Collaborative meaning-making: the emergence of novel languages in humans, machines, and human-machine interactions

Kouwenhoven, T.

3

Computationally Modelling Human Emergent Communication

In this chapter, we study human sequential behaviour by integrating cognitive, evolutionary, and computational approaches. Our work revolves around the emergence of shared vocabularies in the Embodied Communication Game (ECG). Here, participant pairs solve a shared task without access to conventional means of communication, enforcing the emergence of a new communication system. This problem is typically solved by negotiating a shared set of sequential signals that acquire meaning through interactions. Individual differences in Personal Need for Structure (PNS) have been found to influence how this process develops. We trained deep neural networks to mimic the emergence of new communicative systems in humans and used hyperparameter optimisation to approximate latent human cognitive variables in an attempt to explain human behaviour. We demonstrate that models based on bidirectional LSTM networks are better at capturing human behaviour than unidirectional LSTM networks. Suggesting that, in the ECG, human sequence processing is influenced by expected future states. The approximated variables cannot explain the differences in PNS, but we do provide evidence suggesting that random and uncertainty-directed exploration strategies are combined to develop optimal behaviour.

Originally published as: Kouwenhoven, T., Verhoef, T., Raaijmakers, S.A., de Kleijn, R.E. (2023). Modelling Human Sequential Behavior with Deep Learning Neural Networks in Emergent Communication. In M. Goldwater., F. K. Anggoro., B. K. Hayes., & D. C. Ong., editors, *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 44, pages 549-555. Cognitive Science Society.

3.1 Introduction

For communication—between humans or between humans and machines—to be successful, the coordinated actions of all interlocutors must adhere to the *grounding criterion*. Accordingly, interlocutors have to agree on the meaning of the current communicative purposes (Clark and Brennan, 1991). The fulfilment of this criterion relies extensively on the availability of a (partially) shared vocabulary between interlocutors of a conversation (Pickering and Garrod, 2004). Yet, the exact dynamics of how humans or agents settle on an effective grounded shared vocabulary are still unclear (Tylén et al., 2013; Mordatch and Abbeel, 2018). Recent work in computational linguistics started modelling emergent communication setups using multi-agent simulations to understand this process better (e.g. Lazaridou et al., 2018; Chaabouni et al., 2019a, 2020, 2022). However, the findings from these simulations often do not align with the outcomes of similar experiments with humans (Lazaridou et al., 2020; Galke et al., 2022). As such, literature proposes to instil human language patterns in machines by including human feedback in the learning loop instead of only learning from large quantities of data (ter Hoeve et al., 2022; Brandizzi and Iocchi, 2022), or by inducing additional artificial human-like biases into machines (Galke and Raviv, 2025).

The interdisciplinary research presented here attempts to instil such human communicative behaviour in machines, using an experimental setup that allows studying the initial emergence of simple signals where no communication existed before. As such, we explore the grounding problem from an evolutionary perspective, where humans must collaboratively create a novel, shared communication system to play the ECG successfully (Scott-Phillips et al., 2009). This two-player game addresses two fundamental questions in the emergence of languages: how does a signal obtain its communicative intent, and how does this signal obtain its meaning? Most human participants can solve this non-trivial task by establishing an initial convention (i.e., settling on a default behaviour) and collaboratively bootstrapping new signals onwards (Scott-Phillips et al. (2009), Chapter 2). These meaningful signals are subsequently used to play the ECG successfully, creating sequences of communicative behaviour.

Once a communicative system exists, it must be processed by the brain for comprehension and production. However, it is not entirely clear how this happens for human languages. Traditional views see the human brain as a forward-looking prediction machine (e.g. Clark, 2013), but recent findings indicate the importance of backward-looking processes for language comprehension in two self-paced reading and eye-tracking tasks (Onnis et al., 2022). Specifically, context, in the form of preceding words, can be informative for integrating current words. As such, Onnis et al. concluded that both forward *and* backward-looking appear to be important characteristics of language processing. A similar debate exists regarding the processing of everyday sequential actions (De Kleijn et al., 2014). Early accounts suggested that sequential actions are triggered by the perception of motor execution of the previous action (Washburn, 1916). Yet, there is also evidence that anticipated future states also influence subsequent actions and that planning mechanisms play a role in sequential tasks (e.g. Lashley et al., 1951; Cohen and Rosenbaum, 2004; de Kleijn et al., 2018); however, how this happens exactly is hitherto not well understood.

