Proceedings of the Royal Society B. 10.1098/rspb.2021.2266

Supplementary Information for

Cooperation as a signal of time preferences

Julien Lie-Panis and Jean-Baptiste André

Corresponding author: Julien Lie-Panis; Email: `jliep@protonmail.com`

Contents

A The Model . . . 1
A.1 Trust game . . . 1
A.2 Reputation formation . . . 2
A.3 Cooperating with strangers . . . 2
B Discussion of assumptions . . . 2
B.1 Individual discount rate . . . 2
B.2 Initial state . . . 2
B.3 Observation . . . 3
B.4 Binary reputation . . . 3
B.5 Private reputation . . . 3
B.6 Possible equilibria . . . 3
C Evolutionarily stable sets . . . 3
C.1 Methods . . . 4
C.2 Signaler optimal policy set when Choosers discriminate according to reputation . . . 4
C.3 Chooser use of information . . . 5
C.4 Cooperative equilibrium . . . 5
C.5 Other equilibria . . . 6

A. The Model.

A.1. Trust game. We consider a pairwise trust game with two roles, that of Chooser and that of Signaler. The game consists in an asymmetric prisoner's dilemma with two stages: in the first, the Chooser either rejects (R) the Signaler, in which case the interaction ends with both players earning null payoff; or she accepts (A) partnership with her. An accepted Signaler reaps reward r ($r > 0$). In the second stage, that Signaler can either defect (D), keeping r for herself; or prove worthy of the Chooser's trust, by paying cost c to cooperate (C). We assume: $0 < c < r$. When the Signaler defects, the Chooser is harmed, losing h ($h > 0$); and when the Signaler cooperates, she benefits, gaining b, the benefits of cooperation ($b > 0$).

Payoffs for the trust game are summarized below, in Table S1. A specific case to keep in mind is when $h = c$ and $b = r - c$. In such a case, payoffs are symmetric: everything is as if Choosers who play A pay cost c for the Signaler to gain b, thus losing c when the Signaler plays D; and gaining in turn b when the Signaler pays c to play C.

Table S1. Payoffs for the trust game.

		Signaler	
		Cooperate (C)	Defect (D)
Chooser	Accept (A)	$(b, r-c)$	$(-h, r)$
	Reject (R)	$(0, 0)$	$(0, 0)$

Individuals in an infinite population engage in cooperative encounters with strangers throughout their life, as represented by the trust game. Every T_S, an individual is paired up with two individuals she has never encountered before, with whom she plays the trust game; once in each role. The population is progressively renewed: individuals die after a given trust game with probability $\frac{T_S}{T_B}$, at which point they reproduce according to their accumulate payoffs. Individual expected life span is therefore equal to $T_S \times \sum_{i=0}^{\infty}(1 - \frac{T_S}{T_B})^i = T_B$. We assume social time step to be negligible in front of biological time step: $T_S \ll T_B$.

Individuals vary in a hidden quality: their temporal discount rate δ. Individuals are born with a certain δ, which is randomly selected in $]0, \infty[$, depending on a continuous probability distribution which characterizes the population. The probability that δ takes any single positive value is null. An individual's temporal discount rate remains constant throughout her life. At any point in time t, the total payoff an individual can expect to derive from cooperation is equal to her current payoffs in the trust games, plus $\frac{1}{1+\delta}$ times the payoffs she can expect at time $t + T_S$, plus $(\frac{1}{1+\delta})^2$ time the payoffs she can expect at time $t + 2T_S$, etc. (geometric discounting). Since $T_S \ll T_B$, this can be approximated using an infinite sum; individuals engage in a large number of cooperative interactions during their life span. To simplify future calculations, we measure time t in units of T_S from here on ($T_S = 1$).

A.2. Reputation formation. Signaler behavior is observed by the entire population with probability p, and error σ. We assume: $0 < p < 1$ and $0 < \sigma < \frac{1}{2}$. Information is private and binary: when a Signaler is observed, a fraction $1 - \sigma$ of the population makes the observation corresponding to her action (C or D), and a fraction σ wrongly observes the other action. We assume new information replaces old information and that individuals do not forget.

At any point in her life, a Signaler may be in one of three "reputational" *states*, depending on what action she was last observed undertaking (if any). Let $\mathcal{N}$ be the state Signalers are born in, and remain until they are observed for the first time; and let $\mathcal{C}$ ($\mathcal{D}$) be the state attained when a Signaler is last observed playing C (D).

Consider a Chooser-Signaler pair. When the Signaler is in state $\mathcal{N}$, the Chooser has no specific information on which to condition her play. We assume that in this case, there is an exogenous positive chance f that the Chooser accepts the Signaler (see section B.2).

