



Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehonline.org

Third-party monitoring and sanctions aid the evolution of language

Robert Boyd*, Sarah Mathew

School of Human Evolution and Social Change, Arizona State University

ARTICLE INFO

Article history:

Initial receipt 9 March 2015
 Final revision received 18 June 2015
 Available online xxxx

Keywords:

Language
 Third-party sanctions
 Reciprocity
 Honest signaling
 Reputation

ABSTRACT

The control of deception is an important problem in the evolution of all communication systems including human language. A number of authors have suggested that because humans interact repeatedly, reputation can control deception in human language. However, there has been little work on the theory of repeated signaling. This lacuna is important because unlike many others forms of defection, lies are not easily detected, and attempts to determine the truthfulness of signals can lead to false accusations of deception. Here we modify a standard model of animal signaling, the Sir Philip Sidney Game, to allow for repeated interactions between pairs of individuals. We show that unless it is easy to detect lies, communication is unlikely to be evolutionarily stable. However, third-party monitoring of pairwise interactions and sanctioning of dishonesty increases the range of conditions under which cheap talk can evolve, a finding that suggests that cooperation enforced by third-party monitoring and punishment may have predated the evolution of language.

© 2015 Elsevier Inc. All rights reserved.

The control of deception is an important problem in the evolution of communication systems. Maynard Smith and Harper (Maynard Smith & Harper, 2003) define a signal as a behavior that evolved because it affects the behavior of others who have evolved to respond to the signal. When the interest of the individual who generates the signal conflicts with the interest of the receiver, selection will favor deceptive signals that decrease the fitness of receivers unless such deceptive signals entail sufficient cost to deter defection. This problem is particularly acute in the case of human language because words can be recombined to generate an almost unlimited range of meanings; the possibilities for deception are endless (Lachmann, Szamado, & Bergstrom, 2001). The handicap principle is one well-known solution to this problem (Godfray, 1991; Grafen, 1990; Maynard Smith & Harper, 2003; Zahavi, 1975), although its empirical significance is debated (Számadó, 2011). Other solutions include indices, pooling and hybrid equilibria, and repeated interactions (Számadó, 2011; Zollman, 2013).

Of these, repeated interactions coupled with punishment of deception is the most plausible mechanism for maintaining honesty in human language (Lachmann et al., 2001; Scott-Philips, 2014). Contingent behavior of this kind is thought to play a key role in many forms of human cooperation (Axelrod & Hamilton, 1981; Nowak & Sigmund, 1998; Panchanathan & Boyd, 2004; Trivers, 1971; van Veelen, Garca, Rand, & Nowak, 2012). However, the evolution of honest communication differs from other forms of reciprocity. In most forms of reciprocity the failure to provide help is easily detected. The recipient does not receive aid and thus knows that her partner defected and can immediately stop helping the partner. Lies are different. The liar knows the truth, but

the listener can detect the lie only by making use of other evidence—the false signal itself is not sufficient.

There is little formal theory of repeated signaling among known individuals when lies are hard to detect. Silk, Kaldor, and Boyd (2000) study a model of repeated interaction among cercopithecine primates in which individuals signal their intent to peacefully approach with a quiet grunt. In this model, honest signaling is an equilibrium if receivers do not believe signals from particular partners once they have been deceived by that partner. Rich and Zollman (2015) investigate the repeated discrete Sir Philip Sidney game (Maynard Smith, 1991) where individual lies are never detected because the signaler's condition is known only to the signaler. They show that a strategy that keeps track of the rate at which an individual signals need, and stops transferring when that rate is too high can support honest signaling. These two models focus on two ends of an important continuum: in Silk *et al.* lies are always detected immediately, while in the model of Rich and Zollman individual lies are never detected, but dishonest behavior can be detected statistically over the long run. Catteeuw, Han, and Manderick (2014) study a one shot version of the Sir Philip Sidney game, but with punishment. Because it is a one shot model, and signals are costly, and signaler and receiver are related, this model does not capture the key features of the evolution of human language.

