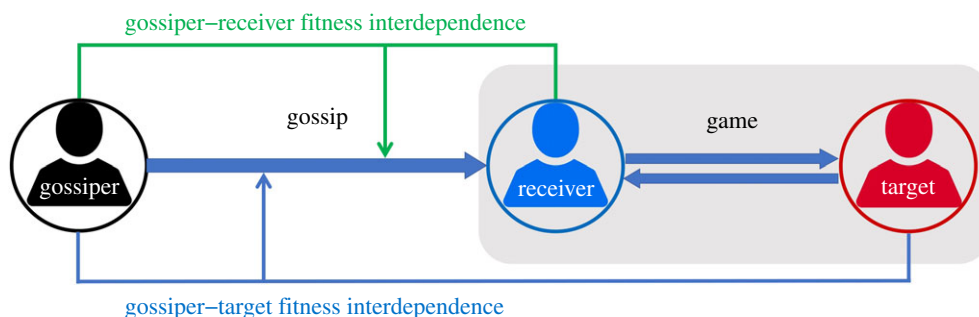
Figure 1. Structure of the sequential interaction between the gossiper, the receiver and the target. First, the gossiper decides to send honest or dishonest gossip to the receiver, then the receiver and the target interact in a two-person game. The gossiper's fitness outcome (and thus their optimal decision about whether to share honest gossip) is influenced by their fitness interdependence with the receiver and the target.

ratio (WTR), which refers to the extent to which a person values another's welfare relative to their own [54], can be considered a proximal mechanism driven by fitness interdependence. Thus, the higher fitness interdependence one has with someone else, the higher one's WTR towards this other person. Accordingly, the condition for altruistic behaviour involving WTR (i.e. $WTR \times b > c$) is the same as that involving the stake (i.e. $sb > c$). Notably, s has a range of values that have different meanings. Positive fitness interdependence ($s > 0$) implies that individuals positively affect one another's survival and reproduction, whereas negative fitness interdependence ($s < 0$) implies that individuals negatively affect one another's survival and reproduction. Of course, there can also be an absence of fitness interdependence ($s = 0$) such that individuals do not have any effect on one another's survival and reproduction.

Notably, it is often challenging to measure how individuals' behaviour affects their own and others' fitness. If this is the case, how do people assess their fitness interdependence with others? Researchers suggest that humans may have evolved to use various situational cues to assess how their fate is evolutionarily intertwined with others [45]. First, group membership may be one such cue for people to assess their fitness interdependence with others. According to the bounded generalized reciprocity perspective [55,56], human groups provide a container for generalized exchange network, in which individuals who behave cooperatively towards ingroup members gain a good reputation, and thus obtain indirect benefits from other group members. Thus, groups are important for individuals' survival and reproductive success. Indeed, numerous studies have documented that group membership plays an important role in fostering individuals' trust, cooperation, and norm enforcement behaviours. For example, people are more likely to trust ingroup members [16], cooperate with ingroup members [16,57], and punish in ways that protect the ingroup victims from norm violations, as well as more harshly punishing norm violators who are members of an outgroup than the ingroup [58–61]. These cooperative interactions create fitness interdependence among each other, in that people have an interest in their group's persistence so that these interactions can continue [62]. Thus, it is plausible that compared with an outgroup member, people may be more likely to treat an ingroup member as someone they have positive fitness interdependence with.

Second, individuals' observable actions towards others in social interactions may also signal their fitness interdependence with others. For instance, when an individual incurs a cost to help another, this helping behaviour can signal that the helper values the recipient, and has enough stake in the welfare of the

recipient who may repay with subsequent trust, implying that the helper has positive fitness interdependence with the recipient [54,63]. Third, partners' emotion expressions can be used to assess one's fitness interdependence with partners. For instance, partners' anger expressions may reflect negative fitness interdependence [64,65], whereas partners' emotions of happiness, forgiveness or gratitude may reflect positive fitness interdependence and a higher likelihood of helping [66–68].

3. Fitness interdependence and gossip strategies: a modelling approach

Drawing on the fitness interdependence framework, we model the gossiper's gossip behaviour towards the receiver in the gossip triad based on the gossiper's fitness interdependence with the target and the receiver (figure 1). We assume that the receiver and the target play a one-shot two-person game. The receiver does not know the target's behavioural type until the gossiper sends them this information. It is in the interest of the receiver to find out the target's type, as the receiver's optimal response may depend on this information. In our theoretical models, we make the following simplifying assumptions: (i) the target is either a cooperating type (always cooperate) or a defecting type (always defect); (ii) the gossiper has accurate knowledge about the target's type that is acquired through experience or direct observation, and always sends gossip that either honestly informs or misleads the receiver about the target's type; (iii) the receiver will trust any gossip that they receive, and (iv) there is no noise (i.e. unintended errors that cause discrepancies between the expected outcomes and actual outcomes; [69]) in the gossip transmission process. This allows us to investigate the conditions under which the gossiper transmits honest or dishonest gossip about the target's type. Later on, we relax the assumption of noise. Modification to the other assumptions is an area that can be addressed in future extensions of our work, but is beyond the scope of the current paper.

