nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-018-0354-z

The signal-burying game can explain why we obscure positive traits and good deeds

Moshe Hoffman[1,3*], Christian Hilbe[2,3*] and Martin A. Nowak[1*]

People sometimes make their admirable deeds and accomplishments hard to spot, such as by giving anonymously or avoiding bragging. Such 'buried' signals are hard to reconcile with standard models of signalling or indirect reciprocity, which motivate costly pro-social behaviour by reputational gains. To explain these phenomena, we design a simple game theory model, which we call the signal-burying game. This game has the feature that senders can bury their signal by deliberately reducing the probability of the signal being observed. If the signal is observed, however, it is identified as having been buried. We show under which conditions buried signals can be maintained, using static equilibrium concepts and calculations of the evolutionary dynamics. We apply our analysis to shed light on a number of otherwise puzzling social phenomena, including modesty, anonymous donations, subtlety in art and fashion, and overeagerness.

Many donors give substantial amounts while purposely withholding their names, including 17 anonymous gifts in the United States of more than US$10 million in 2017[1]. Such anonymous donations are considered particularly virtuous by Maimonides and other religious and philosophical authorities[2]. However, this form of charitable giving is hard to reconcile with standard evolutionary accounts of pro-social behaviour[3–5]. If we give to gain reputational benefits[6,7], why would we ever wish to hide the fact that we gave? If others have an incentive to reward us for giving, why would they ever prefer us to hide our gifts? Similarly, we strive hard to accomplish greatness, which we do partly to attract partners[8,9]. Yet, we sometimes actively hide these accomplishments, and others consider it more commendable when we do so[10]. We see similar puzzles in fashion where people are willing to pay considerable amounts to receive a name brand item, only to make sure the brand is relatively hard to spot[11,12]. Likewise, an artist might put thought and effort into conveying an idea, but then ensure the idea is hard to decipher (for example, by not giving a title or informative description to the abstract painting or musical composition). Finally, when we are interested in someone as a partner, we often subdue our level of interest, and those who play 'hard to get' are seen as more attractive[13,14]. While it is clear that seeming too interested may signal desperation, it is unclear why and when subduing one's interest would be worth the potentially lost opportunity.

Our explanation is based on the intuition that making a positive signal harder to spot can serve as a signal in itself: burying a signal may indicate a lack of interest in those who might have been impressed by the signal but now are less liable to notice it; alternatively, burying may also signal confidence that receivers are liable to find out anyway. As we show below, it is precisely this information that buried signals convey, which is different from the information conveyed by simply choosing a more costly signal, as in classical signalling models[8,9]. We formalize the above intuition using a simple game theory model, which we call the signal-burying game. In so doing, we join a growing body of literature that attempts to explain puzzling aspects of human moral and social behaviours using evolutionary game theory[15–19]. Our model is also closely related to the large body of signalling literature that rationalizes seemingly counter-intuitive behaviours by carefully analysing which information these behaviours convey in a given context[20–25].

The signal-burying game is an asymmetric game between a sender and a receiver (see Fig. 1a). The sender can be one of three different types, referred to as high (h), medium (m) and low (l). The probability that a randomly chosen sender is of a given type is determined by the probability distribution $\mathbf{p}=(p_h, p_m, p_l)$, with $p_h+p_m+p_l=1$. Similarly, there are two different types of receiver, called strongly selective and weakly selective, or in short, strong (s) and weak (w). The probability of each receiver type is given by the probability distribution $\mathbf{q}=(q_s, q_w)$, with $q_s+q_w=1$. Players know their own type and the probability distributions $\mathbf{p}$ and $\mathbf{q}$, but they cannot directly observe the type of their co-player.

In the beginning of the signal-burying game, a sender and a receiver are randomly matched (with their types independently drawn from the respective distributions $\mathbf{p}$ and $\mathbf{q}$). Senders can then decide whether they wish to convey their type by sending a costly signal. Specifically, they can choose among three options. First, they can decide not to send a signal, and hence not to pay any cost. Second, they can send a clear signal at cost $c_i \geq 0$, with the cost depending on the sender's type $i \in \{l, m, h\}$. Clear signals are always observed by the receiver. Third, they can send a buried signal. Such senders pay the same cost c_i to obtain the signal (for example, a university degree), but in addition they make sure the signal is not directly observable. Instead, receivers observe a buried signal only with probability r_i, with $i \in \{l, m, h\}$. If the buried signal is observed, it is tagged as having been buried. Otherwise, if the signal goes unnoticed, the receiver cannot tell whether the sender has sent a buried signal or no signal at all. After the sender has made his decision, the receiver chooses whether or not to accept the sender as a partner, on the basis of the signal she observes. Payoffs for partnering depend on the players' types but are independent of the signal: senders receive a payoff a_{ij}, whereas receivers obtain b_{ij}, with $i \in \{l, m, h\}$ and $j \in \{s, w\}$. In particular, we allow different receivers to value the same sender type differently. Such a heterogeneity could arise, for example, if different receiver

[1]Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, MA, USA. [2]IST Austria, Klosterneuburg, Austria. [3]These authors contributed equally: Moshe Hoffman, Christian Hilbe.
*e-mail: moshehoffman@fas.harvard.edu; christian.hilbe@ist.ac.at; martin_nowak@fas.harvard.edu