Context, in the form of preceding behaviour or incoming signals, and intended future states also play a role in the ECG. Incoming and produced signals (i.e., context) are informative of future behaviour, and anticipated future states can be thought of as desired behaviours by the other (i.e., ending on a specific colour). The behaviours in the ECG are moreover sequential but less complex than everyday actions and can therefore be studied in a relatively controlled manner. As such, investigating this through computational modelling may reveal how sequential processing possibly played a role in shaping human language, what types of agent architectures are required to facilitate natural communication between humans and machines, and contribute to the debate on sequential action processing in humans.

From a computational view, we use behaviour cloning to 1) investigate whether deep learning models can learn the expressed human behaviours during the development of signal–meaning mappings in the ECG; 2) approximate latent human cognitive variables by optimising model parameters that influence learning and exploration (for an overview of similar work, see Schulz and Gershman, 2019); 3) identify the applicability of networks with different processing directions to model human behaviour. We then relate the model parameters with a cognitive measure of Personal Need for Structure (Thompson et al., 1989) and compare the ability to learn human behaviour for models with different processing directions and mechanistic learning preferences. Doing so has the potential to facilitate more natural human-machine interactions through the development of (language) models that possess shared biases, resulting in a more human-like quality. Vice versa, deviations between human and computational biases provide a better understanding of why outcomes of computational simulations might not be as desired. Lastly, a better understanding of the influence of such biases on the emergence of language could steer learning mechanisms in computational simulations of emergent communication and close the gap between evolved human and computational behaviour.

3.2 Background

The origin of language is extensively studied, but the exact dynamics of language emergence remain unknown. One question concerns the origins of the initial signal–meaning mappings in case no prior communication system exists. If neither form nor meaning is known, a possible way to establish this concerns the cooperative process of agreement on the relations between communicative signals and meanings. This process has been studied extensively through laboratory experiments in which participants invent and negotiate novel signals to solve a cooperative task (Steels, 2006; Scott-Phillips and Kirby, 2010; Tylén et al., 2013). These studies show that humans can establish shared conventions and develop communication systems through social coordination. It is, moreover, suggested that in addition to language use, human learning and the transmission of a language affect the emergence of patterns (Kirby et al., 2015; Smith, 2022). A paramount explanation for the highly structured nature of human language is that it emerges due to a human bias for compressible systems, driven by a preference for simplicity (Kemp and Regier, 2012; Kirby et al., 2015; Kirby and Tamariz, 2022).

The Personal Need for Structure Scale is a measure that assesses the presence and degree of a human bias for simplicity (Thompson et al., 1989). This questionnaire quantifies individuals' need for structure (*PNS*), desire for structure and cognitive simplicity (*F1*), and the aim of restructuring an environment into a more manageable and simplified form (*F2*) (Neuberg and Newsom, 1993). Differences in the desire for structure influence how individuals understand and interact with the world (Neuberg and Newsom, 1993) and also affect problem-solving capabilities (Eva et al., 2014; Svecova and Pavlovicova, 2016). Furthermore, *PNS* affects the task progression of participants playing the ECG in that participant pairs who respond differently to a lack of structure are more successful (Chapter 2).