The Chooser may otherwise face one of two informational *events* — indicating the Signaler has behaved in a trustworthy manner, by playing C after a previous partner accepted her (an event we note $\mathcal{T}$); or exploited the trust of that previous partner (event $\mathcal{E}$). Informational events faced by the Chooser correlate with the Signaler's state, but do not coincide given positive noise σ. When the latter is in state $\mathcal{C}$ ($\mathcal{D}$), the former has a $1 - \sigma$ (σ) chance of facing event $\mathcal{T}$, and otherwise faces event $\mathcal{E}$.

A.3. Cooperating with strangers. Mutual cooperation is net beneficial by assumption ($b > 0$ and $r > c$). Choosers stand to gain from accepting a Signaler who subsequently plays C, and to lose from partnering with a Signaler who subsequently plays D. Choosers may condition their play on specific information pertaining to the unfamiliar Signaler: event $\mathcal{T}$ or event $\mathcal{E}$.

When Chooser strategies differentiate between events $\mathcal{T}$ and $\mathcal{E}$, a Signaler's behavior in a given trust game may alter her future payoffs, by leading to a change in state (see section C.2). Signalers may condition their strategy on their temporal discount rate δ, as well as their state.

A fully specified strategy profile in this game therefore involves specifying, throughout an individual's life, whether to play A or R in the Chooser role under events $\mathcal{E}$ and $\mathcal{T}$; and whether to play C or D in the Signaler role, given own personal temporal discount rate δ and current reputational state. Note that we don't allow individuals to condition their play on arbitrary elements which are exogenous to the model (i.e. to change their strategy according to time t — see section B.6). For simplicity, we do not consider mixed strategies either (in which individuals behave probabilistically).

B. Discussion of assumptions.

B.1. Individual discount rate. Variation of time preferences inside a population can originate from two independent sources; evolutionary models show that it is adaptive to be more present-oriented (higher discount rate) when future rewards are more uncertain (1), and when present needs are more pressing (2). An individual's time preferences are thus susceptible to depend on a variety of factors, including mortality and accumulated biological and material capital: when individuals face higher probability of dying, future rewards are more uncertain; and when they have more capital, their present needs are less pressing. Since both mortality and capital tend to increase with age, age should therefore affect time preferences in a complex manner.

In our model, individuals are characterized by differing discount rates, which are exogenous and remain constant throughout their lifetime. We take between-individual differences as a given, and assume we do not need to consider within-individual temporal variability (because older age does not straightforwardly imply higher discounting). Individuals do not have access to any of the factors underlying their discount rate δ; in particular, mortality occurs without memory, and they cannot accumulate capital.

B.2. Initial state. We assumed that Choosers cooperate with Signalers for whom there is no specific information (state $\mathcal{N}$) with non-null probability f. This is a technical assumption, which allows reputation to be established: since a Signaler has a $f \times p$ chance of exiting state $\mathcal{N}$ at every social interaction (they have to be accepted for cooperation and to be observed), they spend on average $T_S \times \frac{1}{fp}$ in initial state $\mathcal{N}$ (an amount of time which is negligible with respect to their life span); after which they may alternate between states $\mathcal{C}$ and $\mathcal{D}$ depending on their strategy, and their partners face event $\mathcal{T}$ or $\mathcal{E}$.

$f > 0$ can be seen as a way to capture the existence of a cooperative past between certain players, outside of the interaction under study. While Choosers cannot gain any privileged information about an individual in the model (they always face new Signalers), they will have played other cooperative games with certain people beforehand and developed specific relations with trusted partners. In the most general sense, each individual Signaler i should face f_i depending on her past outside of the trust game — in any case, so long as $f_i > 0$, reputation for the trust game will be established, and the calculations conducted in section C.2 hold.

B.3. Observation. We rarely observe the cooperative or uncooperative behavior of strangers directly. We may however hear from third-party observers, or from individuals they talked to, etc. Our model can be seen to reflect rapid social transmission of information (gossip) (3). We would obtain the same results were we to assume that Signalers are observed by one or several witnesses with probability p, and that these witnesses are motivated to gossip about Signaler behavior to several acquaintances, who are also motivated to gossip to their acquaintances, etc. — leading all individuals to obtain new information by the next instantiation of the Trust game. σ can be seen to capture the noisiness of the entire social transmission process (even though the population is large, we still assume that $\sigma < \frac{1}{2}$).

Chooser decision in the absence of information arising from the trust game (which corresponds to a partner in state $\mathcal{N}$) is thus kept outside of the model. We only consider strategy given such information — given event $\mathcal{T}$ or given $\mathcal{E}$. Our simplified model allows us to focus on information reliability: section C.3 establishes the conditions under which reputation is reliable enough for cooperation to be established. In contrast, we are not concerned with information availability or consensus formation, which have been studied elsewhere (4–6).

B.4. Binary reputation. In the same spirit, we model trust and cooperation following an asymmetric game, where only Signalers may possess a reputation (and Choosers are only concerned with partner choice). This runs in contrast with most models of cooperation, which involve the framework of indirect reciprocity and a symmetric prisoner's dilemma (e.g. (4, 7, 8)).