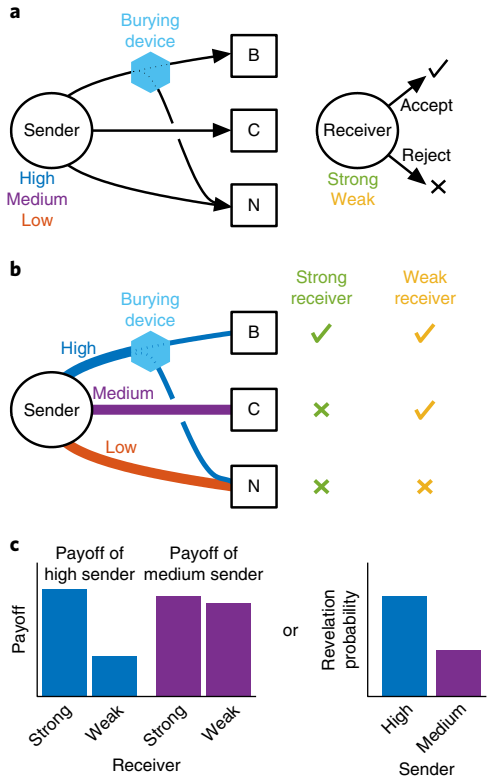

Fig. 1 | The signal-burying game. a, We consider a signalling game between a sender and a receiver. Senders are either of high, medium or low type, whereas receivers are either strongly selective ('strong') or weakly selective ('weak'). Players know their own type but not their co-player's type. To indicate their type, senders can pay a cost to send a signal. If they do, they can additionally choose whether they want to send a clear signal (C) or bury their signal (B). Buried signals become revealed and tagged as being buried with some probability; otherwise, it appears as if the sender has not sent a signal (N). On the basis of the signal they observe, receivers then choose whether or not to accept the sender for some economic interaction. Payoffs for partnering are such that senders always want to interact, whereas strong receivers get a positive payoff only from interacting with high senders, and weak receivers get a positive payoff only from interacting with high or medium senders. **b**, We define a burying equilibrium as an equilibrium in which high senders bury their signal, medium senders send a clear signal, and low senders send no signal. **c**, A burying equilibrium requires that high senders especially value interactions with strong receivers, or that they have a higher revelation probability than medium senders, see conditions (1)–(4) for details.

types have different outside options, or if they are looking for different kinds of partnership.

We assume that senders always wish to partner, $a_{ij} > 0$ for all i and j. Conversely, strong receivers get a positive payoff only from partnering with a high sender (b_{ls}, $b_{ms} < 0 < b_{hs}$), and weak receivers get a positive payoff only from partnering with a high or medium sender ($b_{lw} < 0 < b_{mw}$, b_{hw}). To keep the analysis simple, we additionally assume that the signalling cost c_i for low types is prohibitively high, such that they can always be assumed to send no signal. Moreover, we assume that neither type of receiver would be willing to partner with a mixture of high and low types of sender. This model can also encompass cases in which a sender's fixed trait may serve as a signal. In that case, we simply assume that the signalling cost is zero for individuals who have the trait, whereas it is prohibitively large for individuals who lack it.

In the Supplementary Information, we give a full description of the model, and we provide a complete equilibrium analysis. Furthermore, we consider equilibrium refinements, support our static results with extensive evolutionary simulations and discuss various model extensions. Below we summarize our key insights.

We first ask under which conditions our base model allows for equilibria such that high senders bury their signal, medium senders send a clear signal, and low senders send no signal. For any such equilibrium, it follows from our assumptions that strong receivers accept only those senders who choose to bury (and whose buried signal becomes revealed), whereas weak receivers additionally accept senders with a clear signal (see Fig. 1b and Supplementary Information). For there to be such a burying equilibrium, four conditions need to be met. First, high senders need to prefer sending a buried signal to a clear signal. In equilibrium, burying allows high senders to gain access to some strong receivers (who would have rejected the clear signal, but accept buried signals when they are revealed). However, burying also causes high senders to lose some weak receivers (who would have accepted the clear signal, but now may fail to notice the buried signal). High senders thus prefer to bury if

$$r_h q_s a_{hs} + r_h q_w a_{hw} \geq q_w a_{hw} \quad (1)$$

Conversely, medium senders need to prefer the clear signal over burying

$$r_m q_s a_{ms} + r_m q_w a_{mw} \leq q_w a_{mw} \quad (2)$$

Finally, both sender types need to be willing to pay the cost of the signal in the first place (note that the cost c_i is independent of whether or not the signal is buried)

$$r_h (q_s a_{hs} + q_w a_{hw}) \geq c_h \quad (3)$$

$$q_w a_{mw} \geq c_m \quad (4)$$

Whereas the last two conditions can also be found in standard signalling models[8] (where c_h/r_h needs to be interpreted as the higher cost of the more elaborate signal), the first two conditions represent the key insight of our model.

Combined, the first two conditions require that either high senders especially value partnering with strong receivers, or that it is more likely that their buried signals become revealed (Fig. 1c). Each of these two mechanisms can independently ensure that high senders find it worthwhile to bury: both of them decrease the high sender's opportunity cost $(1 - r_h) q_w a_{hw}$ of burying, while increasing their expected gains $r_h q_s a_{hs}$. The first case is tantamount to saying that high senders need to prefer those receivers who especially care about them. This seems to be a natural assumption if the success and longevity of interactions depends on the parties' shared values and goals[26]. This interpretation is also in line with observations that wealthy consumers low in need for status tend to associate with their own kind and that they pay a premium for quiet goods only they can recognize[11,12]. If this condition is what drives burying, then there is a natural interpretation: sending a costly signal allows one to separate oneself from those with inordinate costs, and burying one's signal allows one to separate oneself from those who benefit inordinately from weak receivers. We note that while this first mechanism is unique to our model, the second mechanism shares similarities to models of counter signalling and strategic disclosure[23–25] (we compare our model to this body of literature in more detail in the discussion).

In signalling games, classical equilibrium concepts often have the problem that they do not constrain the receivers' expectations about behaviours that do not occur in equilibrium. Therefore, we have explored which equilibria additionally satisfy the intuitive criterion[27].

