# Stability of Human-Inspired Agent Societies

Extended Abstract

Joe Collenette
Department of Computer Science
University of Liverpool, UK
J.M.Collenette@liverpool.ac.uk

Katie Atkinson
Department of Computer Science
University of Liverpool, UK

Daan Bloembergen
Intelligent and Autonomous Systems
Centrum Wiskunde & Informatica, Netherlands

Karl Tuyls
Department of Computer Science
University of Liverpool, UK

## KEYWORDS

Agents; OCC; Emotions; Mood; Evolutionarily Stable Strategy

**ACM Reference Format:**
Joe Collenette, Katie Atkinson, Daan Bloembergen, and Karl Tuyls. 2019. Stability of Human-Inspired Agent Societies. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

## 1 INTRODUCTION

Models of emotion, particularly those based on the Ortony, Clore, and Collins (OCC) account of emotions, have been used as part of agents' decision making processes to explore their effects on cooperation within social dilemmas [7, 19, 22]. We analyse two different interpretations of OCC agents. Firstly, Emotional agents that decide their action using only a model of emotions. To analyse the possibility of evolutionary stability of these agents we use the Prisoner's Dilemma game. We contrast the results with the second interpretation of an OCC agent, the Moody agent [7], which additionally uses a psychology-grounded model of mood. Our analysis highlights the different strategies that are needed to achieve success as a society in terms of both stability and cooperation, in the iterated Prisoner's Dilemma. The Emotional agents are better suited playing against a mixed group of agents with differing strategies than the Moody agents are. The Moody agents are more successful than the Emotional agents when only one strategy exists in the society.

## 2 BACKGROUND

The Prisoner's Dilemma is a social dilemma, popularised through the influential Axelrod's tournament [2], where two players pick between cooperating with the other player, or trying to take advantage. A strategy can be described as evolutionarily stable when the majority of agents are using the strategy, and it cannot be invaded by any initially rare strategy [24]. Evolutionary stability in the Prisoner's Dilemma has been extensively analysed [4, 25], with no pure strategy, TIT-FOR-*n*-TATS, or reactive strategies being evolutionarily stable in the iterated version of the Prisoner's Dilemma [5, 9, 20]. Evolutionary stability is an extremely demanding criterion. Our Emotional and Moody agents differ from previous analyses of strategies in this problem set, as they are able to identify individual opponents and change their actions based on the individual, and have a memory extending further than the previous interaction.

*Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.* 

**Table 1: Prisoner's Dilemma payoff matrix. Cooperate ($C$), Defect ($D$).**

| $C, C$ | $D, D$ | $D, C$ | $C, D$ |
|---|---|---|---|
| 3, 3 | 1, 1 | 5, 0 | 0, 5 |

### 2.1 Emotional agents

We look at the intuition behind evolutionary stability of Emotional agents described by Lloyd-Kelly et al [18], as the majority of previous work focussed on experimental studies only [7, 16, 18]. These emotional agents use the OCC model of emotions [22] from the psychology literature. Various agent designers have used this model successfully as part of their agents' decision making process [1, 7, 19, 23]. While the thresholds which trigger emotions can be much larger than those previously defined [17], the thresholds are restricted to lower values to reflect how emotions are short-term [15].

### 2.2 Moody Agents

The Moody agents we will be analysing use the same OCC model as used in the Emotional agents, but in addition they also use a model of mood [6]. Mood is represented as a real number, with lower values representing more depressed moods and higher values representing more positive moods. This reflects how psychologists have represented human mood [3, 8, 10, 12, 13]. The mood value places the mood into one of five possible mood levels (very high, high, neutral, low, very low). The mood value will then affect the action selection of the agent, as inspired by the psychology literature [11, 14]. A full description of the model, and its psychology grounding, is given in Collenette et al. [6].

## 3 OVERVIEW OF EVOLUTIONARY STABILITY ANALYSIS

To analyse whether Emotional and Moody agents can be considered an evolutionarily stable strategy (ESS), we need an opponent strategy that will take the largest advantage of the agents, minimising their payoff. We will use a strategy termed the *Oracle*. The effectiveness of the strategy is achieved by breaking an assumption of the Prisoner's Dilemma, namely that players have no knowledge of the opponent's move, as reflected by the name. Intuitively the Oracle strategy will always cooperate with itself, and when faced with another strategy will choose the worst outcome for the opponent, effectively making it the worst case scenario for the opponent. The Oracle strategy targets the conditions needed to be an ESS, allowing effective analysis of evolutionary stability. The Oracle is the most effective strategy at minimising the payoff of the Emotional agents.

