Master thesis on Cognitive Systems and Interactive Media
Universitat Pompeu Fabra

Modeling social conventions with Sequential Episodic Control

Adriana López Hernández

Supervisor: Ismael Freire González

Co-Supervisor: Adrián Fernández Amil

Host Research Group: SPECS

July 2022

Universitat Pompeu Fabra
Barcelona

Contents

Acknowledgement

I would like to express my sincere gratitude to:

- My supervisors, for their valuable feedback.
- Isa and Diego, for their immense support.
- My family, for their encouragement.

Abstract

Computational models can help bring light to the underlying cognitive mechanisms responsible for the emergence of conventions in human societies. In order to provide meaningful theoretical insight, these models of behavior should aim to resemble human-like performance. Deep Reinforcement Learning (deep RL) algorithms fail to do so; these techniques are data inefficient and require several training instances to approximate the speed of human learning. Episodic Reinforcement Learning (ERL) algorithms, for their part, seek to improve deep RL algorithms by implementing memory buffers that can counter the sample inefficiency problem. Nonetheless, this approach also falls short because it considers memories as isolated, discrete events. On the other hand, episodic control models provide a model-free, non-parametric approach capable of rapid learning, thus resembling more closely human performance. These algorithms introduce a fast memory system inspired by the hippocampus that allows them to search for a solution without having to make strong assumptions about the world. The present work will test the adequacy of one particular episodic control algorithm, the Sequential Episodic Control (SEC) model to simulate human behavioral data in a repeated coordination game. This project will compare the modeled and behavioral data concerning efficiency, fairness, and stability measures to evaluate the model's performance. Finally, given the structure of the model, this project will examine the potential theoretical implications of human conventionalization, as well as the limitations and future work on this approach.

Keywords: Sequential Episodic Control; Episodic Control; Social Conventions; Behavioral Coordination.

Chapter 1

Introduction

1.1 Conventions

Although complex, human social behavior provides societies with guidelines that facilitate interactions between agents. Within a community, different individuals may have conflicting preferences [1], which may force them to either compete or co-operate. On such occasions, they can rely on norms to solve coordination problems, namely, situations where the interests of individuals coincide.

One feature that makes social life complex is that one cannot directly observe other agents' internal mental states —like their goals and intentions. Norms are helpful because they allow agents to reduce uncertainty over future social interactions by providing a common ground. In other words, by establishing a standard that determines which actions are suitable or not, norms help regulate social behavior.

Since the outcome of social interactions depends not only on the behaviors of oneself but also on the actions of others, individuals must act according to their expectations of how different people will react [2]. Thus, norms are widespread behaviors that allow society members to know what to expect when they act a certain way. Furthermore, norms allow agents to anticipate how others might behave in a given situation because they possess an adaptive value; people need to learn what guides behavior in a community to survive.

Conventions emerge as a response to recurrent coordination problems. Specifically, conventions arise when it becomes necessary to regulate the behavior of bilateral individuals or groups in repeated social interactions [3]. Ultimately, when people seek to coordinate, a learning process takes place and gives rise to social order. In everyday life, it is possible to find multiple examples of conventions: from linguistic customs such as the accepted names to give children and pets, to the conceptualization of the notion of fairness [4].

The relationship between social interactions and conventions is bidirectional: they regulate and are regulated by each other. Initially, people may adopt a pre-existing convention to provide them with initial expectations and a code on how to guide their behavior; however, as interactions become more frequent, the original convention may change along with the evolving dynamics between individuals [3]. It follows that, for conventions to be successful, individuals must be competent at learning about the regularities that govern interactions and proficient in shaping and improving existing conventions in ways that benefit everyone.

Consequently, community members will continue to abide according to the established conventions as long as their expectations are maintained. Conventions are self-enforcing in the sense that it is advantageous for individuals to cooperate as long as the behavioral patterns persist and people have reasons to believe others will act consistently. Additionally, coordination problems may have multiple plausible solutions that are chosen arbitrarily. These solutions could be equally probable and acceptable, but one of them will be regularly preferred by individuals. In this manner, we can assert conventions are self-sustaining and arbitrary [5].

1.2 Games and conventions

Game theory is a model that describes social interactions as an analogy of "games," where rational players make strategic decisions based on the other player's plan of action. In this framework, a Nash equilibrium refers to the strategy that is the best response to the strategy of others [6]. Accordingly, in terms of this theoretical

		Player 1	
		Boxing	Ballet
Player 2	Boxing	**(2,1)**	(-1,-1)
	Ballet	(-1,-1)	**(1,2)**

Table 1: "Battle of the Sexes" payoff matrix.

context, the process of conventionalization can be understood as a repeated game with multiple, equally beneficial equilibria.

To reach an equilibrium, the individuals participating in a game of this nature ought to coordinate their behavior. The cognitive processes of the people involved, along with the dynamics of their interaction, will shape the resulting conventions. That is to say, considering the historic payoffs and interplay, players will adhere to one of the equiprobable solutions. The resulting behavioral consistency allows society to deal with recurrent coordination problems efficiently.

A classic example of a coordination game is the "Battle of the Sexes." In this game, Player 1 and Player 2 have two choices: they can either go to a Boxing match (preferred by Player 1) or to the Ballet (preferred by Player 2), but both would rather go out together than attend their favored event. The players must make their decisions simultaneously and cannot communicate with each other before doing so. *Table 1* illustrates the game's payoff matrix [7]. The game has three Nash equilibria: two pure strategies in which both players go to the same event, and one (inefficient) mixed equilibrium where they select at random which event to attend [5]. This game is an example of impure coordination: it is in everyone's best interest to match the other's selection, but doing so is intrinsically unfair.

If the "Battle of the Sexes" were to be played repeatedly, getting to a convention would address the interests of the concerned parties by maximizing their shared payoffs in time. In fact, the emerging conventions would represent additional equilibria;

for example, if the agents alternated their selected events across iterations, they could engage in turn-taking and reach a fair equilibrium after some trial-and-error [8]. Even if the players do not communicate explicitly with each other, they do so by exchanging information in the form of their past patterns of choices; this is what leads the way to coordination [7].

Even though conventions are regularities in behavior, they constrain the agents' actions but without removing their ability to make choices [2]. An individual's reasons for following a convention may not be necessarily rational [9]; consequently, agents have the option to depart from the norm. Nonetheless, if people identify an individual who deviates from an established convention, they could sanction and distrust him. An attitude of non-reciprocity is discouraged because it would jeopardize future collaboration, and, for that reason, cooperation is encouraged [10].

It is also worth noting that conventions —or any norms, for that matter— do not govern all social interactions. In the particular case of coordination problems, they can also be solved through spontaneous coordination [11]. In other words, even if a repeated coordination issue exists, individuals could tackle it from the very beginning of each iteration. Cognitively speaking, it is expensive for an individual to formulate a convention: it requires a learning process and takes up memory capacity; therefore, conventions only arise when it is not efficient to coordinate "on the fly."

For coordination to be feasible, the involved parties must learn about each other's beliefs. Furthermore, they must share their interpretation regarding the expectations each of them holds for the interaction [3]. This means that people should be able to represent conventions mentally; there should be a collectively shared knowledge about the often unspoken agreement they need to observe. Thus, the question of how conventions are created and how individuals conceptualize them can be tackled through research.

1.3 Research on conventions

In the interest of studying dynamic interactions, researchers have developed variations around the design of classical, single-turn (discrete) games. For example, in 2016, Hawkins and Goldstone [5] proposed the "Battle of the Exes," a modification of the "Battle of the Sexes" game. In the new scenario, Player 1 and Player 2 have to choose between two coffee shops: one with an Okay coffee and another one that serves a Great product. The players have to make multiple decisions over time, always simultaneously and without talking to one another. Even though they would both prefer to go to the Great coffee shop, they want to avoid each other above everything else.

With the "Battle of the Exes," Hawkins and Goldstone were interested in studying convention formation in repeated social interactions. The authors carried out a real-life behavioral study in which they paired up the participants in fixed dyads so that they would always interact with the same person. Then, they divided them into distinct experimental conditions in order to analyze two factors that were suspected of having an impact on the emergence of coordination: the difference between the payoffs associated with each choice alternative and the continuity of the interaction within trials.

Hawkins and Goldstone introduced two conditions for the payoff structure: low stakes and high stakes. The "high stakes" condition resulted in a larger gap between the reward obtained from the Great and the Okay alternatives; in other words, there was a marked difference in the quality of the coffee between establishments. In comparison, the opportunity cost of selecting the Okay alternative for the "low stakes" condition was small. *Table 2* shows the payoff matrices for both, the low and the high stakes experimental conditions.

The researchers also laid out two different conditions for the continuity of the interaction: simultaneous (ballistic) and real-time (dynamic) rounds. In the ballistic condition, the participants made their "final" choice before the trial started; this way, they were unaware of the other's selection for the current round. On the other

Low Stakes

Player 2 \ Player 1	Great	Okay
Great	(0,0)	**(1,2)**
Okay	**(2,1)**	(0,0)

High Stakes

Player 2 \ Player 1	Great	Okay
Great	(0,0)	**(1,4)**
Okay	**(4,1)**	(0,0)

Table 2: "Battle of the Exes" payoff matrices for the low and high stakes conditions.

hand, the players that interacted in a dynamic environment could constantly update their choice throughout the trial. In other words, in a real-time round, participants could see each other approaching their currently selected coffee shop and actively adjust their decision until they reached the establishment.

Hawkins and Goldstone found that behavioral coordination was achieved faster in the dynamic condition for the "Battle of the Exes" game. It is worth mentioning that coordination was also achieved in the ballistic condition, even without explicit communication between the players within rounds. Accordingly, they must have been communicating indirectly between trials, extracting information from their shared experience [9]. However, the dynamic condition allowed participants to exchange information in real-time, thus leading to faster convergence.

Nonetheless, coordination in a dynamic environment does not guarantee conventionalization: the within-round dynamic can help avoid ties in real-time, but it does not ensure stability across trials. According to the study's findings, having high stakes facilitated the emergence of a convention in time. This is, when participants had more to lose, it became cost-efficient to come up with a model that prevented them from repeatedly engaging in the same effort to coordinate on each trial (whereas when the opportunity cost was low, players could resort to an "on the fly" solution).

Overall, behavioral research —like the one done by Hawkins and Goldstone— can help answer questions about the emergence of coordination and the conditions that facilitate long-term cooperation. Nonetheless, it does not offer theoretical insight into the cognitive process underlying conventionalization itself. In order to get a sense of the latent components and procedures of this process, it would be relevant to develop computational models of human behavior that help fathom the establishment of a convention [11].

1.4 Modeling conventions

To model conventions is important for more than a couple of reasons: firstly, a coordination model can help us comprehend the underlying nature of convention formation in real-life scenarios. By abstracting the interaction components, one can explicitly study what a coordination problem entails and how an equilibrium is reached. Furthermore, understanding the social dynamics involved in an interaction provides a robust example of organization and adaptability that can be used in the design of models and algorithms, and the insights drawn from modeling real-life interactions can, in turn, improve these algorithms. On top of that, they could be used for the design of artificial-intelligent machines to incorporate them into a human society [11].

Deep Reinforcement Learning (deep RL) techniques are a good candidate for modeling human performance. RL algorithms lead to explicit values that guide behavior: they provide the mapping of a particular action-value couplet considering the initial and successor states, and the immediate or future rewards (state-action-reward mappings) [12]. Combined with deep learning, these algorithms can compile large amounts of information and achieve optimal outcomes.

Accordingly, a deep Q-Learning architecture that uses multi-agent RL can be an effective tool for emulating human interaction in a coordination task, where each involved agent has to make decisions based on historic payoffs and the interaction with others [13]. Moreover, deep RL algorithms have already shown human-level performance in a series of tasks: deep Q-network agents have been able to rival the performance of professional human players in, for instance, multiple Atari 2600 games [14] and even board games like chess, shogi, and Go [15].

Nonetheless, if the objective is to understand the underlying mechanisms responsible for behavioral convergence, deep RL algorithms fail to capture the minutiae. Deep RL focuses on emulating human-level results but does not incorporate an insight into the structure and processes that lead to them. In particular, RL algorithms are very slow learners because they are data inefficient: it can take them up to tens of millions of training iterations in order to achieve their targeted performance (whereas humans are much faster learners) [16].

Sample inefficiency in Deep RL is the result of various factors. Firstly, deep RL algorithms need incremental parameter adjustment in order to refrain from erasing previously learned data upon acquiring new information. Hence, Deep RL only makes small-step adjustments while learning to maximize generalization. Furthermore, Deep RL techniques include a weak inductive bias; this allows them to, ultimately, adapt to more variance and master a more comprehensive range of patterns [17]. However, real-life interactions commonly possess a predictable configuration, and assuming (by default) a stochastic environment can be a relatively slow approach to generate a human-like model for learning [16].

Some researchers have developed alternatives to counter sample inefficiency. For instance, Episodic Reinforcement Learning (ERL) can either exploit previous rewarding outcomes through a value propagation approach or introduce a new memory system to bootstrap learning. An example of an ERL algorithm is the Episodic Memory Deep Q-Network: it adds a memory buffer parallel to a Q-Learning network, thus allowing faster reward propagation that reduces the amount of sampled data [18]. Another ERL algorithm is the Episodic Reinforcement Learning with Associative Memory, which operates with a smaller sample due to its instance-based reasoning model [19]. However, regardless of the improvements, this approach does not fully solve the sample inefficiency problem: ERL algorithms preserve the deep RL system for gradient-based updates and, therefore, continue to require long learning times [20].

1.5 Episodic control models

Another approach that can be useful to study the mechanisms behind convention formation is that of episodic control models. By drawing inspiration from how the hippocampus works, these models mimic the fast learning mechanisms present in humans [21]. These models are non-parametric; therefore, it is unnecessary for them to make strong assumptions about underlying mapping functions present in the world. This way, by exploring data, the algorithms can search for the "best fit" as they gather information directly from the agent's environment. Non-parametric algorithms tackle the sample inefficiency issue present in deep RL, and although it takes them longer to train, they are adequate when there is no prior knowledge about how the world works [22].

One example of an episodic control model is the Model-Free Episodic Control (MFEC). This model assumes an environment with stable states and rewards, where decisions must be made relatively quickly, so planning-based models are too expensive and impractical. Since hippocampal learning is instance-based, this algorithm does not learn until computing an optimal result. Still, it guides the action-in-turn according to the most rewarding state-action couplet experienced so far. To accomplish this, the MFEC updates itself by storing, in a tabular memory, the highest values experienced in a particular state [16].

1.5.1 Sequential Episodic Control

The Sequential Episodic Control (SEC) is another episodic control learning model; it was developed in 2021 by Freire et al. [20]. Like the previous algorithm, it uses tabular memory to save past rewarding states. Nevertheless, in contrast to MFEC and even most ERL algorithms, the SEC model stores complete sequences of state-action couplets instead of discrete events. This way, the SEC model favors the selection of a recurrent pattern of states and actions that is shown to be rewarding.

Episodic control algorithms and, in particular, the SEC model appears to be a suitable approach for modeling behavioral convergence in repeated interactions [20].

Due to the predictable configuration of conventionalization environments, it follows that a potentially competent model is one that can optimize the time needed for achieving coordination. Contrary to model-based techniques, the SEC algorithm uses an episodic-memory-driven approach to bias the agent's action selection towards recently rewarding state-action sequences. Therefore, in a scenario intended to study the formation of conventions, it is expected to resemble human-like behavior more closely than alternative approaches.

1.6 Modeling the Battle of the Exes

The present project will model the Battle of the Exes task proposed by Hawkins and Goldstone to test the performance of the SEC approach in convention formation [3]. However, it is worth noting that this is not the first attempt at modeling this particular repeated coordination game. In 2020, Freire et al. [11] proposed the Control-based Reinforcement Learning (CRL) model, an approach that integrated a lower-level sensorimotor loop (Reactive Layer) and a higher-level learning algorithm (Adaptive Layer.) With this model, they simulated the Battle of the Exes and analyzed whether the modeled data predicted the behavioral results previously obtained by Hawkins and Goldstone.

To assess CRL's performance on this matter, the authors conducted a pairwise comparison between the behavioral and the modeled data. The study relied on Binmore's levels of priority [6] to evaluate CRL's adequacy: they measured the efficiency, fairness, and stability of the results. The first level of priority, efficiency, evaluates whether the players manage to maximize their collective rewards. Fairness shows whether both players' rewards are balanced throughout the game. Lastly, we can understand stability following the notion of arriving at an equilibrium in a game; it concerns the robustness of the arising conventions.

Even though Freire's simulation entailed a data-efficient approach, it failed to predict the human-level performance on the stability measure. This limitation suggests that the CRL model could not capture the complexity underlying human cognitive processes. In other words, the Adaptive Layer failed to abstract the concepts behind the strategies that the players came up with during the task. For example, human players could have decided (inductively) to engage in a "turn-taking" strategy between rounds; nonetheless, the CRL model would not have been able to learn (from scratch) a separate, specific policy for each of the two available alternatives [11].

1.7 Research motivation and objectives

As was previously mentioned, the purpose of this thesis is to assess the performance of the SEC algorithm during a repeated coordination game. Particularly, by implementing this model to simulate the Battle of the Exes scenario, the present work aims to emulate Hawkins and Goldstone's experimental results [3] in terms of Binmore's levels of priority [6]. Consequently, if the model accurately predicts the existing behavioral data, it could provide a comprehensive insight into the underlying theoretical mechanisms responsible for convention emergence in human societies.

The present study will attempt to answer the following questions:

- Can the SEC model achieve human-level performance for Efficiency, Fairness, and Stability?
- If so, what intuition does it provide about the mechanisms underlying human conventionalization?
- What advantages does the SEC algorithm contribute to the modeling of convention formation compared to other approaches (e.g., the CRL model)? What are its limitations?

Given the existing evidence that shows that introducing a sequential inductive bias improves memory efficiency when engaging in the exploration of new environments [21] [20], it is expected that SEC will perform efficiently in the task at hand. Namely, it is hypothesized that the algorithm will be suitable for modeling human-like behavior in the early stages of repeated social interaction, where behavioral coordination does not rely on a preconceived model of the world, but on the dynamics of the interaction itself. In particular, it is expected that implementing a sequential-memory-driven algorithm will address CRL's inability to simulate the stability seen in the behavioral data: by favoring the selection of recurrent state-action pair sequences, the SEC algorithm can optimize the time needed to reach an equilibrium.

1.8 Structure of the report

This thesis project is organized as follows: the first chapter covers the theoretical framework, offering a review of the existing and relevant literature to justify the work. In the second chapter, the methodology is described, detailing the characteristics of the model, the virtual environment for the conventionalization task, and the subsequent implementation of SEC. Then, chapter three reviews and analyzes the gathered data and carries out a pairwise comparison between the preceding behavioral study and the current model's outcome. Finally, chapter four is dedicated to discussing the results and the conclusions on the scope of this work, including recommendations for future work.

Chapter 2

Materials and methods

2.1 Experimental design

In this project, a multi-agent simulation of the Battle of the Exes task [1] will be implemented to evaluate the performance of the Sequential Episodic Control (SEC) model in emulating behavioral data during a repeated coordination task. The experimental design is a 2x1 between-subjects study, where the factor of interest is the continuity of the interaction between the players. Within the Battle of the Exes benchmark, this refers to the dynamic and ballistic versions of the game. Following the methodology done by Freire et al. [11], the condition above is to be played by 50 agents paired in fixed dyads (25 dyads in total). The pairs will interact for 50 trials and will engage in a high stakes situation, where there is a large disparity between the players' payoffs.

The agents' task is to reach one of the two reward spots displayed in the environment, and each round ends with the arrival of an agent to any of them. Additionally, a "tie area" surrounds the reward spots. If a player reaches a particular spot and the other agent is within its tie area, the trial results in a draw, and both agents receive the corresponding payoff. This feature is particularly relevant for the dynamic condition since the players can update their decisions and adjust their trajectories during the

[1] Designed by Hawkins and Goldstone (2019) [3].

round. In this condition, each agent can see their opponent's behavior in real-time and react accordingly. On the contrary, for the ballistic variation, the agents do not have the opportunity to change their course of action once the trial has begun.

2.1.1 Technical setup

To implement the previously described task, we employ the 2D virtual environment developed by Freire et al. [11] for modeling the Control-based Reinforcement Learning (CRL) algorithm. This is a simulated robotic environment, where two mobile machines known as "e-pucks" represent the players. The e-pucks are equipped with six proximity sensors that allow them to perceive the elements in the environment, two wheels (one on each side of the robot) with incorporated motors, and a network of connections that links the sensors with the wheels' motors.

Furthermore, two spots positioned in the environment serve as the two distinct coffee shops, and a circle surrounding each represents the tie area. The reward spots vary in size, depending on whether they embody the high or the small reward, and on every trial, they are randomly allocated to one of two predefined locations. *Figure 1.B* depicts the environment's composition. Moreover, there are two different versions for the virtual environment: one is intended for the ballistic condition, and the second one is for the dynamic interaction within trials.

Each e-puck has three types of sensors which are divided as follows: two sensors for perceiving the high reward spot; two sensors for the low reward spot; and the last two sensors that detect the proximity of the other agent. These three types of sensors are present on both the left side and the right side of the e-puck. *Figure 1.A* illustrates the previous configuration. This arrangement allows the agent to perceive the proximity of different entities and to send this input signal into the wheels' motors through a series of excitatory and inhibitory connections. The motors control the agent's speed according to the input they get from the sensors, weighted by the values of the agent's governing behavioral model.

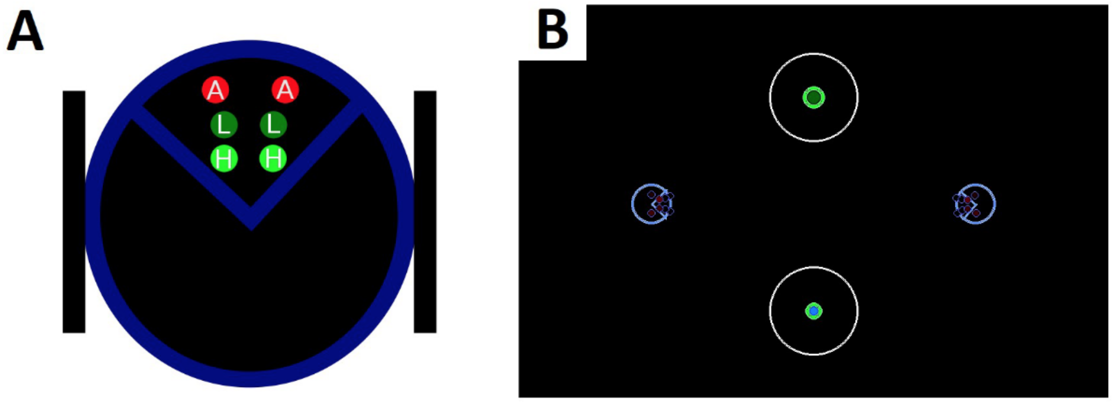


Figure 1: The experimental setup designed by Freire et al., 2020 [11]. *Image A.* Top view of an agent, represented as an "e-puck" in the virtual environment. This particular agent is facing the top of the page. The black rectangles on its sides represent the robot's wheels and the colored circles on the front illustrate the different sensors it possesses: "A" refers to the agent proximity sensors, "L" to the low reward sensors, and "H" to the sensors for the detection of the high reward. *Image B.* Top view of the 2D simulation environment. The blue circles represent the agents (e-pucks) in their initial position, the green spots illustrate the high (large spot) and low (small spot) rewards, and the white circles define the tie area.

2.2 Experimental framework

2.2.1 The SEC model

In order to emulate convention formation, agents are modeled according to the SEC algorithm. This model's approach guides an agent's behavior according to state-action sequences that have been rewarding in recent history. Concretely, the SEC model consists of a short-term memory buffer (E), a long-term episodic memory (EC) component, and an action selection algorithm. *Figure 2* details the aforementioned model's steps.

The SEC model, as proposed by Freire et al. (2021) [20], is structured as follows:

1. First, the model is introduced into an environment that allows it to encounter a series of state-action couplets.

Algorithm 1 – Sequential Episodic Control

E: episodic buffer
EC: episodic memory
for each episode **do**
    Initialize empty E.
    $t = 1$
    while $t < T$ and $r_t = 0$, **do**
        Receive observation o_t from environment.
        Let $s_t = \phi(o_t)$
        Estimate return for each action a via (1)
        Let $a_t \sim \pi\left(\widehat{Q^{EC}}(s_{t,}a)\right)$
        Take action a_t, receive reward r_{t+1}
        Append $(s_{t,}a_t)$ to E
        $++t$
    end while
    if $r_t > 0$ **do**
        Append in $(E,\ r_t)$ in EC
    end if
    Empty E
end for

Figure 2: Sequential Episodic Control Model (SEC) algorithm, proposed by Freire et al., 2021 [20].

2. Then, SEC saves the upcoming input in its short-term memory buffer. In an ordered manner, the model takes in the most recently experienced sequence of state-action pairs.

3. When the short-term memory buffer reaches a total of 50 accumulated couplets, it begins to apply the "first-in, first-out" rule; this means that it begins to drop the oldest couplet in the sequence and updates the buffer with a newly encountered pair.

4. The moment it comes across a reward, the model associates it with the state-action sequence currently saved in the short-term buffer. This sequence-reward pair is then sent into the long-term memory component and saved as a state representation.

5. Now, the action selection algorithm compares each experienced state against all the state representations collected within the long-term memory component

by applying a similarity metric (which gets the Euclidian distance between these states):

$$d(s_t,\ s^{EC}) = \frac{1}{N} \sum_{j=1}^{N} |\ s_{t,j} - s_{i\epsilon EC,j}|$$

6. The model estimates the eligibility scores (G) for the state-action couplets by taking the result of the previous equation and weighting it using a sequential bias (B) to compute the history of recently chosen state-action pairs. This way, the probability of selecting a recurrent sequence increases:

$$G = (1 -\ d(s_t,\ s^{EC}))\ B$$

7. Subsequently, the model selects candidate state-action pairs (C) whose eligibility scores surpassed an absolute and a proportional threshold:

$$C = H\,(G - \theta_{abs})\ \ H(\frac{G}{G_{max}} \theta_{\ prop})$$

8. Next, the model associates the couplet candidates with the reward values assigned to their corresponding stored sequences; these values are normalized to the maximum reward associated with the pairs at hand. In addition, they are weighted using a decay value that depends on how far the couplet is from the end of the sequence:

$$\widehat{Q^{EC}_{a\epsilon A}} = \sum_{i\epsilon C} G_i \frac{r_i}{r_{max}}\ e^{-\delta(i)}\tau$$

This way, the model is biased towards selecting actions that are near bigger rewards.

9. By normalizing the sums of all the relative reward values associated with the candidate state-action couplets, a probability distribution is generated over the discrete action space. And, finally, the model randomly picks an action from this distribution:

$$\widehat{Q^{EC}} = \left[\widehat{Q^{EC}_{a_1}}, \quad \widehat{Q^{EC}_{a_2}}, \ldots, \widehat{Q^{EC}_{a_M}} \right]$$

In summary, the short-term memory buffer transiently stores the most recently experienced sequence of state-action couplets. When the algorithm comes across a reward, it is attached to the current couplet sequence and the pair is sent into the long-term episodic memory to be saved as a state representation. This algorithm will perform randomly in the beginning, but upon gathering observations, it will choose actions more selectively. *Figure 3* illustrates the structure of the SEC model.

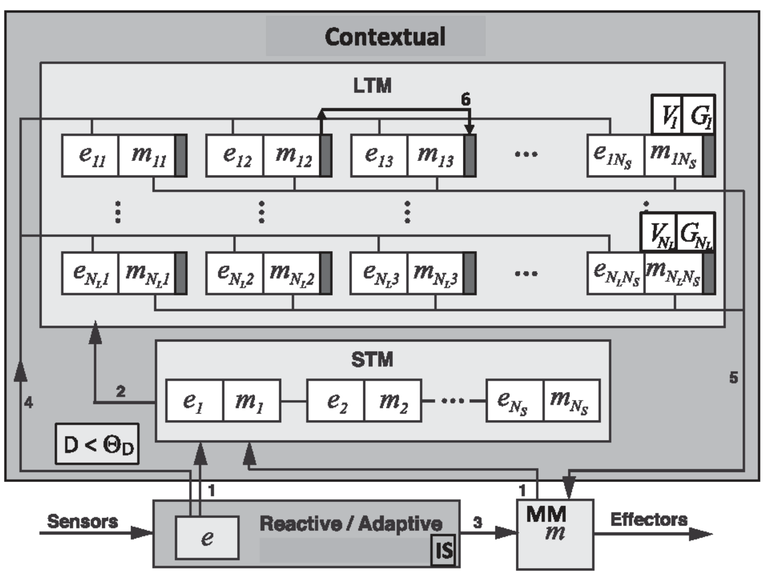


Figure 3: Representation of the Sequential Episodic Control (SEC) model. On the top, the Contextual Layer consists of a short-term memory buffer and a long-term episodic memory component, as well as an action selection algorithm. On the bottom, the Reactive/Adaptive Layer represents the sensorimotor control loop that bridges the Contextual Layer and the environment. In this image, the Sequential Episodic Control Model (SEC) model, proposed by Freire et al., 2021 [20], concerns the Contextual Layer.

2.3 Algorithm and experimental adjustments

Here, we introduce the necessary adjustments in order to adapt the previously described SEC model to the Battle of the Exes task. The first thing to consider is that the structure of the environment is different from the one where Freire et al. (2021) [20] originally implemented the model: the Animal AI testbed [23]. The preceding virtual simulation consisted of a 3D environment where agents received first-person visual input (an 84 by 84 pixels arrangement) at each timestep. Conversely, the technical setup for the simulation at hand is a 2D robotic simulation with two distinct implementations that depend on the continuity of the agents' interactions: the ballistic and the dynamic conditions.

2.3.1 Ballistic condition

For the ballistic condition, just like in the previous implementation, the agents only compute their own payoffs without considering the other player's rewards or the gap between them. However, for this condition, the agents compute both players' actions for each trial. This is, an agent keeps track of its own sequence of selected actions, as well as that of its opponent.

Consequently, we define a trial's "state" as the action sequences that were chosen in the most recently experienced trials by both players; the state is then associated with the agent's selected action for the current trial (namely, the state-action couplet). There are two possible actions an agent can choose on each trial: selecting the low-reward coffee shop is represented with a 0 while setting the course for the high-reward destination is depicted with a 1.

Since states are not continuous real-time occurrences but discrete instances, the short-term memory buffer has a smaller couplet capacity than the original model implementation. Instead of accumulating up to 50 pairs, we designate the length of the buffer to a fixed size that encompasses the ordered lists of the pair of actions selected by the agent and its opponent. In order to determine the appropriate size of a state, we will test different lengths ranging from 2 to 7 trials. This is, we will examine sequences of 6 distinct lengths and choose the value (n) whose result better resembles human behavior. It is worth noting that instead of using the Euclidean distance for comparing the sequence in STM against the ones saved in the LTM, we will use the Hamming distance since it is more accurate for binary vectors.

In addition, given the nature of the task's configuration, every state is also automatically associated with a reward. Hence, unlike the previous implementation, here the algorithm sends a state-action couplet into the long-term memory for each experienced trial. Additionally, the reward associated with the sequence in the short-term memory is equal to the accumulated reward of all the trials encompassed in the state representation and the currently experienced trial.