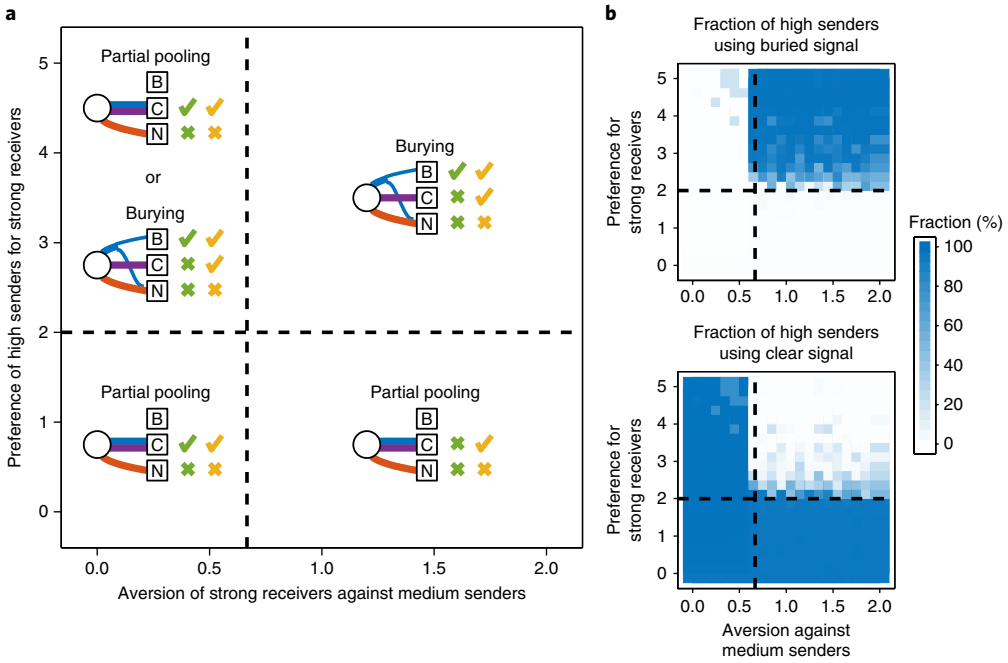

Fig. 2 | Evolutionary simulations are in line with the equilibrium conditions for burying. To explore when a burying equilibrium emerges, we have varied two key parameters, the relative preference of high senders for strong receivers (measured by a_{hs}/a_{hw} on the y axis) and the relative aversion of strong receivers against medium senders (measured by $-b_{ms}/b_{hs}$ on the x axis). **a**, Static equilibrium considerations suggest the existence of four parameter regions. First, if high senders show a low preference for strong receivers, and strong receivers have a low aversion against medium senders, the intuitive criterion[27] predicts a pooling equilibrium; both high and medium senders use a clear signal and both receivers accept this signal. Second, if high senders show a low preference for strong receivers, but strong receivers are strongly averse against medium senders, both sender types use a clear signal, which is accepted only by weak receivers. Third, if high senders highly prefer strong receivers, and strong receivers have a low aversion against medium senders, there are two possible equilibrium outcomes consistent with the intuitive criterion: the pooling equilibrium accepted by both receivers and the burying equilibrium. Fourth, if high senders have a high preference for strong receivers, and strong receivers have a high aversion against medium senders, only the burying equilibrium satisfies the intuitive criterion. **b**, To complement these static predictions, we have considered evolutionary simulations of a pairwise imitation process (for a sample trajectory, see Fig. 3). The simulations agree with the equilibrium predictions. In the only ambiguous case (the third parameter region), where static considerations allow for two equilibria, we observe that the pooling equilibrium is favoured (as it can be more easily reached from the initial population in which no one sends or accepts a signal).

The intuitive criterion imposes a further rationality requirement on receivers: if certain sender types cannot possibly gain from sending a given signal, receivers should assign zero probability that the signal was sent by one of those sender types. We prove in the Supplementary Information that, when the above conditions (1)–(4) are satisfied, the burying equilibrium is the only equilibrium that meets the intuitive criterion if, in addition, strong receivers avoid mixtures of medium and high senders (if $p_h b_{hs} + p_m b_{ms} < 0$). Otherwise there can be another equilibrium that satisfies the intuitive criterion, in which both medium and high senders send clear signals and are accepted by both receivers (Fig. 2a).

This equilibrium analysis is further supported by an evolutionary analysis. Evolutionary analyses are particularly relevant when strategies are not consciously chosen, but instead propagate via learning or evolutionary processes, as is arguably the case for our ideologies, tastes and emotions, including our artistic sense or moral intuitions related to anonymous giving[15,28]. We have therefore simulated the strategy dynamics under a pairwise imitation process[29]. We consider two finite populations of senders and receivers. The proportion of high, medium and low types within the sender population is given by the distribution $\mathbf{p}$, whereas the proportion of strong and weak types in the receiver population is given by $\mathbf{q}$. Initially, senders use no signal and receivers reject everyone. In each iterative step of the simulation, one player is randomly chosen from one of the two populations and given the chance to revise her strategy. When chosen for updating, with probability μ the player adopts a randomly chosen strategy out of the set of all available strategies (corresponding to a mutation in biological models). With the converse probability $1-\mu$, the player considers imitating the strategy of another player of the same type. Imitation events are biased towards strategies that yield higher payoffs (corresponding to selection). The exact revision protocol is provided in the Methods. When simulating this process, we find that in the parameter region in which no other equilibrium satisfies the intuitive criterion, populations quickly settle at the burying equilibrium (Fig. 2b; see also Fig. 3 for representative sample trajectories).

The unique incentive structure in the burying equilibrium enables specific information to be conveyed that is not conveyed by classical costly signals. To formalize this argument, we have extended our base model such that high senders can either distinguish themselves by burying, or by sending an alternative signal that is more costly than the clear signal (Fig. 4, for details see Supplementary Information). Separation through classical costly signalling requires either that, compared to medium senders, high senders have a lower cost of sending the alternative signal, or that they value strong receivers more. In contrast, separation through burying requires that high senders have a higher revelation probability, or that they value strong receivers more relative to the weak receivers (Fig. 4d). Thus, burying is especially useful when one wishes to convey that one's hidden qualities are likely to be revealed anyway, or that one does not particularly care about the weak receivers who may not spot these qualities.

