Journal of Theoretical Biology 395 (2016) 238–244

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Honesty through repeated interactions

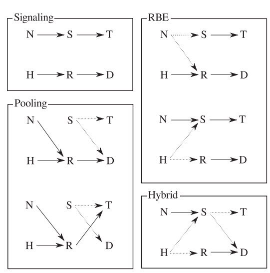
Patricia Rich, Kevin J.S. Zollman *

Department of Philosophy, Baker Hall 135, Carnegie Mellon University, Pittsburgh, PA 15217, USA

HIGHLIGHTS

- We develop a game theoretic model of signaling with repeated interactions.
- Honest signaling can be maintained without signal cost when interactions are repeated.
- This holds even when dishonesty cannot be directly observed.
- Novel tests are needed to determine if this effect accounts for honesty in the wild.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 May 2015
Received in revised form
29 January 2016
Accepted 1 February 2016
Available online 9 February 2016

Keywords:
Handicap theory
Costly signaling
Sir Philip Sidney game
Reputation

ABSTRACT

In the study of signaling, it is well known that the cost of deception is an essential element for stable honest signaling in nature. In this paper, we show how costs for deception can arise endogenously from repeated interactions between individuals. Utilizing the Sir Philip Sidney game as an illustrative case, we show that repeated interactions can sustain honesty with no observable signal costs, even when deception cannot be directly observed. We provide a number of potential experimental tests for this theory which distinguish it from the available alternatives.

1. Introduction

In many cases of signaling in nature, there is honest communication of information between two or more individuals. This occurs even when a first analysis suggests that deception would be fitness enhancing for one of the parties. In order for honest communication to be stable to invasion by dishonesty, there must be some countervailing force which reduces the fitness of deception.

It was originally suggested that the only way to make deception unprofitable would be for the communicating individual to spend a high cost to send the signal – to take on a "handicap" (Zahavi, 1975; Zahavi and Zahavi, 1997). It has since been shown that ubiquitous cost is not necessary to sustain honesty (Hurd, 1995; Számadó, 1999; Lachmann et al., 2001; Számadó, 2011b). Instead, the cost of deception is critical. Honesty can be free, so long as lying is costly. For example, Hurd (1997) showed that reliable communication of fighting ability is possible with very low observed cost so long as the penalty imposed on a weak individual who imitates a strong one is sufficiently high to deter deception – a plausible assumption in animal contests.

The cost of deception – sometimes called "marginal cost" – might not be observed in systems in equilibrium, and therefore could only be found by empirical investigation into how the system behaves outside of its natural state. While the theoretical correctness of this claim has been known for some time, there are relatively few biologically plausible methods for creating marginal cost provided in the literature (for examples, see Lachmann and Bergstrom, 1998; Bergstrom and Lachmann, 1998; Johnstone, 1999; Silk et al., 2000; Számadó, 2008, 2011a; Catteeuw et al.,

* Corresponding author.
E-mail address: kzollman@andrew.cmu.edu (K.J.S. Zollman).

http://dx.doi.org/10.1016/j.jtbi.2016.02.002

2014). This paucity of models makes empirical investigation into marginal cost difficult.

This paper explores the possibility of creating out-of-equilibrium cost, without creating observable costs, in the context of signaling among relatives. We do so by focusing on the possibility that repeated interactions might influence the evaluation of signals. It is plausible that children honestly signal their need to their parents because their signaling habits can be used to condition the parent's response. Signaling thus furnishes children with a kind of "reputation," and a child with a reputation for signaling too much will eventually be ignored by the parent and denied food in a way that harms the child. At the outset, we should be clear that the word "reputation" as we are using it does not suppose there is secondary communication like gossip (as used in Nowak and Sigmund, 1998; Ohtsuki and Iwasa, 2006). Instead we suppose that the parent learns how frequently the child signals and this is what we call the child's reputation. This limited kind of reputation, we argue, could replace direct cost as a mechanism for keeping signaling honest.

We show that this intuitive idea is indeed formally tenable, *even when dishonesty cannot be directly observed*. This restriction distinguishes our model from the few existing models of signaling reputation (Silk et al., 2000; Catteeuw et al., 2014), where dishonesty must be directly observed. In this paper, we augment Maynard Smith's (1991) Sir Philip Sidney game with *reputation-based* strategies and show that pairs of such strategies can constitute equilibria. Most importantly, these equilibria exist when the direct signal cost is too low to function as a traditional handicap. While we do not extend the analysis to other communicative games, these results should generalize to other communicative interactions that feature partial conflict of interest.

In Section 2 we review the Sir Philip Sidney game and present the various equilibria which exist for different parameter settings in this game. We then modify the game in Section 3 and present the central results of the paper. The paper concludes with a discussion of the idealizations and potential empirical tests of the model in Section 4.

