ELSEVIER

Evolution and Human Behavior 30 (2009) 114–123

Evolution and Human Behavior

Do sincere apologies need to be costly? Test of a costly signaling model of apology☆

Yohsuke Ohtsubo*, Esuka Watanabe

Kobe University, Kobe, Japan

Initial receipt 4 April 2008; final revision received 16 September 2008

Abstract

The present study examined a costly signaling model of human apology. The model assumes that an unintentional transgressor is more motivated to restore the relationship with the victim than an intentional transgressor who depreciates the relationship. The model predicts the existence of a separating equilibrium, in which only sincere apologizers will pay a certain cost to restore the relationship, while dishonest apologizers will not. Accordingly, we hypothesized that the receivers of an apology would be sensitive to the cost involved in the apology. Experiments 1 and 2 were vignette experiments, in which participants imagined that they were victims of an interpersonal transgression and received either a costly or no-cost apology. The costliness of the apology was manipulated by the presence of an apology gift in Experiment 1, and by inconvenience voluntarily experienced by the transgressor to make an apology in Experiment 2. In both experiments, participants found the costly apologizer to be more sincere than the no-cost apologizer. Experiment 3 employed a modified dictator game, in which a fictitious partner behaved in an unfair manner and apologized to the participants. The apology cost was manipulated as a fee for sending the apology message. The results of Experiments 1 and 2 were replicated. In addition, when given a chance to send a complaint message to the unfair person, participants in the costly apology condition abstained from doing so. Implications of the study are discussed in relation to applications of the costly signaling theory to interpersonal behavior.

Keywords: Apology; Costly signaling theory; Handicap principle; Sincerity

1. Introduction

The communicative abilities of animals/humans and their evolutionary origins have engaged the interests of many scholars from divergent perspectives (Hauser, 1997). In the animal signal literature, one of the most important issues is the reliability or honesty of signals (Zahavi & Zahavi, 1997). Honesty of signals deserves both theoretical considerations and empirical investigations, as honest communication systems are vulnerable to deceptive signalers and thus unlikely to exist without some mechanisms to keep them honest. Zahavi's (1975) handicap principle (also known as the *costly signaling theory*) explains a mechanism whereby honesty of a signaling system becomes evolutionarily stable: high-quality individuals can credibly communicate their quality by voluntarily accepting some handicaps (or cost) that low-quality individuals cannot bear (see also Grafen, 1990).

Costly signaling theory has been successfully applied to some aspects of human behavior, such as altruistic behavior of hunters (e.g., Gurven, Allen-Arave, Hill, & Hurtado, 2000; Smith & Bliege Bird, 2000; Sosis, 2000), religious behavior (e.g., Irons, 2001; Sosis, 2003; Sosis & Alcorta, 2003) and human courtship behavior (e.g., Griskevicius et al., 2007; Miller, 2000). Although some authors have suggested that costly signaling theory is also applicable to everyday interpersonal communication (Andrews, 2001; Gangestad & Thornhill, 2007), social psychological studies have paid little attention to this theory. Nonetheless, social psychologists have been interested in, and in fact have investigated, deceptive behavior in everyday interpersonal communication (e.g., DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; Ekman, 1985). Having participants keep a

☆ This research was supported by the Japan Society for the Promotion of Science (grant no. 18730398).

* Corresponding author. Faculty of Letters, Department of Psychology, Kobe University, Kobe 657-8501, Japan. Tel.: +81 78 803 5519.

E-mail address: yohtsubo@lit.kobe-u.ac.jp (Y. Ohtsubo).

1090-5138/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2008.09.004

diary of their deceptions, for example, DePaulo et al. revealed that people tend not to tell exploitative lies, while they do often tell white lies and prosocial lies (e.g., lies told in order not to embarrass someone).

The DePaulo et al. finding is somewhat puzzling from the perspective of the evolution of communication because human language does not involve any cost that prevents exploitative deceptions — viz., telling a lie is no more costly than telling the truth (Lachmann, Számadó, & Bergstrom, 2001; Zahavi & Zahavi, 1997). It is known that low-cost signals (or cheap talk) can be honest if signalers and receivers share their interests to a substantial degree or if the signaling game has the coordination game-like incentive structure (Crawford & Sobel, 1982; Farrell & Rabin, 1996; Maynard Smith & Harper, 2003). This explanation might hold in interpersonal communication between relatives or close friends. However, people do not always share their interests. Being deceived and exploited in social exchange is considered to be a serious adaptive problem (Cosmides, 1989). Hence, it is naturally predicted that verbal communication needs to be accompanied by some costly signal when a deceptive incentive is large enough (i.e., when the honesty of low-cost signals is not warranted by shared interests; Zahavi & Zahavi, 1997). As a test case, in the present study, we shall apply the costly signaling theory to human apology.