We examine a revised version of the Sir Philip Sidney game Maynard Smith (1991) in which interactions are repeated, and take place between unrelated individuals. Because only the speaker knows whether he lied or not, the listener has to use other evidence to decide whether her partner lied or not. The listener correctly detects a lie with some probability, and erroneously thinks a true statement is a lie with some probability. Thus the less trusting they are, the more likely they are to detect lies, but also the more likely they are to erroneously perceive a lie. This means there will be a high rate of perception errors (Nowak & Sigmund, 2005)—the speaker spoke the truth but the listener

* Corresponding author. 900 Cady Mall, Arizona State University, AZ, 85287. Tel.: +1 480 965 7671.

E-mail address: robert.t.boyd@gmail.com (R. Boyd).

erroneously believes it is a lie. While there has been some study of the evolution of direct reciprocity when rates of perception errors are very low, little is known about higher rates of perception errors, or the effect of such errors on indirect reciprocity.

1. The model

There is a large population of individuals, and individuals are paired with unrelated partners. One individual is a signaler and the other is a receiver. Signalers are “deserving” with probability p or “undeserving” with probability $1-p$. Receivers can perform a costly action that benefits the signaler. The payoffs are given in Table 1.

Receivers do not know the signaler’s state, but signalers can signal their state at zero cost. After a signal and a helpful act, receivers assess the truthfulness of the signal. The receiver correctly identifies false signals with probability e , and incorrectly identifies a truthful signal as false with probability w . Each time period the interaction continues with probability w . Individuals alternate roles, so if an individual is a receiver in a one time period, she is a signaler in the next time period and so on. The expected number of interactions for each individual in a given social role is $T = \frac{1}{1-w^2}$.

This game structure can represent a number of different situations. For example, mutual aid is common in human societies (Sugiyama, 2004). When an individual is sick or injured, she requests help from others. Later when others are sick or injured, she returns the aid. It also represents many situations in which the signaler has an obligation that may be avoided under some circumstances. In the classic John Hughes film, *Ferris Bueller’s Day Off*, Ferris is obliged to go to high school but feigns illness so that his mother will let him stay home. But instead Ferris ditches class and enjoys a day gadding about Chicago. More generally, the model applies to any circumstance in which the receiver is motivated to perform a costly act benefiting the signaler in one state of the world but not others, and only the signaler knows whether that state of the world is correct.

There are two strategies for each social role: Signalers can be:

- Honest (H) Signals when deserving, does not signal when undeserving.
- Dishonest (D) Always signals.

Receivers can:

- Respond (R) Help a signaler who is in good standing when own standing is good; otherwise do not help. An individual begins in good standing and falls out of good standing if (1) as a signaler she has been identified as giving a false signal, (2) she did not help the last time she received a signal of need from an actor in good standing, or (3) she helped after receiving a signal from someone in bad standing.
- Never Respond (N) Ignore the signal and never help.

Of course, many other strategies are possible. In particular, with these strategies once an individual falls into bad standing she can never get back into good standing. Dealing with this problem is an important, but difficult problem when lies and false accusations of lying cause different actors to have different beliefs about what has occurred. We focus on the strategies described above because they are the simplest that capture the essential features of the problem.

When signals are honest, individuals who respond to the signals of need can resist invasion by those who do not respond when

$$wT_H(a - wc) > c \tag{1}$$

Table 1 Fitness effects during one time period contingent on receiver behavior and signaler state.

	Do not help	Help
Receiver	1	1-c
Undeserving signaler	1-b	1
Deserving signaler	1-a	1

where $T_H = \frac{1}{1-w^2(1-p\varepsilon)}$ is the waiting time until an honest signaler is falsely accused. (See Supplementary Information for proofs, available on the journal’s website at www.ehbonline.org.) The right hand side of (1) is the cost of providing help during the first interaction on which individuals hear a signal. The left hand side is the long-term advantage of receiving help when in need minus the cost of helping. This is multiplied by the number of time periods in which honest players are in good standing. Thus, increasing w makes it more likely that responders are favored. Allowing $\varepsilon > 0$ will reduce the range of conditions under which providing help is favored, but this increment will be small if ε is small. We will assume that (1) is satisfied.