Below, we analyse four possible games that are played between the receiver and the target: a stag-hunt game, a snowdrift game, a helping game and a punishment game (see electronic supplementary material for a detailed description of the games). These four games provide examples for four types of possible outcomes of honest gossip for the fitness of the receiver and the target: (i) mutually beneficial (e.g. stag-hunt game with a cooperating target) [70], (ii) beneficial for the receiver but costly for the target (e.g. snowdrift game with a cooperating target) [71], (iii) beneficial for the target but costly for the receiver (e.g. helping game with a cooperating

Table 1. Conditions for honest gossip across four games with a cooperating or defecting target. In the stag-hunt and snowdrift games, R denotes the benefit of mutual cooperation, P denotes the cost of mutual defection, T is the 'temptation' benefit of defecting against a cooperating player, while S is the 'sucker's pay-off'; in the helping and punishment games, c denotes the cost of helping/punishing the target, b denotes the benefit for the target from helping and γ denotes the fine imposed on the target. V_{gr}, the gossiper's fitness interdependence with the receiver; V_{gt}, the gossiper's fitness interdependence with the target.

game	target type	condition for honest gossip	
stag-hunt game	cooperating	$V_{gr}(R - T) > V_{gt}(S - R)$	inequality (3.3)
	defecting	$V_{gr}(P - S) > V_{gt}(T - P)$	inequality (3.4)
snowdrift game	cooperating	$V_{gr}(T - R) > V_{gt}(R - S)$	inequality (3.5)
	defecting	$V_{gr}(S - P) > V_{gt}(P - T)$	inequality (3.6)
helping game	cooperating	$V_{gt}b > V_{gr}c$	inequality (3.7)
	defecting	$V_{gr}c > V_{gt}b$	inequality (3.8)
punishment game	cooperating	$V_{gr}c > -V_{gt}\gamma$	inequality (3.9)
	defecting	$-V_{gr}c > V_{gt}\gamma$	inequality (3.10)

target) [72], and finally (iv) mutually costly (e.g. punishment game with a defecting target) [73].

(a) The model

The receiver (r) and the target (t) play a two-person game $F = (B, E)$, where E denotes the payoffs from the game and B denotes the strategy set available to the players. The gossiper (g) is not involved in this game, but has a stake in the receiver's and the target's fitness outcomes, which corresponds to the gossiper's fitness interdependence with the receiver and the target. Figure 1 shows the structure of the sequential interaction between the gossiper, the receiver and the target.

Let V_{gr} and V_{gt} denote the fitness interdependence of the gossiper with the receiver and the target, respectively. Accordingly, the gossiper's fitness outcome as a result of the interaction between the receiver and the target can be written as follows:

$$E_g = V_{gr}E_r + V_{gt}E_t. \tag{3.1}$$

Here, E_g, E_r and E_t denote the fitness outcomes for the gossiper, the receiver and the target, respectively, for which we use payoffs in the game as proxy measures. The gossiper's action of sharing honest (H) or dishonest (D) gossip will influence the receiver's behavioural strategy towards the target (B_r), which not only influences the target and receiver's fitness, but also influences the gossiper's fitness via the varying levels of fitness interdependence. Sharing honest gossip will be an equilibrium strategy if the fitness outcome of gossipers sharing honest gossip is higher than the fitness outcome of gossipers sharing dishonest gossip, i.e. $E_g(H) > E_g(D)$.[1] Accordingly, the following condition must hold in the honest gossip equilibrium:

$$V_{gr}E_r(H) + V_{gt}E_t(H) > V_{gr}E_r(D) + V_{gt}E_t(D). \tag{3.2}$$

Here, $V_{gr}E_r(H)$ or $V_{gr}E_r(D)$ describes the influence of the receiver's fitness outcome on the fitness outcome of the gossiper via the fitness interdependence between the gossiper and the receiver, assuming gossip is honest or dishonest; $V_{gt}E_t(H)$ or $V_{gt}E_t(D)$ describes the influence of the target's fitness outcome on the fitness outcome of the gossiper via the fitness interdependence between the gossiper and the

target, assuming gossip is honest or dishonest. We investigate the conditions for honest gossip (based on inequality (3.2)) across four different two-person games (i.e. stag-hunt game, snowdrift game, helping game and punishment game; see electronic supplementary material, figure S1). Table 1 provides an overview of the conditions for honest gossip in these four games with cooperating and defecting targets (details of deriving these conditions can be found in the electronic supplementary material).

