Evolution of Pairwise Commitment and Cooperation

The Anh Han H.ANH@AI.VUB.AC.BE
Artificial Intelligence lab, Vrije Universiteit Brussel & MLG, Université Libre de Bruxelles

Luís Moniz Pereira LMP@FCT.UNL.PT
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Francisco C. Santos FRANCISCOCSANTOS@IST.UTL.PT
INESC-ID & Instituto Superior Técnico, Universidade de Lisboa

Tom Lenaerts TOM.LENAERTS@ULB.AC.BE
MLG, Université Libre de Bruxelles & Artificial Intelligence lab, Vrije Universiteit Brussel

When starting a new collaborative endeavor, it pays to establish upfront how strongly your partners commit to the common goal and what compensation can be expected in case the collaboration is violated. Diverse examples in biological and social contexts have demonstrated the pervasiveness of making prior agreements on posterior compensations, suggesting that this behavior could have been shaped by natural selection (Nesse, 2001; Han, 2013). We discuss here our work in (Han et al., 2013), wherein we analyze the evolutionary relevance of such a commitment strategy in the context of the pairwise one-shot Prisoner's Dilemma (PD). The commitment strategy proposes, prior to any interaction, its co-player to commit to cooperate in the PD, paying a cost to render the commitment deal reliable (e.g. the cost to hire a layer to make a legal contract). Those players that commit and then default (i.e. defects) have to compensate their non-defaulting co-player. Resorting to methods of Evolutionary Game Theory (Sigmund, 2010), we analyze, both mathematically and using numerical simulations, the viability of such a commitment strategy in the co-presence of different free-riding strategies, including the one that commits but then defaults on the commitment, and the one that commits and cooperates only if someone else pays the cost of arranging the commitment (namely, this strategy defects if there is no commitment in place). Our results show that when the cost of arranging a commitment deal is justified with respect to the benefit of cooperation, substantial levels of cooperation can be achieved, even without repeated interactions. On the one hand, commitment proposers can get rid of those individuals that agree to cooperate yet act differently, and, on the other hand, they can maintain a sufficient advantage over those that cooperate only if the commitment is set up by someone else, because a commitment proposer will cooperate with players alike herself, while the latter defect among themselves.

However, when the cost of arranging the commitment goes beyond a certain limit, commitment free-riders pervade and dominate the population. We show in our work that this commitment free-riding issue can be resolved significantly, in case commitment proposers explicitly insist on sharing the cost of arranging the commitment—even if the co-players that before agreed to commit now refuse to share the cost.

Overall, our work demonstrates that the conventional wisdom or common knowledge stating that *good agreements make good friends* provides a highly relevant guideline, which is extendable to *good cost-sharing agreements* lead to even better friends.

Acknowledgments

TAH acknowledges the support from the FWO. TL acknowledges the support from the FWO and FRNS.

References

Han, T. A. (2013). *Intention recognition, commitments and their roles in the evolution of cooperation: From artificial intelligence techniques to evolutionary game theory models*, vol. 9. Springer SAPERE.

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

Nesse, R. M. (2001). *Evolution and the capacity for commitment.* Russell Sage Foundation series on trust. Russell Sage.

Sigmund, K. (2010). *The calculus of selfishness.* Princeton University Press.