2. Handicaps in the Sir Philip Sidney game

The handicap principle was initially formulated by the Zahavis (Zahavi, 1975, Zahavi and Zahavi, 1997) to explain the presence of honesty in situations where there is an incentive for deception. The basic insight was that if signaling carries a cost such that dishonesty is prohibitively expensive but honest signaling worthwhile, signalers do best by signaling honestly. And, if signaling conveys relevant information, receivers do best to make use of the accurate information carried by signals. Mathematical models showing that such a cost structure indeed makes honest signaling evolutionarily stable, e.g. those by Grafen (1990), Godfray (1991) and Maynard Smith (1991), were used to support the Zahavis' claim that the handicap principle is uniquely able to account for reliable signaling in nature.

The Zahavis' description of the principle, and the early models of it, suggest honest signaling in the wild should come with high, observable costs to the signalers. Maynard Smith's (1991) Sir Philip Sidney game provides a relatively tractable example of Grafen's (1990) model of the handicap principle. The game, shown in Fig. 1, involves two players. These players are typically imagined as a chick and a parent, although the model can be interpreted more generally. At the first node of the game some exogenous force (usually called "nature") determines whether the chick is in need of food or not in need of food. Following Bergstrom and Lachmann (1997) we will refer to these states as "needy" – which occurs with probability p – and "healthy" – which occurs with probability $1-p$.

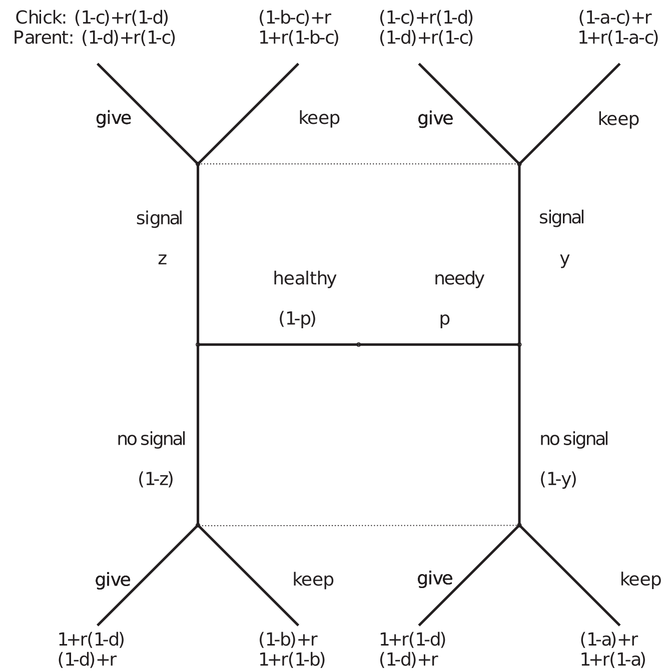


Fig. 1. A game tree illustrating the Sir Philip Sidney game with inclusive fitness. "Nature" determines whether the chick is healthy or needy (at the center node). The chick conditions its behavior on this choice and decides whether to send a costly signal or stay silent. The parent conditions its behavior on the signal, but not on the state of need of the chick. The parent chooses whether or not to donate a resource. Inclusive fitness for each individual is derived from adding r multiplied by the other's individual fitness.

At the second node the chick, conditioning on the decision by nature, either begs for food – signals to the parent – or not. Finally, in response to the signal (but not to the choice by nature) the parent either provides the chick food or keeps the food for itself. Several variations of this game have been proposed where there are more states of need, more signals, and differing amounts of transfer (Johnstone and Grafen, 1992; Bergstrom and Lachmann, 1997, 1998).

Each player's individual fitness is 1 minus the value of any penalty parameters given by the game's outcome: a chick who signals pays a signal cost $0 \le c < 1$; a parent who gives the chick food loses fitness $0 < d < 1$; a chick who does not receive food pays a fitness cost of $0 < a < 1$ if it is needy and $0 < b < 1$ if it is healthy (where $a > b$). The inclusive fitness of each player is determined by their individual fitnesses plus a fraction, r, of the fitness of the other individual. We presume that, at a minimum, the parent wishes to transfer the resource to the needy chick, i.e. $d < ra$.

Depending on the value of the parameters, this game features three types of equilibria, illustrated in Fig. 2. The equilibrium labeled "Signaling" is the state where the neediness of the chick is perfectly communicated to the parent. In addition to the signaling equilibrium, one of the "pooling" equilibria are present. These are equilibria where no information is communicated and the parent responds by either always transferring or never transferring the resource (which parental response is best depends on the underlying probability that the child is needy). The final equilibrium, the hybrid equilibrium, is not critical to our discussion here (for a discussion of the hybrid equilibrium, please see Huttegger and Zollman, 2010; Wagner, 2013; Zollman et al., 2013). We will return to this equilibrium in Section 4.

Holding the parameters a, b, d, and p fixed and allowing r and c to vary defines four regions of interest (Bergstrom and Lachmann, 1997; Huttegger and Zollman, 2010), which are pictured in Fig. 3. In all four areas, at least one of the pooling equilibria exits and is