During the initial trials (when the amount of encountered states is less than n) the empty slots in the short-term memory buffer are filled with -1 values, both for the selected action and the reward value. Even though these rounds' state-action pairs are also stored in the long-term memory, the probability of being picked by the action selection algorithm is low due to the application of the similarity metric, since no posterior sequences include actions equal to -1.

The couplets in the short-term memory update on every trial. However, the short-term memory buffer is not emptied when the sequence is associated with its reward and sent into the long-term memory component. Instead, a copy of the current sequence is sent into the long-term memory and the sequence in the buffer continues to update according to the first-in, first-out rule.

2.3.2 Dynamic condition

For the dynamic condition, the real-time updates are analogous to the ones in the CRL configuration: the state-action couplets are continuous and accumulate within each trial. These updates are registered every 5 timesteps during the simulation. When the round finishes, the sequence stored in the short-term memory buffer is sent into the long-term memory component because it always ends with an agent reaching a reward.

A sequence is composed of the overall state-action couplets experienced within the trial; however, we define the size of the short-term memory buffer as 37 state-action pairs, since 37 is the minimum amount of timesteps needed to reach any given reward (if the agent's behavior is as efficient as possible). This way, all the state representations in the long-term memory possess the same length. Nonetheless, this also means that if an agent takes a long time exploring before reaching a reward, only the last 37 couplets are saved.

Furthermore, reward values saved into the LTM are weighted by an exponential discount associated with the time it takes the agents to arrive at their destination. Needless to say this only applies to positive reward values, so it does not include trials that result in a tie. Thus, if at least one epuck reaches a coffee shop within 37 timesteps, the players will receive the full, corresponding rewards. On the other hand, if it takes them longer to arrive at the reward spots, the value of the payoff starts to decrease following a logarithmic scale. If none of the agents succeeds in reaching either spot in less than 60 timesteps, the round ends with a timeout and starts over.

Additionally, it is worth mentioning that agents can change their course of action at any point during a round in this condition. Therefore, just as with Freire's approach to the Battle of the Exes [11], the agents are equipped with two different layers of behavior: a Reactive Layer, which handles the sensorimotor contingencies of the agent within one round, and an Adaptive Layer, which handles learning across trials.

- The Reactive Layer represents a feedback controller that is implemented as a set of hard-wired sensorimotor control loops and provides the agents with a behavioral baseline. This layer bootstraps learning by allowing the agents to acquire behaviorally relevant information (eg. state-action couplets) in the initial phases of exploration, where the Adaptive Layer hasn't acquired any experience. The Reactive Layer can be conceived as the set of reflexes that are "pre-wired" in living organisms and offer inductive biases to explore the environment [17], thus allowing learning in the Adaptive Layer.

- The Adaptive Layer represents a model-free approach that learns how to exploit rewarding state-action patterns.

More specifically, the Reactive Layer consists of predefined reward-seeking and collision-avoidance behaviors. The reward-seeking conduct is achieved through the agents' proximity sensors, with a crossed excitatory connection combined with a direct inhibitory connection between the reward proximity sensors (s^X) and the wheels' motors (m):

$$\begin{aligned} m_{left} &= f + s^X_{right} - s^X_{left} \\ m_{right} &= f + s^X_{left} - s^X_{right} \end{aligned}$$

where f is a constant forward speed set to 0.3, X can refer to either the high or the low reward, and s^X_{left}, for instance, would be the reward sensor located at the left side of the agent. The sensors are more highly activated when they are closer to the reward spots. Consequently, this configuration causes the more activated sensor to excite the opposite-side motor, thus making the robot turn and approach the input source. Alternatively, if no reward spot is detected, the robot simply advances forward with a speed of f.

As for the collision-avoidance behavior, the combination of the connections between the sensors and the motors is the opposite of the reward-seeking network. What is more, in this configuration the involved sensors are the agent sensors (s^A). As follows, there is a direct excitatory connection and a crossed inhibitory connection between the s^A and the wheels' motors (m):

$$\begin{aligned} m_{left} &= f + s^A_{left} - s^A_{right} \\ m_{right} &= f + s^A_{right} - s^A_{left} \end{aligned}$$

where s^A_{right} refers to the e-puck's right side sensor, which indicates the proximity of the other agent. Just as in the preceding scenario, the sensors are more activated the closer they are to the input source, and if nothing is detected, the agent moves forward with a speed of f. This configuration makes the more activated sensor excite the same-side motor, thus making the robot turn away and avoid the other agent.

The Reactive Layer initializes the e-pucks into exploring the environment in the first trials when they still have no state representations stored in the long-term episodic memory. Essentially, the agents are embedded with the impulse to approximate the rewards present in the environment and the impulse to evade the other player (which is fundamental for the task at hand). In the case of the "other agent"'s sensors, they are turned off entirely once the SEC model takes control in order to avoid them overshadowing the impact of the model and biasing the results against ties. However, they can either be present or absent during the exploration phase. We will be testing for these two possibilities and select the one that results in more human-like behavior.

Moreover, the Reactive Layer interacts with the Adaptive Layer by means of a top-down control mechanism that operates as an inhibitor function: depending on the action selected by the SEC algorithm, either the high reward or the low reward-seeking behavior can be inhibited. For example, if the selected action is to approach the high reward, the low reward-seeking behavior is suppressed.

2.3.3 Experimental modifications

On top of the aforementioned, in order to implement the SEC algorithm into the existing experimental setup we have to apply the following adjustments:

1. The introduction of an exploration phase before the 50 testing trials. In other words, the agents are able to explore the environment by following only the Reactive Layer (before the SEC algorithm is initialized). Different numbers of trials need to be tested to determine the most adequate exploration duration that provides the epucks with a basal knowledge of the technical setup; for

the present implementation, we will assess the values of 20, 30, and 50 trials. During this phase, each agent will choose a random action at the beginning of the trial and store the information provided by their sensors, as well as the reward encountered at the end of each respective trial. For the dynamic condition, this means that the epucks will behave as if they were in the ballistic condition during the exploration trials: the coffee shop selected will not be updated in real-time since only one decision is made at the beginning of the turn.

2. For the purpose of simplification and reducing the number of parameters being tested, the capacity of the long-term memory component is set to equal the total number of trials. Accordingly, this component will be equivalent to 50, plus the number of exploration trials faced by the agent, resulting in a total of 70, 80, or 100 trials. It follows that no forgetting method is applied in the LTM.

3. Saving state-action couplets associated with a zero-value reward into the LTM. In the SEC's previous implementation, a reward that equaled zero was disregarded because it did not provide valuable information to the algorithm. Nonetheless, in the current experimental setup receiving a payoff of zero represents a tie, and thus it distinguishes itself from receiving no reward.

4. The similarity threshold when comparing the STM against the sequences stored in LTM. Given that the configuration of the state-action pairs is different than the one present in the Animal AI testbed, the structural changes in the algorithm need to also adapt to the dimensionality reduction. In other words, the similarity threshold used in the previous experiment may not be adequate for this new implementation; hence, we will examine different threshold values to be surpassed by the candidate state-action pairs' eligibility scores: we will test 0.7, 0.8, and 0.9.

5. The inclusion of negative values associated with a tie. In the present simulation, the spot that delivers a small reward returns a fixed payoff of 1, while

Agent 1

Agent 2		Large	Small
	Large	(0,0)	**(1,4)**
	Small	**(4,1)**	(0,0)

High Stakes, Tie= 0

Agent 1

Agent 2		Large	Small
	Large	(-1,-1)	**(1,4)**
	Small	**(4,1)**	(-1,-1)

High Stakes, Tie= -1

Agent 1

Agent 2		Large	Small
	Large	(-4,-4)	**(1,4)**
	Small	**(4,1)**	(-4,-4)

High Stakes, Tie= -4

Table 3: "Battle of the Exes" payoff matrices for the agents in the virtual simulation, for the high stakes condition with tie values of 0, -1, and -4.

the other reward equals 4 (since the agents are always immersed in the high stakes version of the game). Yet, we will examine increasing the gap between the rewards obtained when the turn ends in a tie and when the players choose different destinations. The tie reward values that we will test are 0, -1, and -4; *Table 3* shows the payoff matrices for these different scenarios.

2.4 Evaluation metrics

The three metrics used to evaluate the SEC model's performance are efficiency, fairness, and stability. The efficiency (E) level is measured by the cumulative sum of the overall rewards that the agents acquired, divided by the largest possible reward that could have been earned by both of them altogether (r_{max}) on each trial:

$$E_t = \frac{\sum_{i=1}^{t} (r_{p_{1i}} + r_{p_{2i}})}{t \times r_{max}}$$

where t refers to the total number of rounds experienced by the agents, i to the current trial, and r_{p1} and r_{p2} to the reward assigned on each trial to Player 1 and Player 2 respectively. If E were to approach 1 it would mean that the agents obtained most of the possible reward they could have acquired; in other words, this result would indicate that they got close to maximizing their overall payoff.

The fairness (F) level is an aggregated value over trials, just like the efficiency metric. It refers to the proportion of times each player received the highest reward on each trial, and then, it is measured by the following normalized payoff ratio:

$$F = \frac{min(h_{p_1}, h_{p_2})}{max(h_{p_1}, h_{p_2})}$$

where h_{p1} refers to the accumulated number of trials in which Player 1 earned the highest reward, and consecutively, h_{p2} concerns the amount of trials in favor of Player 2. If by the end of the simulation F equals 1, it would mean that both agents have obtained the highest reward an equal number of times, and the overall payoff is evenly distributed.

In contrast to the previous metrics, the stability (ST) level regards the predictability of the agents' behaviors. Namely, this metric refers to the persistency of dynamic behavioral patterns throughout the experienced rounds. Following the methodology of Hawkins and Goldstone, ST is measured by computing *"the average surprisal of a Markov Chain encoding the transition probabilities between events on successive rounds"* [5].

First, to measure ST it is necessary to extract the series of the game outcomes belonging to one particular dyad of players; these are encoded as a sequence (S) of states (s). Each s can take up one of the values belonging to the set of outcomes

$\{p1,\ p2,\ d\}$, where *p1* equals "Player 1 wins", *p2* equals "Player 2 wins", and *d* refers to a draw. Now, considering S_m as the set of all the subsequences of length m that conform S, the next phase is to train a Markov Chain of order m for each step t in the time-series:

$$\widehat{P}(X_t|X_{t-1},\ldots X_{t-m}) = \frac{\sum_{s\in S_m} \mathbb{I}_{\{s=X_{t-m}\cdots X_{t-1}X_t\}} + 1/3}{\sum_{s\in S_{m-1}} \mathbb{I}_{\{s=X_{t-m}\cdots X_{t-1}+1\}}}$$

where, for example, Π_{p1} would be the indicator function that equals 1 when *p1* is true. This operation allows keeping a "virtual counter" for the set of outcomes, which converges to the maximum likelihood estimator for each of them.

Following the theory of information proposed by Shannon [24], in order to measure ST we obtain the degree of surprisal for the agents' interactions. Therefore, the next step is to calculate the negative logarithm of the previous probabilities:

$$S(t) = -\log_2[P(x_t|x_{t-1},\ldots,x_{t-m})]$$

where x_t refers to the outcome, and P($x_t|x_{t-1}$,...,x_{t-m}) to the probability of that outcome according to the Markov chain. The surprisals (for the total of trials) integrate a second time-series and finally, to compare conditions at a group level, the last stage is to calculate the mean of the aggregated suprisals for a particular experimental group (eg. for the players in the low stakes and dynamic condition.)

Concretely, measuring ST provides a quantitative metric of how much regularity can exist within the agents' interactions. In this regard, a long period of low surprisals would imply a stable equilibrium. On the contrary, an agent would be "surprised" if it perceives an event that goes against what it has learned (this is, if an improbable event takes place.) During the first trials, when information is yet to be gathered, the degree of uncertainty will always be high, but it is expected to decrease if the players' behaviors start converging. In other words, a higher surprisal level implies lower environmental ST, and contrariwise, a high level of ST entails less uncertainty and, potentially, a robust convention.

2.4.1 Comparison between datasets

After computing the evaluation metrics, we conduct a pairwise comparison of the metrics for the 2x1 conditions between the results of the SEC model and those of the behavioral experiment and an experimental control. If results follow a normal distribution, a one-way ANOVA is applied; if the distribution is non-Gaussian, a Kruskal-Wallis H-test is performed to determine if there is a statistically significant difference between groups. Subsequently, if significant differences are found, we run post-hoc independent tests to compare the human and the control outcomes against the model results, to analyze whether they follow the same tendencies or not. Here, we use an independent samples T-test if the distributions are normal and if they are not, a post-hoc Mann-Whitney U-test. Finally, the results are contrasted against those obtained from the CRL algorithm implementation to see whether they better match the human data.

Chapter 3

Results

In this chapter, we report the results of the SEC model simulations in terms of efficiency, fairness, and stability, across the 2x1 experimental design. The algorithm simulations are measured against the human experimental results from the work done by Hawkins and Goldstein [3] and compared to the results of the CRL modeling done by Freire and collaborators [11]. Additionally, we analyze and interpret the results, and discuss the significance of the human-model comparisons between the SEC and the CRL implementations.

We begin by testing potential parameter values to determine which would make the SEC algorithm better emulate human behavior within the Battle of the Exes framework. The parameters of interest are:

- The value of the payoff associated with a tie, where we test the effect of having reward values of 0, -1, and -4.
- The length of the STM for the ballistic condition, where we examine the following lengths: 2, 3, 4, 5, 6, and 7 components.
- The similarity threshold value that the STM sequence needs to surpass to be considered "similar" to the sequences stored in LTM; the different values to be tested are 0.9, 0.8, and 0.7.

- The number of trials the agents spend exploring and learning about the environment before beginning the 50 experimental trials; here we test whether it is best if epucks spend 20, 30, or 50 trials before accessing the SEC-induced behavior.
- The presence or absence of the "other agent"'s reactive sensors during the dynamic condition's exploration phase.

After fine-tuning SEC's parameters, we select the most adequate values for each experimental condition. [1] Ultimately, for the dynamic condition, the parameter combination that shows the closest resemblance with human data is the one with a similarity threshold of 0.8, 10 exploration trials, a tie reward of -4, and no "avoid epuck" sensors. As for the ballistic condition, we have multiple potential candidates so two options are selected: keeping constant the values of 0.8 for the similarity threshold and 10 exploration trials, alternative *SEC A* has a tie reward of -4 and a STM length of 4, and alternative *SEC B* presents a tie reward of 0 and a STM length of 2.

3.1 Efficiency, fairness, and stability scores

3.1.1 Ballistic condition

The bar charts in *Figure 4* show the efficiency, fairness, and stability mean scores in the Battle of the Exes, for the ballistic condition. These plots illustrate the comparison between the behavioral data from the original Goldstein and Hawkins's experiment [3], the outcome from the randomized control agents, the results from the modeling done by Freire et al. [11] with the CRL algorithm, and the results from the *SEC A* and *SEC B* simulations.

Regarding the efficiency results, the post-hoc Mann-Whitney U-tests show that the difference between the experimental control and the *SEC A* scores is not statistically

[1] For more details on how these values are selected, continue to the following section (*3.2 Parameter fine-tuning*).

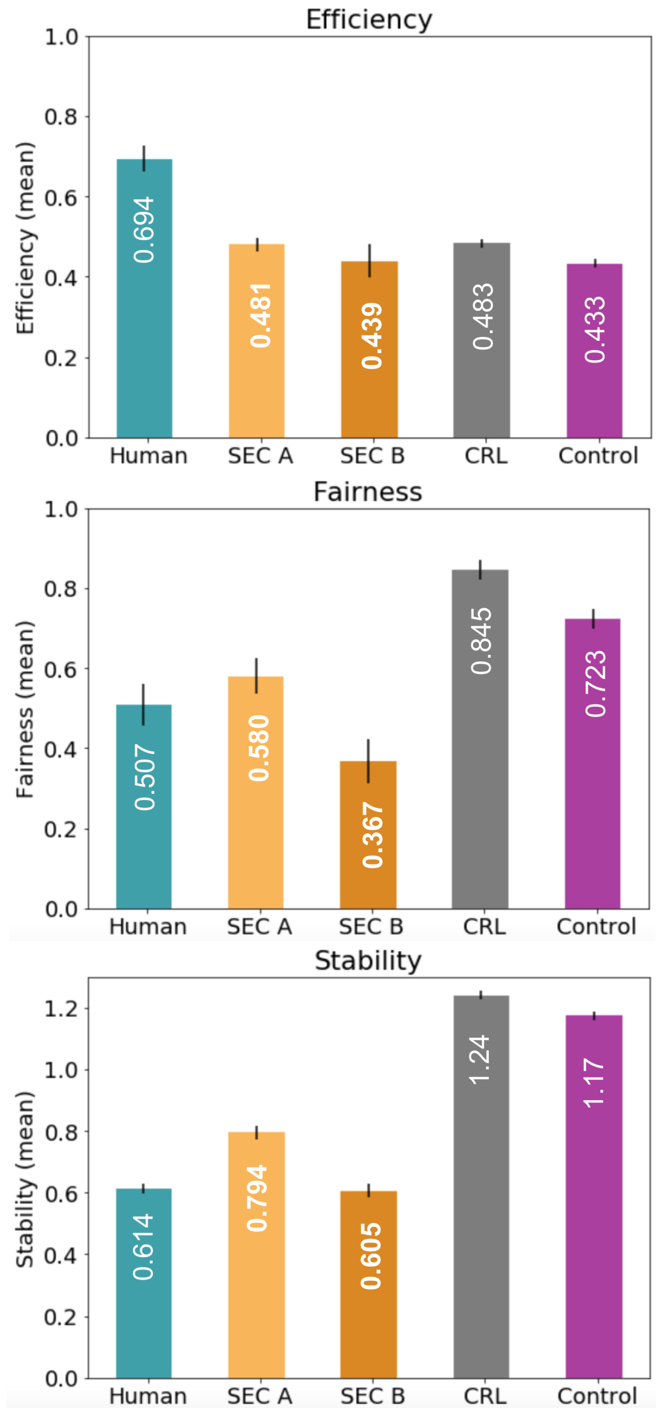


Figure 4: Results of the SEC algorithm compared to behavioral data, an experimental control, and the CRL algorithm in the Battle of the Exes task. These bar plots concern the comparison of the Efficiency, Fairness, and Stability mean scores for the Ballistic Condition. *Human* refers to the results from Goldstein and Hawkins [3]; *Control* refers to the results from a randomized simulation of the task; *CRL* refers to the results from Freire et al. [11]. As for the SEC algorithm results, *SEC A* is defined by the parameter values of 10 exploration trials, a similarity threshold of 0.8, a tie reward of -4, and a STM length of 4 elements; *SEC B* is defined by the parameter values of 10 exploration trials, a similarity threshold of 0.8, a tie reward of 0, and a STM length of 2 elements.

significant ($p=0.150$); neither is there a significant difference between the experimental control and the *SEC B* results ($p=0.270$). On the contrary, when compared against the behavioral data, both *SEC A* ($p=0.00013$) and *SEC B* ($p=7.53e-06$) show a statistically significant difference. In addition, there is as well a statistically significant difference between the experimental control and the human results ($p<0.001$).

As for the fairness metric, post-hoc Mann-Whitney U-tests show a statistically significant difference between the experimental control and both SEC implementations, *SEC A* ($p=3.059e-05$) and *SEC B* ($p=1.144e-05$). In contrast, the differences between human results and the *SEC A* scores ($p=0.223$), as well as human results in comparison to *SEC B* ($p=0.058$), are not statistically significant. Lastly, there is no significant difference between the experimental control and the human results ($p=0.44$).

Lastly, in regards to the stability metric, the post-hoc Mann-Whitney U-tests show a statistically significant difference between the experimental control and the results of both of the SEC simulations, *SEC A* ($p=1.502e-165$) and *SEC B* ($p=2.017e-173$). As for the comparison between the human results and the SEC scores, there is no significant difference between neither for *SEC A* ($p=0.362$) nor for *SEC B* ($p=0.056$). Furthermore, there is as well a statistically significant difference between the experimental control and the human results ($p<0.001$).

3.1.2 Dynamic condition

The bar charts in *Figure 5* show the efficiency, fairness, and stability mean scores in the Battle of the Exes, for the dynamic condition. In these graphs, we can observe the comparison between the behavioral data from the original Goldstein and Hawkins's experiment [3], the outcome from the randomized control agents, the results from the modeling done by Freire et al. [11] with the CRL algorithm, and the results from the *SEC* model simulations.

Regarding the efficiency results, the post-hoc Mann-Whitney U-tests show that the

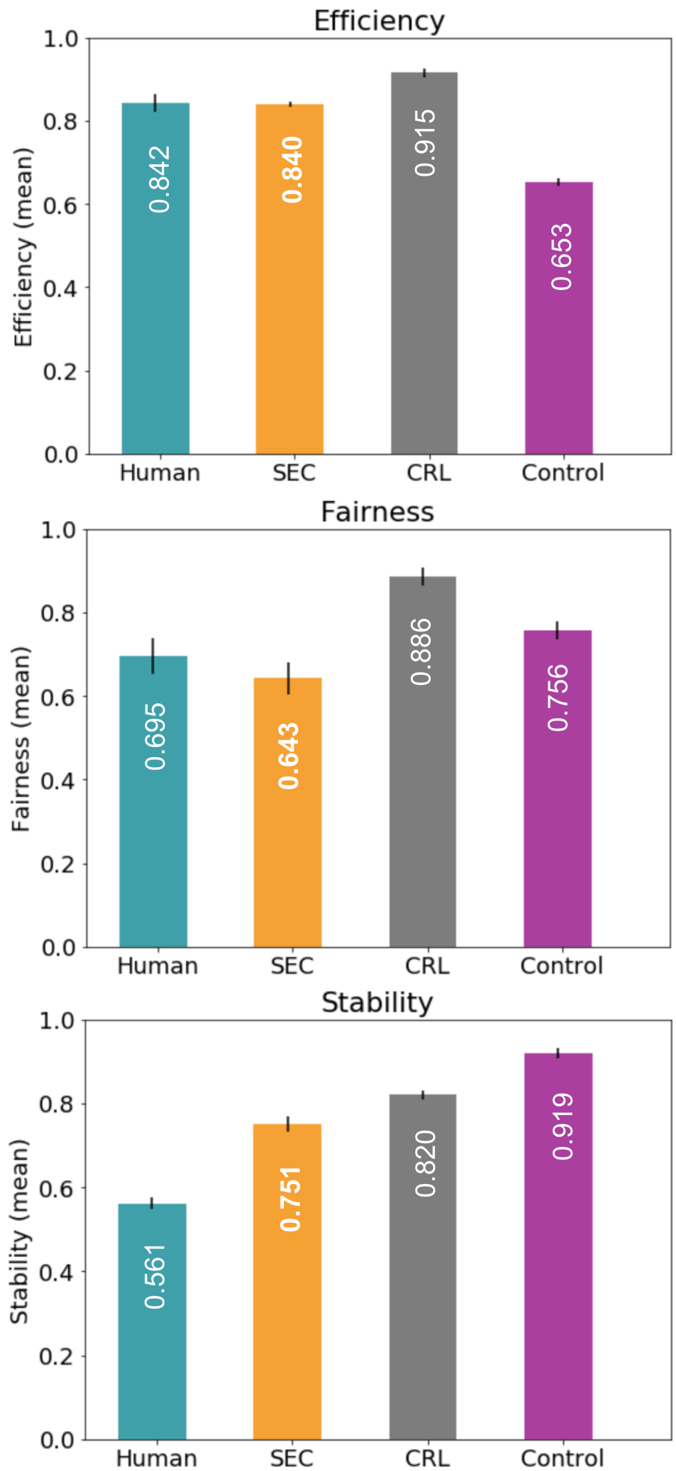


Figure 5: Results of the SEC algorithm compared to behavioral data, an experimental control, and the CRL algorithm in the Battle of the Exes task. These bar plots concern the comparison of the Efficiency, Fairness, and Stability mean scores for the Dynamic Condition. *Human* refers to the results from Goldstein and Hawkins [3]; *Control* refers to the results from a randomized simulation of the task; *CRL* refers to the results from Freire et al. [11]. As for the SEC algorithm results, *SEC* is defined by the parameter values of 10 exploration trials, a similarity threshold of 0.8, a tie reward of -4, and no "avoid epuck" sensors.

difference between the experimental control and the SEC model scores is statistically significant ($p=1.255e\text{-}17$). Similarly, when compared against the behavioral data, the SEC simulations' results show a statistically significant difference ($p=0.0038$). On the other hand, there is no statistically significant difference between the experimental control and the human results ($p=0.26$).

As for the fairness metric, post-hoc Mann-Whitney U-tests show a statistically significant difference between the experimental control and the SEC scores ($p=0.0391$). Also, the difference between human results in comparison to the SEC results is statistically significant ($p=0.0093$). Lastly, there is also a significant difference between the experimental control and the human results ($p=0.04$).

As for the stability scores, the post-hoc Mann-Whitney U-tests show a statistically significant difference between the experimental control and the SEC simulations' results ($p=2.214e\text{-}135$). In regards to the comparison between the human results and the SEC scores, there is as well a significant difference between them ($p=1.532e\text{-}78$). Finally, there is also a statistically significant difference between the experimental control and the human results ($p<0.001$).

Entropy levels through trials

Another element to take into consideration is the evolution of the epucks' experienced entropy throughout a dynamic game. Ideally, if the SEC model is able to learn about the agents' interactions, it should present a reduction in the previously mentioned metric's values while approaching the final turn. In order to look into this, we also measure the agents' mean entropy levels turn by turn. We start by extracting each agent's experienced entropy level for each timestep. Then, for every turn, we then get the mean entropy for all the players in the group, and, finally, we plot the results. *Figure 6* shows the evolution of the entropy levels for the SEC implementation with a similarity threshold of 0.8, a tie reward of -4, and no "avoid epuck" sensors.

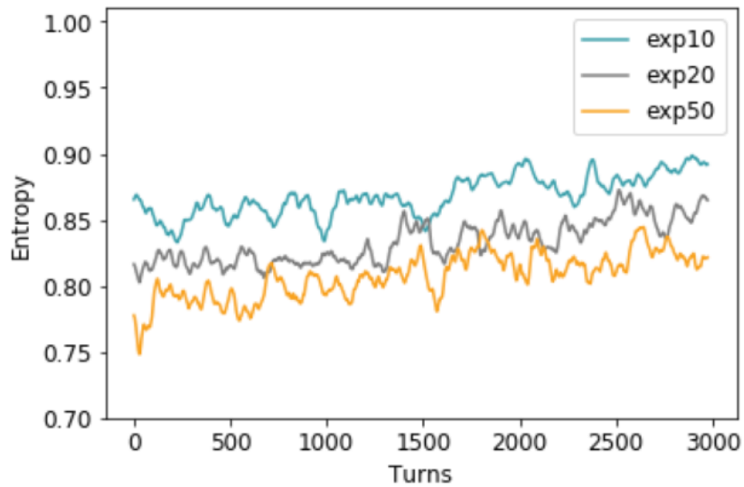


Figure 6: Evolution of the mean entropy levels (turn by turn) for the agents interacting in the Dynamic Condition, while considering a parameter combination of a similarity threshold of 0.8, a tie reward of -4, and no "avoid epuck" sensors. The average entropy value for 10 exploration trials was equal to 0.868; for 20 exploration trials, 0.833; and for 50 exploration trials, 0.807.

3.2 Parameter fine-tuning

In order to select the best SEC parameter values for the foregoing comparisons, we need to test all possible combinations for all of the parameters under consideration (for the ballistic and dynamic conditions). Additionally, these combinations are also contrasted against a randomized control group. *Appendix A* includes an overall comparison of all of the aforementioned combinations in terms of the accumulated reward through trials and the moving average.

Firstly, we look at the graphical comparison of the accumulated reward through trials for each combination. This value comes from the rewards both players earn on each trial and it gives us a preliminary idea of how adequately the model matches human performance. We observe in these graphics that none of the model's variations achieve a performance that maximizes the accumulated reward through a game as well as humans do. The moving average across trials shows the same tendencies.

As we can see from the examples in *Figure 7* and *Figure 8*, the human results outperform all of the model's results, no matter which parameter combination is

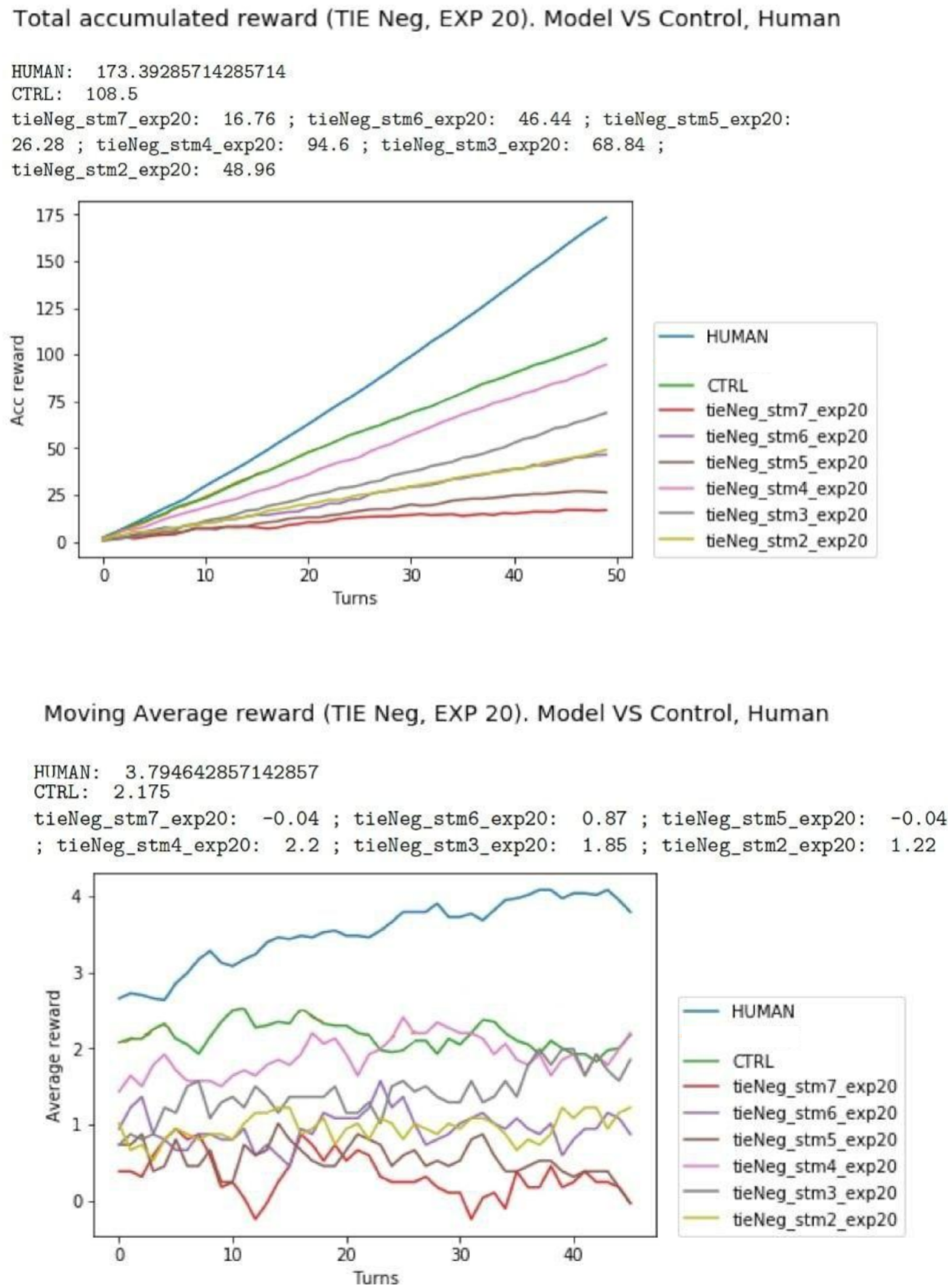


Figure 7: Examples of the Accumulated Reward and the Moving Average comparisons between the human data (blue), the experimental control (green), and the different parameter combinations, for the Ballistic Condition. The top graph shows the Accumulated Reward throughout all the turns in a game, and the bottom one presents the Moving Average across trials. These are the results from running the simulation with the following parameter combination: Ballistic condition, similarity threshold = 0.7, tie reward = -1, exploration = 20 trials, and all of the STM lengths.