If responders are common, honest signalers can resist invasion by rare liars if

$$\frac{pa}{pa + (1-p)b} > \frac{T_L}{T_H} \tag{2}$$

where $T_L = \frac{1}{1-w^2(1-p\varepsilon-(1-p)e)}$ is the waiting time until a liar is exposed.

The left hand side of (2) is the ratio of the incremental fitness effect of signaling only when needed to the incremental fitness of always signaling. If the benefit of lying (b) is large or there are many opportunities for lying ($1-p$), then the ratio is small. When lies are not particularly beneficial or the opportunity to benefit from a lie is rare, the ratio will be close to one. The right hand side gives the ratio of the time until a liar falls into bad standing to the time until an honest signaler falls into bad standing. This ratio is always less than one. Thus the stability of honest signaling depends on the relative difficulty of detecting lies and the propensity to mistake honest signals as lies.

Because identifying lies requires other evidence, it should be thought of as a signal detection problem. After the donor gets the signal she decides whether the signal was truthful. This decision is based on cues that she observes. Ferris fakes a fever, and then joins his friends for a day on the town. If his mother had happened to see him later in the day, she would know he had lied. Of course, such cues can mislead. Ferris might really be home in bed, but his mother sees a Ferris-look-alike driving downtown and concludes that Ferris faked his symptoms. To avoid such mistakes, actors can impose a higher burden of proof before they conclude someone has lied. Ferris’s mother could try to make sure that she got close enough to be sure it was Ferris on the float, but this increased burden of proof will reduce the probability of detecting true lies.

We model this tradeoff using a signal detection model (McNicol, 2005). The cue is a normally distributed random variable, as shown in Fig. 1. Actors infer that a lie has occurred if the cue value is greater than d , and therefore $e = 1 - F(d|lie)$ and $\varepsilon = 1 - F(d|truth)$. As d is increased both e and ε decrease.

This model suggests that the stability of honest signaling is sensitive to the probability that lies are detected. In Fig. 2, we plot $\frac{T_L}{T_H}$ as a function of e for three values of M , the difficulty of distinguishing lies from true

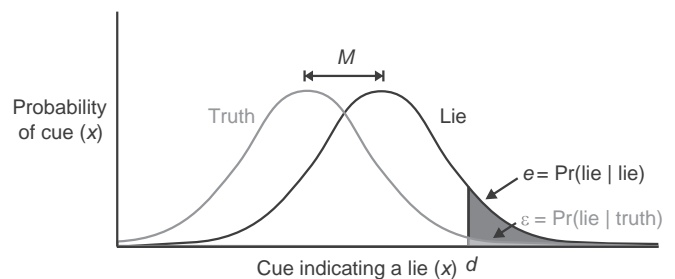


Fig. 1. The Gaussian probability densities of cue values conditioned on a truthful signal (grey) and a lie (black). Actors infer that a lie has occurred if the observed cue value is greater than d . The probability of detecting a lie, e , is always greater than the probability of falsely attributing a lie to an honest signal, ε . Increasing the value of d , decreases both e and ε but decreases the ratio e/ε , so errors become relatively less likely.

statements, and three values of T , the expected length of an interaction. The ratio of fitness effects in (2) necessary to sustain honest signaling increases as the probability of detecting lies decreases. This effect is much greater for longer interactions (larger values of w). Similarly, the greater the relative difficulty of distinguishing lies, the higher the ratio on the right hand side of (2) necessary to support honest signaling, and the magnitude of this effect is greater for longer interactions. Notice, however, that when lies are hard to detect (e is small) the benefit ratio necessary to support honest signaling is always near its maximum value, and therefore honest signaling is unlikely to persist except in special circumstances.

2. Third-party monitoring and punishment

Third party monitoring and punishment can increase the stability of honest communication. Suppose that a number of other individuals can observe the signal and the signaler's subsequent behavior, and that a norm exists that members of the community monitor signals and evaluate their honesty. Then these observations are pooled, and a consensus is reached. Subsequent decisions about whether an individual receives help are based on the community consensus. Individuals who don't participate are punished by withdrawal of future help, thus eliminating the second order free-rider problem (Panchanathan & Boyd, 2004). Such third party monitoring and punishment allows honest signaling to persist under a wider range of conditions for two reasons: First, increasing the number of individuals who attempt to detect deception increases the chance that liars will be exposed and decreases the chance that honest signalers will be falsely accused. Second, increasing the number of individuals who punish liars can make lying less attractive.