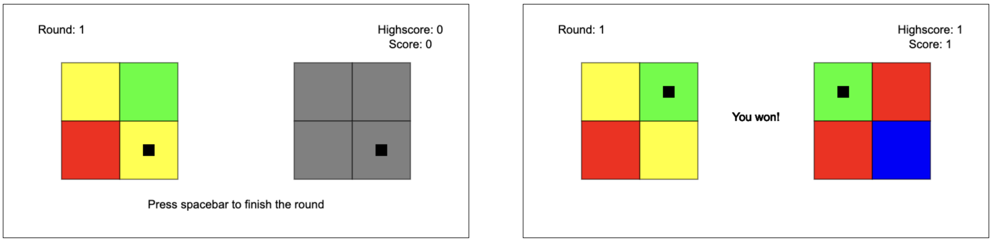
3.2.1 Embodied Communication Game

The ECG is a cooperative two-player game consisting of two 2×2 grid worlds. Each quadrant of the grid has one of four colours. Both players move between the quadrants, using the arrow keys, and share the goal of ending on identically coloured quadrants. When they manage to do so, they score a point. For both grids, the colours and starting positions are determined randomly for each round, with the proviso that there is one overlapping colour such that it is always possible to score a point, i.e., communicate successfully. Players see their own movements and the movements made by their partner, but only see the colours of their quadrants (Figure 3.1a). The colours of both worlds are revealed to both players (Figure 3.1b) when both finish moving. Their goal is to score as many consecutive points as possible, meaning that pairs must find a way to communicate reliably and coordinate behaviours (see Scott-Phillips et al., 2009, for an in-depth explanation).

3.2.2 Modelling Human Behaviours

Our work attempts to model human (sequential) behaviour using computational methods. Similar work by de Kleijn et al. (2018), for example, used reinforcement learning (RL) models to fit human behaviour in a serial reaction time (SRT) task and found that good human performance requires a high learning rate and a low discount factor. Suggesting that low-scoring individuals do not update their action-value function or the expected utility of their actions. Curricularised learning for RL agents in the SRT task showed that similar to infants' curiosity-based learning, exploration can promote robust later learning in virtual agents (de Kleijn et al., 2022).

For textual data, Nikolaus and Fourtassi (2021) evaluated the ability of neural networks to acquire meanings of words and sentences through laboratory tasks that involve cross-situational learning used with children. They showed that neural networks mirror learning patterns of



(a) The view while participants are playing.

(b) The view after both players ended the round.

Figure 3.1: An overview of the two possible game states. While the players are moving, only the participants' own grid is coloured (3.1a). When both players are done, the colours of all quadrants are revealed to both players and feedback is provided (3.1b).

acquiring semantic knowledge in early childhood and suggested that children might use partial representations of sentence structure to guide semantic interpretation. Additionally, language models seem to rely more on word frequency than children, but like children, learn words more slowly when these are part of longer utterances (Chang and Bergen, 2022). These models notably differed from children in the effects of word length, lexical class, and concreteness on learning, emphasising the importance of social, cognitive, and sensorimotor experience in child language development.

3.3 Methods

In this chapter, we attempt to investigate the relationships between computational hyperparameters and cognitive measures through training deep neural networks on human behaviours in the ECG. Specifically, algorithmic hyperparameters are used as a *proxy* of human preferences. We do not claim the existence of exactly these representations in the human brain, but merely use them as another measuring device of human behaviour.[1]

3.3.1 Data

The data used in this chapter was collected for the study described in Chapter 2. Here, we conducted three additional experiments ($N = 46$: 36 females, 10 males; $M_{\text{age}} = 22.2$, $SD_{\text{age}} = 3.53$). Participants received instructions after which they were separated and placed behind two connected computers. This setup ensured that conventional communication was impossible and that the problem of emerging signal–meaning mappings had to be solved by the participants. The game was played for 40 minutes, for an average of 256 rounds, after which participants completed the PNS questionnaire and described the communication systems they attempted to

[1] All code, materials, and data are available on OSF: `https://osf.io/n3uj6/`.

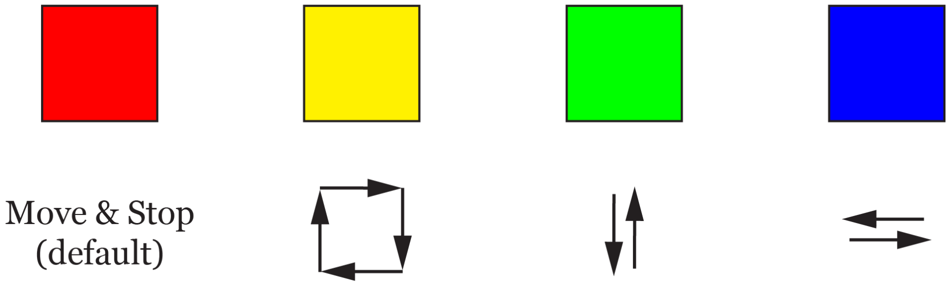


Figure 3.2: An example communication system established by participants. In this system, participants would default to a red quadrant or signal another colour through repetitive movements (displayed by the arrows).

develop. Finally, they were debriefed and allowed to discuss their experience. The Psychology Research Ethics Committee of Leiden University approved this study.