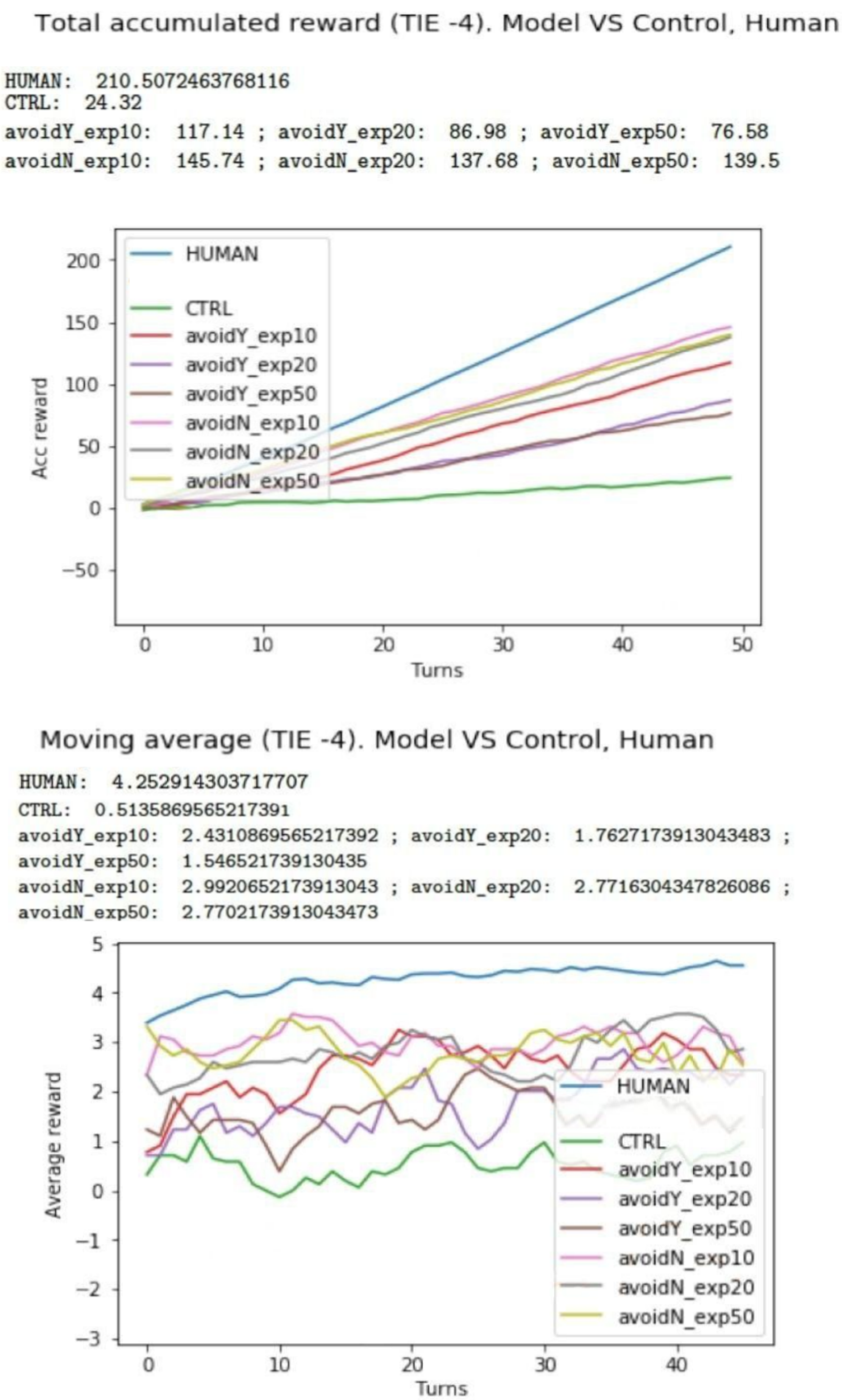


Figure 8: Examples of the Accumulated Reward and the Moving Average comparisons between the human data (blue), the experimental control (green), and the different parameter combinations, for the Dynamic Condition. The top graph shows the Accumulated Reward throughout all the turns in a game, and the bottom one presents the Moving Average across trials. These are the results from running the simulation with the following parameter combination: Dynamic condition, similarity threshold = 0.8, tie reward = -4, all of the "avoid epuck" options, and all of the exploration trials.

selected. In addition, we observe that the control condition always does better than any of the parameter combinations for the ballistic condition, but the model implementations have better results in the dynamic condition; this is true for both, the accumulated reward and the moving average graphs.

It is worth noting that, even though the SEC results do not seem to exactly replicate the behavioral ones, our objective here is to try to identify the parameter combination that emulates them as close as possible. With that in mind, these graphs present us with preliminary indicators that show potential values we could select. For example, in *Figure 7* we can see that the parameter combination that includes a STM length of 4 elements could be the most adequate for the ballistic condition, whereas selecting a STM with a length of 7 components appears to have the worst performance. Likewise, *Figure 8* seems to indicate that defining 10 exploration trials and turning off the avoid "other epuck" sensors might be an adequate choice for the dynamic condition.

3.3 Graphical comparison

In order to define whether SEC can appropriately model human behavior in the task at hand, we evaluate its performance in terms of efficiency, fairness, and stability. To begin with, we compute the evaluation metrics and qualitatively compare the means for all of the parameter combinations discussed in the previous section against the experimental control and the human data. The following sections will illustrate some examples to provide us with a comprehensive idea of which might be the parameter combinations that emulate human results most adequately; to examine the complete set of graphical comparisons for each metric for all the parameter combinations, see *Appendix B*.

3.3.1 Ballistic condition

In the first place, we can look at the particular example of defining a similarity threshold of 0.8 and a STM length of 4, in the ballistic condition. *Figure 9* refers to the metric of efficiency, and we can see that a combination of 20 exploration trials and a tie value of -1 gets closer to human results than the same combination but with 10 exploration trials. Nonetheless, the rest of the parameter combinations also seem to be closer to the experimental control than to the human results; this is true for all of the combinations tested for efficiency in this condition.

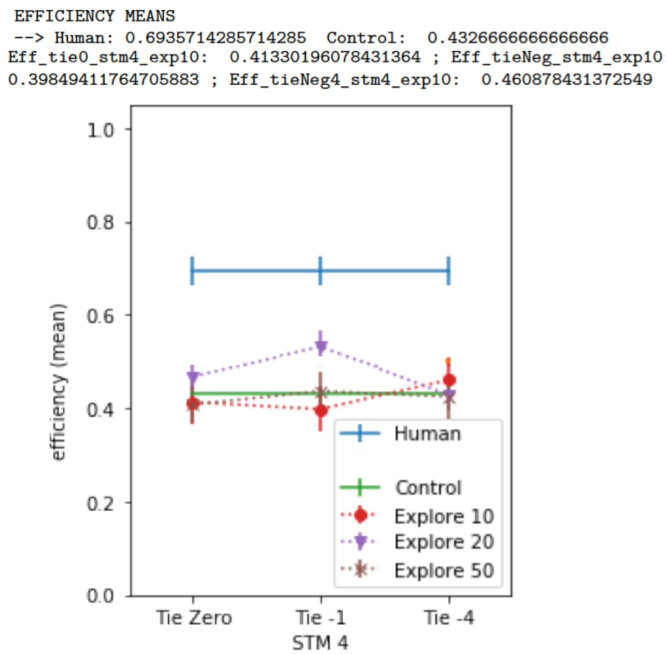


Figure 9: Efficiency results for the Ballistic Condition, when considering a similarity threshold of 0.8 and a STM length of 4 elements.

Regarding the fairness metric for the same exemplified parameter values, *Figure 10* shows how all the model's results seem closer to the human data than to the experimental control except when selecting 10 exploration trials, which seems to have a detrimental impact. However, in contrast to the efficiency scores, the algorithm does not have a consistent outcome when measuring fairness. When defining the similarity threshold as 0.7, for instance, the model obtains lower results than humans (similar to what happens in *Figure 10* when selecting 10 exploration trials); and when defining it as 0.9, the model's results tend to be more similar to the experimental control.

Similarly, the stability results for the ballistic condition seem to show a human-like tendency. As we can see in *Figure 11*, the model appears to approximate human

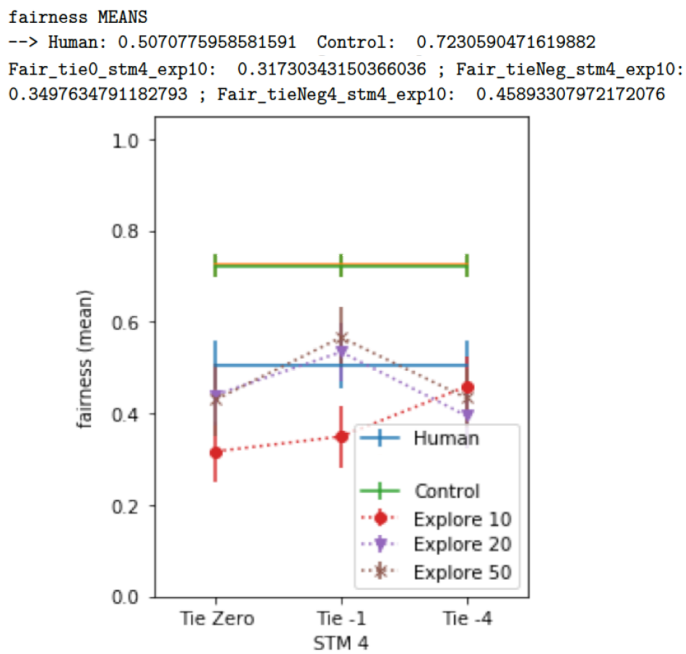


Figure 10: Fairness results for the Ballistic Condition, when considering a similarity threshold of 0.8 and a STM length of 4 elements.

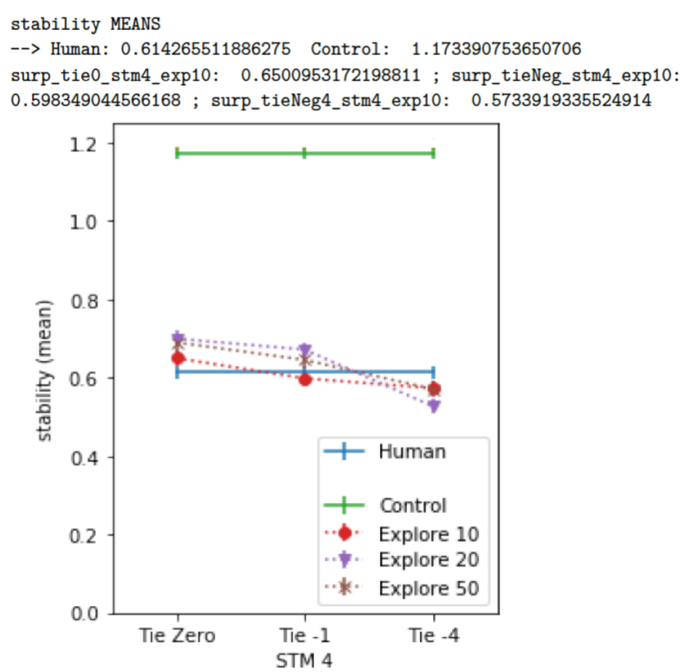


Figure 11: Stability results for the Ballistic Condition, when considering a similarity threshold of 0.8 and a STM length of 4 elements.

data when setting a similarity threshold of 0.8 and a STM length of 4. In general, these results do not generalize to the rest of the parameter combinations: it would appear that setting longer STM lengths and higher similarity thresholds increases the Stability values, sending them closer to the experimental control and away from the human data.

3.3.2 Dynamic condition

In the dynamic condition, the SEC algorithm's outcome for the efficiency metric appears to be closer to the experimental control than to the human data. This can be observed in *Figure 12*, which illustrates the results that arise from a similarity threshold of 0.8 and turning off the epuck's sensor in charge of detecting the other agent. In this example, however, we can also see that efficiency seems to improve when implementing a tie reward of -4; this tendency is also perceived when setting a similarity threshold to 0.7.

The outcomes for the fairness metric in the dynamic condition seem less clear given

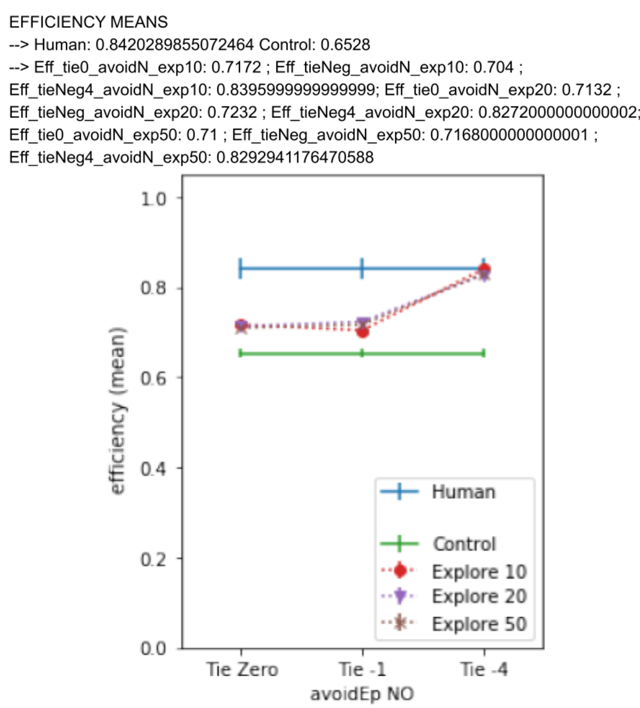


Figure 12: Efficiency results for the Dynamic Condition, when considering a similarity threshold of 0.8 and no "avoid epuck" sensors.

that the behavioral results and the control values are close to one another. As we can see in the example of *Figure 13*, the results from the current set of parameters also surround these values; this also applies to the rest of the combinations for this metric and this condition. It should be noted that there appears to be a slight improvement when setting the tie value to -4 (except when the similarity threshold is set to 0.9).

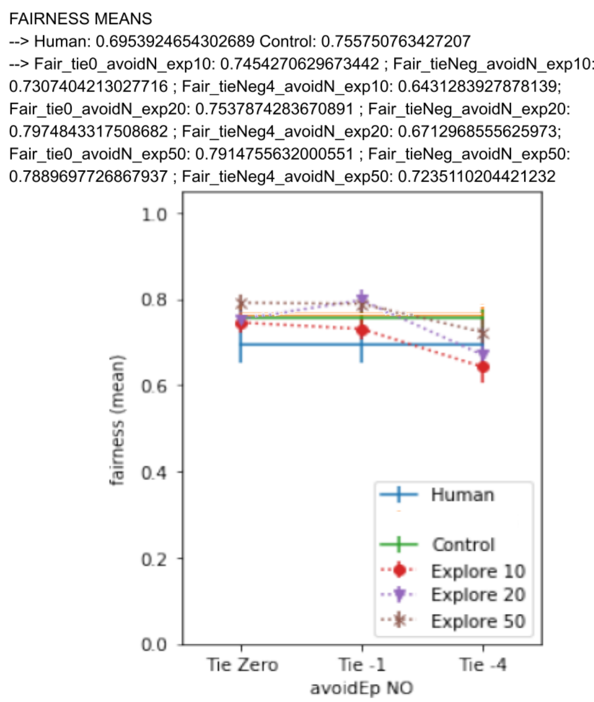


Figure 13: Fairness results for the Dynamic Condition, when considering a similarity threshold of 0.8 and no "avoid epuck" sensors.

Finally, when measuring stability for this condition we can observe in the example of *Figure 14* that the model's results appear to be closer to the experimental control than to the behavioral data, with a slight improvement when setting the tie value to -4. The latter seems slim although, in general, it also seems to be more noticeable when setting 10 exploration trials. The rest of the parameter combinations present a similar tendency as the one present in the current example.

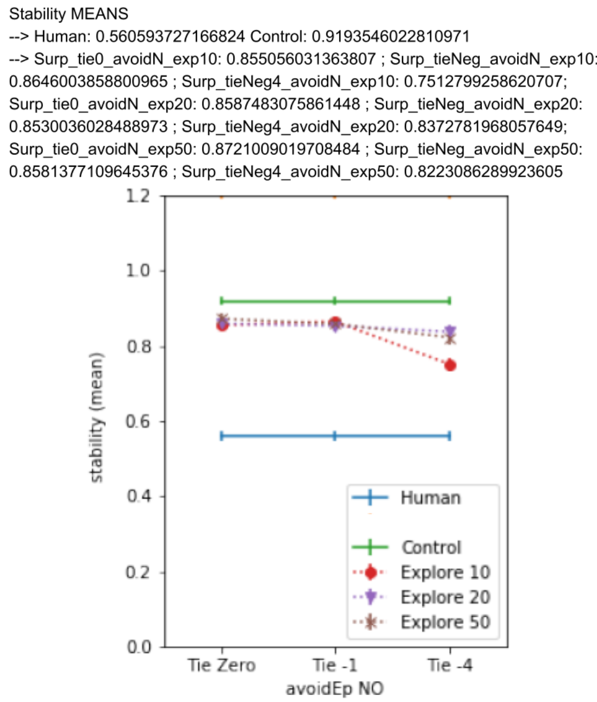


Figure 14: Stability results for the Dynamic Condition, when considering a similarity threshold of 0.8 and no "avoid epuck" sensors.

3.4 Distribution comparison

The previous graphical comparisons provide an overview of which could be the most suitable parameter values for modeling the Battle of the Exes with the SEC algorithm. From these comparisons, we can presume, for example, that setting a similarity threshold of 0.8 and a tie reward of -4 might allow the model to better approach human results. However, with the intention of doing an exhaustive search, we now conduct a pairwise comparison between the model variations' results, the experimental control, and the human data to determine if the distributions are significantly different or not.

We begin by verifying whether the results follow a normal distribution, to determine if we apply a one-way ANOVA or a Kruskal-Wallis H-test to show if there is a statistically significant difference between groups. After running the tests, all of the results indicate differences between the model when compared to the human data and the experimental control ($p<0.05$). Subsequently, we run post-hoc independent tests to compare each parameter combination against the behavioral distribution (an independent samples T-test or a post-hoc Mann-Whitney U-test, depending on whether the distributions are normal or not); we repeat the procedure for the experimental control.

Our aim is to find distributions that show no differences compared to the human data while also being significantly different from the experimental control. *Appendix B* contains the lists with the distributions that showed no difference when compared against the behavioral and the control data. Overall, these comparisons indicate the following:

Efficiency

Regarding the efficiency metric, the results show that human performance is not achieved for either experimental condition, no matter which parameter values are selected. In particular, the results are worse for the ballistic condition since most of the parameter combinations show no differences when compared to the control distribution.

Fairness

For the ballistic condition, the fairness scores show no differences when compares to most of the model variations, especially when the similarity threshold is set to 0.8 and 0.9. On the other hand, the results for the dynamic condition are ambiguous: even though several combinations show no differences when compared against the human data, almost all of them also show no differences when contrasted against the experimental control. Moreover, the behavioral and the control distributions show no differences when compared to one another ($M = 1513.0$; $p = 0.127$).

Stability

For the stability metric in the ballistic condition, several parameter configurations show no differences when compared to the human distribution; these configurations present distinct tie values, number of exploration trials, and STM lengths (with the exception of STM with 6 and 7 components). For the dynamic condition, however, the results show significant differences when compared with the behavioral data for all possible parameter combinations.

Chapter 4

Conclusions and discussion

In the present thesis, we have examined the Sequential Episodic Control (SEC) model's role in forming social conventions in the Battle of the Exes task. For this work, we carried out a 2x1 between-subjects study simulating agents that interacted in a repeated, multi-agent game, and learned following the SEC algorithm. SEC is an episodic control learning model developed by Freire et al. [20]. It uses tabular memory to store previously experienced rewarding states and, contrarily to other approaches, it saves whole sequences of state-action couplets instead of discrete observations. This way, the model favors the selection of a recurrent pattern of states and actions that is shown to be rewarding.

Given that the conventionalization process entails a predictable nature, SEC's ability to potentially identify and engage in behavioral patterns made it a seemingly suitable contender to try to resemble the human data reported by Goldstone and Hawkins [3]. In addition, its model-free approach could allow it to optimize the time needed for achieving coordination, thus tackling the data inefficiency issue faced by alternate models.

Since this work is the first approach to simulating the Battle of the Exes task with the SEC algorithm, the model's parameters had to be adjusted to fit the behavioral data as closely as possible. A total of 5 different parameters were tested, and the resulting combinations that best emulated human data were, for the dynamic condition, a similarity threshold of 0.8, 10 exploration trials, a tie reward of -4, and no "avoid epuck" sensors. As for the ballistic condition, keeping constant the values of 0.8 for the similarity threshold and 10 exploration trials, two options were taken into consideration: alternative *SEC A*, with a tie reward of -4 and a STM length of 4, and alternative *SEC B*, which included a tie reward of 0 and a STM length of 2.

4.1 Efficiency, fairness, and stability performance

After executing the simulations with the ballistic and dynamic implementations of the SEC algorithm, we found that the model's results do not reach human-level performance for all three considered evaluation metrics. Regarding the ballistic condition, the proposed SEC configuration shows statistically significant differences when compared against the behavioral data; however, no differences were found when comparing the fairness and stability scores. On the other hand, the three metrics presented statistically significant differences for the dynamic condition when comparing SEC's results against human performance.

The SEC algorithm showed an improvement with respect to the stability metric, when compared to the Control-based Reinforcement Learning (CRL) [11] previous approach to modeling the task at hand. In particular, in the ballistic condition, it achieved no differences when comparing the distributions of the human and model scores. Furthermore, even though the dynamic condition results did not emulate human data, the SEC algorithm scores got closer to the human results than those obtained by the CRL simulations.

Regarding the efficiency and fairness metrics, SEC only showed no differences when comparing the human and model fairness' distributions in the ballistic condition; this result is consistent with what was achieved by the CRL simulations. However, for the rest of the comparisons, the SEC simulations achieved statistically significant differences between the human and the model scores. Thereupon, in contrast to the CRL results, efficiency levels are not achieved for the SEC implementation in the dynamic condition.

4.2 Mechanisms underlying conventionalization

Ballistic condition

It is worth noting that, for the ballistic condition, several parameter combinations (besides the *SEC A* and *SEC B* configurations) achieved no differences when comparing both, their fairness and stability results, to the human data. This is an indicator of a seeming consistency present in the execution of the ballistic version of the model; however, it is not clear which parameter values have the most weight over the results (in other words, there is no clear indicator of which STM length, which tie value, etc. are responsible for these scores).

Additionally, human efficiency levels could not be emulated and, even more, the model's results showed no statistically significant differences when compared to the experimental control. Although the model is sample-efficient and reaches a fair and stable performance in a few trials, it still does not follow the same mechanisms humans do. It appears that incorporating SEC's sequentiality leads to the emergence of interaction patterns within trials for the ballistic scenario. However, inefficient behavioral patterns, although stable, do not lead to a result as optimal as human performance.

Dynamic condition

For the dynamic condition, the comparison between the SEC model and the human results showed statistically significant differences with the current parameter adjustments (a tie reward equal to -4, no "other agent" sensors, an exploration phase of 10 trials, and a similarity threshold of 0.8). This means that, so far, we have not been able to emulate human behavior for this experimental condition; however, it is worth mentioning that the aforementioned distributions were not that far apart from one another. In addition, as previously stated, the stability metric improved in comparison to the CRL simulation results. Thus, this version of the algorithm appears to draw nearer to an adequate sample-efficient model in a real-time, dynamic environment.

Furthermore, in contrast to the ballistic scenario, for this condition is clearer which parameter values are responsible for best approaching human performance. To begin with, defining 10 as the number of turns for the exploration phase indicates that having too many training experiences might be detrimental to the subsequent algorithm's performance. This could be because spending more turns exploring might quickly flatten the learning curve; hence, our findings indicate that it is better to initialize the testing phase early. As for the similarity threshold, it is worth reminding that for this implementation we used the Hamming distance instead of the Eucledian distance due to the states' reduction; thus, it follows that the original 0.9 value was not necessarily going to be adequate.

Likewise, setting the tie value to -4 improved the overall results. This finding is in line with Goldstone and Hawkins' conclusion about the cost-efficiency of preventing repeated coordination efforts when players are engaged in the high stakes condition [3]. This particular finding is also congruent with what Freire et al. [20] conclude about comparing high and low stakes, since increasing the difference between rewards resulted in better results. Moreover, this could be an indicator that agents are not indifferent but averse to ties.

Regarding the sensors that detect another epuck, we originally thought that the algorithm's performance could be enhanced by including more variance in the epucks' experienced states and behavior during the training turns. It is useful to remember that once the SEC algorithm takes control over the agent, these sensors are turned off entirely: if they were kept active, they would overshadow the effects of the model and bias the results away from ties. Nonetheless, our findings seem to indicate that it is more effective to keep this value the same during the training and testing phases.

Behavioral coordination may surface from within the dynamics of the interaction, but it does not guarantee conventionalization. For this reason, we also measured the mean entropy levels through the dynamic condition trials. We observed that the entropy levels stayed more or less constant throughout the game, with an average value of 0.868 (with values ranging from 0 to 1, inclusive). Despite that this was a reasonable value, the overall entropy tendency did not decrease while approaching the final turn. Learning would imply a reduction in the levels of uncertainty, which could come from the ability to identify behavioral patterns. Therefore, it is possible that the SEC model is not learning from the agents' interactions and, consequently, conventionalization might not be guaranteed for the current implementation.

4.3 Limitations and further research

In general, this work's findings suggest that an episodic control algorithm could potentially succeed in achieving behavioral coordination in repetitive tasks such as the Battle of the Exes. However, more work remains to be done. One extension for this thesis could be testing the current parameter value combination for the low stakes condition from the original Hawkins and Goldstone's experiment. Another evident expansion would be to implement a less altered version of the original SEC model, by limiting the capacity of the long-term memory component and including a forgetting mechanism.

It is worth highlighting that computational power limited the parameter values that could have been explored. For future research, we recommend expanding the parameter search, especially around the values that have shown promise. For instance, we could explore the prospect of "tie aversion" by trying out even lower reward values, or we could test more values close to the similarity threshold of 0.8 for the dynamic condition. This computational power limitation also restricted the number of game simulations; if possible, we also recommend increasing the number of simulations executed for each condition.

Furthermore, we recommend testing different configurations for the SEC model. In the present work, we adapted the algorithm to better fit both, the ballistic and the dynamic conditions. Nonetheless, there is no assurance that the current structure is the most competent for the task at hand. For the ballistic scenario, one possibility could be to define a state as the sequences of past rewards, instead of actions, or possibly to define it as a combination of the two values together. As for the dynamic condition, we suggest beginning by computing information about their opponents into the states of the agents, in order to make this version of the model more social.

Finally, it would be interesting to try combining SEC with a model-based algorithm to test whether it improves its results during the later stages of interaction; or even to test its performance in more complex tasks where, for example, learning needs to be generalized. Likewise, since small group dynamics can be relatively different from those that emerge in larger groups [4], one focus for future research could be to test the SEC model in a task with more agents. In doing so, we could investigate whether the SEC model alone, or SEC enhanced with a model-based algorithm, can manage to emulate human-like performance.

List of Figures

List of Tables

Bibliography

[1] Köster, R. *et al.* Model-free conventions in multi-agent reinforcement learning with heterogeneous preferences (2020). URL `http://arxiv.org/abs/2010.09054`.

[2] Lewis, D. Convention: A philosophical study (1969).

[3] Hawkins, R. X., Goodman, N. D. & Goldstone, R. L. The emergence of social norms and conventions (2019).

[4] Centola, D. & Baronchelli, A. The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 1989–1994 (2015).

[5] Hawkins, R. X. & Goldstone, R. L. The formation of social conventions in real-time environments. *PLoS ONE* **11** (2016).

[6] Binmore, K. G. *Natural justice*, 1–19 (Oxford University Press, New York, 2005).

[7] Luce, R. D. & Raiffa, H. *Games and decisions: Introduction and critical survey* (Courier Corporation, 1989).

[8] Lau, S.-H. P. & Mui, V.-L. Using turn taking to mitigate coordination and conflict problems in the repeated battle of the sexes game. *Theory and Decision* **65**, 153–183 (2008).

[9] Sugden, R. Spontaneous order. *Journal of Economic perspectives* **3**, 85–97 (1989).

[10] Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature* **13**, 1–25 (2002).

[11] Freire, I. T., Moulin-Frier, C., Sanchez-Fibla, M., Arsiwalla, X. D. & Verschure, P. F. Modeling the formation of social conventions from embodied real-time interactions. *PLoS ONE* **15** (2020).

[12] Huys, Q. J., Cruickshank, A. & Seriès, P. Reward-based learning, model-based and model-free. In *Encyclopedia of Computational Neuroscience*, 1–10 (Springer New York, 2014).

[13] Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J. & Graepel, T. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037* (2017).

[14] Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).

[15] Silver, D. *et al.* A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science* **362**, 1140–1144 (2018).

[16] Blundell, C. *et al.* Model-free episodic control (2016). URL `http://arxiv.org/abs/1606.04460`.

[17] Botvinick, M. *et al.* Reinforcement learning, fast and slow. *Trends in cognitive sciences* **23**, 408–422 (2019).

[18] Lin, Z., Zhao, T., Yang, G. & Zhang, L. Episodic memory deep q-networks. *arXiv preprint arXiv:1805.07603* (2018).

[19] Zhu, G., Lin, Z., Yang, G. & Zhang, C. Episodic reinforcement learning with associative memory (2020).

[20] Freire, I. T., Amil, A. F. & Verschure, P. F. Sequential episodic control. *arXiv preprint arXiv:2112.14734* (2021).

[21] Freire, I. T., Amil, A. F., Vouloutsi, V. & Verschure, P. F. M. J. Towards sample-efficient episodic control with dac-ml (2020). URL `http://arxiv.org/abs/2012.13779`.

[22] Brownlee, J. *Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch*, 6–24 (Machine Learning Mastery, 2016).

[23] Beyret, B. *et al.* The animal-ai environment: Training and testing animal-like artificial cognition. *arXiv preprint arXiv:1909.07483* (2019).

[24] Shannon, C. E. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review* **5**, 3–55 (2001).