Our base model can easily be adapted to cover more general scenarios. For example, in many applications, senders have some

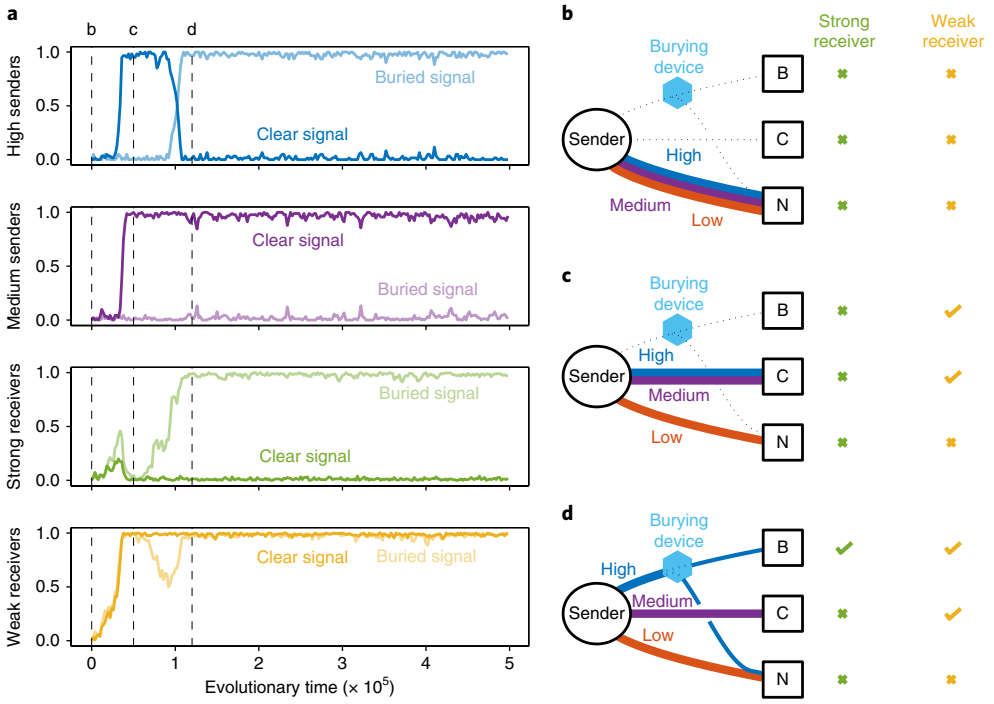

Fig. 3 | Evolutionary dynamics of buried signals. To explore how players learn to bury their signals, we show a representative simulation run for the parameter region in which only the burying equilibrium satisfies the intuitive criterion. **a**, The fraction of senders who use clear or buried signals (top two panels) and the fraction of receivers who accept the respective signal (bottom two panels). **b**–**d**, Stylized snapshots of the population at different points in time. **a**,**b**, Initially, no individual in the sender population sends a signal, and receivers reject everyone. **a**,**c**, Mutations and neutral drift make a substantial fraction of receivers accept clear signals. As a response, high and medium types learn to send a clear signal, which in turn leads strong receivers to reject individuals who send a clear signal. **a**,**d**, Again by mutation and neutral drift, both types of receiver learn to accept buried signals. High-type senders adapt and start using such signals. The resulting burying equilibrium is then stable, and no further change occurs. The protocol and the used parameter values of these simulations are described in the Methods.

discretion about how much they would like to bury their signals (how likely their buried signal will be revealed). To gain traction on this question, we have extended our model by allowing senders to choose between multiple burying devices with different revelation probabilities. If a buried signal becomes revealed, receivers do not learn only that the signal was buried, but also which burying device has been used. Figure 5a illustrates this model extension for the special case where the sender can choose among two revelation probabilities (see Supplementary Information for the general case with arbitrarily many feasible revelation probabilities). By repeating the previous equilibrium analysis for this extended model, we find that high senders tend to be modest, but not too modest (Fig. 5b,c): when given the chance, high senders learn to choose the signal with the highest revelation probability (subject to the constraint that the buried signal still allows them to differentiate themselves from medium senders).

With another model extension, we can capture that some receiver types may decipher buried signals more easily than others (as, for example, when it comes to grasp the true cost of a logo-less designer bag, see Supplementary Fig. 3). In addition, we also characterize the burying equilibrium for cases in which different sender types have a different likelihood to meet a given receiver type. Similarly, our base model can be extended to accommodate for more than three types of sender and two types of receiver (Supplementary Fig. 4). Finally, in the Supplementary Information, we discuss a model extension that suggests an alternative interpretation for burying. While in the model presented herein, burying serves the purpose of impressing certain types of receiver, we can also formulate our main results in terms of a model with only one receiver type. In that model, the sender's payoff depends nonlinearly on the receiver's ex post belief of the sender's type. While the base model suggests that burying occurs when high senders specifically care about strong receivers but not about weak ones, the alternative model suggests that they need to care about some distinctions but not about others: they might care a lot about being seen as high and not as medium, but they might not further bother about being taken for a medium or for a low type.

We can now apply our model to shed light on our motivating puzzles, starting with anonymous donations. While donors may prefer anonymity to avoid being harassed for further donations, this argument alone would not explain why anonymous donors are seen as more virtuous. However, donations are never fully anonymous. These donations are often revealed to the recipient, the inner circle of friends or fellow do-gooders (who correspond to the strongly selective types in our model). These few privy observers, in turn, do not only learn that the donor is generous (sends the costly signal); they are also likely to infer that the generosity was not motivated by immediate fame or the desire for recognition from the masses (that is, the donor does not care about the weak receivers).