## 3.1 Emotional Agents

Statement 1. *Emotional Agents are not an ESS initially*

Emotional agents are initially cooperative. The Oracle agents are able to take advantage of this, defecting against only the Emotional agents. The Oracle will then have a larger payoff than the Emotional agents as they will receive only the 3 payoff or the 5 payoff, while the Emotional agents will receive with a 3 or a 0. The Oracle will have a greater average payoff making Statement 1 valid.

Statement 2. *Emotional agents converge to defection against all Oracle agents given a sufficiently high number of interactions and randomness in pairing.*

The Oracle agents will defect against the Emotional agents. The Emotional agents are guaranteed to change to defection given sufficient time to adjust [7]. This is due to the Oracle agent only defecting against the Emotional agent. If the Emotional agent cooperates then the Oracle will defect to take advantage, if the Emotional agent defects then the Oracle will also defect to protect its payoff. The outcome will always cause the Emotional agent to change to defection as the emotional trigger for defection will always be triggered. Given that the Emotional agents interact with all Oracle agents enough to cause the emotional trigger to fire, then we can guarantee Statement 1.

Statement 3. *Initially cooperative Emotional agents, with fast interactions and slow reproduction are an ESS.*

All initially cooperative Emotional agents will cooperate with each other indefinitely [6]. When two Emotional agents are in mutual cooperation, such as when the two Emotional agents are both initially cooperative, the emotional trigger to switch to defect never gets fired. The Oracle agents also have perfect cooperation among themselves. Given Statement 2 is true, then we now know that the Oracle is always be defecting against all Emotional agents. Therefore the Oracle agents and the Emotional agents will both be receiving the same payoff when interacting with an agent that uses the same strategy or the opposing strategy. As an ESS requires that the Emotional agents are a majority in the society, the expected payoff of the Emotional agents will therefore be higher than the Oracle agent, making Statement 3 valid.

In summary, we can show that with some assumptions, initially cooperative Emotional agents are possibly an ESS. This is significant as no strategy is able to minimise the payoff of the Emotional agents more than the Oracle agent.

## 3.2 Moody Agents

Moody agents have the additional Mood model on top of the emotions which changes how the Moody agents react. We need to analyse each mood level individually in terms of an ESS. We will assume that the Moody agents are initially cooperative and that the mood levels do not change over time.

Statement 4. *Moody agents in an initially very high mood, initially low mood, or initially very low mood are not an ESS*

Moody agents in a very high mood, or a very low mood are functionally equivalent to a full cooperation strategy, or a full defection strategy, both of which are known to not be an ESS [5, 20]. Low moods defect against new opponents, meaning the Moody agents will not cooperate together, lowering their expected payoff below the Oracle agent's.

Statement 5. *Moody agents that are in an initially high mood, or initially neutral mood are functionally equivalent to Moody agents in a neutral mood.*

Neutral moods do not change the action selection; as the Moody agent is initially cooperating there is no change in action when the Moody agent is in a high mood. As there is no change in action selection then the Emotional agents' analysis applies, making these mood levels an ESS. To validate this claim the mood levels should never leave either the high mood or neutral mood levels.

Statement 6. *Moody agents in an initially neutral or initially high mood will move to the very high mood level, when there is a sufficiently small invasion of Oracle agents.*

When two moody agents cooperate together both of their mood values will increase; when they meet an Oracle agent the mood level will go down as they have been taken advantage of. Since there is a majority of Moody agents, the mood values will reach the very high mood levels. Therefore the Moody agents are not an ESS.

In conclusion we can say that overall Moody agents are not an ESS. If the mood level of Moody agents was to stay stable over time, this would go against the design principles of the model [6]. The psychological grounding of the moody model requires that mood levels change over time as per the psychology literature [8, 21].

# 4 CONCLUSION

We have described how Emotional agents that use a model of emotions as part of their decision-making can be considered an ESS when they initially cooperate with new partners and are able to adapt to an invading strategy before reproducing. Moody agents using a simulated model of mood and a model of emotions as their decision-making process, are not an ESS. Some mood levels break the assumption that Moody agents mutually cooperate. We tested both these agents against an Oracle strategy, minimising the expected payoff of the Emotional agents. The Oracle strategy can successfully invade the Moody agents. We aim to now formalise the statements with full proofs, including a proof for the Oracle agent minimising the payoff of the Emotional agents. The majority of the literature concerning these kinds of agents focuses on simulations and observing the effects. We have taken a broader view of human-inspired agents by analysing evolutionary stability in an account that implements both Emotional and Moody agents.