Out of 23 pairs, only 14 managed to create (i.e., reported and demonstrated) a robust communicative system. A Bayesian *t*-test showed that these pairs achieved higher scores than pairs that did not establish a system ($\mathrm{BF}_{10} = 26.73$). A typical system contains sequences of movements (i.e., signals) to indicate different colours (i.e., meanings), an exemplary system is displayed in Figure 3.2. Once established, pairs negotiate which colour is available to both by repeating the sequential moves associated with this colour. We refer the reader to Scott-Phillips et al. and Chapter 2 for a detailed description of the emergence of such communicative behaviour.

A sequence of game states, produced by the movements of each participant, is stored for each round. These game states are a digital representation of the visual environment participants see and are used to train our neural networks. A single state contains the players' position, the position of the other player, the colour of the currently occupied quadrant, and the entire colour layout of the players' grid. This representation reflects the information that a participant sees during the game. A target label—corresponding to arrow keys and the spacebar—is stored for each game state, creating a sequence of state-actions pairs. The target label serves as a class label that is predicted by our deep learning model and is used to compute the prediction loss required to update the model.

3.3.2 The model

We trained a deep neural network—implemented with Long Short Term Memory (LSTM, Hochreiter and Schmidhuber, 1997) cells—on the state–action sequences of each participant. The input data, therefore, differs for each model, but its architecture is generic and fixed (Figure 3.3). The objective of the model is to predict a participant's subsequent move given a particular sequence of states. For unidirectional processing, each state of a sequence is processed

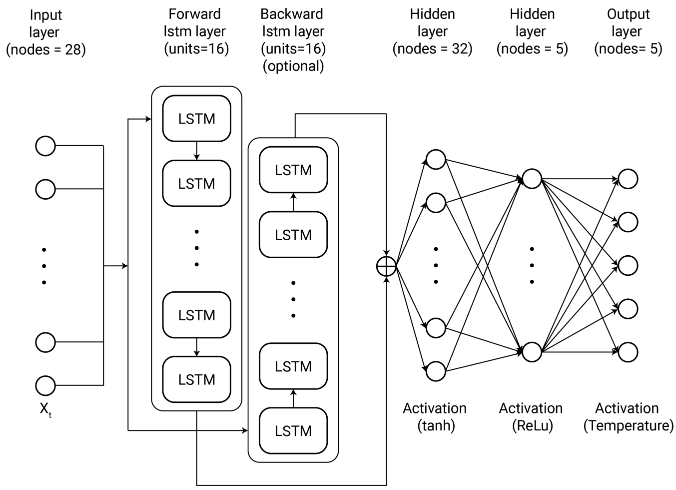


Figure 3.3: The neural network architecture used to model human behaviours. The model input X_t is the state at time t. The output layer uses temperature scaling as an activation function.

chronologically, beginning with the first and ending with the last state[2]. For bidirectional processing, the states are additionally processed in reverse order, thus incorporating (i.e., anticipating) future behaviour to predict a subsequent move. The model output layer computes probabilities for subsequent moves using temperature (τ) scaling. Here, high values of τ cause actions to be approximately equiprobable, and therefore lead to exploratory behaviour. Low values of τ result in greater differences between the probabilities, with higher probabilities for actions with higher expected rewards, and lead to deterministic behaviour. The model learning rate (lr) influences how quickly it updates its predictions, where a high learning rate means quick changes. The Adam optimisation algorithm (Kingma and Ba, 2015) is used to minimise categorical cross-entropy loss.