We have investigated four types of interactions in two-person games between the receiver and the target in the gossip triad. Our models show that the gossiper's fitness interdependence with the target and the receiver differentially affects the gossiper's likelihood to send honest or dishonest gossip across different types of interactions between the receiver and the target: (i) when mutual cooperation is beneficial for both the receiver and a cooperating target (i.e. the stag-hunt game versus a cooperating target), the gossiper will be honest when their overall fitness interdependence with the receiver and the target is positive (figure 2a,e); (ii) when defecting with a cooperating target is beneficial for the receiver, but costly for the target (i.e. the snowdrift game versus a cooperating target), the gossiper will be honest when their fitness interdependence with the receiver is higher than that with the target (figure 2b,f); (iii) when helping a cooperating target is costly for the receiver but beneficial for the target (e.g. the helping game versus a cooperating target), the gossiper will be honest when their fitness interdependence with the target is higher than that with the receiver (figure 2c,g); (iv) when punishing a defecting target decreases the pay-off of both the receiver and the target (e.g. the punishment game versus a defecting target), the gossiper will be honest when their overall fitness interdependence with the receiver and the target is negative (figure 2d,h). The slope of the boundary between the honest and dishonest gossip regions is determined by the marginal cost/benefit of honest gossip for the receiver and the target, respectively (e.g. the values of c and b in the helping game). The slope will be at the main diagonal when the marginal cost and benefit have the same value (e.g. $c = b = 1$). When there is no marginal benefit or cost for the receiver or the target in choosing one action over another action, then the gossiper's fitness interdependence with the receiver or the target does not matter. For example, if there is no
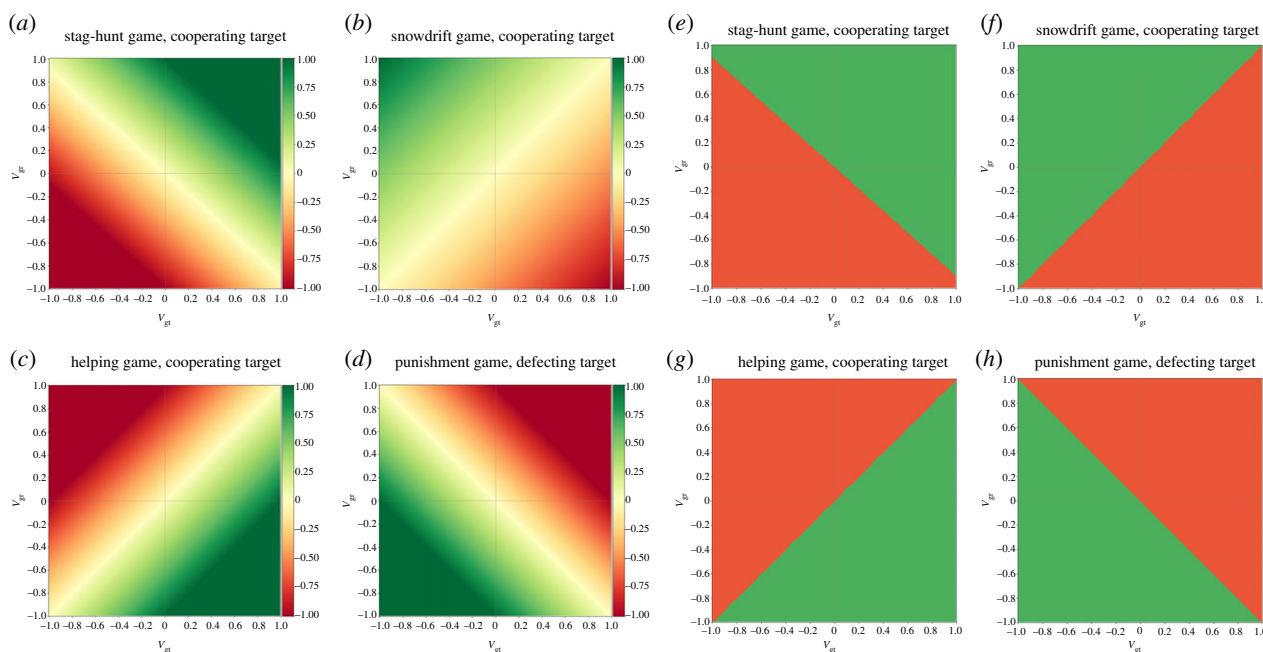
Figure 2. The marginal cost/benefit of honest gossip (*a–d*; darker green: higher marginal benefit of honest gossip, darker red: higher marginal cost of honest gossip; cost and benefit are represented with negative and positive numbers on the right bar) and the predicted behaviour of the gossiper (*e–h*; green area: honest gossip, red area: dishonest gossip) across the four main interaction types as a function of the fitness interdependence between the gossiper and the target (V_{gt}) and between the gossiper and the receiver (V_{gr}). (*a,e*) Mutualism (*receiver/target*: +/+; stag-hunt game with a cooperating target; $T = 0$, $S = 0.1$); (*b,f*) antagonism (*receiver/target*: +/−, snowdrift game with a cooperating target; $T = 1.5$, $S = 0.5$); (*c,g*) antagonism (*receiver/target*: −/+, helping game with a cooperating target; $b = 1$, $c = 1$); (*d,h*) competition (*receiver/target*: −/−, punishment game with a defecting target; $c = 1$, $\gamma = 1$).

marginal cost of being suckered in the snowdrift game (for the target, i.e. $R = S$), then the gossiper's fitness interdependence with the target does not matter (i.e. horizontal slope). If helping is cost-free in the helping game ($c = 0$), then the gossiper's fitness interdependence with the receiver does not matter (i.e. vertical slope, see examples for variations in marginal cost/benefit in the electronic supplementary material, figures S2–S9).

Taken together, our models across four different games illustrate that the gossiper's action will be determined by two broad factors: (i) the gossiper's fitness interdependence with the receiver and the target, and (ii) the marginal cost/benefit (pay-off differences between the two possible options) for the receiver and the target (i.e. game type). This implies that in order to predict the gossiper's behaviour, it is essential to have information about both factors.

However, upon closer investigation (see details in the electronic supplementary material), our models suggest that there is a simple rule that allows the gossiper to make optimal decisions in certain situations without being able to judge the exact parameters of the game between the receiver and the target. Table 2 shows honest and dishonest gossip based on different games and the valence of the gossiper's fitness interdependence with the receiver and the target. The rule is very simple: when there is a perfect match between fitness interdependence and the effect of honest gossip (i.e. V_{gr} and $E_r(H)$ are both positive or negative, and V_{gt} and $E_t(H)$ are both positive or negative), the gossiper should always be honest; when there is a perfect mismatch (i.e. V_{gr} and $E_r(H)$ are opposite in valence, and V_{gt} and $E_t(H)$ are opposite in valence), then the gossiper should always be dishonest. When there is a partial match between fitness interdependence and the effect of honest gossip, the gossiper has to make a choice based on the marginal cost/benefit and their fitness interdependence with the receiver and the target; these are the situations in

which the knowledge (estimation) of marginal cost/benefit is important (table 2). Notably, people can assess fitness interdependence and marginal cost/benefit using various cues (e.g. expression of emotions [45,52,78]), so that while they might not know the exact levels of fitness interdependence or marginal cost/benefit, they probably have some reasonable approximation thereof.