## REFERENCES


- [1] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. 2000. Integrating Models of Personality and Emotions into Lifelike Characters. In *Affective Interactions*, Ana Paiva (Ed.). LNCS, Vol. 1814. Springer, 150–165.
- [2] Robert Axelrod and William Donald Hamilton. 1981. The evolution of cooperation. *Science* 211, 4489 (1981), 1390–1396.
- [3] AC Bilderbeck, ZE Reed, HC McMahon, LZ Atkinson, J Price, JR Geddes, GM Goodwin, and CJ Harmer. 2016. Associations between mood instability and emotional processing in a large cohort of bipolar patients. *Psychological Medicine* (2016), 1–10.
- [4] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. 2015. Evolutionary dynamics of multi-agent learning: a survey. *Journal of Artificial Intelligence Research* 53 (2015), 659–697.
- [5] Robert Boyd and Jeffrey P. Lorberbaum. 1987. No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game. *Nature* 327 (07 05 1987), 58 EP –. http://dx.doi.org/10.1038/327058a0
- [6] Joe Collenette, Katie Atkinson, Daan Bloembergen, and Karl Tuyls. 2016. Modelling Mood in Co-operative Emotional Agents. In *Proc of DARS'16*.
- [7] Joe Collenette, Katie Atkinson, Daan Bloembergen, and Karl Tuyls. 2017. Environmental effects on simulated emotional and moody agents. *The Knowledge Engineering Review* 32 (2017).
- [8] Ed Diener, Randy J Larsen, Steven Levine, and Robert A Emmons. 1985. Intensity and frequency: dimensions underlying positive and negative affect. *Journal of personality and social psychology* 48, 5 (1985), 1253.
- [9] Joseph Farrell and Roger Ware. 1989. Evolutionary stability in the repeated prisoner's dilemma. *Theoretical Population Biology* 36, 2 (1989), 161 – 166. https://doi.org/10.1016/0040-5809(89)90027-0
- [10] Ernst Fehr and Klaus M Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly journal of Economics* 114 (1999), 817–868. Issue 3.
- [11] William E Haley and Bonnie R Strickland. 1986. Interpersonal betrayal and cooperation: Effects on self-evaluation in depression. *Journal of Personality and Social Psychology* 50, 2 (1986), 386.
- [12] Lesley Hepburn and Michael W Eysenck. 1989. Personality, average mood and mood variability. *Personality and Individual Differences* 10, 9 (1989), 975–983.
- [13] Guido Hertel. 1999. Mood effects in social dilemmas: What we know so far. *Resolving social dilemmas: Dynamic, structural, and intergroup aspects* (1999), 227–243.
- [14] Guido Hertel, Jochen Neuhof, Thomas Theuer, and Norbert L Kerr. 2000. Mood effects on cooperation in small groups: Does positive mood simply lead to more cooperation? *Cognition & emotion* 14, 4 (2000), 441–472.
- [15] Dacher Keltner and James J. Gross. 1999. Functional accounts of emotions. *Cognition & Emotion* 13, 5 (1999), 467–480.
- [16] Martyn Lloyd-Kelly. 2014. *Modelling Emotions and Simulating their Effects on Social Interactions in Agent Systems*. Ph.D. Dissertation. University of Liverpool.
- [17] Martyn Lloyd-Kelly, Katie Atkinson, and Trevor Bench-Capon. 2012. Developing co-operation through simulated emotional behaviour. In *13th International Workshop on Multi-Agent Based Simulation*.
- [18] Martyn Lloyd-Kelly, Katie Atkinson, and Trevor Bench-Capon. 2012. Emotion as an Enabler of Co-operation.. In *ICAART (2)*. 164–169.
- [19] Martyn Lloyd-Kelly, Katie Atkinson, and Trevor Bench-Capon. 2014. Fostering co-operative behaviour through social intervention. In *Proc of SIMULTECH'14*. IEEE, 578–585.
- [20] Jeffrey Lorberbaum. 1994. No Strategy is Evolutionarily Stable in the Repeated Prisoner's Dilemma. *Journal of Theoretical Biology* 168, 2 (1994), 117 – 130. https://doi.org/10.1006/jtbi.1994.1092
- [21] John D Mayer and Ellen Hanson. 1995. Mood-congruent judgment over time. *Personality and Social Psychology Bulletin* 21 (1995), 237–237.
- [22] Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The cognitive structure of emotions*. Cambridge university press.
- [23] A. Popescu, J. Broekens, and M. van Someren. 2014. GAMYGDALA: An Emotion Engine for Games. *IEEE Transactions on Affective Computing* 5, 1 (Jan 2014), 32–44. https://doi.org/10.1109/T-AFFC.2013.24
- [24] J Maynard Smith and George R Price. 1973. The logic of animal conflict. *Nature* 246, 5427 (1973), 15.
- [25] Bernhard Thomas. 1985. On evolutionarily stable sets. *Journal of Mathematical Biology* 22, 1 (01 Jun 1985), 105–115. https://doi.org/10.1007/BF00276549