3.3.3 Measures

Game performance was measured by the number of consecutive successful rounds (*high score*). *PNS* and its sub-factors were collected using a 12-statement questionnaire (see Neuberg and Newsom, 1993), here, high values for *PNS*, *F1*, and *F2* correspond to a high need for structure. To obtain participant-specific τ and lr, we performed hyperparameter optimisation on the game data of each participant, resulting in 46 independently trained models. Put differently, an exhaustive grid search was used to optimise model performance using $lr \in \{0.0001, ..., 0.075\}$ and $\tau \in \{0.001, ..., 3.00\}$, with 10 equally spaced steps per parameter, resulting in 100 param-

[2]The backward processing layer is not used for unidirectional networks.

eter settings per participant. Each model was trained independently for five epochs on each parameter combination. We take the learning rate as a *proxy* of the extent to which individuals weigh feedback when updating their estimates and use temperature as an *approximation* of how explorative their behaviour was. The ability of the model to predict human sequential behaviours is reflected in its accuracy (*acc*). Lastly, the categorical cross-entropy loss (*cce*, i.e., negative log-likelihood) explains how likely the model and human would perform the same action in a particular game state. For each model, we used three-fold cross-validation to ensure that the model was not learning the data explicitly but captured the underlying structures of that participant. The cross-validation score (i.e., the average over all folds) described model performance. The parameter combination that resulted in the highest cross-validation score was used as a *proxy* for the latent human cognitive variables.

3.4 Modelling human sequential behaviour

Behaviour cloning was used to explain human behaviour in the ECG on two accounts. Firstly, by comparing PNS measures with the computational parameters. Since Neuberg and Newsom (1993) showed that differences in the need for simple structure influence how individuals understand and interact with the world, the inferred computational parameters, such as learning rate and temperature, may capture these effects as well. Therefore, we sought correspondence between these parameters and the PNS scores of each participant. We hypothesised that learning rate relates to the desire for cognitive simplicity (*F1*) and *high scores* since a desire for structure implies active searching for patterns, which seems crucial to learning signal–meaning mappings in the ECG. Learning these patterns more quickly (i.e., high lr) might result in faster emergence of communicative patterns. Individuals who feel uncomfortable in unstructured environments (i.e., high *F2*) show lower adaptability and flexibility in new environments, preferring to respond with familiar behavioural patterns to counter the uncomfortable feeling (Steinmetz et al., 2011). Since lower values of τ correspond to less exploratory behaviour and a high *lr* corresponds with high adaptability, it was expected for *lr* and τ to correlate negatively with *F2*.

Secondly, we manipulated the sequential processing cells of the models. As argued before, the next move of a signal and the intended finishing colour influence immediate action selection and can therefore be thought of as an anticipated future state. As such, optimisation as described in the previous section is done for the unidirectional (LSTM) and bidirectional LSTM (biLSTM) models. Whereas unidirectional cells process time steps of sequences in a chronological forward manner, bidirectional cells compute inputs forward *and* backwards to make predictions (Schuster and Paliwal, 1997). Note that although the LSTM layer in our model differs for both types, the remaining architecture is identical.

Table 3.1: The average model performance (*acc*) over the cross-validation scores for each participant and the average optimal learning rate and temperature across participants. *Uni* and *bi* correspond to the model types LSTM and biLSTM respectively.

	acc		**cce**		**lr**		τ	
Type	M	SD	M	SD	M	SD	M	SD
Uni	.831	.112	.355	.241	.019	.019	.356	.745
Bi	.972	.055	.084	.153	.039	.020	2.28	.716