As a result, reputation dynamics are particularly simple (for an exhaustive study in the symmetric case, see Ohtsuki and Iwasa (8)). There are only 2 possible states (vs. four in the symmetric case), and only $2^2 = 4$ possible Chooser strategies — which correspond to 4 ways of individually assigning reputation (Choosers need not play the same strategy in principle). In section C.3, we examine the conditions under which "discriminating according to reputation" (to last observation), i.e. playing A given $\mathcal{T}$ and R given $\mathcal{E}$ is advantageous to Choosers.

We can already note this is the only way of collectively assigning reputation which is conducive to cooperation. When Choosers all play one of the 3 other strategies, cooperation is reputation neutral or detrimental: a Signaler who is observed playing C is either as likely (when Choosers always accept or always reject) or less likely (when Choosers play A given $\mathcal{E}$ and R given $\mathcal{T}$) of being rewarded in the future, than a Signaler who is observed playing D. Since cooperation is costly, Signalers all benefit from playing D. We come back to this in Proposition P3, at the end of this document.

B.5. Private reputation. Reputation is also private in our model: every time a Signaler is observed (e.g. playing C), σ percent of Choosers end up with the conflicting piece of information (e.g. $\mathcal{E}$). We could have considered public reputation, whereby the entire population receives the same piece of information, which conflicts with actual Signaler behavior with probability σ.

Such a collective view of reputation would have led to certain technical simplifications: when reputation is public, Signaler state and Chooser information coincide perfectly. In particular, at a cooperative equilibrium where Choosers play play A given $\mathcal{T}$ and R given $\mathcal{E}$, Signalers who are assigned an exploitative reputation are simply never chosen again — which ends up being the case for all individuals under positive noise and infinite social interactions.

The main results should however remain the same. Since we can ignore state $\mathcal{D}$, calculations conducted in section C.2 are greatly simplified: a Signaler in good standing, can either play C and face probability $1 - p\sigma$ of being accepted again, or play D and face smaller probability $1 - p(1 - \sigma)$; as she will in future rounds (as long as she is not assigned exploitative reputation).

Hence she should face a trade-off between:

$$\Pi(C) = \sum_{t=0}^{\infty}(1 - p\sigma)^t \frac{r - c}{(1+\delta)^t} = \frac{1+\delta}{\delta + p\sigma}(r - c)$$

And:

$$\Pi(D) = \sum_{t=0}^{\infty}(1 - p(1-\sigma))^t \frac{r}{(1+\delta)^t} = \frac{1+\delta}{\delta + p(1-\sigma)} r$$

We find $\Pi(C) > \Pi(D) \iff \delta < \hat{\delta} = p[(1-\sigma)(\frac{r}{c} - 1) - \sigma\frac{r}{c}]$, the same result we obtain below when reputation is private (see section C.2 for a complete demonstration in that case).

B.6. Possible equilibria. Our assumptions limit possible Chooser strategies, and therefore limit the set of possible equilibria. Repeated games are generally characterized by a "Folk theorem" (9), whereby numerous Nash equilibria are possible. Since Choosers can only hold one bit of information at a time, they can only engage in one of four simple strategies (they can't engage in a complex "Grim-trigger" strategy, whereby they pass on all Signalers who don't engage in a specific sequence of actions).

We also prevent players from conditioning their play on arbitrary exogenous elements. In theory, Chooser behavior could vary with time: they could cooperate given trustworthy reputation only when the weather is sunny, and reject given exploitative reputation on all days but those marked by the death of a famous pop star. Rejecting such arbitrary scenarios strongly limits the set of feasible outcomes, but does not affect our determination of the main cooperative equilibrium (the calculations conducted in section C.2 remain valid when we allow Signalers to condition their play on time, and the calculations conducted in section C.3 remain valid when we allow Choosers to condition their play on time).

C. Evolutionarily stable sets.

C.1. Methods. In this section, we investigate possible stable endpoints of evolution, by identifying evolutionary stable sets (ES sets) of strategies (10). We reason in terms of strategy sets in order to ignore the effect of meaningless deviations, which do not affect players' expected payoffs at equilibrium. Because calculations are heavy, we start by determining Signaler optimal policy when Choosers discriminate according to reputation (as they do at the main cooperative equilibrium). We then outline the conditions under which Choosers stand to benefit from this discrimination.

We end the section by identifying two strategy sets of interest, and the conditions under which they define a (strict) Nash equilibrium set. Since sets of strict Nash equilibria are ES sets, and since an evolutionary stable strategy (ESS) must be Nash, we are able to deduce the conditions under which cooperation may be favoured by an evolutionary process.

