# Why is it so hard to say sorry?

T. A. Han[a,b], L. M. Pereira[c], F. C. Santos[d] and T. Lenaerts[a,b]

[a] *AI lab, Vrije Universiteit Brussel, Belgium*
[b] *MLG group, Université Libre de Bruxelles, Belgium*
[c] *CENTRIA, Universidade Nova de Lisboa, Portugal*
[d] *GAIPS/INESC-ID & Instituto Superior Técnico, Portugal*

**Abstract**

When making a mistake, individuals are willing to apologize to secure further cooperation, even if the apology is costly. Similarly, individuals arrange commitments to guarantee that an action such as a cooperative one is in the others' best interest, and thus will be carried out to avoid eventual penalties for commitment failure. Hence, both apology and commitment should go side by side in behavioral evolution. Here we discuss our work published in [6], wherein we study the relevance of a combination of those two strategies in the context of the iterated Prisoner's Dilemma (IPD). We show that apologizing acts are rare in non-committed interactions, especially whenever cooperation is very costly, and that arranging prior commitments can considerably increase the frequency of such behavior. In addition, we show that with or without commitments, apology resolves conflicts only if it is sincere, i.e. costly enough. Most interestingly, our model predicts that individuals tend to use much costlier apology in committed relationships than otherwise, because it helps better identify free-riders such as fake committers.

## Summary

Apology is perhaps the most powerful and ubiquitous mechanism for conflict resolution [1, 9], especially among individuals involving in long-term repeated interactions (such as a marriage). An apology can resolve a conflict without having to involve external parties (e.g. teachers, parents, courts), which may cost all sides of the conflict significantly more. Evidence supporting the usefulness of apology abounds, ranging from medical error situations to seller-customer relationships [1]. Apology has been implemented in several computerized systems such as human-computer interaction and online markets so as to facilitate users' positive emotions and cooperation [11, 12].

The iterated Prisoner's Dilemma (IPD) has been the standard model to investigate conflict resolution and the problem of the evolution of cooperation in repeated interaction settings [2, 10]. This IPD game is usually known as a story of tit-for-tat (TFT), which won both Axelrod's tournaments [2]. TFT cooperates if the opponent cooperated in the previous round, and defects if the opponent defected. But if there can be erroneous moves due to noise (i.e. an intended move is wrongly performed), the performance of TFT declines, because an erroneous defection by one player leads to a sequence of unilateral cooperation and defection. A generous version of TFT, which sometimes cooperates even if the opponent defected [8], can deal with noise better, yet not thoroughly. For these TFT-like strategies, apology is modeled implicitly as one or more cooperative acts after a wrongful defection.

In our recent work [6], we describe a model containing strategies that explicitly apologize when making an error between rounds. An apologizing act consists in compensating the co-player an appropriate amount (the higher the more sincere), in order to ensure that this other player cooperates in the next actual round. As such, a population consisting of only apologizers can maintain perfect cooperation. However, other

behaviors that exploit such apology behavior could emerge, such as those that accept apology compensation from others but do not apologize when making mistakes (fake apologizers), destroying any benefit of the apology behavior. Resorting to the evolutionary game theory [10], we show that when the apology occurs in a system where the players first ask for a commitment before engaging in the interaction [4, 5, 7, 3], this exploitation can be avoided. Our results lead to the following conclusions: (i) Apology alone is insufficient to achieve high levels of cooperation; (ii) Apology supported by prior commitment leads to significantly higher levels of cooperation; (iii) Apology needs to be sincere to function properly, whether in a committed relationships or commitment-free ones (which is in accordance with existing experimental studies, e.g. in [9]); (iv) A much costlier apology tends to be used in committed relationships than in commitment-free ones, as it can help better identify free-riders such as fake apologizers: '*commitments bring about sincerity*'.

As apology [11, 12] and commitment [13, 14] have been widely studied in AI and Computer Science, for example, about how these mechanisms can be formalized, implemented, and used to enhance cooperation in human-computer interactions and online market systems [11, 12], as well as general multi-agent systems [14, 13], our study would provide important insights for the design and deployment of such mechanisms; for instance, what kind of apology should be provided to customers when making mistakes, and whether apology can be enhanced when complemented with commitments to ensure better cooperation, e.g. compensation from customers for wrongdoing.

# References


[1] J. Abeler, J. Calaki, K. Andree, and C. Basek. The power of apology. *Economics Letters*, 107(2), 2010.

[2] Robert Axelrod. *The Evolution of Cooperation*. Basic Books, ISBN 0-465-02122-2, 1984.

[3] T. A. Han. *Intention Recognition, Commitments and Their Roles in the Evolution of Cooperation: From Artificial Intelligence Techniques to Evolutionary Game Theory Models*, volume 9. Springer SAPERE series, 2013.

[4] T. A. Han, L. M. Pereira, and F. C. Santos. The emergence of commitments and cooperation. In *AAMAS'2012*, pages 559–566, 2012.

[5] T. A. Han, L. M. Pereira, and F. C. Santos. Intention Recognition, Commitment, and The Evolution of Cooperation. In *Procs of IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE Press, 2012.

[6] T. A. Han, L. M. Pereira, F. C. Santos, and T. Lenearts. Why is it so hard to say sorry: The evolution of apology with commitments in the iterated Prisoner's Dilemma. In *IJCAI'2013*, pages 177–183, 2013.

[7] T.A. Han, L.M. Pereira, F.C. Santos, and T. Lenaerts. Good agreements make good friends. *Scientific reports*, 3(2695), 2013.

[8] M. A. Nowak and K. Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355:250–253, 1992.

[9] Y. Ohtsubo and E. Watanabe. Do sincere apologies need to be costly? test of a costly signaling model of apology. *Evolution and Human Behavior*, 30(2):114–123, 2009.

[10] Karl Sigmund. *The Calculus of Selfishness*. Princeton University Press, 2010.

[11] Jeng-Yi Tzeng. Toward a more civilized design: studying the effects of computers that apologize. *International Journal of Human-Computer Studies*, 61(3):319 – 345, 2004.

[12] S. Utz, U. Matzat, and C. Snijders. On-line reputation systems: The effects of feedback comments and reactions on building and rebuilding trust in on-line auctions. *International Journal of Electronic Commerce*, 13(3):95–118, 2009.

[13] M. Winikoff. Implementing commitment-based interactions. In *AAMAS'2007*, pages 868–875, 2007.

[14] Michael Wooldridge and Nicholas R. Jennings. The cooperative problem-solving process. In *Journal of Logic and Computation*, pages 403–417, 1999.