To illustrate how this might work, we extend the model to allow for third party monitoring and punishment. Each individual is engaged in repeated interactions with n unrelated community members, and each individual has information about whether the signal given in a particular interaction was truthful. In particular, we assume that each community member independently draws a cue from the same distribution as donor in the two-person model. Community members signal their belief about this evidence and reach a consensus about whether the signaler lied. Such signals could be costly, or honesty could be maintained by repeated interaction. To model this, we assume that cue values are averaged, and if the average exceeds the threshold, the signaler is judged a liar by the community, and no community member believes subsequent signals made by the signaler. Thus, the probability that a

lie is detected in any particular interaction is given by $e = 1 - F_n(d|lie)$ where $F_n(d|lie)$ is the cumulative normal with mean μ and variance $1/n$. Similarly, $\varepsilon = 1 - F_n(d|truth)$ where $F_n(d|truth)$ is the cumulative normal with mean $\mu + M$ and variance $1/n$. It is now easier to detect lies and avoid false accusations because the variances are smaller. Notice that the fact that information is aggregated is crucial. If, for example, only a single accusation was sufficient to put a person in bad standing, the result would be very different. Each individual is involved in n interactions during each period, and so the probability that a lie is detected in at least one of these interactions is $1 - (1 - e)^n = 1 - F_n(d|lie)^n$. Similarly, the probability that a truthful signal is mistakenly identified as a lie is $1 - (1 - \varepsilon)^n = 1 - F_n(d|truth)^n$. As is shown in Fig. 3 the combination of these two effects reduces the ratio $\frac{T_L}{T_H}$ on the right hand side of (2) and thereby increases the range of conditions under which honest signaling is evolutionarily stable.

3. Discussion

We have shown that it is difficult to maintain honest communication by direct reciprocity unless it is easy to detect lies, but it is much easier when third-parties monitor communication and punish dishonesty. What is particularly interesting is that honest signaling cannot evolve even under conditions that would allow other forms of reciprocity, such as mutual aid, to evolve. This is because lies are harder to detect than other forms of defection. Models of the evolution of reciprocity assume that defection is easily detected because the recipient does not receive the expected aid. In contrast, the liar knows the truth, but the listener requires other evidence to identify a lie—the lie itself is not enough. Third-party monitoring and punishment of dishonesty makes it easier to detect lies because evidence is pooled from multiple parties, which substantially increases the range of conditions under which honest signaling can be maintained.

These results suggest that third-party monitoring and punishment may have been a precondition for the evolution of human language. This may be one of the reasons why humans, but not other animals, evolved the capacity for language. Many primate species live in permanent social groups and recognize individuals, and reciprocity seems to play some role in social life. But there is little evidence for third-party monitoring and punishment in non-human animal societies. Of course, the evolution of human language entailed other important cognitive adaptations (Bickerton, 2007; Cangelosi & Parisi, 2002; Christiansen & Kirby, 2003; Corballis, 2003; Dunbar, 1998; Fitch, 2010; Fitch, Hauser,