An analogous conclusion holds for modesty. For example, a man who does not draw attention to his substantial wealth when he is first getting to know a potential suitor may signal that he does not need to impress her with this information, because he has many other suitors lined up in case she does not find out (that is, the opportunity cost from missing out on the weak receiver is low), because he is not interested in spending his time with a woman who is sufficiently impressed by wealth alone (that is, the weak receiver) or because he has so many positive attributes that he can afford for one to go unnoticed (which is maybe best reflected in terms of our model by assuming that r is large).

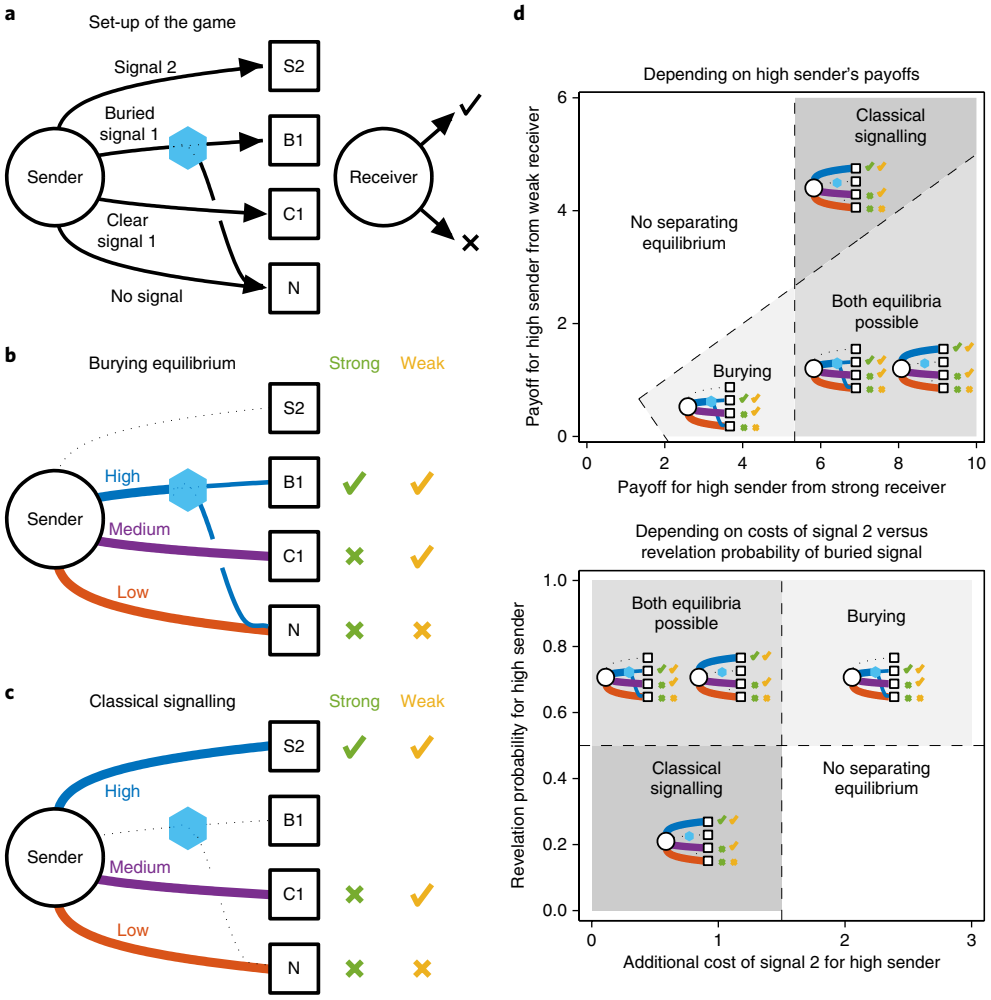

Fig. 4 | A comparison of burying with classical signalling. a, To show that burying conveys different information than standard costly signals, we allow senders either to use a signal S_1 that can be buried or sent clearly, or to send a more costly alternative signal S_2 that is automatically sent clearly. **b**, In this extended model, burying is still a possible equilibrium. **c**, However, this model also allows for a classical signalling equilibrium, in which the high type sends S_2, the medium type sends the clear signal, and the low type sends no signal. **d**, To analyse the conditions that allow for each of these two equilibrium configurations, we have first varied the payoff that high senders get from strong receivers (*x* axis) and from weak receivers (*y* axis). In the right region of the parameter space, high senders get a high absolute payoff from strong receivers; here, classical signalling is an equilibrium. In the region below the main diagonal, high senders get a high relative payoff from strong receivers (compared with the payoff they get from weak receivers); here, burying can occur in equilibrium. In the lower panel, we show that classical signalling is additionally favoured if high senders pay a low cost for the alternative signal S_2, whereas the burying equilibrium is favoured if buried signals of high senders are likely to become revealed.

When researchers brag about their most recent publication, this might show that they have been productive, but it also signals that they do not expect their article to be sufficiently important for their colleagues to take notice on their own. Moreover, we often infer that someone who brags is 'in it for the wrong reasons'. What do we mean by this? In light of our model, bragging may indicate that the primary motive for the accomplishment was merely to impress the weakly selective receivers. For instance, an academic who brags incessantly about her recent publications and awards may come off as being driven by attention and fame, instead of the promotion of science, which may make her less desirable as an editor, dean or collaborator (that is, to selective receivers, compared to the weakly selective public).

Bénabou and Tirole[21] provide an alternative interpretation for such an inference. In their model, players have three different motives to choose a certain action: the intrinsic value they attribute to the action, any extrinsic incentives for taking it (such as subsidies) and the action's reputational value. In their model, increasing the publicity of good deeds generally encourages social behaviour. However, when players differ in the relative weight attributed to these three motives, good actions also increasingly become suspected of being driven only by appearances. In contrast to our model, Bénabou and Tirole[21] treat the extent to which good actions are observed as an exogenous parameter, not as an option that players can use strategically. As a consequence, they do not address why senders would create a signal that is specifically targeted at some receiver types at the risk of losing others.