What happens when the estimation of the marginal cost/benefit (e.g. game parameters) is not perfect? In order to answer this question, we add noise to the estimation of the relevant parameters (S, T, c, b and γ) in the models. We assume that noise is drawn from a normal distribution (f) with a mean μ and standard deviation σ (see details in the electronic supplementary material, figures S10–S18). Figure 3 shows the effect of noise on the decisions made by the gossiper (figure 3*a–d*) and the probability of making a mistake (i.e. being dishonest when the optimal decision is to be honest and *vice versa*; figure 3*e–h*). Noise creates a fuzzy edge between areas for honest and dishonest gossip, and the probability of making a mistake is the highest near the edge of these areas. Moving further away from the edge, this probability drops to zero. How fast this probability drops to zero depends on two factors (see further details in the electronic supplementary material): (i) the magnitude of noise, i.e. the higher the σ, the larger is the area affected by noise; (ii) the incentive structure of the game, i.e. the higher the marginal benefit to be honest (or the marginal cost to be dishonest), the less important is noise (see examples in the electronic supplementary material). Note that the quadrants that include the edges are ones with partial match between fitness interdependence and marginal cost/benefit, and the effect of noise is the most prominent in these quadrants (given reasonable values). This effect is in line with our previous conclusion: gossipers can make optimal decisions in the quadrants of perfect match or perfect mismatch
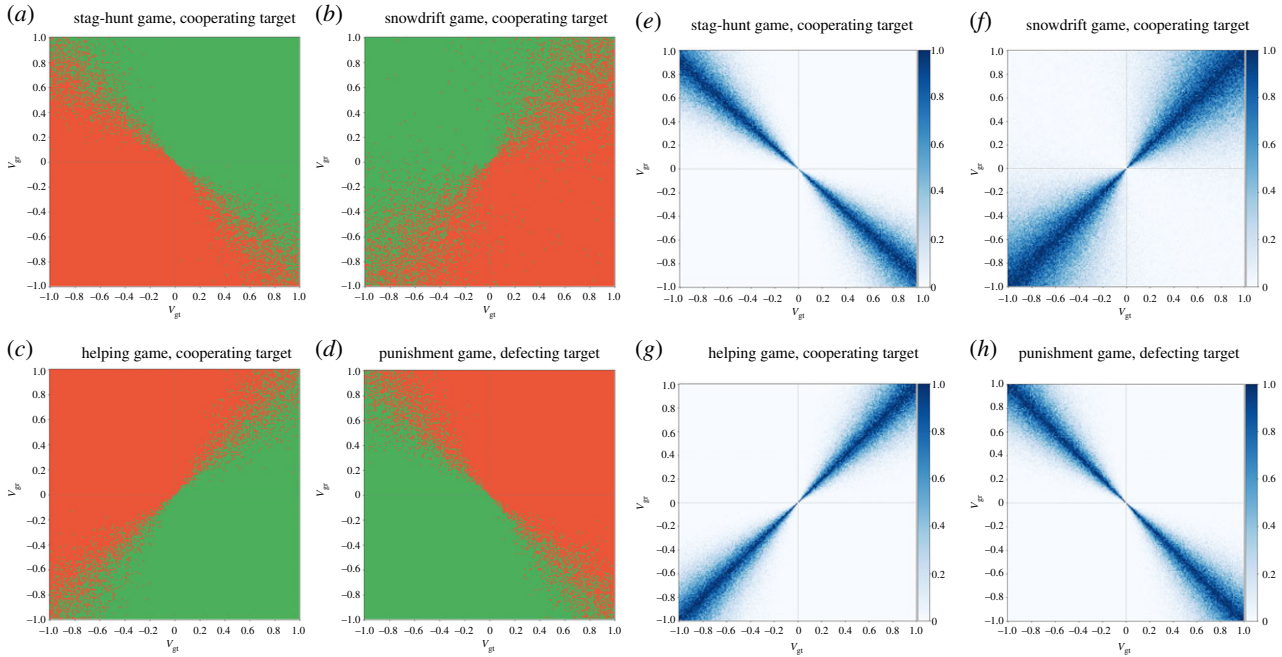
Figure 3. The predicted behaviour of the gossiper under noise (*a*–*d*; green area: honest gossip, red area: dishonest gossip) and the probability of making a mistake (*e*–*h*; $\mu = 0$; $\sigma = 0.25$; see further details in the electronic supplementary material). (*a,e*) Mutualism (*receiver/target*: +/+; stag-hunt game with a cooperating target; $T = 0$, $S = 0.1$); (*b,f*) antagonism (*receiver/target*: +/−, snowdrift game with a cooperating target; $T = 1.5$, $S = 0.5$); (*c,g*) antagonism (*receiver/target*: −/+, helping game with a cooperating target; $b = 1$; $c = 1$); (*d,h*) competition (*receiver/target*: −/−, punishment game with a defecting target; $c = 1$, $\gamma = 1$).

Table 2. Summary of predictions of the gossiper's behaviour depending on the game type and fitness interdependence. V_{gr}, the gossiper's fitness interdependence with the receiver; V_{gt}, the gossiper's fitness interdependence with the target; $E_r(H)$, the fitness outcome of honest gossip for the receiver; $E_t(H)$, the fitness outcome of honest gossip for the target. +/−, positive/negative.

game type	fitness interdependence		effect of honest gossip		gossiper's expected behaviour	
	V_{gr}	V_{gt}	$E_r(H)$	$E_t(H)$		
stag-hunt game[a]	+	+	+	+	honest	
	+	−	+	+	honest if inequality (3.3) holds	dishonest otherwise
	−	+	+	+	honest if inequality (3.3) holds	dishonest otherwise
	−	−	+	+	dishonest	
snowdrift game[a]	+	+	+	−	honest if inequality (3.5) holds	dishonest otherwise
	+	−	+	−	honest	
	−	+	+	−	dishonest	
	−	−	+	−	honest if inequality (3.5) holds	dishonest otherwise
helping game[a]	+	+	−	+	honest if inequality (3.7) holds	dishonest otherwise
	+	−	−	+	dishonest	
	−	+	−	+	honest	
	−	−	−	+	honest if inequality (3.7) holds	dishonest otherwise
punishment game[b]	+	+	−	−	dishonest	
	+	−	−	−	honest if inequality (3.10) holds	dishonest otherwise
	−	+	−	−	honest if inequality (3.10) holds	dishonest otherwise
	−	−	−	−	honest	

[a]Versus a cooperating target.
[b]Versus a defecting target.

even if they do not know the exact values of game parameters. Similar conclusions can be drawn when there is noise in estimating fitness interdependence (see electronic supplementary material, figure S19) or in estimating both game parameters and fitness interdependence (see electronic supplementary material, figure S20).