C.2. Signaler optimal policy set when Choosers discriminate according to reputation. Let us consider a Signaler of discount rate δ. Since $T_B \gg T_S$, her payoffs from any point in time t can be approximated using the infinite sum $\sum_{t'=t}^{\infty}(\frac{1}{1+\delta})^{\frac{t'-t}{T_S}}\pi(x'_t, a'_t)$ — where $\pi(x'_t, a'_t)$ is her expected payoff for the Trust game conducted at time t', in future state x'_t, when choosing action $a'_t = a(x'_t)$ (as per her strategy).

Let us assume that Choosers discriminate according to reputation: when in the role of Chooser, all individuals in the population play A given $\mathcal{T}$ and R given $\mathcal{E}$. The probability that a partner will place her trust in our Signaler is a function of state: $F_N = f$ percent of Choosers cooperate with a Signaler in state $\mathcal{N}$ by hypothesis; and that fraction jumps to $F_C = 1 - \sigma$ for a Signaler in state $\mathcal{C}$, and to $F_D = \sigma$ for a Signaler in state $\mathcal{D}$.

In a given state $\mathcal{X}$, the Signaler can expect payoff $\pi(\mathcal{X}, C) = F_X \times (r - c)$ when she plays C, and $\pi(\mathcal{X}, D) = F_X \times r$ when she plays D. When she is not observed, her state remains the same in the next Trust game. When she is observed, with probability $F_X \times p$ (her partner first has to play A for observation to be possible), her state changes to $\mathcal{C}$ when she plays C and $\mathcal{D}$ when she plays D.

A Signaler's future state can therefore be described as a function of her current state and action, without reference to time t. Her optimal policy can be obtained following Bellmann's principle (11), by defining the value function V:

$$V(\mathcal{X}) = \max_{a\in\{C,D\}}\{\pi(\mathcal{X}, a) + \frac{1}{1+\delta}V(\mathcal{X}')\}$$

$$V(\mathcal{X}) = \max\{F_X(r-c) + \frac{1}{1+\delta}[F_X pV(\mathcal{C}) + (1 - F_X p)V(\mathcal{X})], \quad F_X r + \frac{1}{1+\delta}[F_X pV(\mathcal{D}) + (1 - F_X p)V(\mathcal{X})]\}$$

$$V(\mathcal{X}) = F_X \times \max\{(r-c) + \frac{pV(\mathcal{C})}{1+\delta},\ r + \frac{pV(\mathcal{D})}{1+\delta}\} + \frac{(1 - F_X p)V(\mathcal{X})}{1+\delta}$$

$$V(\mathcal{X}) = \frac{F_X(1+\delta)}{F_X p + \delta} \times \max\{(r-c) + \frac{pV(\mathcal{C})}{1+\delta},\ r + \frac{pV(\mathcal{D})}{1+\delta}\}$$

Since $F_X > 0$ for any state $\mathcal{X}$ in which the Signaler may find herself, her optimal policy in that state is determined by the comparison between two expressions which do not depend on $\mathcal{X}$. There are therefore two possibilities: either it pays more to play C now, in which case it will always pay more to play C (whatever the attained state) and $V(\mathcal{C})$ can be calculated assuming the Signaler always plays C and therefore remains in state $\mathcal{C}$:

$$V(\mathcal{C}) = \sum_{t'=t}^{\infty}(\frac{1}{1+\delta})^{\frac{t'-t}{T_S}} F_C(r-c) = \frac{1+\delta}{\delta}F_C(r-c)$$

Or it pays more to play D now, in which case the optimal policy is to always play D and:

$$V(\mathcal{D}) = \sum_{t'=t}^{\infty}(\frac{1}{1+\delta})^{\frac{t'-t}{T_S}} F_D r = \frac{1+\delta}{\delta}F_D r$$

Our Signaler's optimal policy is thus determined by the comparison:

$$(r-c) + \frac{pF_C(r-c)}{\delta} > r + \frac{pF_D r}{\delta}$$

$$\delta < \hat{\delta} = p[F_C(\frac{r}{c} - 1) - F_D\frac{r}{c}] = \frac{p \times \hat{\beta}}{c} \qquad [1]$$

Where $\hat{\beta} = F_C(r-c) - F_D r$ is the benefit of consistently playing C instead of D. Optimal policy is to always cooperate if $\delta < \hat{\delta}$, and to always defect if $\delta > \hat{\delta}$. Note that restricting Signalers to stationary strategies is unnecessary here: when Choosers always discriminate according to reputation, Signaler future state is only a function of current state and action, and Bellman's principle can be applied.

Note also that the above formulation defines an optimal set of Signaler strategies. Signalers whose discount rate is precisely equal to $\hat{\delta}$ are indifferent between playing C and D following the above equation. (Since discount rates are continuously distributed in the population, this happens with null probability.) In addition, Signalers whose discount rate is smaller (larger) than $\hat{\delta}$ and who always cooperate (defect) never reach reputational state $\mathcal{D}$ ($\mathcal{C}$) — and are therefore indifferent between playing C and D given that unattained state.