Appendix A

Parameter fine-tuning

BALLISTIC CONDITION, SIMILARITY THRESHOLD = 0.7

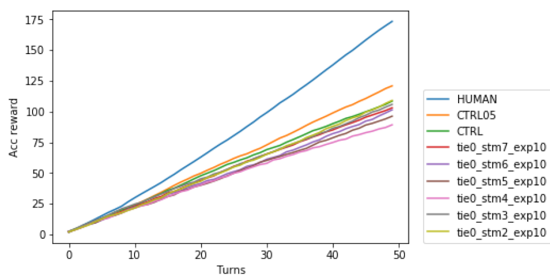
3 Model data VS Control (randomized) and Human data

3.1 Accumulated reward through trials

3.1.1 Tie reward = 0

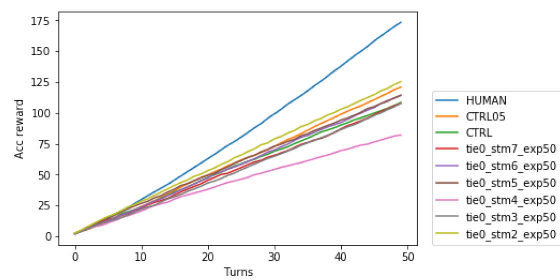
```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tie0_stm7_exp10:  103.0 ; tie0_stm6_exp10:  101.4 ; tie0_stm5_exp10:  96.2 ;
tie0_stm4_exp10:  89.2 tie0_stm3_exp10:  105.8 ; tie0_stm2_exp10:  109.0
```

Total accumulated reward (TIE 0, EXP 10). Model VS Control, Human



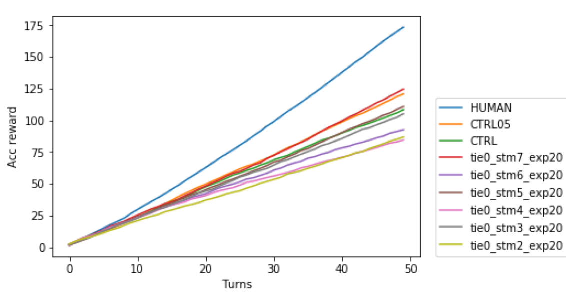

```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tie0_stm7_exp50:  107.6 ; tie0_stm6_exp50:  114.2 tie0_stm5_exp50:  114.2 ;
tie0_stm4_exp50:  82.0 ; tie0_stm3_exp50:  107.6 ; tie0_stm2_exp50:  125.4
```

Total accumulated reward (TIE 0, EXP 50). Model VS Control, Human




```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tie0_stm7_exp20:  124.6 ; tie0_stm6_exp20:  92.6 ; tie0_stm5_exp20:  111.0 ;
tie0_stm4_exp20:  84.6 ; tie0_stm3_exp20:  105.2 ; tie0_stm2_exp20:  87.0
```

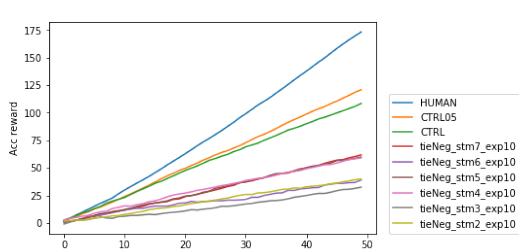
Total accumulated reward (TIE 0, EXP 20). Model VS Control, Human



3.1.2 Tie reward = -1

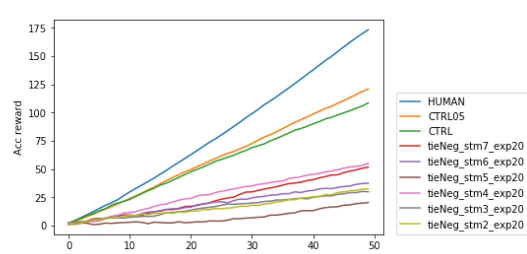
```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tieNeg_stm7_exp10:  61.84 ; tieNeg_stm6_exp10:  38.88 ; tieNeg_stm5_exp10:
59.88 ; tieNeg_stm4_exp10:  59.04 tieNeg_stm3_exp10:  32.44 ; tieNeg_stm2_exp10:
39.72
```

Total accumulated reward (TIE Neg, EXP 10). Model VS Control, Human



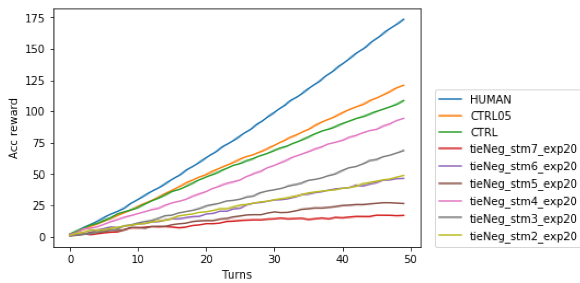

```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tieNeg_stm7_exp50:  51.76 ; tieNeg_stm6_exp50:  37.48 ; tieNeg_stm5_exp50:  20.4
; tieNeg_stm4_exp50:  55.12 ; tieNeg_stm3_exp50:  29.92 ; tieNeg_stm2_exp50:
32.44
```

Total accumulated reward (TIE Neg, EXP 50). Model VS Control, Human



```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tieNeg_stm7_exp20:  16.76 ; tieNeg_stm6_exp20:  46.44 ; tieNeg_stm5_exp20:
26.28 ; tieNeg_stm4_exp20:  94.6 ; tieNeg_stm3_exp20:  68.84 ;
tieNeg_stm2_exp20:  48.96
```

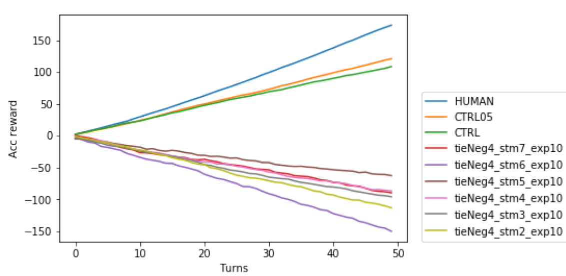
Total accumulated reward (TIE Neg, EXP 20). Model VS Control, Human



3.1.3 Tie reward = -4

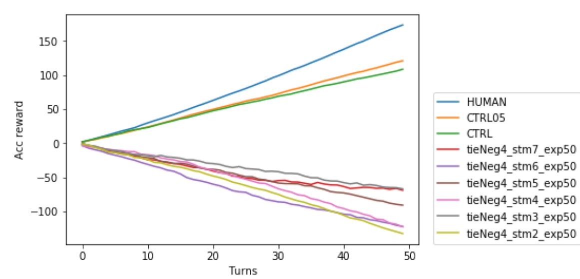
```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tieNeg4_stm7_exp10:  -89.56 ; tieNeg4_stm6_exp10:  -149.88 ; tieNeg4_stm5_exp10:
-62.52 ; tieNeg4_stm4_exp10:  -86.44 tieNeg4_stm3_exp10:  -95.8 ;
tieNeg4_stm2_exp10:  -112.96
```

Total accumulated reward (TIE Neg4, EXP 10). Model VS Control, Human



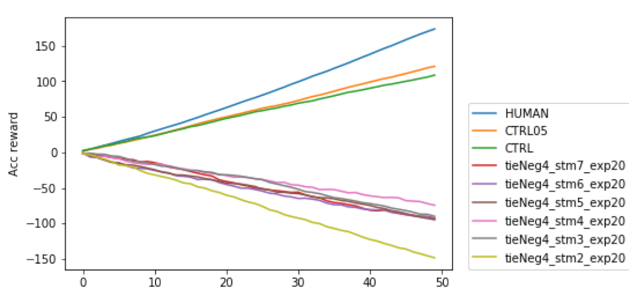

```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tieNeg4_stm7_exp50:  -68.76 ; tieNeg4_stm6_exp50:  -121.8 tieNeg4_stm5_exp50:
-90.6 ; tieNeg4_stm4_exp50:  -122.32 ; tieNeg4_stm3_exp50:  -66.68 ;
tieNeg4_stm2_exp50:  -132.2
```

Total accumulated reward (TIE Neg4, EXP 50). Model VS Control, Human




```
HUMAN:  173.39285714285714
CTRL05:  120.9
CTRL:  108.5
tieNeg4_stm7_exp20:  -92.68 ; tieNeg4_stm6_exp20:  -93.72 ; tieNeg4_stm5_exp20:
-94.76 ; tieNeg4_stm4_exp20:  -74.48 ; tieNeg4_stm3_exp20:  -90.08 ;
tieNeg4_stm2_exp20:  -148.32
```

Total accumulated reward (TIE Neg4, EXP 20). Model VS Control, Human

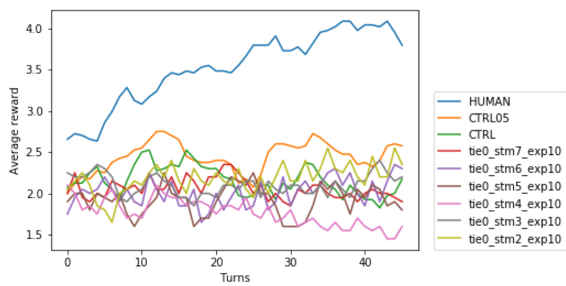


3.2 Moving Average

3.2.1 Tie reward = 0

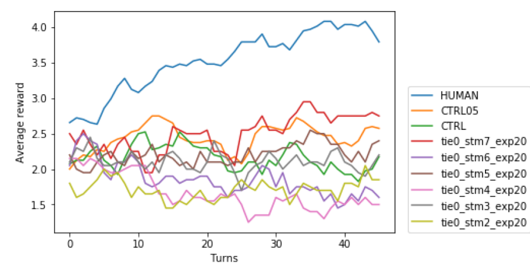
```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tie0_stm7_exp10:  1.9 ; tie0_stm6_exp10:  2.3 ; tie0_stm5_exp10:  1.8 ;
tie0_stm4_exp10:  1.6 tie0_stm3_exp10:  2.2 ; tie0_stm2_exp10:  2.35
```

Moving Average reward (TIE 0, EXP 10). Model VS Control, Human



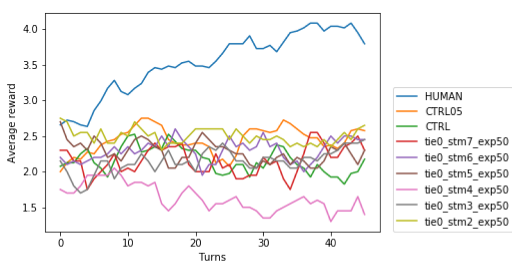

```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tie0_stm7_exp20:  2.75 ; tie0_stm6_exp20:  1.6 ; tie0_stm5_exp20:  2.4 ;
tie0_stm4_exp20:  1.5 ; tie0_stm3_exp20:  2.2 ; tie0_stm2_exp20:  1.85
```

Moving average reward (TIE 0, EXP 20). Model VS Control, Human




```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tie0_stm7_exp50:  2.3 ; tie0_stm6_exp50:  2.45 tie0_stm5_exp50:  2.3 ;
tie0_stm4_exp50:  1.4 ; tie0_stm3_exp50:  2.45 ; tie0_stm2_exp50:  2.65
```

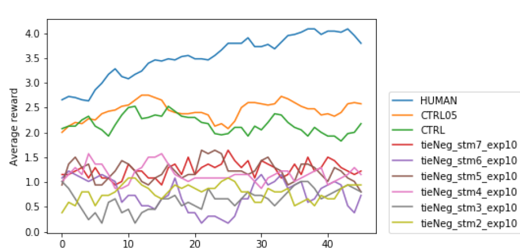
Average reward (TIE 0, EXP 50). Model VS Control, Human



3.2.2 Tie reward= -1

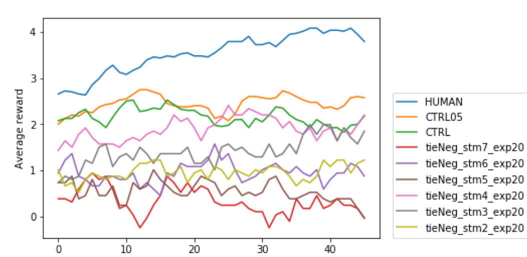
```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tieNeg_stm7_exp10:  1.22 ; tieNeg_stm6_exp10:  0.73 ; tieNeg_stm5_exp10:  0.8 ;
tieNeg_stm4_exp10:  1.15 tieNeg_stm3_exp10:  0.8 ; tieNeg_stm2_exp10:  0.94
```

Moving Average reward (TIE Neg, EXP 10). Model VS Control, Human



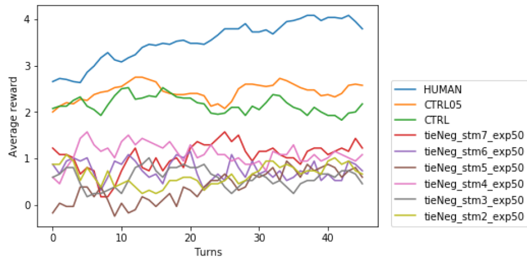

```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tieNeg_stm7_exp20:  -0.04 ; tieNeg_stm6_exp20:  0.87 ; tieNeg_stm5_exp20:  -0.04
; tieNeg_stm4_exp20:  2.2 ; tieNeg_stm3_exp20:  1.85 ; tieNeg_stm2_exp20:  1.22
```

Moving Average reward (TIE Neg, EXP 20). Model VS Control, Human



```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tieNeg_stm7_exp50:  1.22 ; tieNeg_stm6_exp50:  0.73 ; tieNeg_stm5_exp50:  0.59 ;
tieNeg_stm4_exp50:  1.08 ; tieNeg_stm3_exp50:  0.45 ; tieNeg_stm2_exp50:  0.66
```

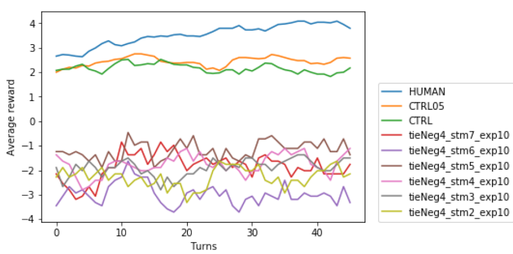
Moving Average reward (TIE Neg, EXP 50). Model VS Control, Human



3.2.3 Tie reward = -4

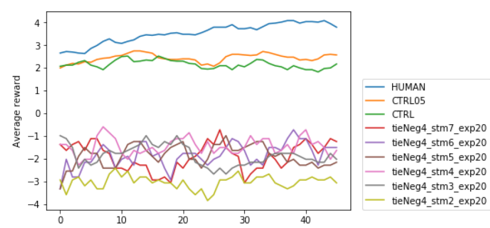
```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tieNeg4_stm7_exp10:  -1.76 ; tieNeg4_stm6_exp10:  -3.32 ; tieNeg4_stm5_exp10:
-1.37 ; tieNeg4_stm4_exp10:  -1.11 tieNeg4_stm3_exp10:  -1.5 ;
tieNeg4_stm2_exp10:  -2.15
```

Total accumulated reward (TIE Neg4, EXP 10). Model VS Control, Human



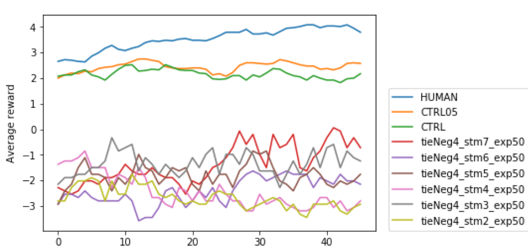

```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tieNeg4_stm7_exp20:  -1.24 ; tieNeg4_stm6_exp20:  -1.5 ; tieNeg4_stm5_exp20:
-2.15 ; tieNeg4_stm4_exp20:  -1.63 ; tieNeg4_stm3_exp20:  -2.02 ;
tieNeg4_stm2_exp20:  -3.06
```

Moving average reward (TIE Neg4, EXP 20). Model VS Control, Human




```
HUMAN:  3.794642857142857
CTRL05:  2.575
CTRL:  2.175
tieNeg4_stm7_exp50:  -0.72 ; tieNeg4_stm6_exp50:  -2.15 tieNeg4_stm5_exp50:
-1.76 ; tieNeg4_stm4_exp50:  -2.8 ; tieNeg4_stm3_exp50:  -1.24 ;
tieNeg4_stm2_exp50:  -2.93
```

Moving average reward (TIE Neg4, EXP 50). Model VS Control, Human



BALLISTIC CONDITION, SIMILARITY THRESHOLD = 0.8

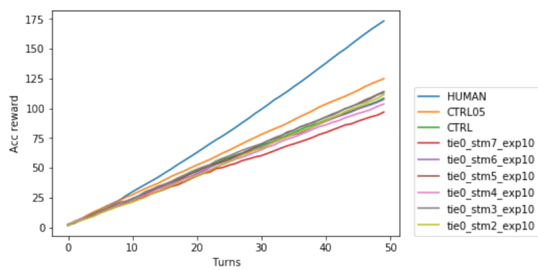
3 Model data VS Control (randomized) and Human data

3.1 Accumulated reward through trials

3.1.1 Tie reward = 0

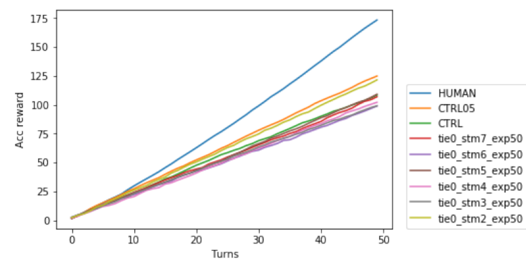
```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tie0_stm7_exp10:  96.8 ; tie0_stm6_exp10:  107.4 ; tie0_stm5_exp10:  114.0 ;
tie0_stm4_exp10:  103.6 tie0_stm3_exp10:  113.0 ; tie0_stm2_exp10:  111.4
```

Total accumulated reward (TIE 0, EXP 10). Model VS Control, Human



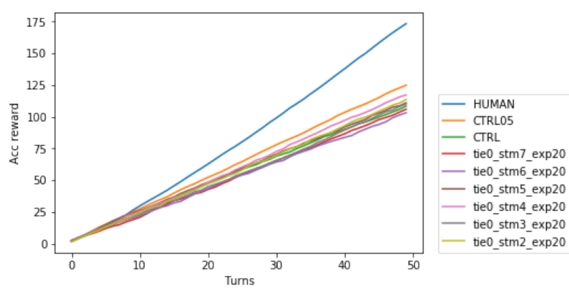

```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tie0_stm7_exp50:  107.0 ; tie0_stm6_exp50:  99.0 tie0_stm5_exp50:  109.2 ;
tie0_stm4_exp50:  102.2 ; tie0_stm3_exp50:  99.2 ; tie0_stm2_exp50:  121.6
```

Total accumulated reward (TIE 0, EXP 50). Model VS Control, Human




```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tie0_stm7_exp20:  106.0 ; tie0_stm6_exp20:  103.2 ; tie0_stm5_exp20:  111.0 ;
tie0_stm4_exp20:  117.2 ; tie0_stm3_exp20:  110.2 ; tie0_stm2_exp20:  113.6
```

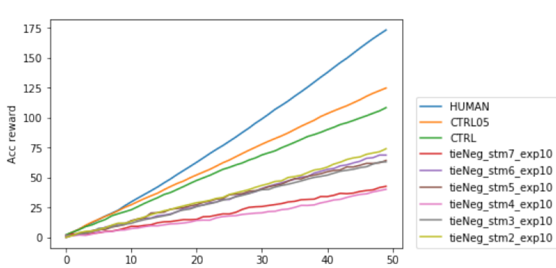
Total accumulated reward (TIE 0, EXP 20). Model VS Control, Human



3.1.2 Tie reward= -1

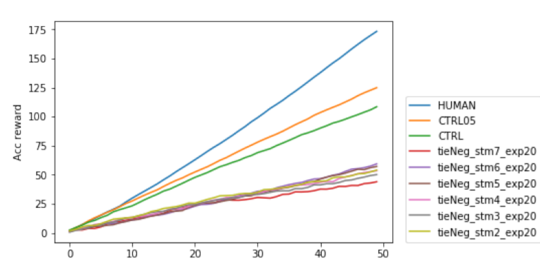
```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tieNeg_stm7_exp10:  42.8 ; tieNeg_stm6_exp10:  68.84 ; tieNeg_stm5_exp10:  64.36
; tieNeg_stm4_exp10:  40.28 tieNeg_stm3_exp10:  63.24 ; tieNeg_stm2_exp10:
74.16
```

Total accumulated reward (TIE Neg, EXP 10). Model VS Control, Human



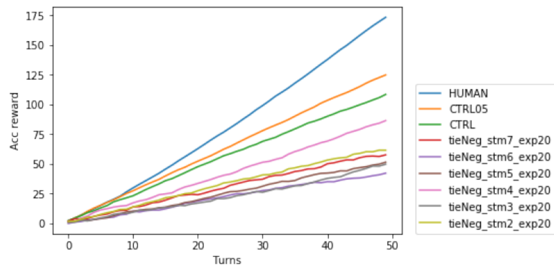

```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tieNeg_stm7_exp50:  43.92 ; tieNeg_stm6_exp50:  59.32 ; tieNeg_stm5_exp50:
57.08 ; tieNeg_stm4_exp50:  53.44 ; tieNeg_stm3_exp50:  50.08 ;
tieNeg_stm2_exp50:  54.0
```

Total accumulated reward (TIE Neg, EXP 50). Model VS Control, Human



```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tieNeg_stm7_exp20:  57.64 ; tieNeg_stm6_exp20:  42.24 ; tieNeg_stm5_exp20:
51.48 ; tieNeg_stm4_exp20:  86.48 ; tieNeg_stm3_exp20:  49.8 ;
tieNeg_stm2_exp20:  61.56
```

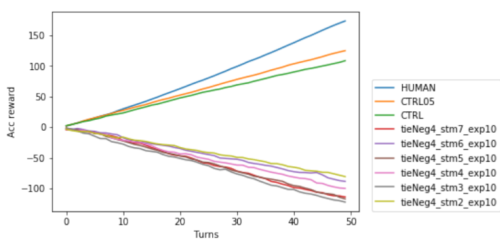
Total accumulated reward (TIE Neg, EXP 20). Model VS Control, Human



3.1.3 Tie reward = -4

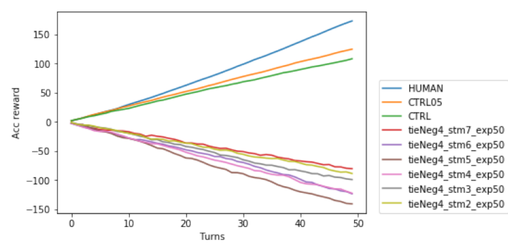
```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tieNeg4_stm7_exp10:  -114.0 ; tieNeg4_stm6_exp10:  -88.52 ; tieNeg4_stm5_exp10:
-117.12 ; tieNeg4_stm4_exp10:  -99.96 tieNeg4_stm3_exp10:  -122.32 ;
tieNeg4_stm2_exp10:  -80.72
```

Total accumulated reward (TIE Neg4, EXP 10). Model VS Control, Human



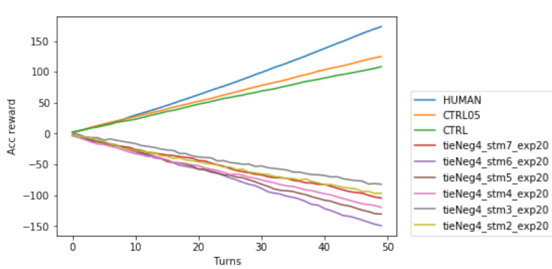

```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tieNeg4_stm7_exp50:  -80.2 ; tieNeg4_stm6_exp50:  -123.36 tieNeg4_stm5_exp50:
-140.52 ; tieNeg4_stm4_exp50:  -122.32 ; tieNeg4_stm3_exp50:  -98.92 ;
tieNeg4_stm2_exp50:  -88.52
```

Total accumulated reward (TIE Neg4, EXP 50). Model VS Control, Human




```
HUMAN:  173.39285714285714
CTRL05:  124.9
CTRL:  108.5
tieNeg4_stm7_exp20:  -104.64 ; tieNeg4_stm6_exp20:  -149.36 ;
tieNeg4_stm5_exp20:  -130.64 ; tieNeg4_stm4_exp20:  -119.72 ;
tieNeg4_stm3_exp20:  -82.28 ; tieNeg4_stm2_exp20:  -96.84
```

Total accumulated reward (TIE Neg4, EXP 20). Model VS Control, Human

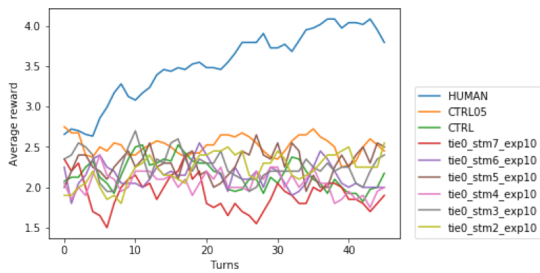


3.2 Moving Average

3.2.1 Tie reward = 0

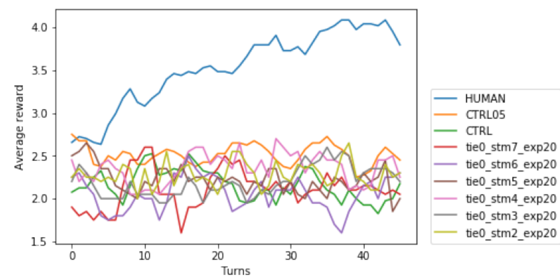
```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tie0_stm7_exp10:  1.9 ; tie0_stm6_exp10:  2.0 ; tie0_stm5_exp10:  2.5 ;
tie0_stm4_exp10:  2.0 tie0_stm3_exp10:  2.4 ; tie0_stm2_exp10:  2.55
```

Moving Average reward (TIE 0, EXP 10). Model VS Control, Human



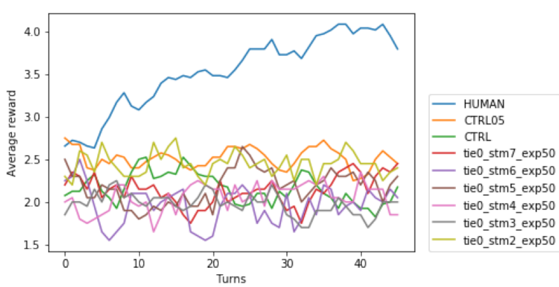

```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tie0_stm7_exp20:  2.05 ; tie0_stm6_exp20:  2.3 ; tie0_stm5_exp20:  2.0 ;
tie0_stm4_exp20:  2.25 ; tie0_stm3_exp20:  2.2 ; tie0_stm2_exp20:  2.3
```

Moving average reward (TIE 0, EXP 20). Model VS Control, Human




```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tie0_stm7_exp50:  2.45 ; tie0_stm6_exp50:  2.05 tie0_stm5_exp50:  2.3 ;
tie0_stm4_exp50:  1.85 ; tie0_stm3_exp50:  2.0 ; tie0_stm2_exp50:  2.4
```

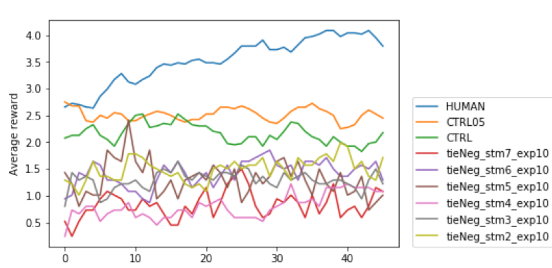
Average reward (TIE 0, EXP 50). Model VS Control, Human



3.2.2 Tie reward= -1

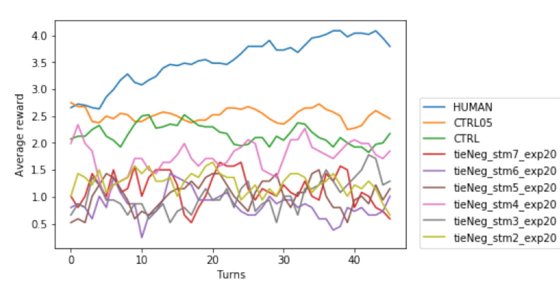
```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tieNeg_stm7_exp10:  1.08 ; tieNeg_stm6_exp10:  1.29 ; tieNeg_stm5_exp10:  1.01 ;
tieNeg_stm4_exp10:  1.08 tieNeg_stm3_exp10:  1.22 ; tieNeg_stm2_exp10:  1.71
```

Moving Average reward (TIE Neg, EXP 10). Model VS Control, Human



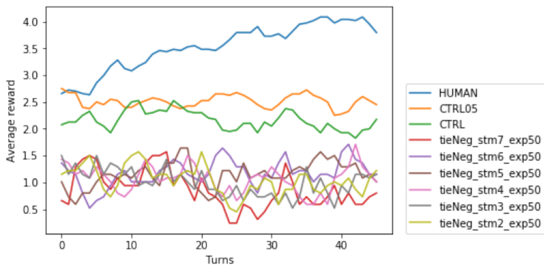

```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tieNeg_stm7_exp20:  0.59 ; tieNeg_stm6_exp20:  1.01 ; tieNeg_stm5_exp20:  1.15 ;
tieNeg_stm4_exp20:  1.85 ; tieNeg_stm3_exp20:  1.29 ; tieNeg_stm2_exp20:  0.66
```

Moving Average reward (TIE Neg, EXP 20). Model VS Control, Human



```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tieNeg_stm7_exp50:  0.8 ; tieNeg_stm6_exp50:  1.15 ; tieNeg_stm5_exp50:  1.01 ;
tieNeg_stm4_exp50:  1.01 ; tieNeg_stm3_exp50:  1.15 ; tieNeg_stm2_exp50:  1.22
```

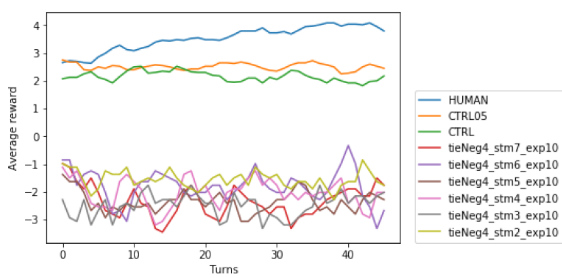
Moving Average reward (TIE Neg, EXP 50). Model VS Control, Human



3.2.3 Tie reward = -4

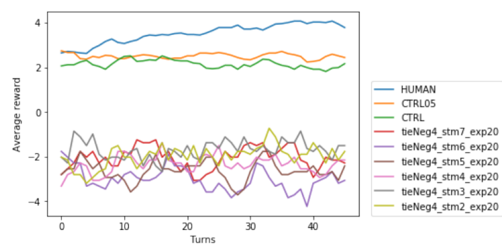
```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tieNeg4_stm7_exp10:  -1.76 ; tieNeg4_stm6_exp10:  -2.67 ; tieNeg4_stm5_exp10:
-2.28 ; tieNeg4_stm4_exp10:  -2.02 tieNeg4_stm3_exp10:  -2.02 ;
tieNeg4_stm2_exp10:  -1.76
```