(b) Overview of related literature and implications for future research

To our knowledge, there are no extant models that have tested the optimal gossip strategies in situations varying in fitness interdependence. Our models provide a functional explanation for why people tend to share honest or dishonest information, which we believe will benefit future empirical research on the proximate causes of honest and dishonest gossip behaviour. Although we are not aware of any study that has directly tested the effect of fitness interdependence on honest or dishonest gossip behaviour, some studies do suggest negative or positive fitness interdependence between the gossiper and the target. For instance, gossiping about rivals is indicative of negative fitness interdependence, whereas gossiping about friends or loved ones is a strong indication of positive fitness interdependence [36]. Strong negative emotions (e.g. anger) also suggest negative fitness interdependence [64,65]. The few studies that could be evaluated based on our criteria show a pattern consistent with our predictions (see electronic supplementary material, table S1).

Negative fitness interdependence between the gossiper and the target (e.g. the gossiper and the target being rivals or competitors) makes the gossiper more likely to dishonestly describe a target's action as bad when it is good (e.g. describing a cooperative person as a free rider), and to honestly describe the target's action as bad when it is indeed bad (e.g. describing a free rider as they are) [79,80]. Similar observations were made when there was potential negative interdependence between the gossiper and the receiver (e.g. they were rivals): the gossiper was more likely to be dishonest if the effect of honest gossip was positive for the receiver [81]. By contrast, positive honest gossip (gossip that benefits the target) is more likely to be shared about targets with whom the gossiper has positive fitness interdependence (e.g. friends or lovers, [79]). In line with our predictions, negative emotions indicative of potential negative fitness interdependence are more likely to lead to honest gossip when honest gossip negatively affects the target [22].

Another potential indicator of fitness interdependence is group membership. A recent study [36] found that gossipers were more likely to share dishonest gossip with outgroup receivers (i.e. receivers with potentially negative fitness interdependence) when honest gossip positively affected the receivers (i.e. it would have increased the receivers' pay-off, see [36]). Conversely, gossipers were more likely to share honest gossip with ingroup receivers when honest gossip had positive consequences for the receivers [36].

All in all, our modelling results suggest that the optimal gossip strategy varies according to the marginal costs and benefits resulting from honest gossip and the gossiper's fitness interdependence with the other actors in the gossip triad. A case in point is that people adopt a dishonest strategy to share only cooperative reputation information about related others (e.g. children) to protect their outcomes, while they adopt honest strategies about unrelated others (e.g. sharing non-cooperative information).

One novelty of our model is that it can also be used as a baseline against which to evaluate broader proximate mechanisms, often taking the forms of general principles of human behaviour. Notably, many societies have developed strong norms, sometimes even formal rules, for how to deal with (non-public) information, such that honest strategies may sometimes be adopted regardless of the optimal strategies in a certain situation. An example is the norm or cultural wisdom 'honesty is the best policy' (Benjamin Franklin). We recommend and envision research programmes that provide critical tests of normative influences on human behaviour, which might (or might not) overrule the pressure of fitness interdependence. Future research can use insights from our models to test predictions about when gossip is likely to be honest or dishonest. While it may not be possible to estimate the exact values of fitness interdependence, proxies such as rivalry, competition, friendship, group membership, negative or positive emotions could be used to predict the valence (negative versus positive) of fitness interdependence [45].

4. Conclusion

A growing body of research has begun to highlight the importance of gossip honesty in promoting and sustaining cooperation. Yet, the field is still in the early stages of understanding the situational underpinnings of individuals' strategies to share honest or dishonest gossip. Here, we draw on recent theoretical developments on fitness interdependence to build formal models to predict when people are likely to share honest or dishonest gossip about a target's behavioural type (i.e. cooperating versus defecting type) depending on the levels of fitness interdependence between the gossiper and the other two parties in the gossip triad, as well as the type of games that the receiver is about to play with the target. We show that honesty is determined by the marginal cost/benefit resulting from honest or dishonest gossip. These findings are consistent with the results of earlier work on signalling games [41,82]. We also show that there is a simple 'matching rule' between the valence of fitness interdependence and the effect of honest gossip: gossipers should always be honest when there is a perfect match and they should be dishonest when there is a perfect mismatch. This simple rule allows gossipers to make an optimal decision even if there is noise in estimating fitness interdependence and the effect of honest gossip. Our models are guided by the scientific principle of parsimony, but we believe that future research can extend our models by relaxing some assumptions and include complexity that is part of interdependence in everyday life. In addition to noise, social network structures may also be important in guiding individuals' gossip behaviour. Further research can address these questions by exploring how cues of fitness interdependence between the gossiper and the target and between the gossiper and the receiver might independently or jointly affect the gossiper's strategies across various interaction contexts using laboratory experiments and scenario studies. More importantly, future theoretical work can extend our models by taking into account other factors, such as social network structures [83].

Endnote

[1]Although there seems to be no difference between our approach (fitness maximization) and classic notions of rationality (e.g. the expected utility theory), these two approaches are different. Expected utility theory prescribes a rational choice that maximizes the expected utility. Yet, humans often encounter environmental, cognitive and biological constraints that restrict information processing [74]. Despite these constraints, humans have been successful in decision-making, because the human mind is equipped with a collection of specialized information processing mechanisms shaped through natural selection to encourage behaviours that have adaptively maximized fitness in the ancestral environment [75,76]. More importantly, the design of these mechanisms is ecologically rational rather than classically rational [77]. For example, humans may have evolved specialized psychological mechanisms to infer their fitness interdependence with others from other situational cues (rivalry, group membership, etc.) [45,52].