A more precise definition of this optimal set is therefore: (i) if $\delta < \hat{\delta}$, play C in states $\mathcal{N}$ and $\mathcal{C}$, and C or D in unattained state $\mathcal{D}$; (ii) if $\delta = \hat{\delta}$, play any strategy; (iii) if $\delta > \hat{\delta}$, play D in states $\mathcal{N}$ and $\mathcal{D}$, and C or D in unattained state $\mathcal{C}$. We refer to this set as "throughout one's life, cooperate if $\delta < \hat{\delta}$ and defect if $\delta > \hat{\delta}$" from here on. When Choosers discriminate according to reputation, any two strategies strategies in this set yield identical payoffs on average, and any strategy not in the set can be expected to yield a strictly inferior payoff.

C.3. Chooser use of information. Let us consider a Chooser at a certain point in time t, who possesses specific information $\omega \in \{\mathcal{T}, \mathcal{E}\}$ on her prospective Signaler partner. (Since all Signalers are born in state $\mathcal{N}$, Choosers never possess specific information at time $t = 0$, when the population's first trust game occurs.) If she accepts, she can expect payoff $P_t(C|\omega) \times b + P_t(D|\omega) \times (-h) = P_t(C|\omega)(b+h) - h$. In contrast, passing given ω yields certain null payoff. Accepting the Signaler given informational event ω is beneficial on average if and only if the above expression is positive, which is equivalent to:

$$P_t(C|\omega) > \frac{h}{b+h}$$

Accepting a Signaler given ω is beneficial on average if and only if that event is a sufficiently good predictor of the Signaler's cooperation at time t. By assumption, Choosers cannot take time into account. When Signaler strategy only depends on individual discount rate δ (as it does following the above optimal policy), that assumption proves to be unnecessary: in such a case, the fraction $P(C)$ of trustworthy Signalers is constant, as are predictive values $P_t(C|\mathcal{T})$ and $P_t(C|\mathcal{E})$ (see under).

Let us therefore assume that Signalers play a stationary strategy, and consider the constant fraction $\tau = P(C)$. Following Bayes' rule, $P_t(C|\mathcal{T})$ is equal to $\frac{P_t(\mathcal{T}|C)}{P_t(\mathcal{T})} \times \tau$. Both events $\mathcal{T}|C$ and $\mathcal{T}$ require that the Signaler has exited state $\mathcal{N}$ (i.e. has been observed at least once), a possibility whose probability does not depend on Signaler strategy, and which is positive, and simplifies in the above expression*. We deduce that our Chooser stands to gain from accepting given $\mathcal{T}$ if and only if:

$$\frac{1-\sigma}{\tau(1-\sigma)+(1-\tau)\sigma} \times \tau > \frac{h}{b+h}$$

$$\tau > \underline{\tau} = \frac{\sigma h}{\sigma h + (1-\sigma)b} \tag{2a}$$

Note that, since $\sigma > 0$, the denominator of the above expression is positive whatever the fraction τ of cooperative Signalers (when $t > 0$, $\mathcal{T}$ is a non-trivial event). The obtained lower bound $\underline{\tau}$ is also positive: accepting given trustworthy reputation can only be (strictly) worthwhile if some individuals actually cooperate. In addition it is an increasing function of σ: the larger the error, the larger the minimum fraction of trustworthy Signalers. We can simplify the above expression when errors tend towards 0 while remaining positive:

$$\tau > \underline{\tau} = \frac{h}{b}\sigma + o(\sigma), \qquad \sigma \to 0^+ \tag{2a'}$$

We proceed similarly for event $\mathcal{E}$. Following Bayes' rule, $P_t(C|\mathcal{E})$ is equal to $\frac{P_t(\mathcal{E}|C)}{P_t(\mathcal{E})} \times \tau$, which yields an expression which does not depend on time t, and whose denominator is positive. Choosers stand to lose from accepting given $\mathcal{E}$ if and only if:

$$\frac{\sigma}{\tau\sigma+(1-\tau)(1-\sigma)} \times \tau < \frac{h}{b+h}$$

$$\tau < \overline{\tau} = 1 - \frac{\sigma b}{\sigma b + (1-\sigma)h} \tag{2b}$$

The obtained upper bound is smaller than 1: rejecting given exploitative reputation can only be (strictly) worthwhile if some individuals actually defect. $\overline{\tau}$ is a decreasing function of σ: the larger the error, the larger the minimum fraction of exploitative Signalers. We can simplify the above expression when errors tend towards 0 while remaining positive:

$$\tau < \overline{\tau} = 1 - \frac{b}{h}\sigma + o(\sigma),\ \sigma \to 0^+ \tag{2b'}$$

When Signalers play a stationary strategy such as the above optimal policy, discriminating according to reputation is strictly beneficial if and only if the fraction of trustworthy signalers τ verifies: $0 < \underline{\tau} < P(C) < \overline{\tau} < 1$. (A rapid calculation shows that $\underline{\tau}$ is always smaller than $\overline{\tau}$ when $\sigma < \frac{1}{2}$).