Others have attempted to explain modesty as a 'counter signal'[23]. Classic examples of counter signals include Mark Zuckerberg's and Steve Jobs's casual attire, who did not find it necessary to conform to the typical fashion habits of managers to impress people (which made them even more impressive for some observers). In a similar model of strategic disclosure, Harbaugh and To[24] present data suggesting that faculty in more prestigious universities tend to avoid mentioning their titles in their voicemails, and that they would actively substitute 'instructor' for 'professor' in course syllabi.

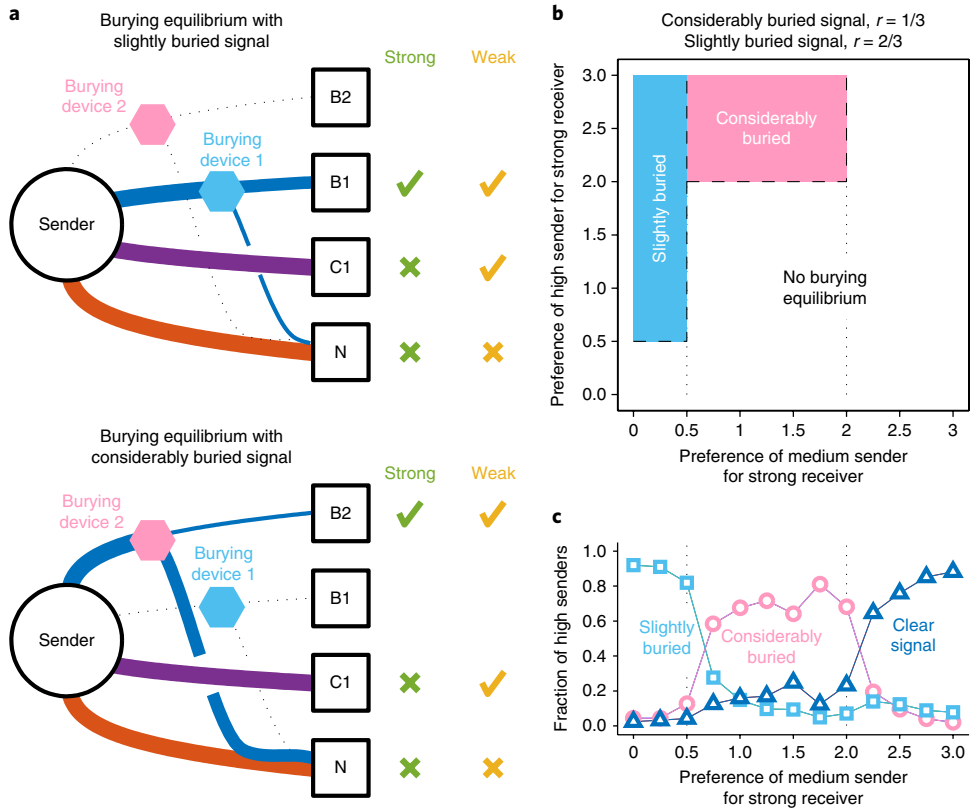

Fig. 5 | Burying equilibria in a model with multiple burying devices. In many applications, senders do not choose only whether or not to bury their signal, but also how much they would like to bury it. **a**, In the simplest case, we can model such a scenario by allowing senders to choose between two possible burying devices. We assume that the second device has a lower revelation probability. Hence, we say that signals are 'considerably buried' with device 2, and 'slightly buried' with device 1. If a buried signal becomes revealed, the receiver learns which burying device has been used. **b**, Whether a burying equilibrium exists, and which burying device will be used, again depends to which extent high and medium senders prefer interactions with strong receivers. As in the baseline model, high senders need to derive a high payoff from partnering with strong receivers, (that is, the value of a_{hs}/a_{hw} depicted on the y axis needs to be high). If medium senders do not particularly value interactions with strong receivers (that is, if the value of a_{ms}/a_{mw} on the x axis is low), it suffices for high senders to weakly bury their signal. As medium senders become more interested in strong receivers, high senders are forced to bury their signal considerably. **c**, Evolutionary simulations confirm these static predictions. We have fixed the high senders' preference for strong receivers (at $a_{hs}/a_{hw}=3$), and varied the medium senders' preference for strong receivers. Whenever slightly buried signals suffice to achieve separation, high senders learn to use them. Simulation results are averaged over 15 individual simulation runs, with each simulation run having 5×10^6 time periods.

In counter signalling models, higher types find it easier to distinguish themselves from lower types than medium types do (for example, because receivers obtain some noisy private information about the sender's type, in addition to the sender's publicly sent signal[23,24]). As a consequence, medium types may have more of an incentive to send the public signal. In our model, senders bury because they are confident their signal will be seen anyway, whereas with a counter signal, senders are confident their qualities shine through even when they do not send a signal at all. Counter signalling models are thus unable to explain why individuals would be willing to pay the cost of a signal without revealing it. In addition, counter signalling is usually not interpreted as a way to get access to certain receivers at the cost of losing others (although an appropriate modification of these models might yield such a result).

Turning now to our third application, what might artists be signalling by purposely leaving it open which messages are hidden in their work? Our model allows for several interpretations: the artist might be signalling that she does not care what her average contemporary (the weak receivers) thinks of her work; she might be sufficiently confident in her reputation as a good artist that art critics will scavenge to find the buried meaning (which may be approximated by having a larger r); or there may be so many buried insights that some are bound to be spotted even if they are not pointed out (which again may be approximated by assuming that r is large). In fashion, likewise, subtlety is often appreciated and actively sought out[11,12]. Wearing an expensive handbag with a large brand symbol on it may signal wealth, but also that you want all observers to notice that you are wealthy, and not just those who themselves are wealthy and sophisticated enough to know the subtle signals of expense.