References

1. Komorita SS, Parks CD. 1995 Interpersonal relations: mixed-motive interaction. *Annu. Rev. Psychol.* **46**, 183–207. (doi:10.1146/annurev.ps.46.020195.001151)

2. 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)

3. Rand DG, Nowak MA. 2013 Human cooperation. *Trends Cogn. Sci.* **17**, 413–425. (doi:10.1016/j.tics.2013.06.003)

4. Dores Cruz TD, Nieper AS, Testori M, Martinescu E, Beersma B. 2021 An integrative definition and framework to study gossip. *Group Org. Manag.* **46**, 252–285. (doi:10.1177/1059601121992887)

5. Dunbar RIM. 2004 Gossip in evolutionary perspective. *Rev. Gen. Psychol.* **8**, 100–110. (doi:10.1037/1089-2680.8.2.100)

6. Foster EK. 2004 Research on gossip: taxonomy, methods, and future directions. *Rev. Gen. Psychol.* **8**, 78–99. (doi:10.1037/1089-2680.8.2.78)

7. Semmann D, Krambeck HJ, Milinski M. 2005 Reputation is valuable within and outside one's own social group. *Behav. Ecol. Sociobiol.* **57**, 611–616. (doi:10.1007/s00265-004-0885-3)

8. Wittek R, Wielers R. 1998 Gossip in organizations. *Comput. Math. Org. Theory* **4**, 189–204. (doi:10.1023/A:1009636325582)

9. Wu J, Balliet D, Van Lange PAM. 2016 Reputation, gossip, and human cooperation. *Social Pers. Psychol. Compass* **10**, 350–364. (doi:10.1111/spc3.12255)

10. Diekmann A, Jann B, Przepiorka W, Wehrli S. 2014 Reputation formation and the evolution of cooperation in anonymous online markets. *Am. Sociol. Rev.* **79**, 65–85. (doi:10.1177/0003122413512316)

11. Fehr D, Sutter M. 2019 Gossip and the efficiency of interactions. *Games Econ. Behav.* **113**, 448–460. (doi:10.1016/j.geb.2018.10.003)

12. Jiao R, Przepiorka W, Buskens V. 2021 Reputation effects in peer-to-peer online markets: a meta-analysis. *Social Sci. Res.* **95**, 102522. (doi:10.1016/j.ssresearch.2020.102522)

13. Milinski M. 2016 Reputation, a universal currency for human social interactions. *Phil. Trans. R. Soc. B* **371**, 20150100. (doi:10.1098/rstb.2015.0100)

14. Piazza J, Bering JM. 2008 Concerns about reputation via gossip promote generous allocations in an economic game. *Evol. Hum. Behav.* **29**, 172–178. (doi:10.1016/j.evolhumbehav.2007.12.002)

15. Feinberg M, Cheng JT, Willer R. 2012 Gossip as an effective and low-cost form of punishment. *Behav. Brain Sci.* **35**, 25. (doi:10.1017/S0140525X11001233)

16. Romano A, Balliet D, Wu J. 2017 Unbounded indirect reciprocity: is reputation-based cooperation bounded by group membership? *J. Exp. Soc. Psychol.* **71**, 59–67. (doi:10.1016/j.jesp.2017.02.008)

17. Smith ER, Collins EC. 2009 Contextualizing person perception: distributed social cognition. *Psychol. Rev.* **116**, 343–364. (doi:10.1037/a0015072)

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

19. Giardini F. 2012 Deterrence and transmission as mechanisms ensuring reliability of gossip. *Cogn. Process.* **13**, 465–475. (doi:10.1007/s10339-011-0421-0)

20. Beersma B, Van Kleef GA. 2011 How the grapevine keeps you in line: gossip increases contributions to the group. *Social Psychol. Pers. Sci.* **2**, 642–649. (doi:10.1177/1948550611405073)

21. Ellingsen T, Johannesson M. 2008 Anticipated verbal feedback induces altruistic behavior. *Evol. Hum. Behav.* **29**, 100–105. (doi:10.1016/j.evolhumbehav.2007.11.001)

22. Feinberg M, Willer R, Stellar J, Keltner D. 2012 The virtues of gossip: reputational information sharing as prosocial behavior. *J. Pers. Social Psychol.* **102**, 1015–1030. (doi:10.1037/a0026650)

23. Feinberg M, Willer R, Schultz M. 2014 Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* **25**, 656–664. (doi:10.1177/0956797613510184)

24. Samu F, Számadó S, Takács K. 2020 Scarce and directly beneficial reputations support cooperation. *Scient. Rep.* **10**, 11486. (doi:10.1038/s41598-020-68123-x)

25. Sommerfeld RD, Krambeck HJ, Semmann D, Milinski M. 2007 Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl Acad. Sci. USA* **104**, 17 435–17 440. (doi:10.1073/pnas.0704598104)

26. Wu J, Balliet D, Van Lange PAM. 2015 When does gossip promote generosity? Indirect reciprocity under the shadow of the future. *Social Psychol. Pers. Sci.* **6**, 923–930. (doi:10.1177/1948550615595272)

27. Wu J, Balliet D, Van Lange PAM. 2016 Reputation management: why and how gossip enhances generosity. *Evol. Hum. Behav.* **37**, 193–201. (doi:10.1016/j.evolhumbehav.2015.11.001)

28. Barclay P. 2012 Harnessing the power of reputation: strengths and limits for promoting cooperative behaviors. *Evol. Psychol.* **10**, 868–883. (doi:10.1177/147470491201000509)