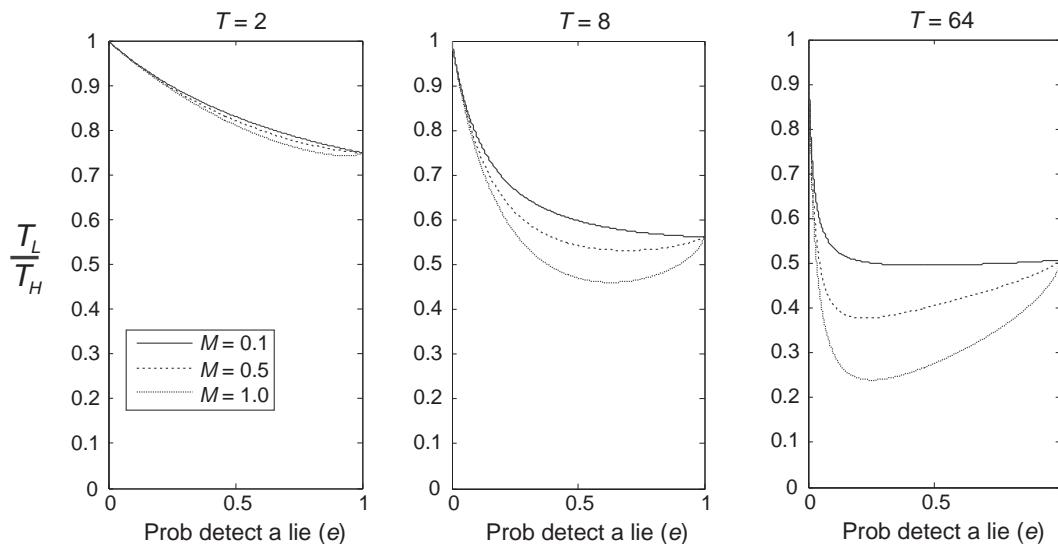


Fig. 2. The ratio of the waiting times until D/R and H/R players fall into bad standing, $\frac{T_L}{T_H}$, as a function of the probability of detecting a lie, e . The benefit ratio, $\frac{pa}{pa+(1-p)b}$, must be greater than $\frac{T_L}{T_H}$ for honest signaling to resist invasion by rare liars. These results suggest that (1) if e is small and therefore lies are hard to detect, selection will favor honest signaling only when it is highly beneficial compared to lying ($pa \approx pa + (1 - p)b$).

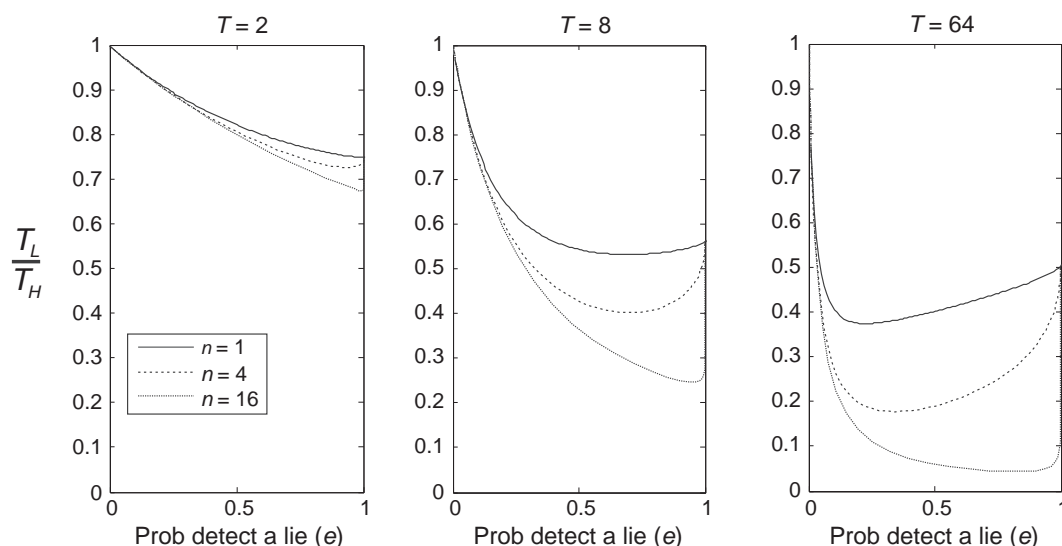


Fig. 3. The ratio of the waiting times until D/R and H/R players fall into bad standing, $\frac{T_L}{T_H}$, as a function of the probability of detecting a lie, e when honest communication is a norm enforced by mutual monitoring and sanctions in a community of n individuals. The ratio, $\frac{pa}{pa+(1-p)b}$, must be greater than $\frac{T_L}{T_H}$ for honest signaling to resist invasion by rare liars. These results indicate that normative sanctions increase the range of conditions under which honest signaling is evolutionarily stable.

