A Rational Theory of Apologies

Benjamin Ho
Stanford GSB

Abstract

Apologies are a previously unstudied social institution integral in the maintenance of relationships within society. Their application ranges from organizational design to political systems to legal settings and beyond. This paper formulates a game theoretic costly signaling model using rational agents that serves as a framework for understanding the use of apologies in these settings. By adding a multi-dimensional type-space, I find a cheap talk equilibrium where apologies arise as a trade off between revelation of preference alignment and ability. I finally consider evolutionary implications and offer a positive explanation of the fundamental attribution error from

Introduction: Why Rational Apologies?

Apologies are an important unstudied institution largely ignored by most of social science (psychology, sociology, political science) and entirely ignored by economics.

My approach is to develop a rational choice framework to understand apologies in a principal-agent context. I avoid behavioral/psychological assumptions and primarily limit my analysis to preference for consumption, as this allows applicability to potentially highly rational actors such as politicians and governments. When I allow psychological concepts such as remorse, I provide an evolutionary justification.

In this paper, I argue that apologies exist to maintain relationships. However, if apologies help relations, talk is cheap, why is it that not everybody apologizes?

Applications

Organizations: The prevalence of apologies in various organizational settings is indicative of differences in task assignment, in risk taking, in turnover, and in organizational design.

Politics: There is a stylized fact that politicians never apologize. Consider, Bush on Iraq, Clinton on Lewinsky and Berlusconi on Germany.

Government: An apology by a government is important either between the government and its people (e.g. South African apartheid, Japanese interment, or the United States civil war), or between governments in international relations.

Law: In recent decades in the US, apologies have become increasingly important in litigation damages. Also, California and Texas and others have passed laws to prevent apologies from being used as evidence to encourage their use. Apologies are especially relevant in medical malpractice as a vicious circle has arisen. Doctors are afraid to apologize because of the large risk of lawsuits. Patients are more likely to sue due to anger for not receiving an apology.

Corporate Governance: CEOs are expected to be responsible to shareholders.

Interdisciplinary Theory

Remorse/Guilt
Psychologists beginning with Freud believe that apologies serve as "negative affect alleviation." Harmful actions create painful guilt that can only be eliminated by an apology.

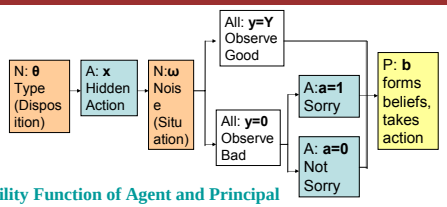
Sympathy
An apology is a demonstration of sympathy

Justification/Excuses
Attribution Theory (Ross,1993; etc.) says that people misattribute the cause of outcomes to the person rather than the environment. An apology shifts the attribution back to the environment.

Relationship/Shame
Apologies are crucial for the maintenance of relationships (Tavuchis, 2001; etc.). Also, relationships are crucial for group formation and norm enforcement (Kandori, 1992; Kandel and Lazear, 1992; etc.)

Pride/Status
Apologies are difficult because they precipitate a loss of pride

Basic Model



Utility Function of Agent and Principal

$$U_A = u(y,x,\theta) - k(\omega)\mathbf{1}_a + v(b,\theta)$$

$$U_p = y$$

Description of Variables

$x \in X$, a compact set of policy choices for the agent
$\omega \in \Omega$, a compact set of possible environmental factors with probability density $F(\omega)$
$y \in \{0,Y\}$, two possible outcomes of the interaction
$\theta \in \{\theta_B, \theta_G\}$ measure of preference congruence between agent and principal
$\phi(x,\omega)$ probability of success given the agent's action and the environment
$k(\omega)$ the cost of apology
$b(y,a) = \Pr(\theta = \theta_G; y, a)$ the principal's belief of the agent's type given the outcome and whether an apology is tendered
$v(b,\theta)$ the continuation value for the agent

Overview of the Model

There are two types of agents, (θ_B, θ_G), where the good type is more aligned with the principal. The principal would like to maintain relations with the good type. The agent chooses an action x (for example effort), before observing the environment, ω. The action and the environment jointly determines the probability of success. If the outcome is failure, the agent can apologize at some personal cost $k(\omega)$. The principal forms beliefs based on the outcome and the apology. The continuation value is here given exogenously though it can be easily endogenized. Essentially a Spence signaling model.

Model Concepts

Remorse/Guilt:
An agent feels remorse if he makes a mistake. In other words, there is a difference between the *ex post* optimal x (after observing ω) and the *ex ante* optimal x.

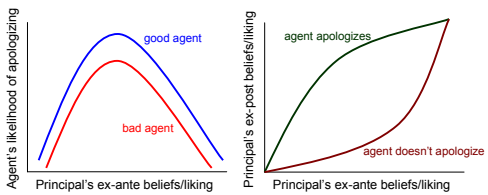
Sympathy:
An agent is sympathetic if preferences are more aligned (an agent demonstrates sympathy by appearing to be θ_G rather than θ_B).