Total accumulated reward (TIE Neg4, EXP 10). Model VS Control, Human



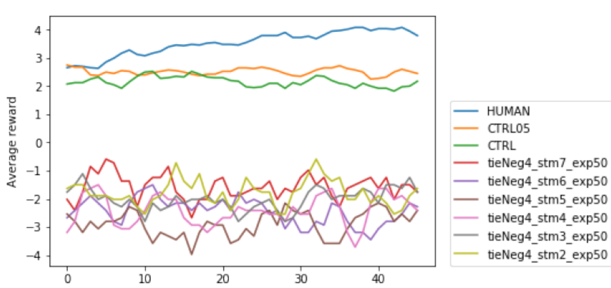

```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tieNeg4_stm7_exp20:  -2.28 ; tieNeg4_stm6_exp20:  -3.06 ; tieNeg4_stm5_exp20:
-2.41 ; tieNeg4_stm4_exp20:  -2.15 ; tieNeg4_stm3_exp20:  -1.5 ;
tieNeg4_stm2_exp20:  -1.76
```

Moving average reward (TIE Neg4, EXP 20). Model VS Control, Human




```
HUMAN:  3.794642857142857
CTRL05:  2.45
CTRL:  2.175
tieNeg4_stm7_exp50:  -1.76 ; tieNeg4_stm6_exp50:  -2.28 tieNeg4_stm5_exp50:
-2.41 ; tieNeg4_stm4_exp50:  -2.41 ; tieNeg4_stm3_exp50:  -1.76 ;
tieNeg4_stm2_exp50:  -1.63
```

Moving average reward (TIE Neg4, EXP 50). Model VS Control, Human



BALLISTIC CONDITION, SIMILARITY THRESHOLD = 0.9

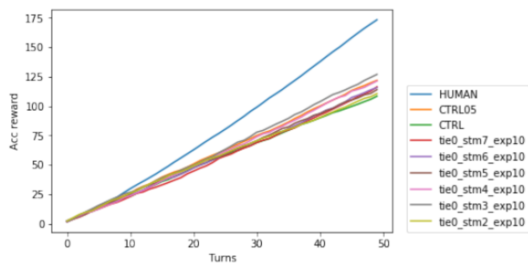
3 Model data VS Control (randomized) and Human data

3.1 Accumulated reward through trials

3.1.1 Tie reward = 0

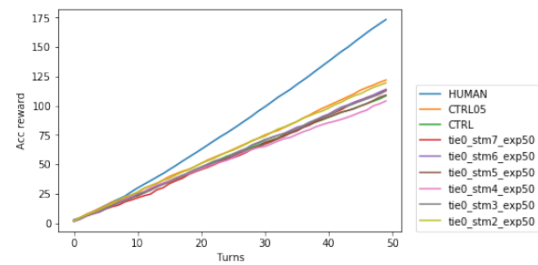
```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tie0_stm7_exp10:  116.4 ; tie0_stm6_exp10:  115.8 ; tie0_stm5_exp10:  114.0 ;
tie0_stm4_exp10:  121.4 tie0_stm3_exp10:  127.0 ; tie0_stm2_exp10:  110.6
```

Total accumulated reward (TIE 0, EXP 10). Model VS Control, Human



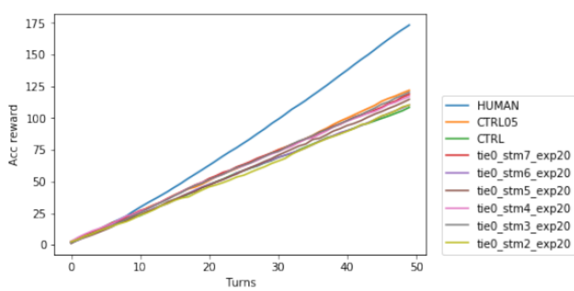

```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tie0_stm7_exp50:  113.2 ; tie0_stm6_exp50:  114.0 tie0_stm5_exp50:  109.0 ;
tie0_stm4_exp50:  104.0 ; tie0_stm3_exp50:  112.6 ; tie0_stm2_exp50:  119.4
```

Total accumulated reward (TIE 0, EXP 50). Model VS Control, Human




```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tie0_stm7_exp20:  118.8 ; tie0_stm6_exp20:  110.4 ; tie0_stm5_exp20:  114.8 ;
tie0_stm4_exp20:  117.0 ; tie0_stm3_exp20:  120.4 ; tie0_stm2_exp20:  110.4
```

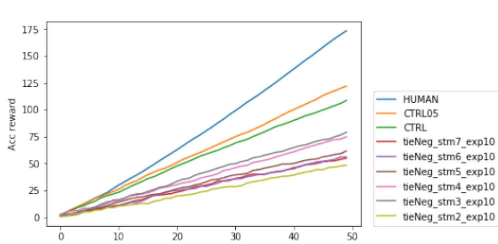
Total accumulated reward (TIE 0, EXP 20). Model VS Control, Human



3.1.2 Tie reward= -1

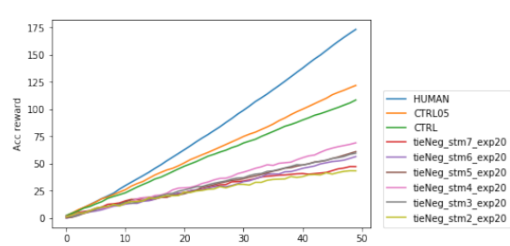
```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tieNeg_stm7_exp10:  54.84 ; tieNeg_stm6_exp10:  55.96 ; tieNeg_stm5_exp10:
61.56 ; tieNeg_stm4_exp10:  74.72 tieNeg_stm3_exp10:  78.92 ; tieNeg_stm2_exp10:
48.4
```

Total accumulated reward (TIE Neg, EXP 10). Model VS Control, Human



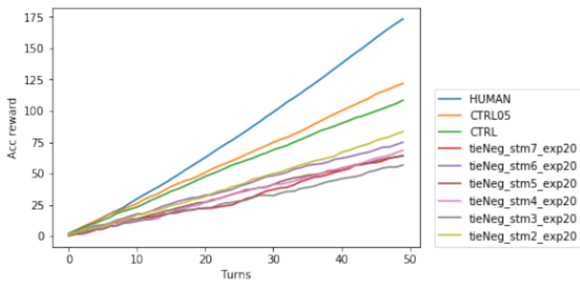

```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tieNeg_stm7_exp50:  47.28 ; tieNeg_stm6_exp50:  56.52 ; tieNeg_stm5_exp50:  61.0
; tieNeg_stm4_exp50:  69.12 ; tieNeg_stm3_exp50:  59.6 ; tieNeg_stm2_exp50:
43.36
```

Total accumulated reward (TIE Neg, EXP 50). Model VS Control, Human



```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tieNeg_stm7_exp20:  64.08 ; tieNeg_stm6_exp20:  75.0 ; tieNeg_stm5_exp20:  64.36
; tieNeg_stm4_exp20:  68.56 ; tieNeg_stm3_exp20:  56.8 ; tieNeg_stm2_exp20:
83.4
```

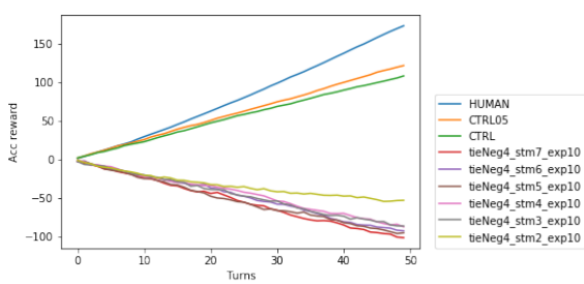
Total accumulated reward (TIE Neg, EXP 20). Model VS Control, Human



3.1.3 Tie reward = -4

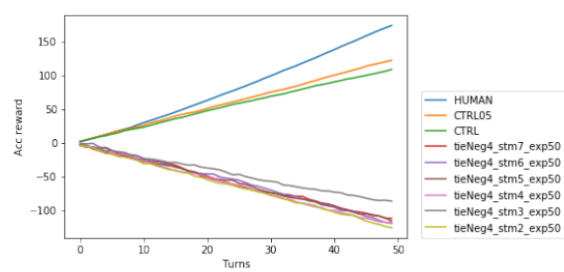
```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tieNeg4_stm7_exp10:  -101.0 ; tieNeg4_stm6_exp10:  -92.16 ; tieNeg4_stm5_exp10:
-94.76 ; tieNeg4_stm4_exp10:  -86.96 tieNeg4_stm3_exp10:  -85.92 ;
tieNeg4_stm2_exp10:  -52.64
```

Total accumulated reward (TIE Neg4, EXP 10). Model VS Control, Human



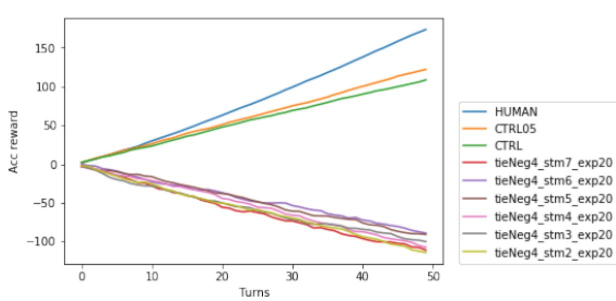

```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tieNeg4_stm7_exp50:  -111.4 ; tieNeg4_stm6_exp50:  -118.16 tieNeg4_stm5_exp50:
-115.56 ; tieNeg4_stm4_exp50:  -119.2 ; tieNeg4_stm3_exp50:  -85.92 ;
tieNeg4_stm2_exp50:  -125.44
```

Total accumulated reward (TIE Neg4, EXP 50). Model VS Control, Human




```
HUMAN:  173.39285714285714
CTRL05:  121.9
CTRL:  108.5
tieNeg4_stm7_exp20:  -110.88 ; tieNeg4_stm6_exp20:  -89.04 ; tieNeg4_stm5_exp20:
-90.6 ; tieNeg4_stm4_exp20:  -107.24 ; tieNeg4_stm3_exp20:  -99.96 ;
tieNeg4_stm2_exp20:  -114.0
```

Total accumulated reward (TIE Neg4, EXP 20). Model VS Control, Human

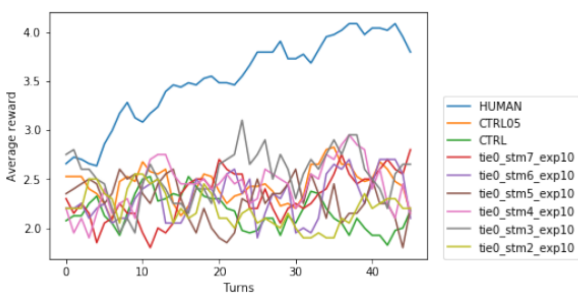


3.2 Moving Average

3.2.1 Tie reward = 0

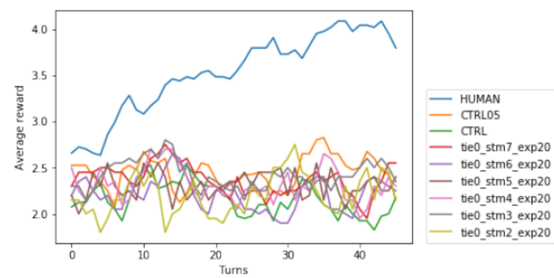
```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tie0_stm7_exp10:  2.8 ; tie0_stm6_exp10:  2.1 ; tie0_stm5_exp10:  2.2 ;
tie0_stm4_exp10:  2.15 tie0_stm3_exp10:  2.65 ; tie0_stm2_exp10:  2.2
```

Moving Average reward (TIE 0, EXP 10). Model VS Control, Human



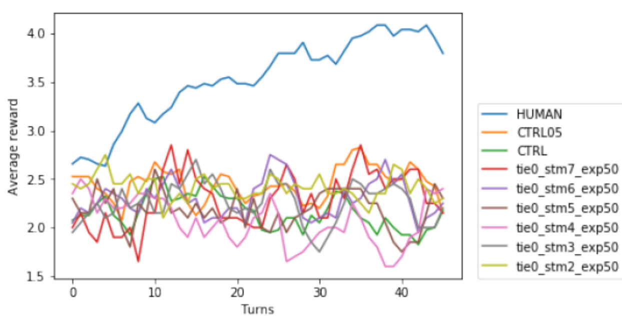

```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tie0_stm7_exp20:  2.55 ; tie0_stm6_exp20:  2.25 ; tie0_stm5_exp20:  2.4 ;
tie0_stm4_exp20:  2.3 ; tie0_stm3_exp20:  2.35 ; tie0_stm2_exp20:  2.15
```

Moving average reward (TIE 0, EXP 20). Model VS Control, Human




```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tie0_stm7_exp50:  2.15 ; tie0_stm6_exp50:  2.25 tie0_stm5_exp50:  2.3 ;
tie0_stm4_exp50:  2.4 ; tie0_stm3_exp50:  2.2 ; tie0_stm2_exp50:  2.3
```

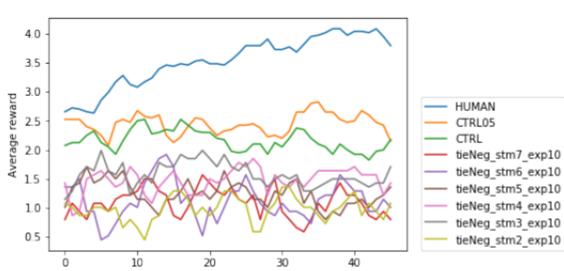
Average reward (TIE 0, EXP 50). Model VS Control, Human



3.2.2 Tie reward= -1

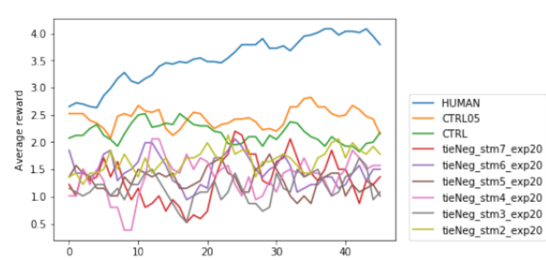
```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tieNeg_stm7_exp10:  0.8 ; tieNeg_stm6_exp10:  1.01 ; tieNeg_stm5_exp10:  1.36 ;
tieNeg_stm4_exp10:  1.43 tieNeg_stm3_exp10:  1.71 ; tieNeg_stm2_exp10:  1.08
```

Moving Average reward (TIE Neg, EXP 10). Model VS Control, Human



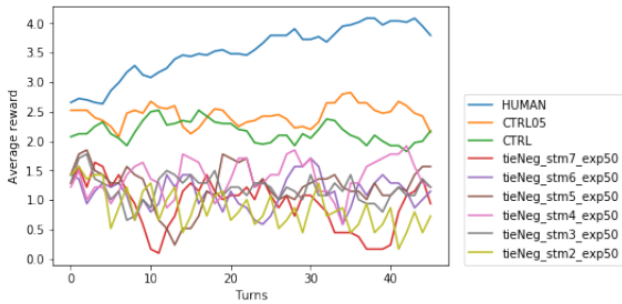

```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tieNeg_stm7_exp20:  1.36 ; tieNeg_stm6_exp20:  1.5 ; tieNeg_stm5_exp20:  1.01 ;
tieNeg_stm4_exp20:  1.57 ; tieNeg_stm3_exp20:  1.08 ; tieNeg_stm2_exp20:  1.78
```

Moving Average reward (TIE Neg, EXP 20). Model VS Control, Human



```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tieNeg_stm7_exp50:  0.94 ; tieNeg_stm6_exp50:  1.15 ; tieNeg_stm5_exp50:  1.57 ;
tieNeg_stm4_exp50:  1.22 ; tieNeg_stm3_exp50:  1.22 ; tieNeg_stm2_exp50:  0.73
```

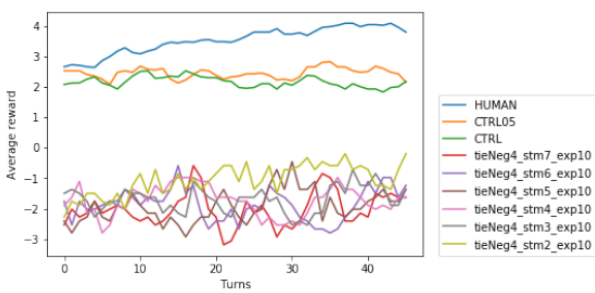
Moving Average reward (TIE Neg, EXP 50). Model VS Control, Human



3.2.3 Tie reward = -4

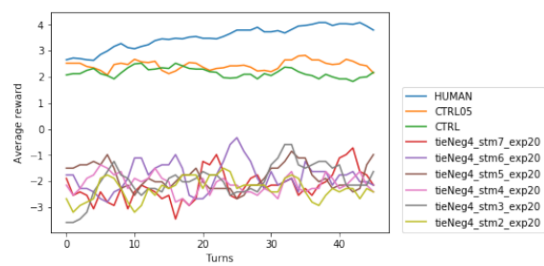
```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tieNeg4_stm7_exp10:  -1.63 ; tieNeg4_stm6_exp10:  -1.24 ; tieNeg4_stm5_exp10:
-1.37 ; tieNeg4_stm4_exp10:  -1.63 tieNeg4_stm3_exp10:  -1.37 ;
tieNeg4_stm2_exp10:  -0.2
```

Total accumulated reward (TIE Neg4, EXP 10). Model VS Control, Human



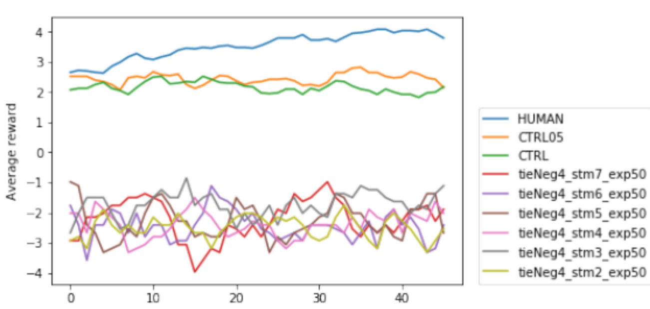

```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tieNeg4_stm7_exp20:  -2.15 ; tieNeg4_stm6_exp20:  -2.15 ; tieNeg4_stm5_exp20:
-0.98 ; tieNeg4_stm4_exp20:  -2.41 ; tieNeg4_stm3_exp20:  -1.63 ;
tieNeg4_stm2_exp20:  -2.41
```

Moving average reward (TIE Neg4, EXP 20). Model VS Control, Human




```
HUMAN:  3.794642857142857
CTRL05:  2.15
CTRL:  2.175
tieNeg4_stm7_exp50:  -1.89 ; tieNeg4_stm6_exp50:  -2.41 tieNeg4_stm5_exp50:
-2.67 ; tieNeg4_stm4_exp50:  -2.02 ; tieNeg4_stm3_exp50:  -1.11 ;
tieNeg4_stm2_exp50:  -2.54
```

Moving average reward (TIE Neg4, EXP 50). Model VS Control, Human



DYNAMIC CONDITION, SIMILARITY THRESHOLD = 0.7

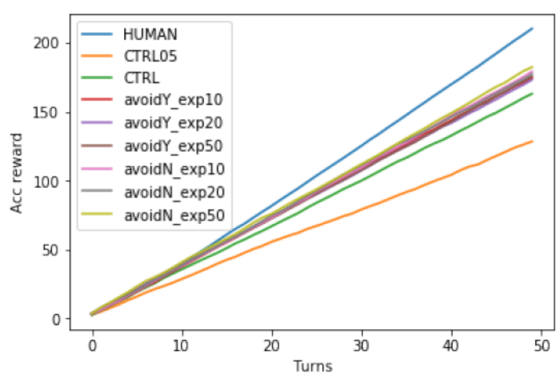
3 Model data VS Control (randomized) and Human data

3.1 Accumulated reward through trials

3.1.1 Tie reward = 0

```
HUMAN:  210.5072463768116
CTRL05:  128.6
CTRL:  163.2
avoidY_exp10:  174.8 ; avoidY_exp20:  173.4 ; avoidY_exp50:  175.8
avoidN_exp10:  179.3 ; avoidN_exp20:  177.4 ; avoidN_exp50:  182.5
```

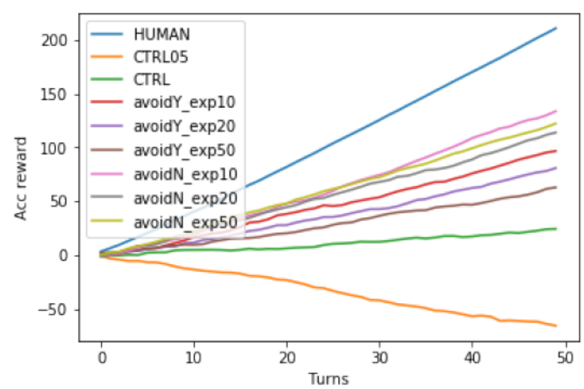
Total accumulated reward (TIE 0). Model VS Control, Human



3.1.3 Tie reward -4

```
HUMAN:  210.5072463768116
CTRL05:  -65.64
CTRL:  24.32
avoidY_exp10:  96.6 ; avoidY_exp20:  80.74 ; avoidY_exp50:  62.8
avoidN_exp10:  133.52 ; avoidN_exp20:  113.76 ; avoidN_exp50:  122.08
```

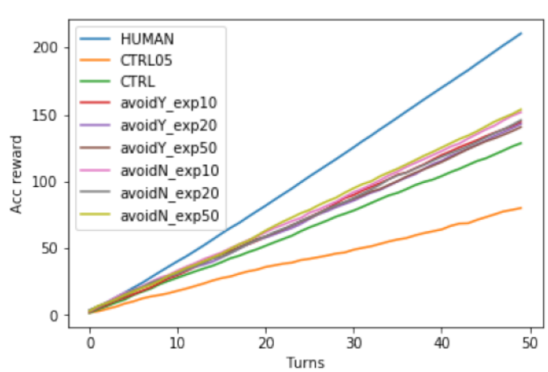
Total accumulated reward (TIE -4). Model VS Control, Human



3.1.2 Tie reward= -1

```
HUMAN:  210.5072463768116
CTRL05:  80.04
CTRL:  128.48
avoidY_exp10:  144.3 ; avoidY_exp20:  142.9 ; avoidY_exp50:  140.52
avoidN_exp10:  151.58 ; avoidN_exp20:  145.7 ; avoidN_exp50:  153.68
```

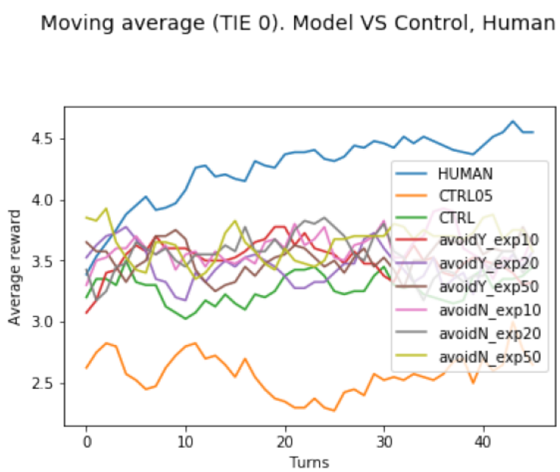
Total accumulated reward (TIE -1). Model VS Control, Human



3.2 Moving Average

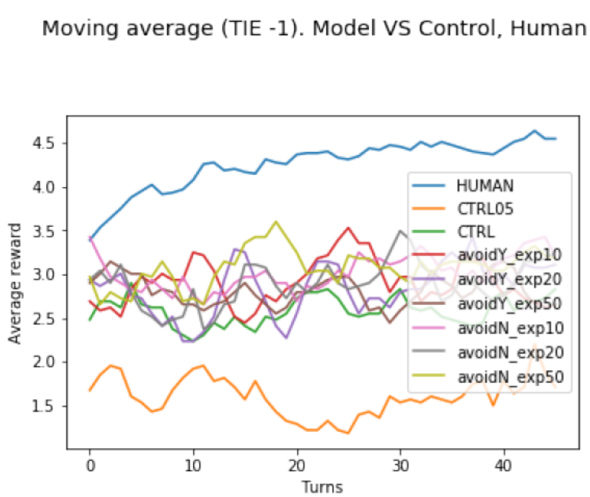
3.2.1 Tie reward = 0

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  2.5766304347826092
CTRL:  3.2744565217391304
avoidY_exp10:  3.510869565217391 ; avoidY_exp20:  3.4570652173913046 ;
avoidY_exp50:  3.527173913043478
avoidN_exp10:  3.5983695652173906 ; avoidN_exp20:  3.5385869565217387 ;
avoidN_exp50:  3.642934782608696
```



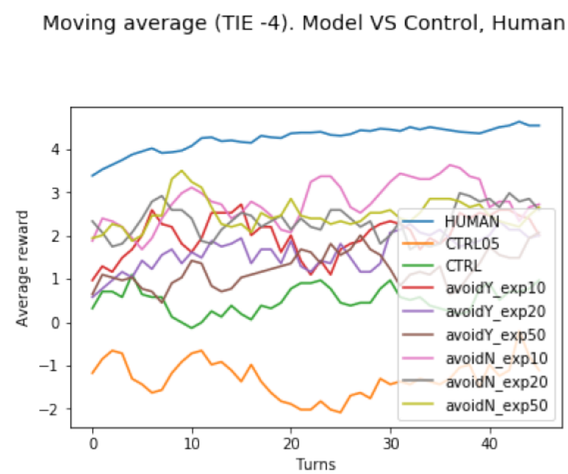
3.2.2 Tie reward = -1

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  1.607282608695652
CTRL:  2.5842391304347823
avoidY_exp10:  2.905326086956522 ; avoidY_exp20:  2.8383695652173913 ;
avoidY_exp50:  2.8109782608695655
avoidN_exp10:  3.020217391304348 ; avoidN_exp20:  2.8870652173913043 ;
avoidN_exp50:  3.071195652173913
```



3.2.3 Tie reward = -4

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  -1.3007608695652173
CTRL:  0.5135869565217391
avoidY_exp10:  2.0100000000000007 ; avoidY_exp20:  1.6341304347826087 ;
avoidY_exp50:  1.2356521739130435
avoidN_exp10:  2.775869565217391 ; avoidN_exp20:  2.346304347826087 ;
avoidN_exp50:  2.493260869565217
```



DYNAMIC CONDITION, SIMILARITY THRESHOLD = 0.8

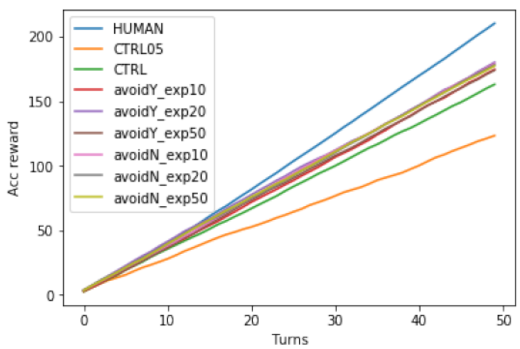
3 Model data VS Control (randomized) and Human data

3.1 Accumulated reward through trials

3.1.1 Tie reward = 0

```
HUMAN:  210.5072463768116
CTRL05:  123.5
CTRL:  163.2
avoidY_exp10:  174.8 ; avoidY_exp20:  180.3 ; avoidY_exp50:  174.2
avoidN_exp10:  179.3 ; avoidN_exp20:  178.3 ; avoidN_exp50:  177.5
```

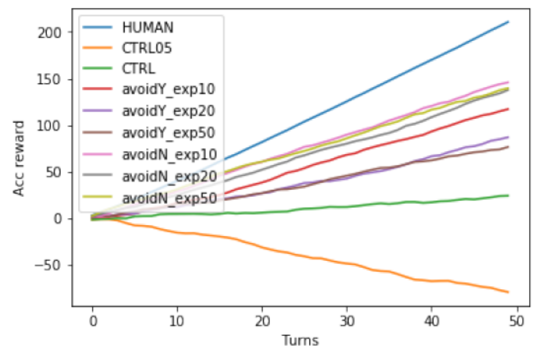
Total accumulated reward (TIE 0). Model VS Control, Human



3.1.3 Tie reward -4

```
HUMAN:  210.5072463768116
CTRL05:  -78.9
CTRL:  24.32
avoidY_exp10:  117.14 ; avoidY_exp20:  86.98 ; avoidY_exp50:  76.58
avoidN_exp10:  145.74 ; avoidN_exp20:  137.68 ; avoidN_exp50:  139.5
```

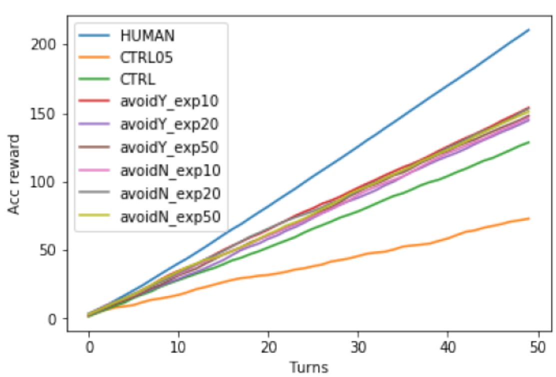
Total accumulated reward (TIE -4). Model VS Control, Human



3.1.2 Tie reward= -1

```
HUMAN:  210.5072463768116
CTRL05:  72.9
CTRL:  128.48
avoidY_exp10:  153.82 ; avoidY_exp20:  144.58 ; avoidY_exp50:  147.94
avoidN_exp10:  146.4 ; avoidN_exp20:  153.12 ; avoidN_exp50:  150.88
```

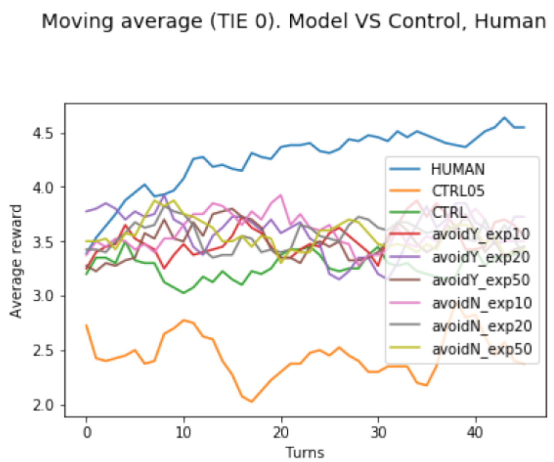
Total accumulated reward (TIE -1). Model VS Control, Human



3.2 Moving Average

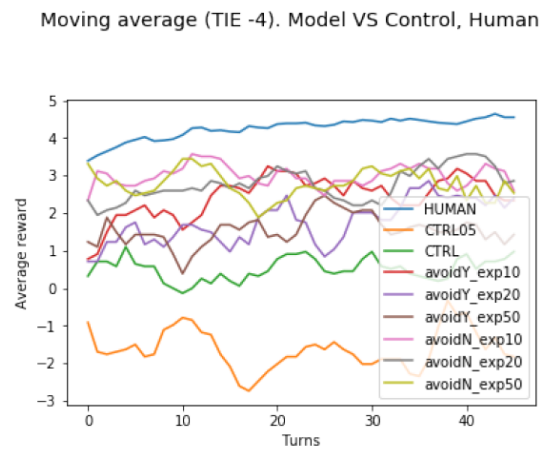
3.2.1 Tie reward = 0

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  2.452717391304348
CTRL:  3.2744565217391304
avoidY_exp10:  3.5211956521739136 ; avoidY_exp20:  3.595652173913044 ;
avoidY_exp50:  3.4940217391304347
avoidN_exp10:  3.589673913043478 ; avoidN_exp20:  3.5608695652173914 ;
avoidN_exp50:  3.552717391304349
```