Finally, turning to overeagerness, when we are interested in someone as a partner, we are often advised to 'play hard to get' or 'seem disinterested'[13,14]. One could interpret such behaviours as part of a negotiation, where people should understate their true interest to increase their bargaining power in the later relationship. However, such an intuition cannot explain why overeagerness is often seen as unbecoming; if it was only about bargaining, receivers should happily accept senders who fail to play down their interest. Alternatively, and more in line with our model, overeagerness might be taken as a cue that the sender's mate value is low, and that the sender is in need of this partnership. However, the question remains when it is worth being honest or deceptive about the fact that one actually benefits from this partnership; an equilibrium analysis is needed to know precisely when it is worth hiding one's interest, and how that behaviour is sustained and interpreted in equilibrium. By understating their interest, senders may indicate that they have many other potential suitors, or that they are confident that even

their subtle signals will suffice. But if everyone plays down their interest as adviced, how can it convey this information? It must be that there is an opportunity cost to playing down one's interest that depends on one's type, which our model helps elucidate: the cost is the lost relationships with weakly selective receivers. This cost is worth bearing when there is a high chance that one's subtle signals will be noted eventually, or when one is not particularly interested in weakly selective receivers anyway.

Methods

Static analysis. To explore under which condition individuals bury their signals, we have characterized all perfect Bayesian Nash equilibria (PBNE) of the signal-burying game (for all details, see Supplementary Information). The PBNE is the standard way game theorists solve signalling games. In a PBNE, the strategy of each type of each player is specified in such a way that no player can gain, in expected value, given her preferences and her information, and given that the other players act as specified. A PBNE can be interpreted as a necessary condition for a strategy profile to be sensible—if it is not a PBNE, then some type of player could benefit from deviating. Equivalently, if strategies are learned or evolved, a mutation or experimentation that leads her to behave differently would succeed and propagate.

Evolutionary simulations. We have modelled the evolution of strategies using a stochastic imitation process. There are two populations, a sender population of size N_S and a receiver population of size N_R. Each of these populations is divided into smaller subpopulations: for senders, there is a subpopulation of high-type senders of size $p_h N_S$, of medium senders of $p_m N_S$, and of low senders with size $p_l N_S$ (the proportions p_h, p_m and p_l are constant in time and satisfy $p_h + p_m + p_l = 1$). Similarly, the receiver population consists of strong receivers of size $q_s N_R$ and of weak receivers of size $q_w N_R$ (again with q_s and q_w being constant and $q_s + q_w = 1$). Within each population, individuals choose among the strategies described in the main text. Senders can thus either send no signal, send a clear signal or send a buried signal (yielding 3 possible strategies), whereas receivers need to decide whether or not to accept each signal type (yielding $2^3 = 8$ possible strategies). To calculate the players' payoffs in each time step, all individuals of the sender population are matched with all individuals of the receiver population, playing the game according to their predefined strategy.

We employ a simple pairwise comparison process[29] to model how the players' strategies change over time. In each time step, some individual i is chosen randomly from one of the two populations (with all individuals having the same probability to be chosen). This player is then given the chance to update her strategy. With probability $\mu > 0$ (the mutation rate), the player adopts a random strategy out of the set of available strategies. With probability $1 - \mu$, player i instead considers imitating a co-player. To this end, the player randomly chooses some other individual j from the same subpopulation. If player i's payoff in that period is π_i and if player j's payoff is π_j, we assume that i adopts j's strategy with probability $\rho = [1 + \exp(\beta(\pi_i - \pi_j))]^{-1}$. The parameter $\beta > 0$ corresponds to the strength of selection. If $\beta \to 0$, called the limit of weak selection, then $\rho \to 1/2$, regardless of the payoffs, and strategy updating essentially occurs at random. As β increases, strategy updating increasingly favours those strategies that lead to a higher payoff, $\pi_j > \pi_i$. Taken together, these two elementary updating processes of imitation and mutation give rise to an ergodic stochastic process.

We have analysed this stochastic process with computer simulations. These simulations were typically run for at least 2×10^6 time steps for each parameter combination. Unless stated otherwise, we have used population sizes $N_S = N_R = 300$, with $p_h = 0.2$, $p_m = 0.3$, $p_l = 0.5$ and $q_s = q_w = 0.5$. The costs of the signal were $c_h = c_m = 1$ and $c_l = 100$ (incorporating our assumption that signals are too costly for low senders). We considered the case of equal revelation probabilities for all senders, $r_h = r_m = r_l = 1/3$. The payoffs for partnering were

$$a_{hs} = 12, a_{hw} = 3, a_{ms} = 4, a_{mw} = 4, a_{ls} = 1, a_{lw} = 1,$$
$$b_{hs} = 6, b_{hw} = 6, b_{ms} = -10, b_{mw} = 4, b_{ls} = -10, b_{lw} = -10$$

These parameters have been chosen as they satisfy the restrictions for a burying equilibrium, as stated in conditions (1)–(4). Other parameters might affect the quantitative outcomes, but all simulations we have performed were in good qualitative agreement with the results predicted from our static equilibrium analysis, and they exhibit the same comparative statics (as an example, see Supplementary Fig. 1). Such an agreement between evolutionary results and equilibrium predictions is, in general, not guaranteed. In signalling games, as in any game in which different strategies may be indistinguishable along the equilibrium path, neutral drift can play an important role for the evolutionary dynamics. For reasonable population sizes and selection strengths, even non-equilibrium states can be reached rather frequently[30]. Moreover, analytical results for stochastic population dynamics can often be obtained only under rather restrictive assumptions, such as rare mutations, large populations or strong selection[31,32]. Our simulations can thus serve as a robustness check when these conditions are not satisfied. For the figures in the main text, we have used a strength of selection $\beta = 1$ and a mutation rate $\mu = 0.02$ throughout. However, our qualitative results are robust with respect to changes in these parameters, as shown in Supplementary Fig. 2.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The MATLAB algorithm that has been used to simulate the evolutionary dynamics of the baseline model (as shown in Figs. 2 and 3) is provided in the Supplementary Appendix. The simulations for the various model extensions (as discussed in Supplementary Section 3) require only minor modifications of this baseline algorithm. The corresponding MATLAB files are available from the corresponding authors upon request.