C.4. Cooperative equilibrium.

Proposition 1 (P1) *The strategy set in which, throughout their life, (i) Choosers accept given $\mathcal{T}$ and reject given $\mathcal{E}$, and (ii) Signalers of discount rate δ cooperate if $\delta < \overset{*}{\delta} = p[(1-\sigma)(\frac{r}{c}-1) - \sigma\frac{r}{c}]$ and defect if $\delta > \overset{*}{\delta}$* (**Conditional Trust and Trustworthiness**)

a) is a set of strict Nash equilibria iff $\frac{\sigma h}{\sigma h+(1-\sigma)b} < P(\delta < \overset{*}{\delta}) < 1 - \frac{\sigma b}{\sigma b+(1-\sigma)h}$

*Any given Signaler of age t_S has a probability $1-(1-fp)^{t_S}$ of having exited state $\mathcal{N}$. As long as $t > 0$, Choosers cannot be certain they will encounter a newborn Signaler, since the population is progressively renewed from time $t = 1$. The probability they face $\omega \in \{\mathcal{T}, \mathcal{E}\}$ is therefore positive, and may depend on time t (particularly for the first generation of Choosers).

b) is Nash iff $\frac{\sigma h}{\sigma h+(1-\sigma)b} \leq P(\delta < \overset{*}{\delta}) \leq 1 - \frac{\sigma b}{\sigma b+(1-\sigma)h}$

The value of trustworthy reputation is then: $\overset{*}{\rho} = p \times \overset{*}{\beta} = p \times [(1-\sigma)(r-c) - \sigma r]$.

Proof of P1-a): Let us assume individuals all play according to the above strategy set. We prove that all deviations available to Signalers and Choosers are detrimental if and only if $\frac{\sigma h}{\sigma h+(1-\sigma)b} < P(\delta < \overset{*}{\delta}) < 1 - \frac{\sigma b}{\sigma b+(1-\sigma)h}$.

i. The proportion of trustworthy Signalers (who play strategy C at a given point in time) is stable and equal to $\tau = P(\delta < \overset{*}{\delta})$. Choosers are therefore in the situation described in section C.3: deviation to playing R given $\mathcal{T}$ is detrimental iff $P(\delta < \overset{*}{\delta}) > \underline{\tau} = \frac{\sigma h}{\sigma h+(1-\sigma)b}$, and deviation to playing A given $\mathcal{E}$ is detrimental iff $P(\delta < \overset{*}{\delta}) < \overline{\tau} = 1 - \frac{\sigma b}{\sigma b+(1-\sigma)h}$.

ii. Choosers discriminate according to reputation: Signalers are in the situation described in section C.2. Following the previous calculations, Signalers are playing their optimal policy: deviation to any strategy outside the set is detrimental.

iii. Note that deviations inside the optimal policy set are meaningless. Whether a Signaler plays C or D given an unattained state does not affect her payoffs, or that of other players (her Chooser partners). Since $P(\delta = \overset{*}{\delta}) = 0$, the possibility of a Signaler being born with quality precisely equal to $\overset{*}{\delta}$ can be neglected; and Signaler behavior given that improbable eventuality does not affect other players' payoffs either.

iv. Signalers cooperate when $\delta < \overset{*}{\delta} \iff \frac{1}{\delta} \times (p \times \overset{*}{\beta}) > c$. Everything is as if trustworthy Signalers are those who can afford to pay c, the costs of cooperation, in order to gain $\overset{*}{\beta}$ their entire future life, with probability p. Indeed, since $\sum_{t'=t}^{\infty} (\frac{1}{1+\delta})^{\frac{t'-t}{T_S}} = \frac{1}{\delta}$, an individual's social future may be represented by a single trust game whose payoffs are discounted with rate $\frac{1}{\delta}$. Since optimal Signaler policy does not depend on (attained) state, the value of establishing and maintaining a trustworthy reputation appear equal, and can be captured by $\overset{*}{\rho}$.

Proof of P1-b): following the calculations conducted in section C.3, Choosers stand to gain from deviation to playing $R|\mathcal{T}$ iff $P(\delta < \overset{*}{\delta}) < \frac{\sigma h}{\sigma h+(1-\sigma)b}$, and to playing $A|\mathcal{E}$ iff $P(\delta < \overset{*}{\delta}) > 1 - \frac{\sigma b}{\sigma b+(1-\sigma)h}$. There are no profitable deviations available to Signalers.