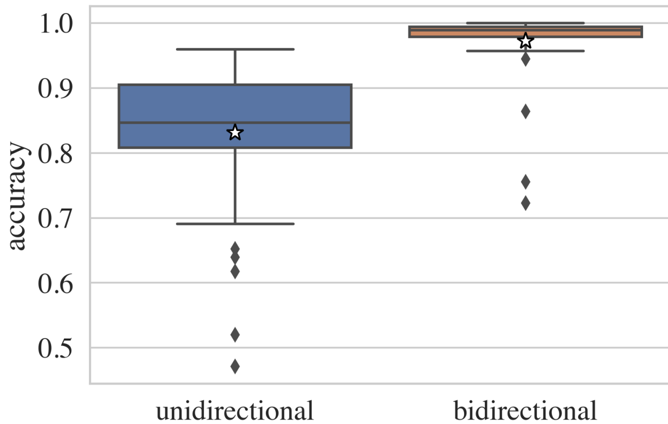
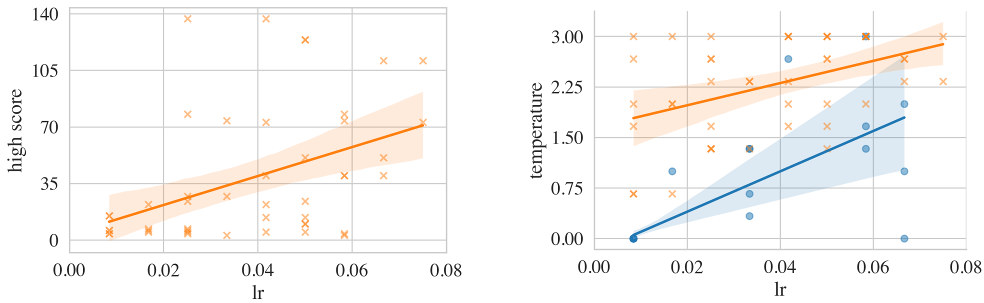


Figure 3.4: BiLSTM models show greater and more robust accuracy than LSTM models. Stars indicate mean accuracy.

3.4.1 Results

Statistical analyses were done using *R* 4.0.5 (R Core Team, 2023) and the BayesFactor 0.9.12-4.3 package (Morey et al., 2018). First, we consider the overall performance of both network types. The mean accuracy (*acc*) over all independently trained models shows that both network types can learn to predict subsequent moves relatively well (Table 3.1).

Comparison between the two network types with a Bayesian *t*-test on *acc* and *cce* with network type as a predictor revealed a large performance difference ($\mathrm{BF}_{10}acc = 6.63e + 11, d = 1.66$ and $\mathrm{BF}_{10}cce = 1.50e11, d = -1.59$). Indicating that bidirectional sequence processing can better capture the human behaviour in the ECG than unidirectional sequence processing (Figure 3.4). This result is robust when controlled for the number of parameters between the two network architectures. Optimal learning rate and temperature were higher for biLSTM networks when compared to LSTM networks ($\mathrm{BF}_{10}lr = 5.85e3, d = .790$ and $\mathrm{BF}_{10}\tau = 3.46e + 14, d = 2.00$). Since the learning rate was taken as a proxy for the extent to which individuals update their estimates, a higher learning rate implies flexible behaviour. Therefore, this result



(a) Relationship between learning rate and high score.

(b) Relationship between learning rate and temperature.

Figure 3.5: Relationships between learning rate, high score, and temperature. Each point corresponds to one participant. Note: darker marks denote overlapping data points, and the shaded area is the 95% confidence interval. Blue is used for *unidirectional* networks and orange is used for *bidirectional networks.*

suggests that bidirectional processing requires more flexibility toward updating behaviour policies. Additionally, it implies that explorative behaviour might complement updating these policies. We can assume that a higher learning rate translates to better learning in humans since learning is required to play the ECG successfully and learning rates were significantly higher for pairs that managed to establish a communicative system compared to those that did not ($\mathrm{M}_{successful} = .047, \mathrm{M}_{unsuccessful} = .025, \mathrm{BF}_{10} = 556, d = 1.39$).

We now consider the relationships between model parameters, cognitive measures, and high scores as described earlier. Successful participants (i.e., those with a high score) performed complex and structured sequences in order to communicate. Nevertheless, we find that for LSTM networks, but not for biLSTM networks, *high score* negatively influences *acc* ($\mathrm{BF}_{10} = 3.07, r = -.346, r^2 = .120$). This suggests that unidirectional processing is able to learn simpler human behaviour relatively well but has difficulties capturing more elaborate behaviours. This finding may explain the difference observed in Figure 3.4.