29. Dores Cruz TD *et al.* 2021 Gossip and reputation in everyday life. *Phil. Trans. R. Soc. B* **376**, 20200301. (doi:10.1098/rstb.2020.0301)

30. Hilbe C, Schmid L, Tkadlec J, Chatterjee K, Nowak MA. 2018 Indirect reciprocity with private, noisy, and incomplete information. *Proc. Natl Acad. Sci. USA* **115**, 12 241–12 246. (doi:10.1073/pnas.1810565115)

31. Paolucci M. 2000 False reputation in social control. *Adv. Complex Syst.* **03**, 39–51. (doi:10.1142/s0219525900000042)

32. Kelley HH, Stahelski AJ. 1970 Social interaction basis of cooperators' and competitors' beliefs about others. *J. Pers. Soc. Psychol.* **16**, 66–91. (doi:10.1037/h0029849)

33. Fonseca MA, Peters K. 2018 Will any gossip do? Gossip does not need to be perfectly accurate to promote trust. *Games Econ. Behav.* **107**, 253–281. (doi:10.1016/j.geb.2017.09.015)

34. Számadó S, Szalai F, Scheuring I. 2016 Deception undermines the stability of cooperation in games of indirect reciprocity. *PLoS ONE* **11**, e0147623. (doi:10.1371/journal.pone.0147623)

35. Giardini F, Vilone D. 2016 Evolution of gossip-based indirect reciprocity on a bipartite network. *Scient. Rep.* **6**, 37931. (doi:10.1038/srep37931)

36. Peters K, Fonseca MA. 2020 Truth, lies, and gossip. *Psychol. Sci.* **31**, 702–714. (doi:10.1177/0956797620916708)

37. Godfray HC. 1991 Signalling of need by offspring to their parents. *Nature* **352**, 328–330. (doi:10.1038/352328a0)

38. Grafen A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546. (doi:10.1016/S0022-5193(05)80088-8)

39. Maynard SJ, Harper D. 2003 *Animal signals*. Oxford, UK: Oxford University Press.

40. Maynard SJ, Harper DG. 1995 Animal signals: models and terminology. *J. Theor. Biol.* **177**, 305–311. (doi:10.1111/mec.14451)

41. Bergstrom CT, Számadó S, Lachmann M. 2002 Separating equilibria in continuous signalling games. *Phil. Trans. R. Soc. Lond. B* **357**, 1595–1606. (doi:10.1098/rstb.2002.1068)

42. Nöldeke G, Samuelson L. 1999 How costly is the honest signaling of need? *J. Theor. Biol.* **197**, 527–539. (doi:10.1006/jtbi.1998.0893)

43. Számadó S, Czégel D, Zachar I. 2019 One problem, too many solutions: how costly is honest signalling of need? *PLoS ONE* **14**, e0208443. (doi:10.1371/journal.pone.0208443)

44. Ohtsuki H, Iwasa Y, Nowak MA. 2009 Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature* **457**, 79–82. (doi:10.1038/nature07601)

45. Aktipis A *et al.* 2018 Understanding cooperation through fitness interdependence. *Nat. Hum. Behav.* **2**, 429–431. (doi:10.1038/s41562-018-0378-4)

46. Brown SL, Brown RM. 2006 Selective investment theory: recasting the functional significance of close relationships. *Psychol. Inq.* **17**, 1–29. (doi:10.1207/s15327965pli1701_01)

47. Cronk L, Steklis D, Steklis N, van den Akker OR, Aktipis A. 2019 Kin terms and fitness interdependence. *Evol. Hum. Behav.* **40**, 281–291. (doi:10.1016/j.evolhumbehav.2018.12.004)

48. Roberts G. 2005 Cooperation through interdependence. *Anim. Behav.* **70**, 901–908. (doi:10.1016/j.anbehav.2005.02.006)

49. Giardini F, Wittek RPM. 2019 Silence is golden. Six reasons inhibiting the spread of third-party gossip. *Front. Psychol.* **10**, 1120. (doi:10.3389/fpsyg.2019.01120)

50. Kelley HH, Holmes JG, Kerr NL, Reis HT, Rusbult CE, Van Lange PAM. 2003 *An atlas of interperonsal relations*. Cambridge, UK: Cambridge University Press.

51. Van Lange PAM, Rusbult CE. 2012 Interdependence theory. In *Handbook of theories of social psychology* (eds PAM Van Lange, AW Kruglanski, ET Higgins), pp. 251–272. Thousand Oaks, CA: Sage Publications.

52. Balliet D, Tybur JM, Van Lange PAM. 2017 Functional interdependence theory: an evolutionary account of social situations. *Pers. Social Psychol. Rev.* **21**, 361–388. (doi:10.1177/1088868316657965)

53. Tinbergen N. 1963 On aims and methods of ethology. *Z. Tierpsychol.* **20**, 410–433. (doi:10.1111/j.1439-0310.1963.tb01161.x)

54. Delton AW, Robertson TE. 2016 How the mind makes welfare tradeoffs: evolution, computation, and emotion. *Curr. Opin. Psychol.* **7**, 12–16. (doi:10.1016/j.copsyc.2015.06.006)

55. Yamagishi T, Jin N, Kiyonari T. 1999 Bounded generalized reciprocity: ingroup boasting and ingroup favoritism. *Adv. Group Process.* **16**, 161–197.