Data availability. The raw data generated by the MATLAB programs, which were used to generate the figures of our evolutionary simulations, are available as Supplementary data files.

Received: 6 February 2018; Accepted: 24 April 2018;
Published online: 28 May 2018

References

1. *Big Charitable Gifts: Where Donors Have Given $1 Million or More* (The Chronicle of Philanthropy, accessed 4 April 2018); http://philanthropy.com/factfile/gifts
2. Maimonides, M. *The Mishneh Torah* (Rambam/Maimonides and Moznaim Publishers, New York, 1998).
3. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
4. Rand, D. G. & Nowak, M. A. Human cooperation. *Trends Cogn. Sci.* **117**, 413–425 (2012).
5. Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. *Nat. Human Behav.* http://dx.doi.org/10.1038/s41562-018-0320-9 (2018); erratum http://dx.doi.org/10.1038/s41562-018-0342-3 (2018).
6. Sigmund, K. *The Calculus of Selfishness* (Princeton Univ. Press, Princeton, NJ, 2010).
7. Ohtsuki, H. & Iwasa, Y. How should we define goodness? Reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**, 107–120 (2004).
8. Spence, M. Job market signaling. *Q. J. Econ.* **87**, 355–374 (1973).
9. Grafen, A. Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546 (1990).
10. Banerjee, R. The development of an understanding of modesty. *Br. J. Dev. Psychol.* **18**, 499–517 (2000).
11. Berger, J. & Ward, M. Subtle signals of inconspicuous consumption. *J. Consum. Res.* **37**, 555–569 (2010).
12. Han, Y. J., Nunes, J. C. & Drèze, X. Signaling status with luxury goods: the role of brand prominence. *J. Mark.* **74**, 15–30 (2010).
13. Whitchurch, E., Wilson, T. D. & Gilbert, D. T. "He loves me, he loves me not…" Uncertainty can increase romantic attraction. *Psychol. Sci.* **22**, 172–175 (2010).
14. Bar-Anan, Y., Wilson, T. D. & Gilbert, D. T. The feeling of uncertainty intensifies affective reactions. *Emotion* **9**, 123–127 (2009).
15. Pinker, S. *How the Mind Works* (Norton and Company, New York, NY, 1997).
16. DeScioli, P. & Kurzban, R. Mysteries of morality. *Cognition* **112**, 281–299 (2009).
17. Johnson, D. D. P. & Fowler, J. H. The evolution of overconfidence. *Nature* **477**, 317–320 (2011).
18. Hoffman, M., Yoeli, E. & Nowak, M. A. Cooperate without looking: Why we care what people think and not just what they do. *Proc. Natl Acad. Sci. USA* **112**, 1727–1732 (2015).
19. Hoffman, M., Yoeli, E. & Navarrete, C. D. in *The Evolution of Morality* (eds Shackelford, T. K. & Hansen, R. D.) 289–316 (Springer, New York, NY, 2016).
20. Spence, M. Signaling in retrospect and the informational structure of markets. *Am. Econ. Rev.* **92**, 434–458 (2002).
21. Bénabou, R. & Tirole, J. Incentives and prosocial behavior. *Am. Econ. Rev.* **96**, 1652–1678 (2006).
22. Holmström, B. Managerial incentive problems: a dynamic perspective. *Rev. Econ. Stud.* **66**, 169–182 (1999).
23. Feltovich, N., Harbaugh, R. & To, T. To cool for school? Signalling and countersignalling. *RAND J. Econ.* **33**, 630–649 (2002).
24. Harbaugh, R., & To, T. *False Modesty: When Disclosing Good News Looks Bad* Working Paper http://dx.doi.org/10.2139/ssrn.777924 (2005).
25. Carbajal, J. C., Hall, J. & Li, H. *Inconspicuous Conspicuous Consumption* Working Paper no. 38 (Peruvian Economic Association, 2015).
26. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
27. Cho, I.-K. & Kreps, D. M. Signaling games and stable equilibria. *Q. J. Econ.* **102**, 179–221 (1987).

28. Henrich, J. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton Univ. Press, Princeton, NJ, 2016).
29. Traulsen, A. & Hauert, C. in *Reviews of Nonlinear Dynamics and Complexity* (ed. Schuster, H. G.) 25–61 (Wiley-VCH, Weinheim, 2009).
30. Veller, C. & Hayward, L. K. Finite-population evolution with rare mutations in asymmetric games. *J. Econ. Theory* **162**, 93–113 (2016).
31. Kandori, M., Mailath, G. J. & Rob, R. Learning, mutation, and long run equilibria in games. *Econometrica* **61**, 29–56 (1993).
32. Fudenberg, D., Nowak, M. A., Taylor, C. & Imhof, L. A. Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theor. Popul. Biol.* **70**, 352–363 (2006).

Acknowledgements

We thank B. Burum, J. Jordan and E. Yoeli for insightful discussions and constructive feedback, and A. Ferdowsian for his help with setting up the simulations. This work was supported by a grant from the John Templeton Foundation and by the Office of Naval Research Grant N00014-16-1-2914 (M.A.N.). C.H. acknowledges generous support from the ISTFELLOW programme and by the Schrödinger scholarship of the Austrian Science Fund (FWF) J3475. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

All authors contributed to all aspects of this research programme. If some authors contributed more to some aspects, they chose to bury this signal.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-018-0354-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.H. or C.H. or M.A.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.