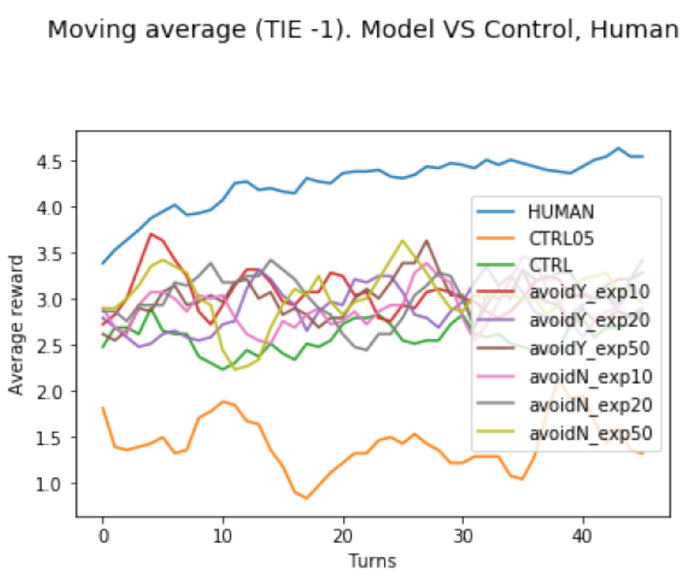
3.2.3 Tie reward = -4

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  -1.622934782608696
CTRL:  0.5135869565217391
avoidY_exp10:  2.4310869565217392 ; avoidY_exp20:  1.7627173913043483 ;
avoidY_exp50:  1.546521739130435
avoidN_exp10:  2.9920652173913043 ; avoidN_exp20:  2.7716304347826086 ;
avoidN_exp50:  2.7702173913043473
```



3.2.2 Tie reward = -1

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  1.433804347826087
CTRL:  2.5842391304347823
avoidY_exp10:  3.0841304347826095 ; avoidY_exp20:  2.9022826086956526 ;
avoidY_exp50:  2.9745652173913046
avoidN_exp10:  2.9144565217391305 ; avoidN_exp20:  3.04304347826087 ;
avoidN_exp50:  3.0247826086956517
```



DYNAMIC CONDITION, SIMILARITY THRESHOLD = 0.9

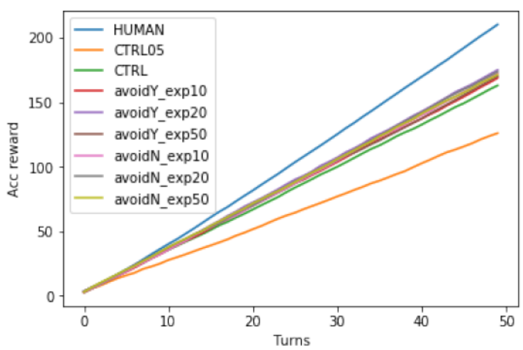
3 Model data VS Control (randomized) and Human data

3.1 Accumulated reward through trials

3.1.1 Tie reward = 0

```
HUMAN:  210.5072463768116
CTRL05:  126.2
CTRL:  163.2
avoidY_exp10:  169.0 ; avoidY_exp20:  175.2 ; avoidY_exp50:  170.4
avoidN_exp10:  172.6 ; avoidN_exp20:  173.9 ; avoidN_exp50:  171.6
```

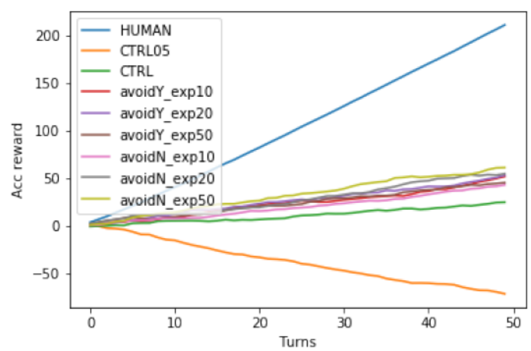
Total accumulated reward (TIE 0). Model VS Control, Human



3.1.3 Tie reward -4

```
HUMAN:  210.5072463768116
CTRL05:  -71.88
CTRL:  24.32
avoidY_exp10:  51.36 ; avoidY_exp20:  52.66 ; avoidY_exp50:  44.86
avoidN_exp10:  42.52 ; avoidN_exp20:  54.22 ; avoidN_exp50:  60.72
```

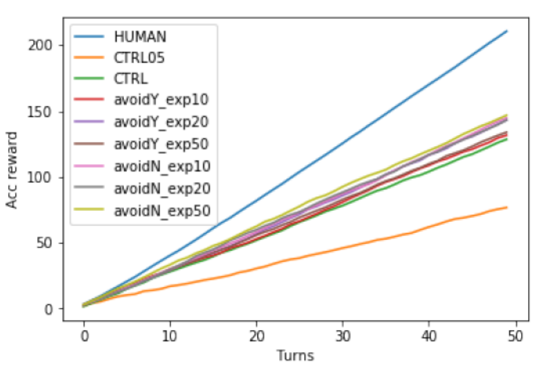
Total accumulated reward (TIE -4). Model VS Control, Human



3.1.2 Tie reward= -1

```
HUMAN:  210.5072463768116
CTRL05:  76.68
CTRL:  128.48
avoidY_exp10:  131.56 ; avoidY_exp20:  143.04 ; avoidY_exp50:  133.8
avoidN_exp10:  144.58 ; avoidN_exp20:  143.18 ; avoidN_exp50:  146.68
```

Total accumulated reward (TIE -1). Model VS Control, Human

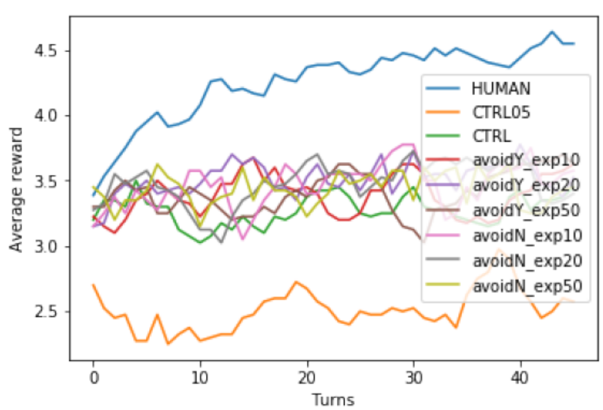


3.2 Moving Average

3.2.1 Tie reward = 0

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  2.511413043478261
CTRL:  3.2744565217391304
avoidY_exp10:  3.3880434782608706 ; avoidY_exp20:  3.5103260869565216 ;
avoidY_exp50:  3.4070652173913043
avoidN_exp10:  3.4624999999999995 ; avoidN_exp20:  3.4804347826086954 ;
avoidN_exp50:  3.435326086956522
```

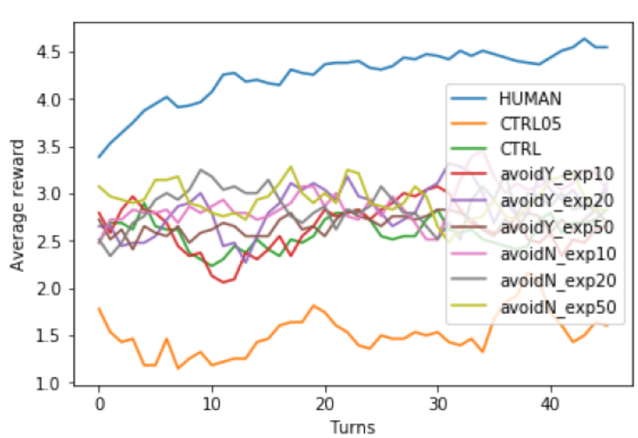
Moving average (TIE 0). Model VS Control, Human



3.2.2 Tie reward = -1

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  1.515978260869565
CTRL:  2.5842391304347823
avoidY_exp10:  2.6344565217391307 ; avoidY_exp20:  2.8589130434782613 ;
avoidY_exp50:  2.670978260869565
avoidN_exp10:  2.9106521739130433 ; avoidN_exp20:  2.8665217391304343 ;
avoidN_exp50:  2.933478260869566
```

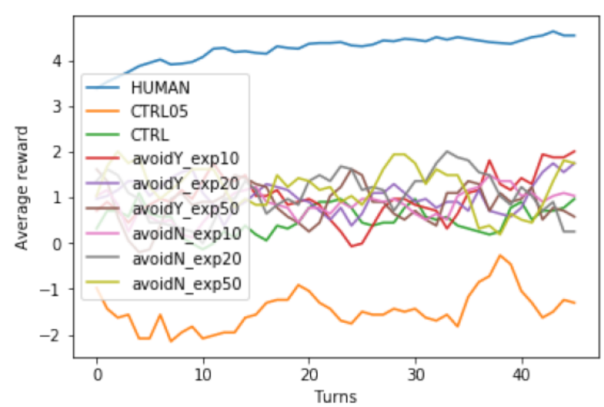
Moving average (TIE -1). Model VS Control, Human



3.2.3 Tie reward = -4

```
Average through trials
HUMAN:  4.252914303717707
CTRL05:  -1.4703260869565218
CTRL:  0.5135869565217391
avoidY_exp10:  1.0123913043478259 ; avoidY_exp20:  1.0420652173913043 ;
avoidY_exp50:  0.8626086956521738
avoidN_exp10:  0.8498913043478259 ; avoidN_exp20:  1.0632608695652173 ;
avoidN_exp50:  1.2370652173913044
```

Moving average (TIE -4). Model VS Control, Human



Appendix B

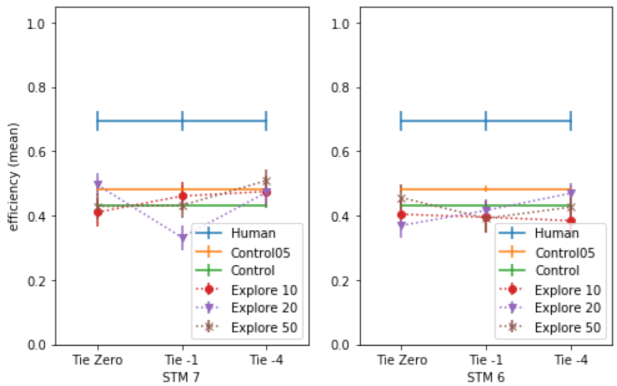
Efficiency, fairness, and stability results

BALLISTIC CONDITION, SIMILARITY THRESHOLD = 0.7

5 EFFICIENCY

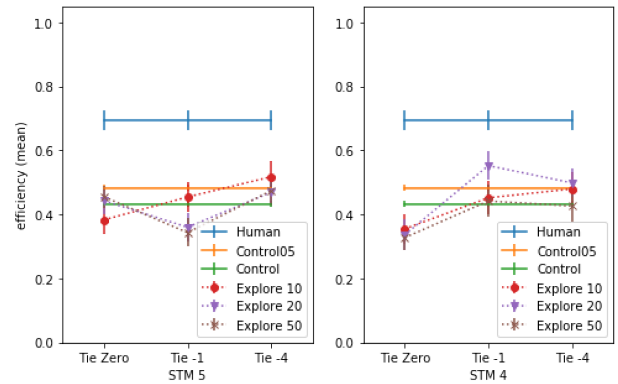
STM: 7, 6

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4836
--> Eff_tie0_stm7_exp10:  0.4108078431372549 ; Eff_tieNeg_stm7_exp10:
0.4611764705882353 ; Eff_tieNeg4_stm7_exp10:  0.4754509803921569 ;
Eff_tie0_stm6_exp10:  0.4047529411764706 ; Eff_tieNeg_stm6_exp10:
0.3956392156862746 ; Eff_tieNeg4_stm6_exp10:  0.3848
```



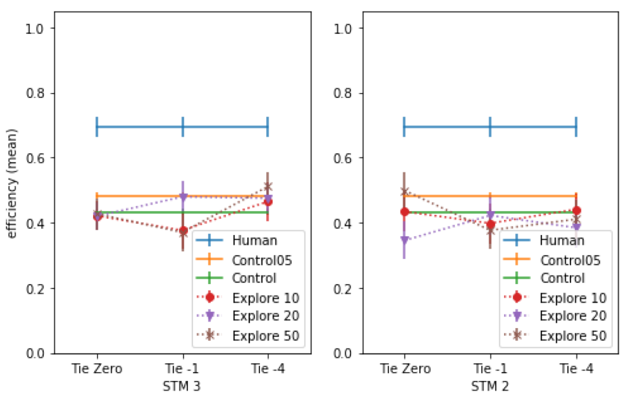
STM: 5, 4

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4836
--> Eff_tie0_stm5_exp10:  0.3834980392156862 ; Eff_tieNeg_stm5_exp10:
0.45491764705882354 ; Eff_tieNeg4_stm5_exp10:  0.517929411764706 ;
Eff_tie0_stm4_exp10:  0.355278431372549 ; Eff_tieNeg_stm4_exp10:
0.45226666666666665 ; Eff_tieNeg4_stm4_exp10:  0.48012549019607836
```



STM: 3, 2

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4836
--> Eff_tie0_stm3_exp10:  0.42158431372549027 ; Eff_tieNeg_stm3_exp10:
0.3768156862745098 ; Eff_tieNeg4_stm3_exp10:  0.4653019607843137 ;
Eff_tie0_stm2_exp10:  0.43439999999999995 ; Eff_tieNeg_stm2_exp10:
0.3983686274509804 ; Eff_tieNeg4_stm2_exp10:  0.4412078431372549
```



5.2 Statistical tests. Are the distributions different?

5.2.2 Human data VS. each condition

No differences: []

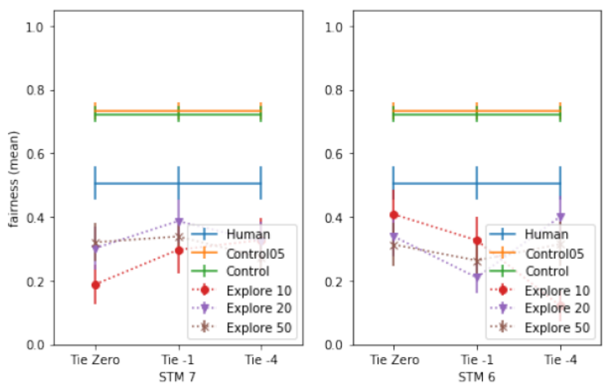
5.2.4 Control data VS. each condition

No differences: ['Eff_tie0_stm7_exp10', 'Eff_tie0_stm6_exp10', 'Eff_tie0_stm4_exp10', 'Eff_tie0_stm3_exp10', 'Eff_tie0_stm2_exp10', 'Eff_tieNeg_stm7_exp10', 'Eff_tieNeg_stm6_exp10', 'Eff_tieNeg_stm5_exp10', 'Eff_tieNeg_stm4_exp10', 'Eff_tieNeg_stm3_exp10', 'Eff_tieNeg_stm2_exp10', 'Eff_tieNeg4_stm6_exp10', 'Eff_tieNeg4_stm3_exp10', 'Eff_tieNeg4_stm2_exp10', 'Eff_tie0_stm6_exp20', 'Eff_tie0_stm5_exp20', 'Eff_tie0_stm3_exp20', 'Eff_tieNeg_stm6_exp20', 'Eff_tieNeg_stm5_exp20', 'Eff_tieNeg_stm3_exp20', 'Eff_tieNeg_stm2_exp20', 'Eff_tieNeg4_stm7_exp20', 'Eff_tieNeg4_stm6_exp20', 'Eff_tieNeg4_stm5_exp20', 'Eff_tieNeg4_stm3_exp20', 'Eff_tieNeg4_stm2_exp20', 'Eff_tie0_stm7_exp50', 'Eff_tie0_stm6_exp50', 'Eff_tie0_stm5_exp50', 'Eff_tie0_stm3_exp50', 'Eff_tie0_stm2_exp50', 'Eff_tieNeg_stm7_exp50', 'Eff_tieNeg_stm6_exp50', 'Eff_tieNeg_stm5_exp50', 'Eff_tieNeg_stm4_exp50', 'Eff_tieNeg_stm3_exp50', 'Eff_tieNeg_stm2_exp50', 'Eff_tieNeg4_stm6_exp50', 'Eff_tieNeg4_stm5_exp50', 'Eff_tieNeg4_stm4_exp50', 'Eff_tieNeg4_stm2_exp50', 'Eff_tieNeg4_stm2_exp50']

6 FAIRNESS

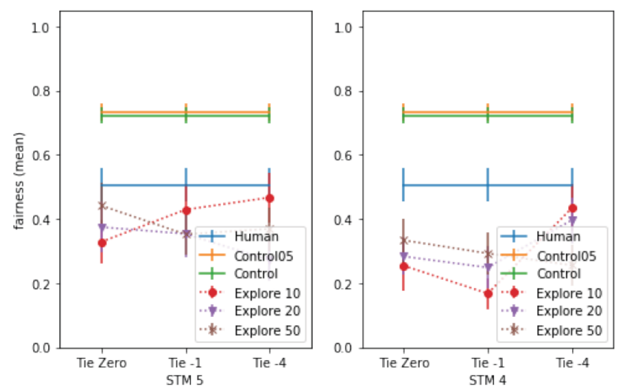
STM: 7, 6

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7313342933296493
--> Fair_tie0_stm7_exp10:  0.18860395477899183 ; Fair_tieNeg_stm7_exp10:
0.29723519920201846 ; Fair_tieNeg4_stm7_exp10:  0.3279387084650242 ;
Fair_tie0_stm6_exp10:  0.40881834843040926 ; Fair_tieNeg_stm6_exp10:
0.3269156010230179 ; Fair_tieNeg4_stm6_exp10:  0.12240318272236429
```



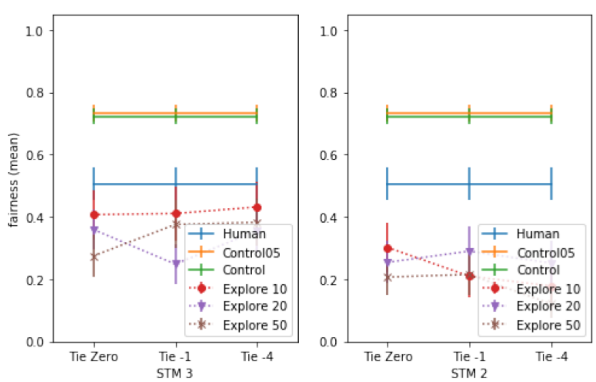
STM: 5, 4

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7313342933296493
--> Fair_tie0_stm5_exp10:  0.3293475321211895 ; Fair_tieNeg_stm5_exp10:
0.429519341832843 ; Fair_tieNeg4_stm5_exp10:  0.467310445979862 ;
Fair_tie0_stm4_exp10:  0.2549966039643789 ; Fair_tieNeg_stm4_exp10:
0.16830664778490867 ; Fair_tieNeg4_stm4_exp10:  0.4343276805036396
```



STM: 3, 2

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7313342933296493
--> Fair_tie0_stm3_exp10:  0.40770770739780027 ; Fair_tieNeg_stm3_exp10:
0.41125128179926945 ; Fair_tieNeg4_stm3_exp10:  0.4320979977680745 ;
Fair_tie0_stm2_exp10:  0.30231272438245116 ; Fair_tieNeg_stm2_exp10:
0.21017603027833204 ; Fair_tieNeg4_stm2_exp10:  0.17860010640411772
```



6.2 Statistical tests. Are the distributions different?

6.2.2 Human data VS. each condition

No differences: ['Fair_tie0_stm6_exp10', 'Fair_tie0_stm3_exp10', 'Fair_tieNeg_stm5_exp10', 'Fair_tieNeg_stm3_exp10', 'Fair_tieNeg4_stm5_exp10', 'Fair_tieNeg4_stm4_exp10', 'Fair_tieNeg4_stm3_exp10', 'Fair_tie0_stm5_exp20', 'Fair_tieNeg_stm7_exp20', 'Fair_tieNeg4_stm6_exp20', 'Fair_tieNeg4_stm4_exp20', 'Fair_tie0_stm5_exp50', 'Fair_tieNeg4_stm5_exp50', 'Fair_tieNeg4_stm3_exp50']

6.2.4 Control data VS. each condition

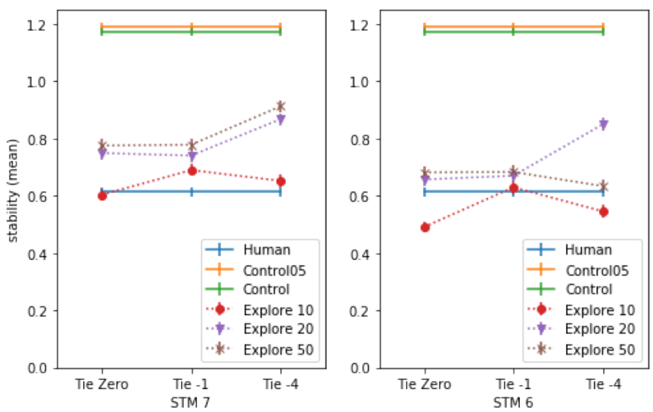
No differences: []

7 Stability

Especially in ballistic conditions, people use the random assignment of the payoffs to coordinate. One player will always go top and the other will always go bottom. The stability of this pattern isn't captured by the outcomes, so we need to check the direction sequence as well.

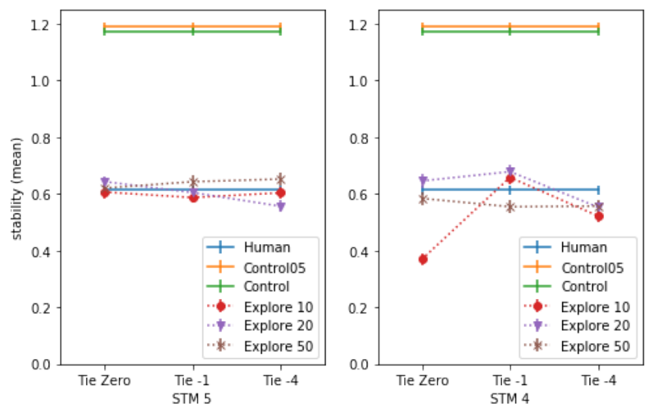
STM: 7, 6

stability MEANS
--> Human: 0.614265511886275 Control: 1.173390753650706 Control05: 1.1892249874503698
--> surp_tie0_stm7_exp10: 0.6037552872122885 ; surp_tieNeg_stm7_exp10: 0.6896346377362855 ; surp_tieNeg4_stm7_exp10: 0.6521868543984393 ; surp_tie0_stm6_exp10: 0.4906606847493709 ; surp_tieNeg_stm6_exp10: 0.629833029465428 ; surp_tieNeg4_stm6_exp10: 0.545714908832942



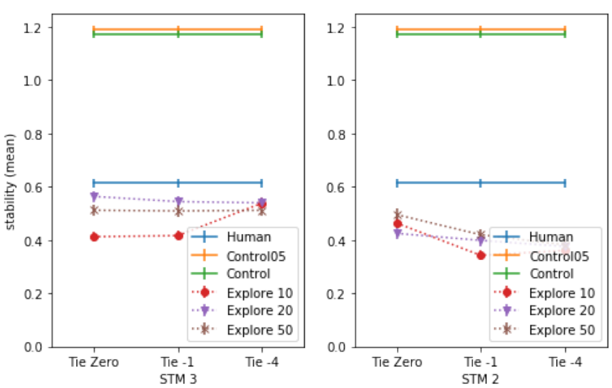
STM: 5, 4

stability MEANS
--> Human: 0.614265511886275 Control: 1.173390753650706 Control05: 1.1892249874503698
--> surp_tie0_stm5_exp10: 0.6065774015823695 ; surp_tieNeg_stm5_exp10: 0.5871729621079608 ; surp_tieNeg4_stm5_exp10: 0.6038880117596466 ; surp_tie0_stm4_exp10: 0.36950646889842753 ; surp_tieNeg_stm4_exp10: 0.658286404699041 ; surp_tieNeg4_stm4_exp10: 0.521823613976765



STM: 3, 2

stability MEANS
--> Human: 0.614265511886275 Control: 1.173390753650706 Control05:1.1892249874503698
--> surp_tie0_stm3_exp10: 0.4127712200412373 ; surp_tieNeg_stm3_exp10: 0.4167325125621432 ; surp_tieNeg4_stm3_exp10: 0.53859644487595458 ; surp_tie0_stm2_exp10: 0.4639668259569625 ; surp_tieNeg_stm2_exp10: 0.34286422626087065 ; surp_tieNeg4_stm2_exp10: 0.36080770166356474



7.3 Statistical tests. Are the distributions different?

7.3.2 Human data VS. each condition

No differences: ['surp_tieNeg_stm5_exp10', 'surp_tie0_stm3_exp20', 'surp_tieNeg_stm5_exp20', 'surp_tieNeg_stm3_exp20', 'surp_tieNeg4_stm4_exp20', 'surp_tieNeg4_stm3_exp20', 'surp_tie0_stm4_exp50', 'surp_tieNeg_stm5_exp50', 'surp_tieNeg_stm4_exp50', 'surp_tieNeg4_stm4_exp50']

7.3.4 Control data VS. each condition

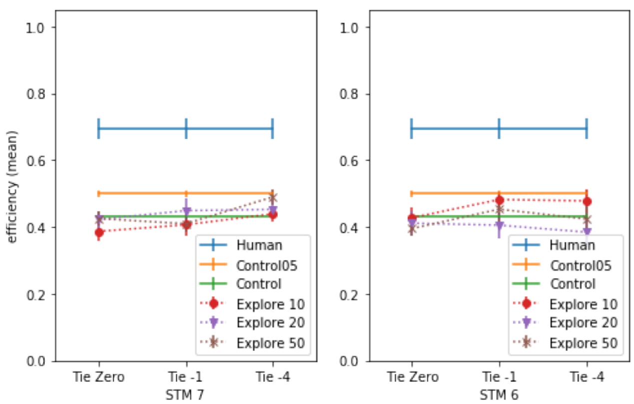
No differences: []

BALLISTIC CONDITION, SIMILARITY THRESHOLD = 0.8

5 EFFICIENCY

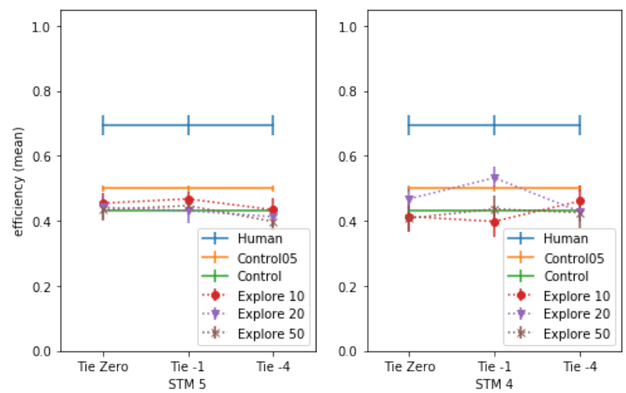
STM: 7, 6

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4996
--> Eff_tie0_stm7_exp10:  0.3871999999999993 ; Eff_tieNeg_stm7_exp10:
0.4068705882352941 ; Eff_tieNeg4_stm7_exp10:  0.43954509803921565 ;
Eff_tie0_stm6_exp10:  0.42792156862745095 ; Eff_tieNeg_stm6_exp10:
0.48239999999999994 ; Eff_tieNeg4_stm6_exp10:  0.4785254901960784
```



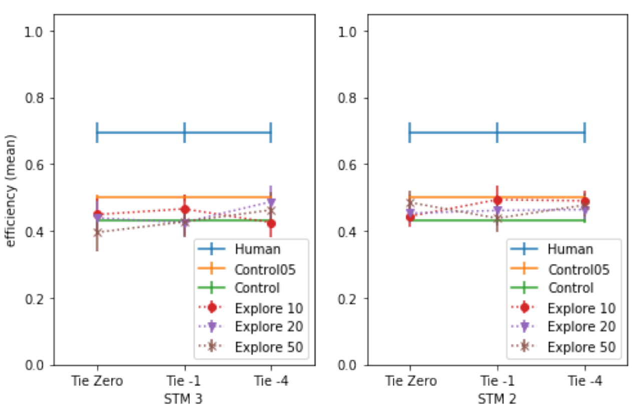
STM: 5, 4

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4996
--> Eff_tie0_stm5_exp10:  0.45465098039215684 ; Eff_tieNeg_stm5_exp10:
0.4677490196078431 ; Eff_tieNeg4_stm5_exp10:  0.43430588235294126 ;
Eff_tie0_stm4_exp10:  0.41330196078431364 ; Eff_tieNeg_stm4_exp10:
0.39849411764705883 ; Eff_tieNeg4_stm4_exp10:  0.460878431372549
```



STM: 3, 2

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4996
--> Eff_tie0_stm3_exp10:  0.44956862745098036 ; Eff_tieNeg_stm3_exp10:  0.4664 ;
Eff_tieNeg4_stm3_exp10:  0.42577254901960787 ; Eff_tie0_stm2_exp10:
0.4445176470588235 ; Eff_tieNeg_stm2_exp10:  0.49396078431372553 ;
Eff_tieNeg4_stm2_exp10:  0.49058823529411766
```



5.2 Statistical tests. Are the distributions different?

5.2.2 Human data VS. each condition

No differences: []

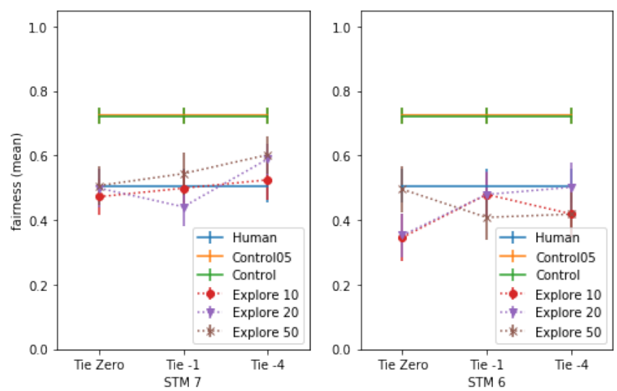
5.2.4 Control data VS. each condition

No differences: ['Eff_tie0_stm3_exp10', 'Eff_tieNeg_stm6_exp10', 'Eff_tieNeg_stm5_exp10', 'Eff_tieNeg_stm3_exp10', 'Eff_tieNeg_stm2_exp10', 'Eff_tieNeg4_stm6_exp10', 'Eff_tieNeg4_stm4_exp10', 'Eff_tieNeg4_stm2_exp10', 'Eff_tie0_stm5_exp20', 'Eff_tie0_stm4_exp20', 'Eff_tie0_stm3_exp20', 'Eff_tie0_stm2_exp20', 'Eff_tieNeg_stm7_exp20', 'Eff_tieNeg_stm4_exp20', 'Eff_tieNeg_stm2_exp20', 'Eff_tieNeg4_stm7_exp20', 'Eff_tieNeg4_stm3_exp20', 'Eff_tieNeg4_stm2_exp20', 'Eff_tie0_stm2_exp50', 'Eff_tieNeg_stm6_exp50', 'Eff_tieNeg_stm4_exp50', 'Eff_tieNeg_stm3_exp50', 'Eff_tieNeg_stm2_exp50', 'Eff_tieNeg4_stm7_exp50', 'Eff_tieNeg4_stm4_exp50', 'Eff_tieNeg4_stm3_exp50', 'Eff_tieNeg4_stm2_exp50', 'Eff_tieNeg4_stm2_exp50']

6 FAIRNESS

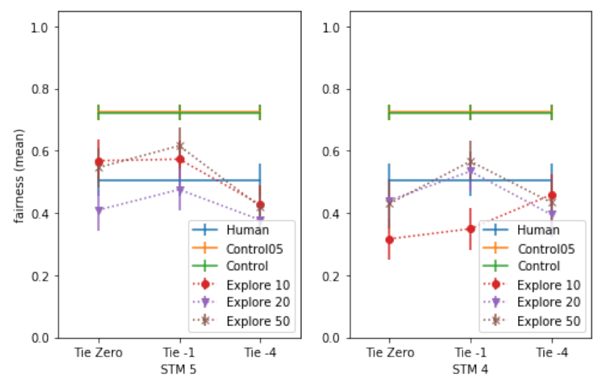
STM: 7, 6

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7235714383655559
--> Fair_tie0_stm7_exp10:  0.4726500640742127 ; Fair_tieNeg_stm7_exp10:
0.4992516343305817 ; Fair_tieNeg4_stm7_exp10:  0.5249805380228721 ;
Fair_tie0_stm6_exp10:  0.34666353965951485 ; Fair_tieNeg_stm6_exp10:
0.47981759468601576 ; Fair_tieNeg4_stm6_exp10:  0.4195949779121135
```