1.1. Costly signaling model of apology

The apology-making context can be considered one of the situations where the reliability of signals becomes a crucial concern. If a victim unwittingly forgives a transgressor varnishing over his or her exploitative intent with verbal apology (e.g., saying "I am sorry"), he or she may be subject to similar transgressions again in the future. In this section, we shall first develop a formal model of costly apology. In the following sections, we shall briefly review previous studies related to the idea of costly apology and provide an overview of the present study.

As in the standard model of the signaling game, we assume two players: a message sender (S) and a message receiver (R). There is asymmetric information between S and R: S has private information to which R cannot have direct access. In the apology-making context, given some transgression committed by the apologizer, the private information is whether the transgression was committed accidentally. To make the model more concrete, suppose that cooperative S, denoted as S_C, has accidentally committed a personal transgression against R, and obtained some benefit, b_e from it. S_C sincerely feels sorry for it and says "I am sorry" to R for her wrongdoing (henceforth, we shall use feminine pronouns for S and masculine pronouns for R). Alternatively, the private information can be defined as whether S sincerely repented her capriciously committed transgression. In either case, it is expected that S's sincerity is correlated with the likelihood of her future cooperation.

Receiving S's apologetic statement, such as "I am sorry," R needs to be cautious because he expects to receive a similar apology not only from S_C but also from an exploitative S, denoted as S_E, who does not sincerely feel sorry. S_E might also say "I am sorry," expecting that she will be forgiven and can exploit R again. Although S personally knows which type (S_C or S_E) she actually is, R does not. In some instances, R may infer from circumstantial evidence that S committed the transgression accidentally (Malle & Knobe, 1997) and assumes that S is S_C. In other instances, R may suspect that S did it with exploitative intent and erroneously assume S, who is in fact S_C, being S_E. In the latter instances, because merely saying "I am sorry" will not work, S_C and R need some costly signaling system that prevents S_E from producing the deceptive signal.

Assume that both S_C and R will gain the benefit of b_c from one round of cooperative interaction, and S will gain the benefit of b_e ($>b_c$) from one round of exploitation regardless of whether it was intentionally committed or not. Their interaction will be repeated with the probability of w if R decides to continue the relationship with his current partner. However, he will terminate the relationship when he thinks that the likelihood of his current partner being S_E is too high to justify the continuation of the relationship. Not to lose the potentially beneficial relationship with R, S_C has to somehow prove her true identity to R. If she successfully convinces him that she is S_C, her net benefit from interactions with R is $b_e+b_c\times w/(1-w)$. Here, $w/(1-w)$ is the expected number of future interactions.

Suppose that S_C pays the cost of a ($\geq b_e$) in making her apology. By definition, S_E is not willing to pay any cost greater than the benefit from the one-shot exploitation, b_e. On the other hand, S_C has an incentive to pay it if she is better off by paying the cost a to assure the future benefit of $b_c w/(1-w)$ than keeping the benefit from one-shot exploitation, i.e., $b_e \leq b_e - a + b_c w/(1-w)$. The model can be summarized in the following inequality:

$$b_e \leq a \leq b_c w(1-w).$$

When the above inequality holds, S_C will make the costly apology while S_E will not. Accordingly, R can be assured that anyone paying the apology cost of a is not S_E.

The present model assumes that S_C's paying a will be offset by the benefit from repeat interactions. A similar idea was proposed by an economist, Nelson (1974), to explain the utility of dissipative advertisements, whose cost, he supposed, is offset by repeat purchases only when the advertiser produces a high-quality good (see also Gintis, 2000, chapter 13). More relevant to the apology-making context, McElreath and Boyd (2007) proposed a similar explanation for why the contrite tit-for-tat (CTFT) strategy works in the noisy repeated prisoner's dilemma game, in which players sometimes defect by mistake. CTFT accepts the partner's defection once when it has accidentally committed a defection in the previous round. McElreath