56. Yamagishi T, Mifune N. 2008 Does shared group membership promote altruism?: Fear, greed, and reputation. *Ration. Soc.* **20**, 5–30. (doi:10.1177/1043463107085442)

57. Balliet D, Wu J, De Dreu CKW. 2014 Ingroup favoritism in cooperation: a meta-analysis. *Psychol. Bull.* **140**, 1556–1581. (doi:10.1037/a0037737)

58. Bernhard BH, Fehr E, Fischbacher U. 2006 Group affiliation and altruistic norm enforcement. *Am. Econ. Rev.* **96**, 217–221. (doi:10.1257/000282806777212594)

59. Delton AW, Krasnow MM. 2017 The psychology of deterrence explains why group membership matters for third-party punishment. *Evol. Hum. Behav.* **38**, 734–743. (doi:10.1016/j.evolhumbehav.2017.07.003)

60. Jordan JJ, McAuliffe K, Warneken F. 2014 Development of in-group favoritism in children's third-party punishment of selfishness. *Proc. Natl Acad. Sci. USA* **111**, 12 710–12 715. (doi:10.1073/pnas.1402280111)

61. Martin JW, Young L, McAuliffe K. 2020 The impact of group membership on punishment versus partner choice. *PsyArXiv*. (doi:10.31234/osf.io/5qr32)

62. Barclay P. 2020 Reciprocity creates a stake in one's partner, or why you should cooperate even when anonymous. *Proc. R. Soc. B* **287**, 20200819. (doi:10.1098/rspb.2020.0819rspb20200819)

63. 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)

64. Gerpott FH, Balliet D, Columbus S, Molho C, de Vries RE. 2018 How do people think about interdependence? A multidimensional model of subjective outcome interdependence. *J. Pers. Social Psychol.* **115**, 716–742. (doi:10.1037/pspp0000166)

65. Pietroni D, Van Kleef GA, De Dreu CKW, Pagliaro S. 2008 Emotions as strategic information: effects of other's emotional expressions on fixed-pie perception, demands, and integrative behavior in negotiation. *J. Exp. Social Psychol.* **44**, 1444–1454. (doi:10.1016/j.jesp.2008.06.007)

66. McCullough ME, Kurzban R, Tabak BA. 2013 Cognitive systems for revenge and forgiveness. *Behav. Brain Sci.* **36**, 1–15. (doi:10.1017/S0140525X11002160)

67. Smith A, Pedersen EJ, Forster DE, McCullough ME, Lieberman D. 2017 Cooperation: the roles of interpersonal value and gratitude. *Evol. Hum. Behav.* **38**, 695–703. (doi:10.1016/j.evolhumbehav.2017.08.003)

68. van Doorn EA, Heerdink MW, van Kleef GA. 2012 Emotion and the construal of social situations: inferences of cooperation versus competition from expressions of anger, happiness, and disappointment. *Cogn. Emot.* **26**, 442–461. (doi:10.1080/02699931.2011.648174)

69. Van Lange PAM, Ouwerkerk JW, Tazelaar MJA. 2002 How to overcome the detrimental effects of noise in social interaction: the benefits of generosity. *J. Pers. Social Psychol.* **82**, 768–780. (doi:10.1037/0022-3514.82.5.768)

70. Fang C, Kimbrough SO, Pace S, Valluri A, Zheng Z. 2002 On adaptive emergence of trust behavior in the game of stag hunt. *Group Decis. Negot.* **11**, 449–467. (doi:10.1023/A:1020639132471)

71. Hauert C, Doebeli M. 2004 Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Science* **428**, 643–646. (doi:10.1038/nature02360)

72. Nowak MA, Sigmund K. 1998 Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577. (doi:10.1038/31225)

73. Fehr E, Fischbacher U. 2004 Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87. (doi:10.1016/S1090-5138(04)00005-4)

74. Todd PM, Gigerenzer G. 2007 Environments that make us smart: ecological rationality. *Curr. Dir. Psychol. Sci.* **16**, 167–171. (doi:10.1111/j.1467-8721.2007.00497.x)

75. Cosmides L, Tooby J. 2013 Evolutionary psychology: new perspectives on cognition and motivation. *Annu. Rev. Psychol.* **64**, 201–229. (doi:10.1146/annurev.psych.121208.131628)

76. Delton AW, Krasnow MM, Cosmides L, Tooby J. 2011 Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc. Natl Acad. Sci. USA* **108**, 13 335–13 340. (doi:10.1073/pnas.1102131108)

77. Cosmides L, Tooby J. 1994 Better than rational: evolutionary psychology and the invisible hand. *Am. Econ. Rev.* **84**, 327–332. (doi:10.2307/2117853)

78. Murnighan JK, Wang L. 2016 The social world as an experimental game. *Organ. Behav. Hum. Decis. Process.* **136**, 80–94. (doi:10.1016/j.obhdp.2016.02.003)

79. McAndrew FT, Bell EK, Garcia CM. 2007 Who do we tell and whom do we tell on? Gossip as a strategy for status enhancement. *J. Appl. Social Psychol.* **37**, 1562–1577. (doi:10.1111/j.1559-1816.2007.00227.x)

80. Wyckoff JP, Asao K, Buss DM. 2019 Gossip as an intrasexual competition strategy: predicting information sharing from potential mate versus competitor mating strategies. *Evol. Hum. Behav.* **40**, 96–104. (doi:10.1016/j.evolhumbehav.2018.08.006)

81. Fonseca MA, Peters K. 2021 Is it costly to deceive? People are adept at detecting gossipers' lies but may not reward honesty. *Phil. Trans. R. Soc. B* **376**, 20200304. (doi:10.1098/rstb.2020.0304)

82. Lachmann M, Számadó S, Bergstrom CT. 2001 Cost and conflict in animal signals and human language. *Proc. Natl Acad. Sci. USA* **98**, 13 189–13 194. (doi:10.1073/pnas.231216498)

83. Rauwolf P, Mitchell D, Bryson JJ. 2015 Value homophily benefits cooperation but motivates employing incorrect social information. *J. Theor. Biol.* **367**, 246–261. (doi:10.1016/j.jtbi.2014.11.023)