In our model, cooperation is therefore stabilized by variation of individual time preferences. Following proposition P1, cooperation is strict Nash and therefore ESS if the fraction of individuals whose discount rate is inferior to $\overset{*}{\delta}$ exceeds $\underline{\tau} > 0$ and the fraction of individuals whose discount rate is superior to $\overset{*}{\delta}$ exceeds $1 - \overline{\tau} > 0$. When in contrast $P(\overset{*}{\delta}) < \underline{\tau}$ or $P(\overset{*}{\delta}) > \overline{\tau}$, the above strategy set is not Nash, and therefore not ESS. (In either equality case, it is also not ESS, since rare mutants playing $R|\mathcal{T}$ or $A|\mathcal{E}$ when in the Chooser role perform as well the resident against the resident, and against themselves).

To take an extreme illustrative example, were all individuals to possess the same discount rate δ_0, $P(\delta < \overset{*}{\delta})$ would have to be equal to 0 or 1 — meaning that Choosers would stand to gain from acceptance given exploitative reputation or rejection given trustworthy reputation. Cooperation is also impossible if $\overset{*}{\delta} \leq 0 \iff \overset{*}{\rho} \leq 0$, in which case $P(\delta < \overset{*}{\delta})$ has to be null. A necessary condition for the above equilibrium is therefore:

$$\overset{*}{\rho} > 0 \iff \sigma < \frac{\frac{r}{c} - 1}{2\frac{r}{c} - 1} < \frac{1}{2} \qquad [2]$$

We deduce an upper bound on error σ, which is more restrictive than our initial assumption ($\sigma < \frac{1}{2}$). The above equation underscores the importance of Chooser discrimination according to reputation: were Choosers to treat Signalers identically whatever their reputation (e.g. always accept), cooperation would yield no relative benefit to trustworthy Signalers in the future. Mathematically, if we assume $F_C = F_D$ in the calculations performed in section C.2, we obtain $\hat{\beta} = 0$, and therefore $\hat{\rho} = 0$. (Even better, if Choosers accept given $\mathcal{E}$ and reject given $\mathcal{T}$, we obtain $\hat{\beta} < 0$). The two below propositions show that our model admits only one other Nash equilibrium, in which individuals trivially do not engage in cooperation.

C.5. Other equilibria.

Proposition 2 (P2) *The strategy set in which, throughout their lives, (i) Choosers reject given $\mathcal{E}$ and $\mathcal{T}$ and (ii) Signalers defect* (**Pooling with Rejection**) *is always a set of strict Nash equilibria.*
The value of trustworthy reputation in such a situation is: $\overset{*}{\rho}_0 = 0$.

Proof of P2-a): Let us assume that individuals all play according to the above strategy profile. We prove that there are no profitable deviations for either role.

i. Since Signalers always exploit their partners, acceptance given either $\mathcal{T}$ or $\mathcal{E}$ is always detrimental to Choosers (both events occur since $\sigma > 0$).

ii. Since Choosers do not use past behavior, a Signaler's future payoffs are unaffected by her current actions, precluding any profitable deviation to a more cooperative strategy outside of the set. (The set of Signaler strategies defined by (ii) includes playing either strategy C or D given unattained state $\mathcal{C}$.)

iii. A trustworthy Signaler would pay c to gain no future benefits: $\overset{*}{\rho}_0 = 0$.

Pooling with Rejection is always an ES set in our model. The evolution of cooperation from non-cooperation raises a bootstrapping problem (12). A rare mutant of initial frequency $\mu \ll 1$ who plays a strategy in the Conditional Trust and Trustworthiness set loses: (i) h with probability $1-\mu$, every time she faces event $\mathcal{T}$ when in the Chooser role (which occurs with probability $\sigma + O(\mu)$), and (ii) is eventually accepted in the Signaler role (after an average of $\frac{1}{f}$ iterations of the trust game), at which point everything is as if she pays c to gain $\mu \times \overset{*}{\rho}$ her whole life if her discount rate is smaller than $\overset{*}{\delta}$.

Going beyond the confines of our model, one may find reasons to (moderately) put into doubt the evolutionary stability of this trivial non-cooperative equilibrium. To begin, when error σ is sufficiently small (so that losing h with probability σ one's whole life is not overly costly), the above equilibrium may be invaded (by mutants playing Conditional Trust and Trustworthiness) due to stochastic effects (13).

Another possibility is to add an additional, unrealized informational event $\mathcal{G}$ (or, equivalently to assume that σ is in fact null in this situation). When a trust game is never played, there may be no reason to consider positive reputation for that game. We can imagine the following scenario (which, once again, cannot occur in our model as it stands): (i) first, Choosers may deviate to playing A given null event $\mathcal{G}$ (or null event $\mathcal{T}$ when $\sigma = 0$), without any impact on their payoff. This would not be a meaningless deviation: when a fraction μ of Choosers play $A|\mathcal{G}$, (ii) Signalers benefit from deviation to C given sufficiently small discount rate $\delta < \mu \times \overset{*}{\delta}$ (Mathematically, F_C and F_D are multiplied by μ in the demonstration of section C.2.) As long as $P(\delta < \overset{*}{\delta}) \geq \underline{\tau}$, (iii) this second advantageous Signaler deviation does not make the original Chooser deviation disadvantageous as long as the fraction of trustworthy Signalers exceeds $\underline{\tau}$. Hence, following this hypothetical scenario, Pooling with Rejection is subject to indirect invasion arising from neutral mutations when $P(\delta < \overset{*}{\delta} \geq \underline{\tau})$ (as per the definition of indirect invasion introduced by Jordan et al. (14)).