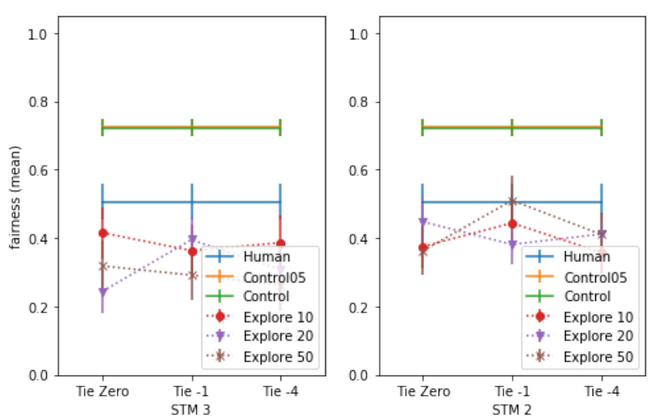
STM: 5, 4

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7235714383655559
--> Fair_tie0_stm5_exp10:  0.5683157044194196 ; Fair_tieNeg_stm5_exp10:
0.5734048701496676 ; Fair_tieNeg4_stm5_exp10:  0.42772570017693357 ;
Fair_tie0_stm4_exp10:  0.31730343150366036 ; Fair_tieNeg_stm4_exp10:
0.3497634791182793 ; Fair_tieNeg4_stm4_exp10:  0.45893307972172076
```



STM: 3, 2

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7235714383655559
--> Fair_tie0_stm3_exp10:  0.4157944407959887 ; Fair_tieNeg_stm3_exp10:
0.3637345649519563 ; Fair_tieNeg4_stm3_exp10:  0.3865841919376446 ;
Fair_tie0_stm2_exp10:  0.37563487824599684 ; Fair_tieNeg_stm2_exp10:
0.44431536733332394 ; Fair_tieNeg4_stm2_exp10:  0.3583057530704589
```



6.2 Statistical test. Are the distributions different?

6.2.2 Human data VS. each condition

No differences: ['Fair_tie0_stm7_exp10', 'Fair_tie0_stm3_exp10', 'Fair_tie0_stm2_exp10', 'Fair_tieNeg_stm7_exp10', 'Fair_tieNeg_stm6_exp10', 'Fair_tieNeg_stm5_exp10', 'Fair_tieNeg_stm4_exp10', 'Fair_tieNeg_stm2_exp10', 'Fair_tieNeg4_stm7_exp10', 'Fair_tieNeg4_stm6_exp10', 'Fair_tieNeg4_stm5_exp10', 'Fair_tieNeg4_stm4_exp10', 'Fair_tieNeg4_stm3_exp10', 'Fair_tie0_stm7_exp20', 'Fair_tie0_stm5_exp20', 'Fair_tie0_stm4_exp20', 'Fair_tie0_stm2_exp20', 'Fair_tieNeg_stm7_exp20', 'Fair_tieNeg_stm6_exp20', 'Fair_tieNeg_stm5_exp20', 'Fair_tieNeg_stm4_exp20', 'Fair_tieNeg_stm3_exp20', 'Fair_tieNeg4_stm7_exp20', 'Fair_tieNeg4_stm6_exp20', 'Fair_tieNeg4_stm5_exp20', 'Fair_tieNeg4_stm4_exp20', 'Fair_tieNeg4_stm2_exp20', 'Fair_tie0_stm7_exp50', 'Fair_tie0_stm6_exp50', 'Fair_tie0_stm5_exp50', 'Fair_tie0_stm4_exp50', 'Fair_tie0_stm2_exp50', 'Fair_tieNeg_stm7_exp50', 'Fair_tieNeg_stm6_exp50', 'Fair_tieNeg_stm5_exp50', 'Fair_tieNeg_stm4_exp50', 'Fair_tieNeg_stm2_exp50', 'Fair_tieNeg4_stm7_exp50', 'Fair_tieNeg4_stm6_exp50', 'Fair_tieNeg4_stm5_exp50', 'Fair_tieNeg4_stm4_exp50', 'Fair_tieNeg4_stm2_exp50', 'Fair_tieNeg4_stm2_exp50']

6.2.4 Control data VS. each condition

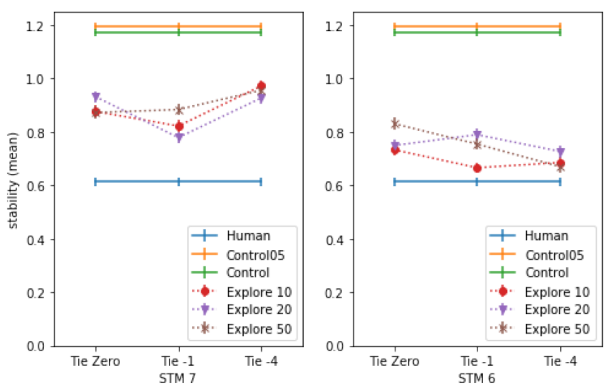
No differences: ['Fair_tieNeg_stm5_exp50']

7 Stability

Especially in ballistic conditions, people use the random assignment of the payoffs to coordinate. One player will always go top and the other will always go bottom. The stability of this pattern isn't captured by the outcomes, so we need to check the direction sequence as well.

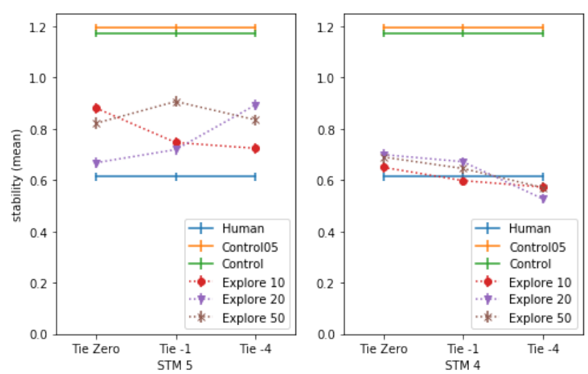
STM: 7, 6

```
stability MEANS
--> Human: 0.614265511886275  Control:  1.173390753650706  Control05:
1.1946734534118921
--> surp_tie0_stm7_exp10:  0.8766053168111169 ; surp_tieNeg_stm7_exp10:
0.8220459132634025 ; surp_tieNeg4_stm7_exp10:  0.9720519140181452 ;
surp_tie0_stm6_exp10:  0.7334570951592904 ; surp_tieNeg_stm6_exp10:
0.6661036679043778 ; surp_tieNeg4_stm6_exp10:  0.6855283757196574
```



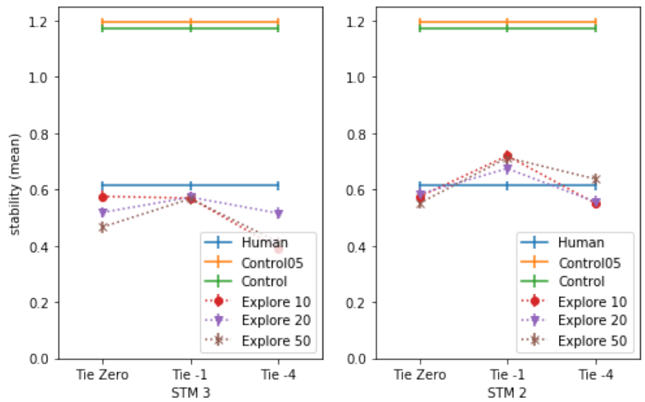
STM: 5, 4

```
stability MEANS
--> Human: 0.614265511886275  Control:  1.173390753650706  Control05:
1.1946734534118921
--> surp_tie0_stm5_exp10:  0.881064152181087 ; surp_tieNeg_stm5_exp10:
0.7467425598499248 ; surp_tieNeg4_stm5_exp10:  0.7241784622265914 ;
surp_tie0_stm4_exp10:  0.6500953172198811 ; surp_tieNeg_stm4_exp10:
0.598349044566168 ; surp_tieNeg4_stm4_exp10:  0.5733919335524914
```



STM: 3, 2

```
stability MEANS
--> Human: 0.614265511886275  Control:  1.173390753650706  Control05:
1.1946734534118921
--> surp_tie0_stm3_exp10:  0.5756405620927081 ; surp_tieNeg_stm3_exp10:
0.569397631612808 ; surp_tieNeg4_stm3_exp10:  0.391535786624145 ;
surp_tie0_stm2_exp10:  0.5742955275908538 ; surp_tieNeg_stm2_exp10:
0.7195270148217874 ; surp_tieNeg4_stm2_exp10:  0.5514905606351554
```



7.3 Statistical tests. Are the distributions different?

7.3.3 Human data VS. each condition

No differences: ['surp_tie0_stm3_exp10', 'surp_tie0_stm2_exp10', 'surp_tieNeg_stm4_exp10', 'surp_tieNeg_stm3_exp10', 'surp_tieNeg4_stm4_exp10', 'surp_tie0_stm2_exp20', 'surp_tieNeg_stm3_exp20', 'surp_tie0_stm2_exp50', 'surp_tieNeg_stm3_exp50']

7.3.4 Control data VS. each condition

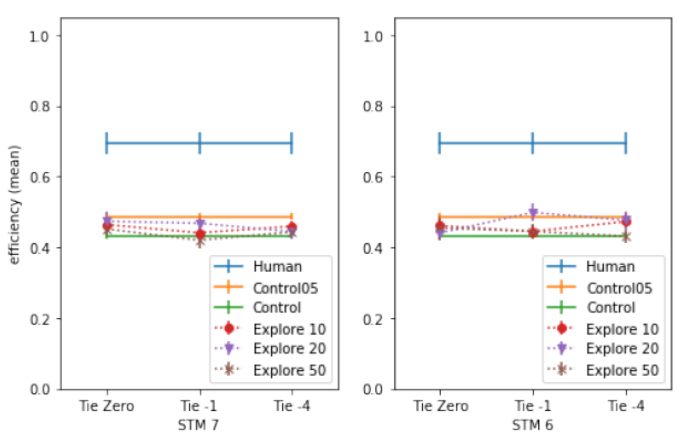
No differences: []

BALLISTIC CONDITION, SIMILARITY THRESHOLD = 0.9

5 EFFICIENCY

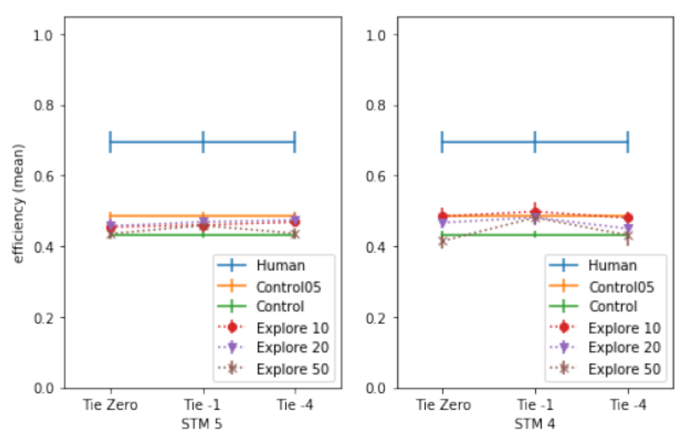
STM: 7, 6

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4876
--> Eff_tie0_stm7_exp10:  0.46432941176470593 ; Eff_tieNeg_stm7_exp10:
0.4412235294117647 ; Eff_tieNeg4_stm7_exp10:  0.45956078431372555 ;
Eff_tie0_stm6_exp10:  0.4612392156862745 ; Eff_tieNeg_stm6_exp10:
0.4452705882352941 ; Eff_tieNeg4_stm6_exp10:  0.47298823529411765
```



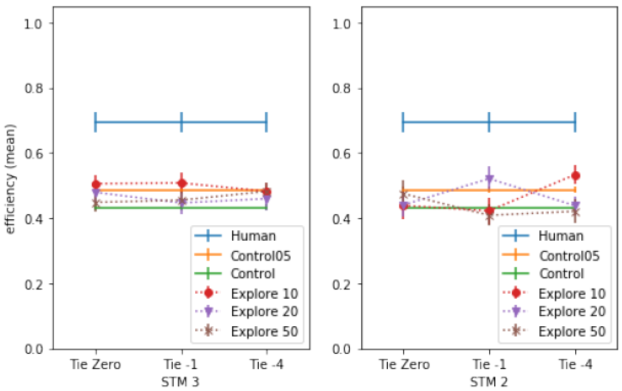
STM: 5, 4

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4876
--> Eff_tie0_stm5_exp10:  0.4545098039215686 ; Eff_tieNeg_stm5_exp10:
0.45990588235294116 ; Eff_tieNeg4_stm5_exp10:  0.4687372549019607 ;
Eff_tie0_stm4_exp10:  0.48484705882352946 ; Eff_tieNeg_stm4_exp10:
0.49816470588235295 ; Eff_tieNeg4_stm4_exp10:  0.48026666666666673
```



STM: 3, 2

```
EFFICIENCY MEANS
--> Human: 0.6935714285714285  Control:  0.4326666666666666  Control05:  0.4876
--> Eff_tie0_stm3_exp10:  0.5062117647058824 ; Eff_tieNeg_stm3_exp10:
0.508407843137255 ; Eff_tieNeg4_stm3_exp10:  0.48276078431372554 ;
Eff_tie0_stm2_exp10:  0.43920000000000003 ; Eff_tieNeg_stm2_exp10:
0.42283921568627447 ; Eff_tieNeg4_stm2_exp10:  0.5333803921568627
```



5.2 Statistical tests. Are the distributions different?

5.2.2 Human data VS. each condition

No differences: []

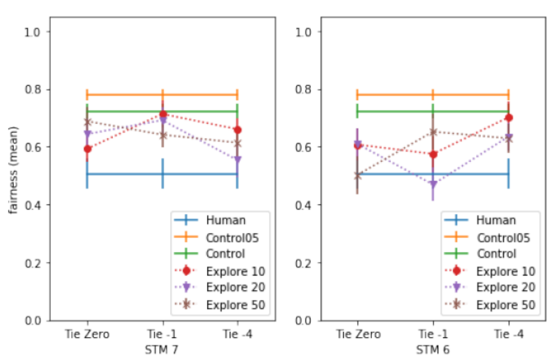
5.2.4 Control data VS. each condition

No differences: ['Eff_tie0_stm7_exp10', 'Eff_tie0_stm6_exp10','Eff_tie0_stm2_exp10', 'Eff_tieNeg_stm7_exp10', 'Eff_tieNeg_stm6_exp10', 'Eff_tieNeg_stm2_exp10', 'Eff_tieNeg4_stm7_exp10', 'Eff_tie0_stm7_exp20', 'Eff_tie0_stm6_exp20', 'Eff_tie0_stm5_exp20', 'Eff_tie0_stm3_exp20', 'Eff_tie0_stm2_exp20', 'Eff_tieNeg_stm5_exp20', 'Eff_tieNeg_stm3_exp20', 'Eff_tieNeg4_stm7_exp20', 'Eff_tieNeg4_stm6_exp20', 'Eff_tieNeg4_stm4_exp20', 'Eff_tieNeg4_stm3_exp20', 'Eff_tieNeg4_stm2_exp20', 'Eff_tie0_stm7_exp50', 'Eff_tie0_stm6_exp50', 'Eff_tie0_stm5_exp50', 'Eff_tie0_stm4_exp50', 'Eff_tie0_stm3_exp50', 'Eff_tie0_stm2_exp50', 'Eff_tieNeg_stm7_exp50', 'Eff_tieNeg_stm6_exp50', 'Eff_tieNeg_stm5_exp50', 'Eff_tieNeg_stm3_exp50', 'Eff_tieNeg_stm2_exp50', 'Eff_tieNeg4_stm7_exp50', 'Eff_tieNeg4_stm6_exp50', 'Eff_tieNeg4_stm5_exp50', 'Eff_tieNeg4_stm4_exp50', 'Eff_tieNeg4_stm3_exp50', 'Eff_tieNeg4_stm2_exp50', 'Eff_tieNeg4_stm2_exp50']

6 FAIRNESS

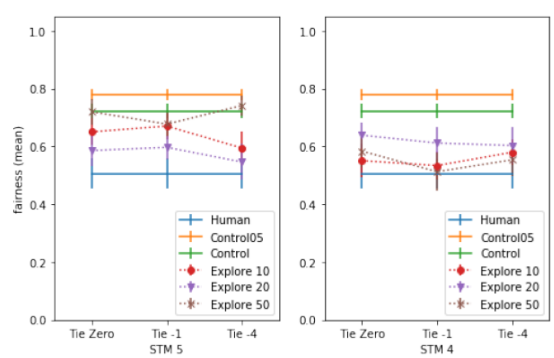
STM: 7, 6

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7798477323330265
--> Fair_tie0_stm7_exp10:  0.5927041315843039 ; Fair_tieNeg_stm7_exp10:
0.7133001998001999 ; Fair_tieNeg4_stm7_exp10:  0.6605497769058141 ;
Fair_tie0_stm6_exp10:  0.6072537816709643 ; Fair_tieNeg_stm6_exp10:
0.5750184402800502 ; Fair_tieNeg4_stm6_exp10:  0.7009107239166064
```



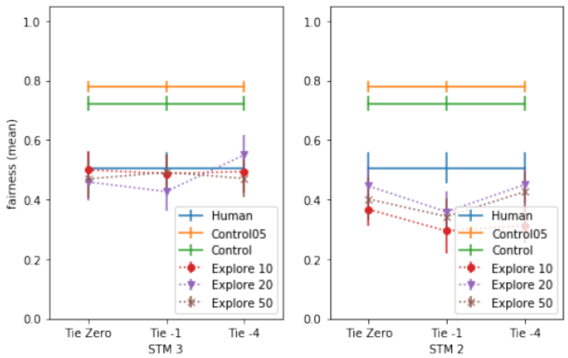
STM: 5, 4

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7798477323330265
--> Fair_tie0_stm5_exp10:  0.6509528979792139 ; Fair_tieNeg_stm5_exp10:
0.6708314648314648 ; Fair_tieNeg4_stm5_exp10:  0.5949103726598006 ;
Fair_tie0_stm4_exp10:  0.5512175145116323 ; Fair_tieNeg_stm4_exp10:
0.5341276370688134 ; Fair_tieNeg4_stm4_exp10:  0.580145792838053
```



STM: 3, 2

```
fairness MEANS
--> Human: 0.5070775958581591  Control:  0.7230590471619882  Control05:
0.7798477323330265
--> Fair_tie0_stm3_exp10:  0.5003582691817986 ; Fair_tieNeg_stm3_exp10:
0.4873474282855088 ; Fair_tieNeg4_stm3_exp10:  0.4945872789423873 ;
Fair_tie0_stm2_exp10:  0.3672752626000304 ; Fair_tieNeg_stm2_exp10:
0.2948770534436733 ; Fair_tieNeg4_stm2_exp10:  0.31379307359307357
```



6.2 Statistical tests. Are the distributions different?

6.2.2 Human data VS. each condition

No differences: ['Fair_tie0_stm7_exp10', 'Fair_tie0_stm6_exp10','Fair_tie0_stm4_exp10', 'Fair_tie0_stm3_exp10', 'Fair_tie0_stm2_exp10', 'Fair_tieNeg_stm6_exp10', 'Fair_tieNeg_stm5_exp10', 'Fair_tieNeg_stm4_exp10', 'Fair_tieNeg_stm3_exp10', 'Fair_tieNeg4_stm7_exp10', 'Fair_tieNeg4_stm5_exp10', 'Fair_tieNeg4_stm4_exp10', 'Fair_tieNeg4_stm3_exp10', 'Fair_tie0_stm7_exp20', 'Fair_tie0_stm6_exp20', 'Fair_tie0_stm5_exp20', 'Fair_tie0_stm4_exp20', 'Fair_tie0_stm3_exp20', 'Fair_tie0_stm2_exp20', 'Fair_tieNeg_stm7_exp20', 'Fair_tieNeg_stm6_exp20', 'Fair_tieNeg_stm5_exp20', 'Fair_tieNeg_stm4_exp20', 'Fair_tieNeg_stm3_exp20', 'Fair_tieNeg_stm2_exp20', 'Fair_tieNeg4_stm7_exp20', 'Fair_tieNeg4_stm6_exp20', 'Fair_tieNeg4_stm5_exp20', 'Fair_tieNeg4_stm4_exp20', 'Fair_tieNeg4_stm3_exp20', 'Fair_tieNeg4_stm2_exp20', 'Fair_tie0_stm6_exp50', 'Fair_tie0_stm4_exp50', 'Fair_tie0_stm3_exp50', 'Fair_tie0_stm2_exp50', 'Fair_tieNeg_stm7_exp50', 'Fair_tieNeg_stm6_exp50', 'Fair_tieNeg_stm5_exp50', 'Fair_tieNeg_stm4_exp50', 'Fair_tieNeg_stm3_exp50', 'Fair_tieNeg4_stm7_exp50', 'Fair_tieNeg4_stm6_exp50', 'Fair_tieNeg4_stm4_exp50', 'Fair_tieNeg4_stm3_exp50', 'Fair_tieNeg4_stm2_exp50', 'Fair_tieNeg4_stm2_exp50']

6.2.4 Control data VS. each condition

No differences: ['Fair_tieNeg_stm7_exp10', 'Fair_tieNeg_stm5_exp10', 'Fair_tieNeg4_stm7_exp10', 'Fair_tieNeg4_stm6_exp10', 'Fair_tie0_stm7_exp20', 'Fair_tie0_stm4_exp20', 'Fair_tieNeg_stm7_exp20', 'Fair_tieNeg4_stm6_exp20', 'Fair_tie0_stm7_exp50', 'Fair_tie0_stm5_exp50', 'Fair_tieNeg_stm7_exp50', 'Fair_tieNeg_stm6_exp50', 'Fair_tieNeg_stm5_exp50', 'Fair_tieNeg4_stm6_exp50', 'Fair_tieNeg4_stm5_exp50']

7 Stability

Especially in ballistic conditions, people use the random assignment of the payoffs to coordinate. One player will always go top and the other will always go bottom. The stability of this pattern isn't captured by the outcomes, so we need to check the direction sequence as well.

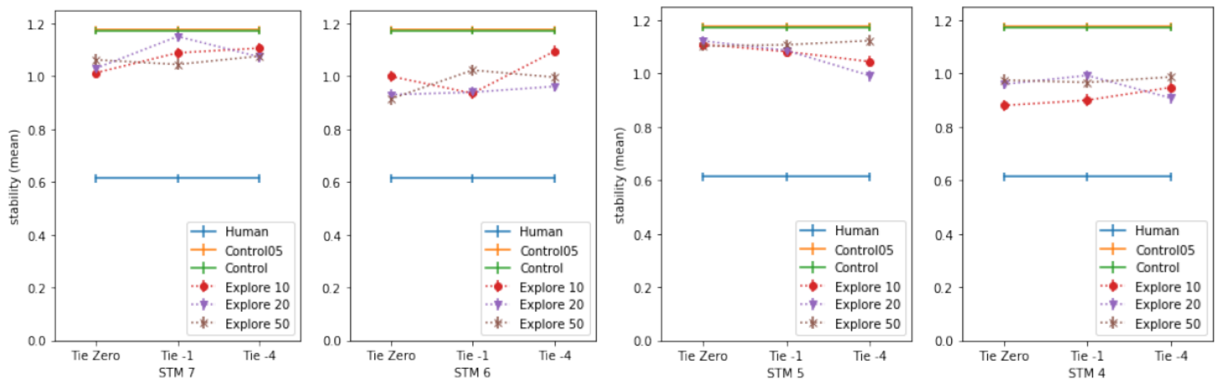
STM: 7, 6

```
stability MEANS
--> Human: 0.614265511886275  Control:  1.173390753650706  Control05:
1.1778293533837618
--> surp_tie0_stm7_exp10:  1.0129136603820583 ; surp_tieNeg_stm7_exp10:

1.0886449498301964 ; surp_tieNeg4_stm7_exp10:  1.1064387525080992 ;
surp_tie0_stm6_exp10:  0.9999068201338194 ; surp_tieNeg_stm6_exp10:
0.935448872577806 ; surp_tieNeg4_stm6_exp10:  1.0944603938894815
```

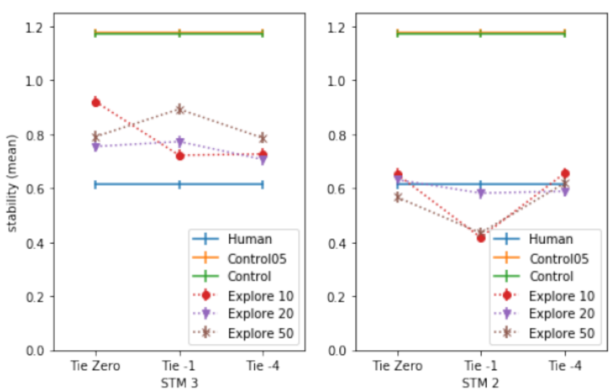
STM: 5, 4

```
stability MEANS
--> Human: 0.614265511886275  Control:  1.173390753650706  Control05:
1.1778293533837618
--> surp_tie0_stm5_exp10:  1.1088700055244105 ; surp_tieNeg_stm5_exp10:
1.0819875537944539 ; surp_tieNeg4_stm5_exp10:  1.044314452880202 ;
surp_tie0_stm4_exp10:  0.8802703900993712 ; surp_tieNeg_stm4_exp10:
0.8999878808167251 ; surp_tieNeg4_stm4_exp10:  0.9477094892029033
```



STM: 3, 2

```
stability MEANS
--> Human: 0.614265511886275  Control:  1.173390753650706  Control05:
1.1778293533837618
--> surp_tie0_stm3_exp10:  0.9211653742343269 ; surp_tieNeg_stm3_exp10:
0.7222257152478428 ; surp_tieNeg4_stm3_exp10:  0.7268239559616158 ;
surp_tie0_stm2_exp10:  0.6523309540176306 ; surp_tieNeg_stm2_exp10:
0.4178800214761433 ; surp_tieNeg4_stm2_exp10:  0.655667443908527
```



7.3 Statistical tests. Are the distributions different?

7.3.2 Human data VS. each condition

No differences: ['surp_tieNeg_stm2_exp20', 'surp_tieNeg4_stm2_exp20', 'surp_tie0_stm2_exp50', 'surp_tieNeg4_stm2_exp50', 'surp_tieNeg4_stm2_exp50']

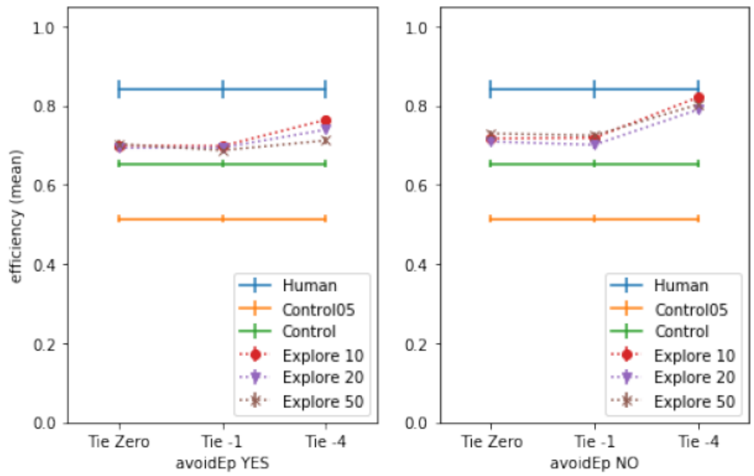
7.3.4 Control data VS. each condition

No differences:['surp_tieNeg_stm7_exp20']

DYNAMIC CONDITION, SIMILARITY THRESHOLD = 0.7

5 EFFICIENCY

```
EFFICIENCY MEANS
--> Human: 0.8420289855072464  Control:  0.6528  Control05:  0.5144000000000001
--> Eff_tie0_avoidY_exp10:  0.6991999999999998 ; Eff_tieNeg_avoidY_exp10:  0.698
; Eff_tieNeg4_avoidY_exp10:  0.764 ; Eff_tie0_avoidN_exp10:  0.7172 ;
Eff_tieNeg_avoidN_exp10:  0.7188 ; Eff_tieNeg4_avoidN_exp10:  0.8208
--> Eff_tie0_avoidY_exp20:  0.6933098039215687 ; Eff_tieNeg_avoidY_exp20:
0.6940000000000001 ; Eff_tieNeg4_avoidY_exp20:  0.7393254901960783 ;
Eff_tie0_avoidN_exp20:  0.7092941176470589 ; Eff_tieNeg_avoidN_exp20:
0.7011372549019608 ; Eff_tieNeg4_avoidN_exp20:  0.7897568627450982
--> Eff_tie0_avoidY_exp50:  0.7032 ; Eff_tieNeg_avoidY_exp50:  0.6872 ;
Eff_tieNeg4_avoidY_exp50:  0.7120000000000002 ; Eff_tie0_avoidN_exp50:
0.7296941176470589 ; Eff_tieNeg_avoidN_exp50:  0.7247999999999999 ;
Eff_tieNeg4_avoidN_exp50:  0.8028392156862745
```



5.2 Statistical tests. Are the distributions different?

5.2.2 Human data VS. each condition

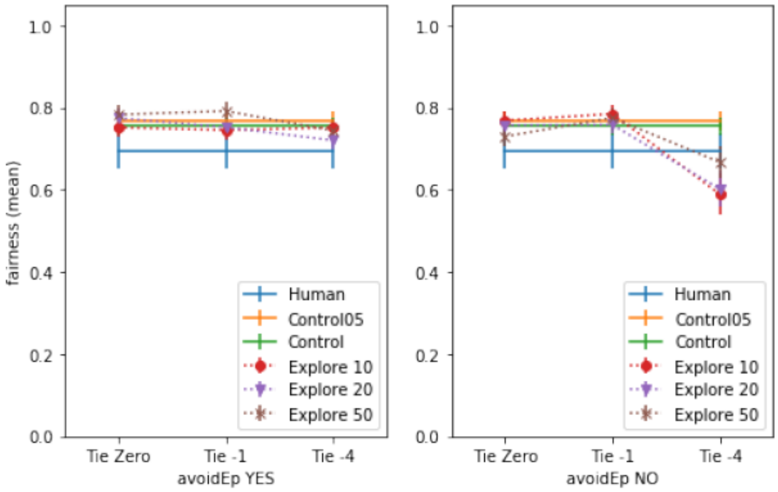
No differences: []