Justification/Excuses:
An apology with a particular form of $k(\omega)$. Or, an apology is able to shift the principal's beliefs from the agent's type to the environment.

Relationship/Shame:
How much the principal likes the agent and how likely she is to continue the relationship is captured by the principal's beliefs, b. Shame can be captured by $k(\omega)$.

Pride/Status:
The status loss can be captured by $k(\omega)$. See also extension.

Results (Preliminary)



Under certain regularity conditions we obtain the following results.

- A lack of apology implies either a low type or bad environment
- Good types always apologize more often than bad types
- "Love means never having to say you're sorry" (Love Story, 1972)
- Apologies are more frequent when type is uncertain (b is near 0.5)
- Apologies are more effective when type is uncertain
- Apologies devalue with use

Empirical Evidence

- Apologies always help regardless of prior belief or liking (ex post b increases regardless of ex ante b) (Bennett and Dewberry, 1994)
- Apologies lead to impression improvement (b increases) (Ohbuchi, et al., 1989; Bennett and Earwalker, 2001; etc.)
- Apologies more effective when agent less guilty (more effective when effort is high) (Bennett and Earwalker, 2001)
- Forgiveness occurs more often when principal has higher empathy/liking (higher b) (McCullough, 1997)
- Apologies more prevalent in Japan (θ valued over η) (Ohbuchi et al., 1989; Tavuchis, 1992)
- Apologies more common by women (women have lower status cost) (Gallup, 1989)
- Apologies increase impression, decreases status (see extension) (Tiedens, 2001)

Extension: Cheap Talk & 2D Typespace

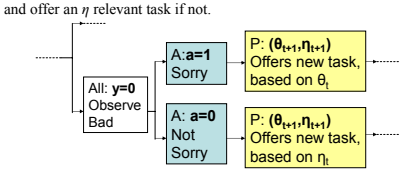
Although it is quite plausible that in interpersonal relationships, it is possible for social systems to develop a $k(\omega)$ cost to maintain relationships, in highly rational environments such as politics or international relations this becomes less plausible. However, we can drop the cost of apology and have cheap talk apologies if we modify the actions of the principal. To illustrate, consider the following experimental result:

Tiedens (2001) constructs two videos featuring Bill Clinton. In one, Clinton appears apologetic about the Lewinsky affair. In the other, he appears angry. She shows each video to two separate subject groups. Those who saw angry Clinton liked him less, and lamented that he did not apologize. Those who saw apologetic Clinton liked him more. However, when asked about Clinton's leadership ability, status, and crucially, electability, the angry Clinton fared much better.

To achieve this cheap talk equilibrium, add another dimension of type. In addition to $\theta \in \{\theta_B, \theta_G\}$ representing preference alignment, let there also be another dimension of type, call it *ability*, $\eta \in \{\eta_H, \eta_L\}$. A high ability worker can accomplish tasks with less effort than low ability workers.

In addition, instead of only one task let there be many tasks. An individual agent has a different θ_i and a different η_i for each task, i. The principal still only has a prior belief on each type, but the principal does know the correlation between tasks. For example, the principal knows that in some tasks, θ is important (friendship, trust, marrying your daughter), and in other tasks, η is important (fixing your car, running the country).

Then, a cheap-talk apology equilibrium is achieved if the principal plays the following strategy: offer a θ relevant task in the next period if an apology is tendered, and offer an η relevant task if not.



Extension: Evolution & Attribution

In most interpersonal situations, it is quite believable that the cause of an apology is some emotional psychic reaction we call remorse. This model provides insight into how this social institution of remorse might have emerged. It is sensible that parents would instill remorse in their children. Those with a sense of remorse are better capable of maintaining relationships that are crucial in our modern embedded society (Sen, 1977; Granovetter, 1985; etc.).

This model also provides a positive, evolutionary justification of the **Fundamental Attribution Error** (FAE) (Ross, 1977), a psychology theory as central to modern social psychology theory and is well-established by numerous experimental findings. Briefly, the theory states that when an outcome is the product of both an agent's *disposition* or type, and the *situation* or environment, observers tend to over attribute the cause of the outcome to the agent's disposition rather than acknowledging the impact of the situation.

The most famous experiment (Darley-Batson, 1973) of the FAE has an injured man on the street and various passerby's who walk by. Sometimes the passerby stops and acts as a Good Samaritan. Sometimes, the passerby callously hurries by. Observers assume that the difference in behavior can be explained by disposition. In fact, the experiment is constructed so that the variable that explains nearly all the variance in behavior is whether the passerby was in a hurry.

In this model, in the absence of an apology, it is rational for the principal to believe the agent was at fault rather than the environment, even if this often leads to misattribution. The intuition is that the environment has no incentive to correct a misattribution, the agent does.