Proposition 3 (P3) *A Nash equilibrium of this game is either Conditional Trust and Trustworthiness or Pooling with Rejection.*

Proof of P3): Let us assume we are at a Nash equilibrium. We prove we are either at Conditional Trust and Trustworthiness or Pooling with Rejection.

i. Since Choosers are at equilibrium, Signalers' prospects depend solely on their state. We can introduce F_C (F_D), a Signaler's chance of facing a cooperative partner when in state $\mathcal{C}$ ($\mathcal{D}$), which remains constant. (F_C and F_D depend on the strategy or strategies played by the Chooser population.) Signaler optimal policy is thus obtained as in section C.2; optimal policy will be to always play C if one's discount rate δ is smaller than $\frac{\hat{\rho}}{c}$, and to always play D when δ is larger than $\frac{\hat{\rho}}{c}$, $\hat{\rho}$ being a function of F_C and F_D and the game parameters (one of these two conditions may be impossible).

ii. If $\hat{\rho} \leq 0$, optimal Signaler policy is to always play D, hence optimal Chooser strategy is to always reject prospective partners. We are thus in the Pooling with Rejection equilibrium.

iii. If $\hat{\rho} > 0$, optimal policy is to always play C for a fraction $f(C) = P(\delta < \frac{\hat{\rho}}{c})$ of Signalers, and to always defect for others. We show by contraposition that $\underline{\tau} \leq f(C) \leq \overline{\tau}$. Indeed, let us assume that $f(C) < \underline{\tau}$. In such a case, Choosers earn greater payoff when they reject given $\mathcal{T}$ than when they cooperate given that event: at a Nash equilibrium, Choosers would therefore play the former, and $\hat{\rho}$ would have to be negative, which is not the case (replace F_C with 0 in equation (1) to see this). An analogous reasoning can be made when that fraction exceeds $\overline{\tau}$: Choosers must therefore discriminate according to reputation, and we must be in the Conditional Trust and Trustworthiness equilibrium (with $\hat{\rho} = \overset{*}{\rho}$).

References

1. TW Fawcett, JM McNamara, AI Houston, When is it adaptive to be patient? A general framework for evaluating delayed rewards. *Behav. Process.* **89**, 128–136 (2012).
2. H Mell, N Baumard, JB André, Time is money. Waiting costs explain why selection favors steeper time discounting in deprived environments. *Evol. Hum. Behav.* **42**, 379–387 (2021).
3. F Giardini, D Vilone, Evolution of gossip-based indirect reciprocity on a bipartite network. *Sci. Reports* **6**, 37931 (2016) Number: 1 Publisher: Nature Publishing Group.
4. MA Nowak, K Sigmund, Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998) Number: 6685 Publisher: Nature Publishing Group.
5. M Hoffman, E Yoeli, CD Navarrete, Game Theory and Morality in *The Evolution of Morality*, Evolutionary Psychology, eds. TK Shackelford, RD Hansen. (Springer International Publishing, Cham), pp. 289–316 (2016).
6. B Burum, MA Nowak, M Hoffman, An evolutionary explanation for ineffective altruism. *Nat. Hum. Behav.* **4**, 1245–1257 (2020).

7. O Leimar, P Hammerstein, Evolution of cooperation through indirect reciprocity. *Proc. Royal Soc. London. Ser. B: Biol. Sci.* **268**, 745–753 (2001) Publisher: Royal Society.
8. H Ohtsuki, Y Iwasa, The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **239**, 435–444 (2006).
9. D Fudenberg, E Maskin, The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica* **54**, 533–554 (1986) Publisher: [Wiley, Econometric Society].
10. B Thomas, On evolutionarily stable sets. *J. Math. Biol.* **22**, 105–115 (1985).
11. R Bellman, *Dynamic programming.* (Dover Publications, Mineola, N.Y), Dover ed edition, (2003).
12. JB André, Contingency in the Evolutionary Emergence of Reciprocal Cooperation. *The Am. Nat.* **185**, 303–316 (2015).
13. MA Nowak, A Sasaki, C Taylor, D Fudenberg, Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).
14. JJ Jordan, M Hoffman, P Bloom, DG Rand, Third-party punishment as a costly signal of trustworthiness. *Nature* **530**, 473–476 (2016).