5.2.4 Control data VS. each condition

No differences: []

6 FAIRNESS

```
FAIRNESS MEANS
--> Human: 0.6953924654302689  Control:  0.755750763427207  Control05:
0.7684366808801792
--> Fair_tie0_avoidY_exp10:  0.7519136824500735 ; Fair_tieNeg_avoidY_exp10:
0.7456164893556063 ; Fair_tieNeg4_avoidY_exp10:  0.7517677458677932 ;
Fair_tie0_avoidN_exp10:  0.7689827931372756 ; Fair_tieNeg_avoidN_exp10:
0.7850053816882528 ; Fair_tieNeg4_avoidN_exp10:  0.5893840401948024
--> Fair_tie0_avoidY_exp20:  0.774880226196083 ; Fair_tieNeg_avoidY_exp20:
0.7524611019818787 ; Fair_tieNeg4_avoidY_exp20:  0.7206273978291344 ;
Fair_tie0_avoidN_exp20:  0.7573901931162869 ; Fair_tieNeg_avoidN_exp20:
0.7585419446168067 ; Fair_tieNeg4_avoidN_exp20:  0.602218287636386
--> Fair_tie0_avoidY_exp50:  0.7831818619147264 ; Fair_tieNeg_avoidY_exp50:
0.7920117982218197 ; Fair_tieNeg4_avoidY_exp50:  0.7441907792615696 ;
Fair_tie0_avoidN_exp50:  0.7299301101221009 ; Fair_tieNeg_avoidN_exp50:
0.7755079120219911 ; Fair_tieNeg4_avoidN_exp50:  0.6673682522517231
```



6.2 Statistical tests. Are the distributions different?

6.2.2 Human data VS. each condition

No differences: [CTRL, tie0_avoidY_exp10, tie0_avoidN_exp10, tieNeg_avoidY_exp10, tieNeg_avoidN_exp10, tieNeg4_avoidY_exp10, tie0_avoidN_exp20, tieNeg_avoidY_exp20, tieNeg_avoidN_exp20, tie0_avoidN_exp50, tieNeg_avoidY_exp50, tieNeg_avoidN_exp50, tieNeg4_avoidY_exp50]

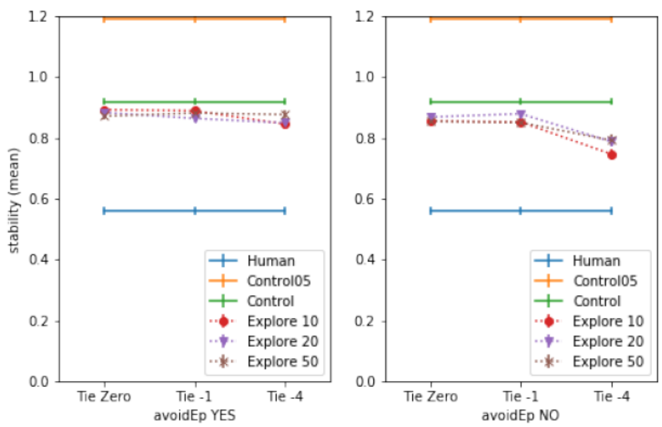
6.2.4 Control data VS. each condition

No differences: [tie0_avoidY_exp10 , tie0_avoidN_exp10 , tieNeg_avoidY_exp10 , tieNeg_avoidN_exp10 , tieNeg4_avoidY_exp10 , tieNeg4_avoidN_exp10 , tie0_avoidY_exp20 , tie0_avoidN_exp20 , tieNeg_avoidY_exp20 , tieNeg_avoidN_exp20 , tieNeg4_avoidY_exp20 , tie0_avoidY_exp50 , tie0_avoidN_exp50 , tieNeg_avoidY_exp50 , tieNeg_avoidN_exp50 , tieNeg4_avoidY_exp50 , tieNeg4_avoidN_exp50]

8 Stability

Especially in ballistic conditions, people use the random assignment of the payoffs to coordinate. One player will always go top and the other will always go bottom. The stability of this pattern isn't captured by the outcomes, so we need to check the direction sequence as well.

```
Stability MEANS
--> Human: 0.560593727166824  Control:  0.9193546022810971  Control05:
1.1919621990583782
--> Surp_tie0_avoidY_exp10:  0.8925986044473517 ; Surp_tieNeg_avoidY_exp10:
0.8888832802060835 ; Surp_tieNeg4_avoidY_exp10:  0.8451587474183844 ;
Surp_tie0_avoidN_exp10:  0.8547486420379161 ; Surp_tieNeg_avoidN_exp10:
0.8520952302565852 ; Surp_tieNeg4_avoidN_exp10:  0.7465333939448513
--> Surp_tie0_avoidY_exp20:  0.8833916177230178 ; Surp_tieNeg_avoidY_exp20:
0.863928862594518 ; Surp_tieNeg4_avoidY_exp20:  0.8490542610062994 ;
Surp_tie0_avoidN_exp20:  0.8683641456266773 ; Surp_tieNeg_avoidN_exp20:
0.8790536912131036 ; Surp_tieNeg4_avoidN_exp20:  0.7875718336078943
--> Surp_tie0_avoidY_exp50:  0.8712587576275208 ; Surp_tieNeg_avoidY_exp50:
0.8815890240530705 ; Surp_tieNeg4_avoidY_exp50:  0.8764761296305218 ;
Surp_tie0_avoidN_exp50:  0.8544600156569562 ; Surp_tieNeg_avoidN_exp50:
0.8503881623690641 ; Surp_tieNeg4_avoidN_exp50:  0.7917642622672574
```



8.2 Statistical tests. Are the distributions different?

8.2.2 Human data VS. each condition

No differences: []

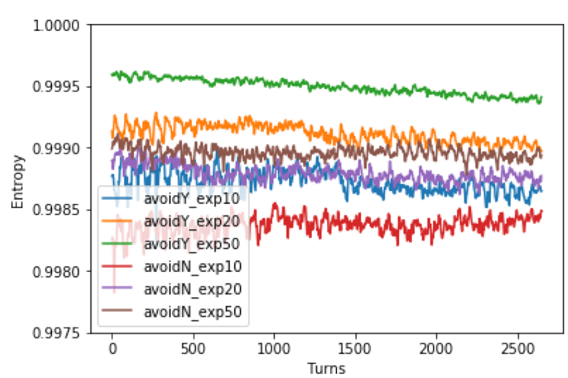
8.2.4 Control data VS. each condition

No differences: []

9 Entropy level through trials

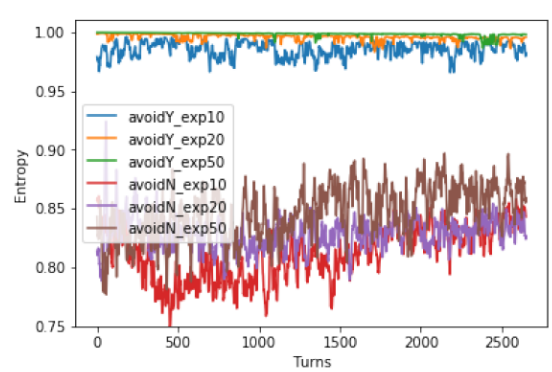
```
Average through trials
avoidY_exp10:  0.9987241649803376 ; avoidY_exp20:  0.9991122677167523 ;
avoidY_exp50:  0.9994936765798744
avoidN_exp10:  0.9983688338451505 ; avoidN_exp20:  0.9987934874255257 ;
avoidN_exp50:  0.9989582410137154
```

Entropy average (TIE 0)

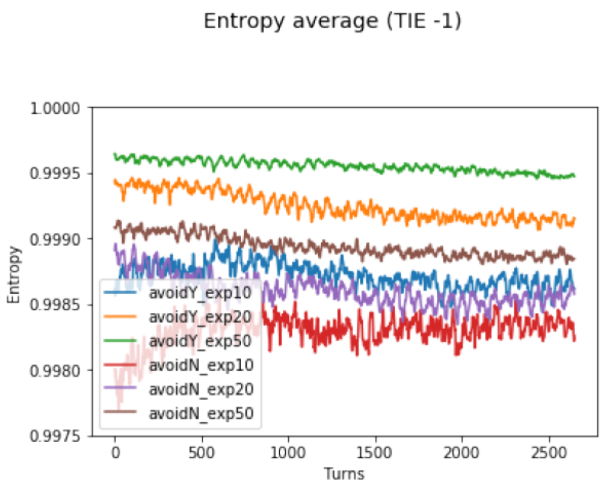



```
Average through trials
avoidY_exp10:  0.9853769856677598 ; avoidY_exp20:  0.9956899405355188 ;
avoidY_exp50:  0.9982585565697834
avoidN_exp10:  0.8114993014501183 ; avoidN_exp20:  0.8283353201431537 ;
avoidN_exp50:  0.8467784729643727
```

Entropy average (TIE -4)



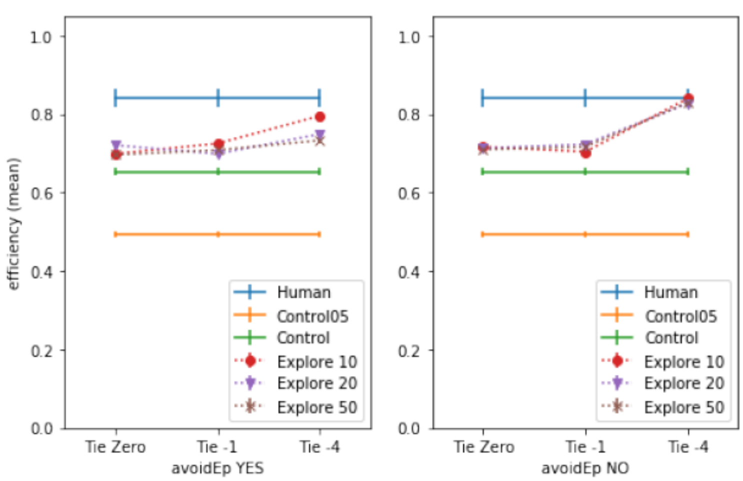
```
Average through trials
avoidY_exp10:  0.9987168749920785 ; avoidY_exp20:  0.999248245273553 ;
avoidY_exp50:  0.9995506338178117
avoidN_exp10:  0.9982959667974377 ; avoidN_exp20:  0.9986046484819712 ;
avoidN_exp50:  0.9989397995389879
```



DYNAMIC CONDITION, SIMILARITY THRESHOLD = 0.8

5 EFFICIENCY

```
EFFICIENCY MEANS
--> Human: 0.8420289855072464  Control:  0.6528  Control05:  0.494
--> Eff_tie0_avoidY_exp10:  0.6992 ; Eff_tieNeg_avoidY_exp10:  0.7252 ;
Eff_tieNeg4_avoidY_exp10:  0.7956 ; Eff_tie0_avoidN_exp10:  0.7172 ;
Eff_tieNeg_avoidN_exp10:  0.704 ; Eff_tieNeg4_avoidN_exp10:  0.8395999999999999
--> Eff_tie0_avoidY_exp20:  0.7209098039215687 ; Eff_tieNeg_avoidY_exp20:
0.6987999999999999 ; Eff_tieNeg4_avoidY_exp20:  0.7492 ; Eff_tie0_avoidN_exp20:
0.7132 ; Eff_tieNeg_avoidN_exp20:  0.7232 ; Eff_tieNeg4_avoidN_exp20:
0.8272000000000002
--> Eff_tie0_avoidY_exp50:  0.6968 ; Eff_tieNeg_avoidY_exp50:
0.7083999999999999 ; Eff_tieNeg4_avoidY_exp50:  0.7332 ; Eff_tie0_avoidN_exp50:
0.71 ; Eff_tieNeg_avoidN_exp50:  0.7168000000000001 ; Eff_tieNeg4_avoidN_exp50:
0.8292941176470588
```



5.2 Statistical tests. Are the distributions different?

5.2.2 Human data VS. each condition

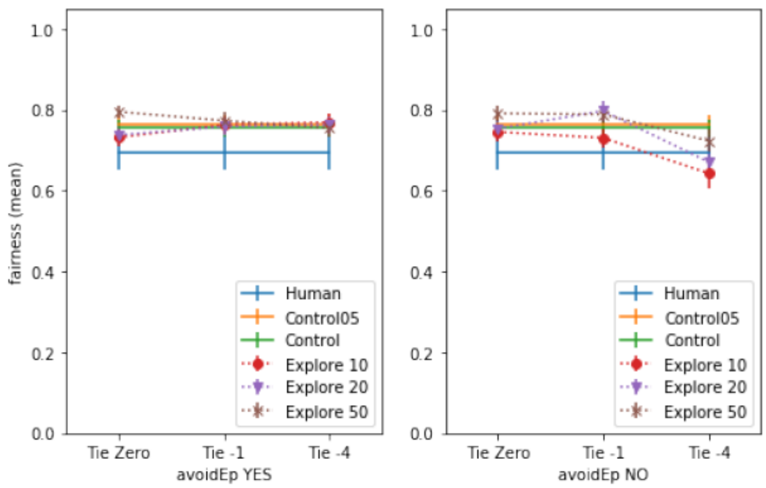
No differences: []

5.2.4 Control data VS. each condition

No differences: []

6 FAIRNESS

```
FAIRNESS MEANS
--> Human: 0.6953924654302689  Control:  0.755750763427207  Control05:
0.7639548849842968
--> Fair_tie0_avoidY_exp10:  0.7320508367953689 ; Fair_tieNeg_avoidY_exp10:
0.7631316855253051 ; Fair_tieNeg4_avoidY_exp10:  0.7692845925416983 ;
Fair_tie0_avoidN_exp10:  0.7454270629673442 ; Fair_tieNeg_avoidN_exp10:
0.7307404213027716 ; Fair_tieNeg4_avoidN_exp10:  0.6431283927878139
--> Fair_tie0_avoidY_exp20:  0.7381125864436336 ; Fair_tieNeg_avoidY_exp20:
0.7586502029991665 ; Fair_tieNeg4_avoidY_exp20:  0.7635387165891943 ;
Fair_tie0_avoidN_exp20:  0.7537874283670891 ; Fair_tieNeg_avoidN_exp20:
0.7974843317508682 ; Fair_tieNeg4_avoidN_exp20:  0.6712968555625973
--> Fair_tie0_avoidY_exp50:  0.7953333475613392 ; Fair_tieNeg_avoidY_exp50:
0.7735149363581185 ; Fair_tieNeg4_avoidY_exp50:  0.7547020879578088 ;
Fair_tie0_avoidN_exp50:  0.7914755632000551 ; Fair_tieNeg_avoidN_exp50:
0.7889697726867937 ; Fair_tieNeg4_avoidN_exp50:  0.7235110204421232
```



6.2 Statistical tests. Are the distributions different?

6.2.2 Human data VS. each condition

No differences: [CTRL, tie0_avoidY_exp10 , tie0_avoidN_exp10 , tieNeg_avoidY_exp10 , tieNeg_avoidN_exp10 , tieNeg4_avoidY_exp10 , tie0_avoidY_exp20 , tie0_avoidN_exp20 , tieNeg_avoidY_exp20 , tieNeg_avoidN_exp20 , tieNeg4_avoidY_exp20 , tie0_avoidY_exp50 , tie0_avoidN_exp50 , tieNeg_avoidY_exp50 , tieNeg_avoidN_exp50 , tieNeg4_avoidY_exp50 , tieNeg4_avoidN_exp50]

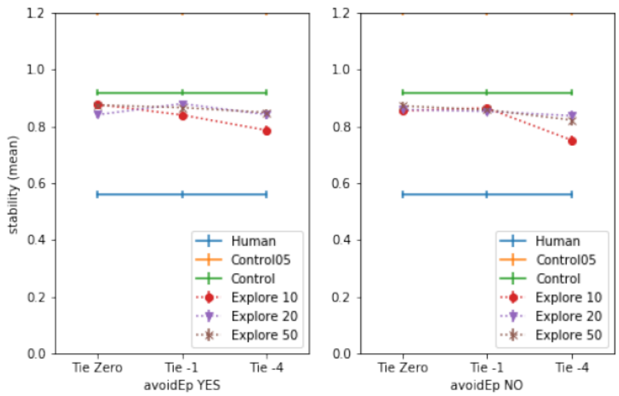
6.2.4 Control data VS. each condition

No differences: [tie0_avoidY_exp10 , tie0_avoidN_exp10 , tieNeg_avoidY_exp10 , tieNeg_avoidN_exp10 , tieNeg4_avoidY_exp10 , tie0_avoidY_exp20 , tie0_avoidN_exp20 , tieNeg_avoidY_exp20 , tieNeg_avoidN_exp20 , tieNeg4_avoidY_exp20 , tie0_avoidY_exp50 , tie0_avoidN_exp50 , tieNeg_avoidY_exp50 , tieNeg_avoidN_exp50 , tieNeg4_avoidY_exp50 , tieNeg4_avoidN_exp50]

8 Stability

Especially in ballistic conditions, people use the random assignment of the payoffs to coordinate. One player will always go top and the other will always go bottom. The stability of this pattern isn't captured by the outcomes, so we need to check the direction sequence as well.

```
Stability MEANS
--> Human: 0.560593727166824  Control:  0.9193546022810971  Control05:
1.2059977839158174
--> Surp_tie0_avoidY_exp10:  0.8774272953231588 ; Surp_tieNeg_avoidY_exp10:
0.8399623075033492 ; Surp_tieNeg4_avoidY_exp10:  0.7860440273712826 ;
Surp_tie0_avoidN_exp10:  0.855056031363807 ; Surp_tieNeg_avoidN_exp10:
0.8646003858800965 ; Surp_tieNeg4_avoidN_exp10:  0.7512799258620707
--> Surp_tie0_avoidY_exp20:  0.8408880568340382 ; Surp_tieNeg_avoidY_exp20:
0.8797226108168673 ; Surp_tieNeg4_avoidY_exp20:  0.8412806601008157 ;
Surp_tie0_avoidN_exp20:  0.8587483075861448 ; Surp_tieNeg_avoidN_exp20:
0.8530036028488973 ; Surp_tieNeg4_avoidN_exp20:  0.8372781968057649
--> Surp_tie0_avoidY_exp50:  0.8747459932505334 ; Surp_tieNeg_avoidY_exp50:
0.8660417991112224 ; Surp_tieNeg4_avoidY_exp50:  0.8485454717316218 ;
Surp_tie0_avoidN_exp50:  0.8721009019708484 ; Surp_tieNeg_avoidN_exp50:
0.8581377109645376 ; Surp_tieNeg4_avoidN_exp50:  0.8223086289923605
```



8.2 Statistical tests. Are the distributions different?

8.2.2 Human data VS. each condition

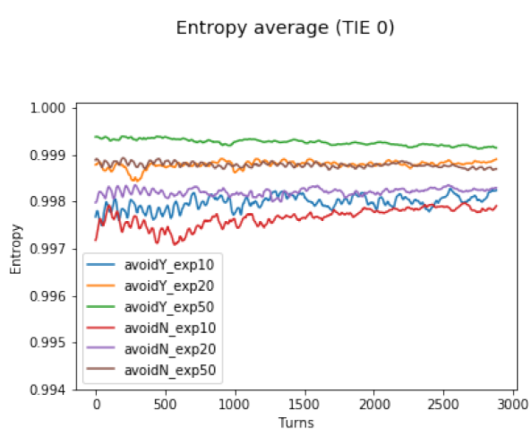
No differences: []

8.2.4 Control data VS. each condition

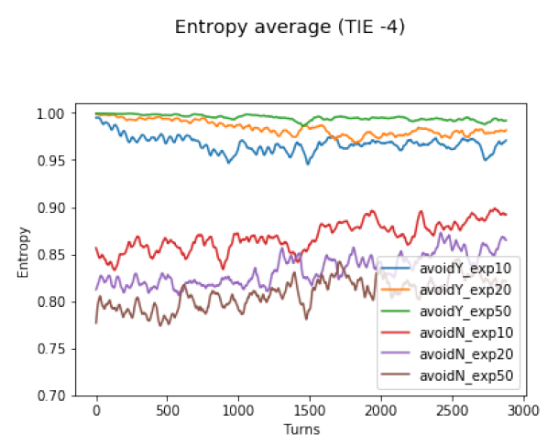
No differences: []

9 Entropy level through trials

```
Average through trials
avoidY_exp10:  0.997978443578964 ; avoidY_exp20:  0.9987874515249865 ;
avoidY_exp50:  0.999261021576852
avoidN_exp10:  0.9976577542023384 ; avoidN_exp20:  0.998206520900848 ;
avoidN_exp50:  0.998778960303053
```

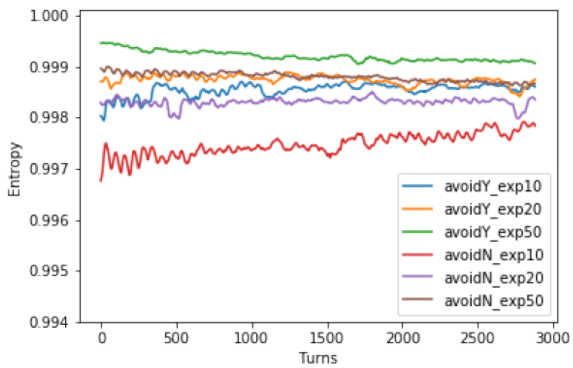



```
Average through trials
avoidY_exp10:  0.9670985286615371 ; avoidY_exp20:  0.9838200506977252 ;
avoidY_exp50:  0.9950155608572868
avoidN_exp10:  0.8684419927459963 ; avoidN_exp20:  0.8334928788847565 ;
avoidN_exp50:  0.8086362165681777
```



```
Average through trials
avoidY_exp10:  0.9985364052082073 ; avoidY_exp20:  0.9987377466832706 ;
avoidY_exp50:  0.9991996106059529
avoidN_exp10:  0.9974663095303776 ; avoidN_exp20:  0.9983009935845008 ;
avoidN_exp50:  0.9987948401427651
```

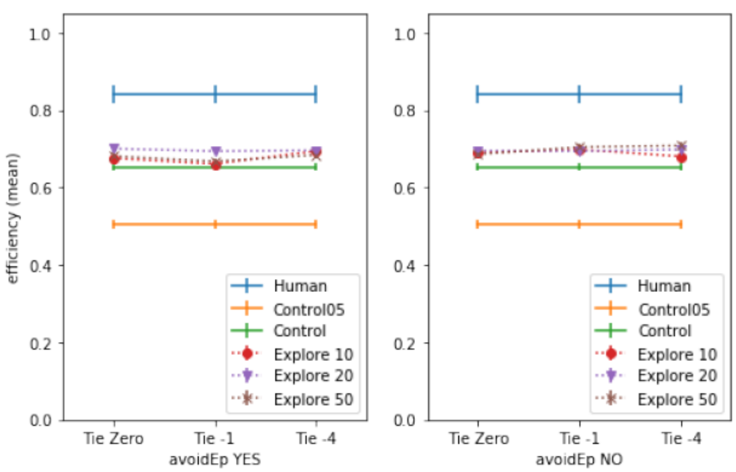
Entropy average (TIE -1)



DYNAMIC CONDITION, SIMILARITY THRESHOLD = 0.9

5 EFFICIENCY

```
EFFICIENCY MEANS
--> Human: 0.8420289855072464  Control:  0.6528  Control05:  0.5048
--> Eff_tie0_avoidY_exp10:  0.6759999999999999 ; Eff_tieNeg_avoidY_exp10:
0.6616000000000001 ; Eff_tieNeg4_avoidY_exp10:  0.6944 ; Eff_tie0_avoidN_exp10:
0.6904000000000002 ; Eff_tieNeg_avoidN_exp10:  0.6988 ;
Eff_tieNeg4_avoidN_exp10:  0.6808
--> Eff_tie0_avoidY_exp20:  0.7007999999999999 ; Eff_tieNeg_avoidY_exp20:
0.6944000000000001 ; Eff_tieNeg4_avoidY_exp20:  0.6964 ; Eff_tie0_avoidN_exp20:
0.6956 ; Eff_tieNeg_avoidN_exp20:  0.6945333333333332 ;
Eff_tieNeg4_avoidN_exp20:  0.6985176470588236
--> Eff_tie0_avoidY_exp50:  0.6816000000000001 ; Eff_tieNeg_avoidY_exp50:
0.6680000000000001 ; Eff_tieNeg4_avoidY_exp50:  0.6844 ; Eff_tie0_avoidN_exp50:
0.6864 ; Eff_tieNeg_avoidN_exp50:  0.7047999999999999 ;
Eff_tieNeg4_avoidN_exp50:  0.7088
```



5.2 Statistical tests. Are the distributions different?

5.2.2 Human data VS. each condition

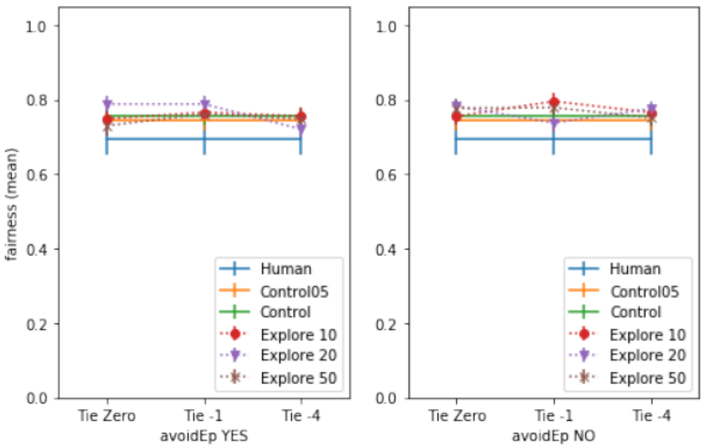
No differences: []

5.2.4 Control data VS. each condition

No differences: [tie0_avoidY_exp10, tieNeg_avoidY_exp10, tieNeg_avoidY_exp50, tieNeg_avoidY_exp50]

6 FAIRNESS

```
FAIRNESS MEANS
--> Human: 0.6953924654302689  Control:  0.755750763427207  Control05:
0.7432581138331911
--> Fair_tie0_avoidY_exp10:  0.7485465405744851 ; Fair_tieNeg_avoidY_exp10:
0.7660019750631206 ; Fair_tieNeg4_avoidY_exp10:  0.7561410073952611 ;
Fair_tie0_avoidN_exp10:  0.7553790929311187 ; Fair_tieNeg_avoidN_exp10:
0.7955890713764182 ; Fair_tieNeg4_avoidN_exp10:  0.7658560416677592
--> Fair_tie0_avoidY_exp20:  0.7883171516480507 ; Fair_tieNeg_avoidY_exp20:
0.788415479458547 ; Fair_tieNeg4_avoidY_exp20:  0.7211866189719802 ;
Fair_tie0_avoidN_exp20:  0.7825660721399715 ; Fair_tieNeg_avoidN_exp20:
0.7377560708432023 ; Fair_tieNeg4_avoidN_exp20:  0.7759685503156899
--> Fair_tie0_avoidY_exp50:  0.7302805239480162 ; Fair_tieNeg_avoidY_exp50:
0.7615282872763557 ; Fair_tieNeg4_avoidY_exp50:  0.7479429687720036 ;
Fair_tie0_avoidN_exp50:  0.7771128326998229 ; Fair_tieNeg_avoidN_exp50:
0.7788792517483861 ; Fair_tieNeg4_avoidN_exp50:  0.7513882931242938
```



6.2 Statistical tests. Are the distributions different?

6.2.2 Human data VS. each condition

No differences: [tie0_avoidN_exp10 , tieNeg_avoidY_exp10 , tieNeg_avoidN_exp10 , tieNeg4_avoidY_exp10 , tieNeg4_avoidN_exp10 , tie0_avoidY_exp20 , tie0_avoidN_exp20 , tieNeg_avoidY_exp20 , tieNeg_avoidN_exp20 , ieNeg4_avoidN_exp20 , tie0_avoidY_exp50 , tie0_avoidN_exp50 , tieNeg_avoidY_exp50 , tieNeg_avoidN_exp50 , tieNeg4_avoidY_exp50 , tieNeg4_avoidN_exp50]

6.2.4 Control data VS. each condition

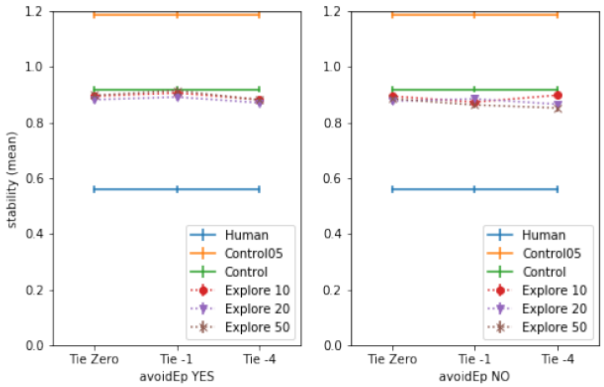
No differences: [tie0_avoidY_exp10 , tie0_avoidN_exp10 , tieNeg_avoidY_exp10 , tieNeg_avoidN_exp10 , tieNeg4_avoidY_exp10 , tieNeg4_avoidN_exp10 , tie0_avoidY_exp20 , tie0_avoidN_exp20 , tieNeg_avoidY_exp20 , tieNeg_avoidN_exp20 , tieNeg4_avoidY_exp20 , tieNeg4_avoidN_exp20 , tie0_avoidY_exp50 , tie0_avoidN_exp50 , tieNeg_avoidY_exp50 , tieNeg_avoidN_exp50 , tieNeg4_avoidY_exp50 , tieNeg4_avoidN_exp50]

8 Stability

Especially in ballistic conditions, people use the random assignment of the payoffs to coordinate. One player will always go top and the other will always go bottom. The stability of this pattern isn't captured by the outcomes, so we need to check the direction sequence as well.

```
Stability MEANS
--> Human: 0.560593727166824  Control:  0.9193546022810971  Control05:
1.1863097782831769
--> Surp_tie0_avoidY_exp10:  0.8942548964220283 ; Surp_tieNeg_avoidY_exp10:
0.9063272579531665 ; Surp_tieNeg4_avoidY_exp10:  0.8819783980260968 ;
Surp_tie0_avoidN_exp10:  0.8944613838405538 ; Surp_tieNeg_avoidN_exp10:
0.8731439629567188 ; Surp_tieNeg4_avoidN_exp10:  0.8979939543036595
--> Surp_tie0_avoidY_exp20:  0.8819407932857645 ; Surp_tieNeg_avoidY_exp20:
0.8915055588778631 ; Surp_tieNeg4_avoidY_exp20:  0.870854895528409 ;
Surp_tie0_avoidN_exp20:  0.8769405410144947 ; Surp_tieNeg_avoidN_exp20:
0.8847466888065838 ; Surp_tieNeg4_avoidN_exp20:  0.8657674029508096
--> Surp_tie0_avoidY_exp50:  0.8975765354673321 ; Surp_tieNeg_avoidY_exp50:
0.9148389214985698 ; Surp_tieNeg4_avoidY_exp50:  0.88084024535563 ;
```

```
Surp_tie0_avoidN_exp50:  0.8852827830488105 ; Surp_tieNeg_avoidN_exp50:
0.8630983171814012 ; Surp_tieNeg4_avoidN_exp50:  0.8521428662067669
```



8.2 Statistical tests. Are the distributions different?

8.2.2 Human data VS. each condition

No differences: []

8.2.4 Control data VS. each condition

No differences: [tieNeg_avoidY_exp50]

9 Entropy level through trials