& Chomsky, 2005; Hauser, Chomsky, & Fitch, 2002; Hurford, 2007; Jackendoff & Pinker, 2005; Nowak, Komarova, & Niyogi, 2001; Nowak & Krakauer, 1999; Pinker & Bloom, 1990; Számadó & Szathmáry, 2006). These cognitive requirements may also have precluded evolution of language in other animals.

These findings have some implications for understanding the origins of large-scale cooperation in humans. Humans are the only vertebrate species that cooperate in large groups of genetically unrelated individuals, and it is unclear why this capacity evolved in humans but not other species. Language has held an important place in accounting for the unusual level of human cooperation because it allows large numbers of people to coordinate joint enterprises (Smith, 2010). Furthermore, language allows people to create social contracts that would be impossible to specify without words (for e.g. “I will lend you my boat for a day, if you give me 10% of your day’s catch”) which enables people to establish myriad mutually beneficial social contracts (Cosmides, 1989; Pinker & Bloom, 1990). However, the current results suggest that some capacity for large-scale cooperation may have been present before the capacity for language evolved. Third-party monitoring coupled with sanctioning of defectors is a key ingredient of mechanisms underpinning large-scale cooperation such as warfare (Mathew & Boyd, 2011). The fact that language may depend on the same mechanisms runs counter to the idea that language was the critical trigger that facilitated the evolution of large-scale cooperation in humans. Instead our findings suggest that language may require a capacity for large-scale cooperation, and that some other factor must have catalyzed the highly cooperative trajectory of humans, such as culture (Boyd, Richerson, & Henrich, 2011; Mathew, Boyd, & van Veelen, 2013), intelligence (Pinker, 2010), theory of mind (Tomasello & Carpenter, 2007; van Schaik & Burkart, 2010), a dietary shift to large game (Kaplan, Hill, Lancaster, & Hurtado, 2000), altricial offspring (Hill, Barton, & Hurtado, 2009; Hrdy, 2009), or pair bonding (Chapais, 2008).

These results also speak to the relative role of indirect and direct reciprocity in sustaining human cooperation. Indirect reciprocity is widely acknowledged to be foundational to human cooperation: it is described as the basis of human moral systems (Alexander, 1987); models show that indirect reciprocity is evolutionarily stable (Leimar & Hammerstein, 2001; Nowak & Krakauer, 1999; Nowak & Sigmund, 1998; Ohtsuki, Iwasa, & Nowak, 2009; Panchanathan & Boyd, 2003); experiments have demonstrated that people are more willing to cooperate with those who cooperate with others (Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009; Ule, Schram, Riedl, & Cason, 2009). Yet, it is not obvious

why indirect reciprocity is needed for pairwise cooperation, especially because the conditions for direct reciprocity to evolve are more permissive. The results described here suggest an answer to this puzzle. If preventing dishonest signaling of need among a pair of individuals requires third party monitoring, then humans were already equipped with a psychology for paying attention to what others do even when it does not concern them directly. Such a pre-adaptation may alleviate the initial disadvantages that indirect reciprocity has over direct reciprocity, and may have enabled indirect reciprocity’s role as a cornerstone of human social exchange.

Acknowledgments

This research was made possible through the support of a grant (ID: 48952) from the John Templeton Foundation to the Institute of Human Origins at Arizona State University. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

References

- Alexander, R. D. (1987). *The biology of moral systems*. Aldine de Gruyter.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.
- Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, *117*, 510–526.
- Boyd, R., Richerson, P., & Henrich, J. (2011). Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology*, *65*, 431–444.
- Cangelosi, A., & Parisi, D. (Eds.). (2002). *Simulating the evolution of language*. Springer.
- Catteeuw, D., Han, The Anh, & Manderick, B. (2014). The evolution of honest signaling by social punishment. *GECCO '14. Proceedings of the 2014 conference on genetic and evolutionary computation* (pp. 153–160).
- Chapais, B. (2008). *Primeval kinship: How pair-bonding gave birth to human society*. Harvard University Press.
- Christiansen, M., & Kirby, S. (Eds.). (2003). *Language evolution*. Oxford University Press.
- Corballis, M. C. (2003). *From hand to mouth: The origins of language*. Princeton University Press.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.
- Dunbar, R. (1998). *Grooming, gossip and the evolution of language*. Harvard University Press.
- Fitch, T. (2010). *The evolution of language*. Cambridge University Press.
- Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*, 179–210 (discussion 211–25).
- Godfray, H. C. J. (1991). Signalling of need by offspring to their parents. *Nature*, *352*, 328–330.
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science (New York, N.Y.)*, 298, 1569–1579.
- Hill, K., Barton, M., & Hurtado, A. M. (2009). The emergence of human uniqueness: Characters underlying behavioral modernity. *Evolutionary Anthropology*, 18, 187–200.
- Hrdy, S. (2009). *Mothers and others: The evolutionary origins of mutual understanding*. Cambridge, MA: Harvard University Press.
- Hurford, J. R. (2007). *The origins of meaning: Language in the light of evolution*. Oxford University Press.
- Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (reply to Fitch, Hauser, and Chomsky). *Cognition*, 97, 211–225.
- Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, 9, 156–185.
- Lachmann, M., Szamado, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13189–13194.
- Leimar, O., & Hammerstein, P. (2001). The evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London B*, 268, 745–753.
- Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, 108, 1091–1094.
- Mathew, S., Boyd, R., & van Veelen, M. (2013). Human cooperation among kin and close associates may require enforcement of norms by third parties. In P. J. Richerson, & M. Christiansen (Eds.), *Cultural Evolution, Strüngmann Forum Report 12, series editor J. Lupp* (pp. 45–60). Cambridge, MA: MIT Press (chapter 3).
- Maynard Smith, J. (1991). Honest signalling: The Philip Sidney game. *Animal Behaviour*, 42, 1034–1035.
- Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford University Press.
- McNicol, D. (2005). *A primer of signal detection theory*. Psychology Press.
- Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science (New York, N.Y.)*, 291, 114–118.
- Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96, 8028–8033.
- Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393, 673–676.
- Nowak, M. A., & Sigmund, K. (2005). The evolution of indirect reciprocity. *Nature*, 437, 1291–1298.
- Ohtsuki, H., Iwasa, Y., & Nowak, M. A. (2009). Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature*, 457, 79–82.
- Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology*, 224, 115–126.
- Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432, 499–502.
- Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality and language. *Proceedings of the National Academy of Sciences*, 107, 8993–8999.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707–727.
- Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, 325, 1272–1275.
- Rich, P., & Zollman, K. J. S. (2015). *Honesty through repeated interaction*. Carnegie Mellon Tech Report #CMU-PHIL-192.
- Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Palgrave/Macmillan.
- Silk, J., Kaldor, E., & Boyd, R. (2000). Cheap talk when interests conflict. *Animal Behaviour*, 59, 423–432.
- Smith, E. A. (2010). Communication and collective action: Language and the evolution of human cooperation. *Evolution and Human Behavior*, 31, 231–245.
- Sugiyama, L. S. (2004). Illness, injury, and disability among Shiwiar forager-horticulturalists: Implications of health-risk buffering for the evolution of human life history. *American Journal of Physical Anthropology*, 123, 371–389.
- Számádó, S. (2011). The cost of honesty and the fallacy of the handicap principle. *Animal Behaviour*, 81, 3–10.
- Számádó, S., & Szathmáry, E. (2006). Selective scenarios for the emergence of natural language. *Trends in Ecology & Evolution*, 21, 555–561.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10, 121–125.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–57.
- Ule, A., Schram, A., Riedl, A., & Cason, T. (2009). Indirect punishment and generosity toward strangers. *Science*, 326, 1701–1704.
- van Schaik, C. P., & Burkart, J. M. (2010). *Mind the gap: Cooperative breeding and the evolution of our unique features*. Berlin & Heidelberg: Springer-Verlag, 477–496 (chapter Cooperativ).
- van Veelen, M., Garca, J., Rand, D. G., & Nowak, M. A. (2012). Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 9929–9934.
- Zahavi, A. (1975). Mate selection – A selection for a handicap. *Journal of Theoretical Biology*, 53, 205–214.
- Zollman, K. (2013). Finding alternatives to handicap theory. *Biological Theory*, 8, 127–132.