Bayesian regression showed that for biLSTM networks, there is a positive linear relationship between learning rate and high score (Figure 3.5a $\mathrm{BF}_{10} = 12.8, r^2 = .183$), confirming our hypothesis and suggesting that participants who adopt new behaviours faster are more successful in creating new signal–meaning mappings in the ECG. We moreover find that regardless of processing directionality, temperature, and learning rate are related (Figure 3.5b, $\mathrm{BF}_{10} biLSTM = 28.1, r = .452, r^2 = .204$ and $\mathrm{BF}_{10} LSTM = 1.40e7, r = .772, r^2 = .597$), suggesting that participants who explored more also adapted new behaviours faster. Surprisingly, we did not find a relation between exploration and *high score*. A relationship was expected since explorative behaviour may lead to new conventions in the ECG. Lastly, learning rate or temperature cannot explain *PNS*, *F1*, or *F2* for LSTM and biLSTM networks. Thereby also rejecting the remaining hypotheses.

3.5 Discussion

In this chapter, we modelled human sequential behaviour in the Embodied Communication Game with deep neural networks and investigated possible relationships between human cognitive preferences and computational parameters. Specifically, we looked at relationships between participants' personal need for structure, learning rate, and temperature parameters. Though we showed that current deep neural networks can learn the behaviour associated with creating signal–meaning mappings, we did not find any correspondences between cognitive and computational variables. As such, *PNS*, used here as a human bias for structure (Kirby and Tamariz, 2022), cannot be captured with this setup. Further research should investigate how parameters of various network architectures may correspond to cognitive measures or look at different games that investigate emergent communication (e.g. Galantucci, 2005; Steels and Loetzsch, 2012; Mordatch and Abbeel, 2018). The ability to capture human biases, such as the human bias for compressible and simple systems (Kemp and Regier, 2012; Kirby et al., 2015), in computational systems is insightful for simulations of emergent communication as they are then closer to human experiments. Furthermore, playing these collaborative games between humans and machines might also result in shared grounded vocabularies that are adapted to the biases of humans *and* computers, ultimately resulting in better conversational AI (Chapter 1).

Manipulation of the processing directionality of action sequences showed that participants' behaviour was explained better by biLSTM models than by LSTM models. This thereby provides additional arguments for the bidirectional processing of sequential actions in humans (Lashley et al., 1951; Cohen and Rosenbaum, 2004; Onnis et al., 2022). For communicative purposes in the ECG, integrating current actions is dependent on the preceding shared context (i.e., the negotiations of signals and intended final colours), and must be taken into account when deciding what moves to take next. The difficulties for LSTM networks to learn more complex behaviours performed by more successful participants also indicate that unidirectional processing is insufficient to capture more elaborate human behaviour. Although additional analysis is needed to support this, these findings suggest that the effect of a backward-looking mechanism found by Onnis et al. (2022) in a self-paced reading task might originate in the very early stage of forming signalling conventions. To verify this, simulations of emergent communication with deep learning agents should look at the effect of processing directionality of network architectures on the structure of emergent communicative protocols. Integrating bidirectional networks may close the current gap between human experiments and simulations.

We demonstrated that for biLSTM networks, the learning rate has a positive influence on high scores and is positively correlated with temperature (Figure 3.5b). This seems to support the recent view which suggests that humans combine random and uncertainty-directed exploration strategies to develop optimal behaviour (Jepma et al., 2016; Schulz and Gershman, 2019). An explanation for this could be that explorative behaviour in the ECG led to the emergence of new signals, which need to be learned quickly (i.e., require a high learning rate) to be useful. In other words, the correlation between learning rate and temperature likely reflects the fact that participants who are more explorative benefit from higher learning rates (i.e., there is no benefit to explorative behaviour if you do not use the explored options to update expected values). However, a more in-depth analysis is required to strengthen this link further. For optimal behaviour, learning rate and explorative behaviour would be expected to decrease over time as strategies are learned and exploration becomes less necessary, instead exploiting the knowledge gathered thus far. However, literature on how learning rate and temperature parameters develop with age and experience has yielded conflicting results (Nussenbaum and Hartley, 2019). Games like the ECG could be extended over time to investigate the dynamic nature of the temperature and learning rate parameters.

Lastly, we acknowledge that the ECG is a highly simplified setup, thereby limiting the generalisability to real-world processing (Nastase et al., 2020). It also goes without saying that these models are mere approximations of the human brain and do not capture its breadth, but we can nevertheless use them as a proxy to mimic human processes. These findings must therefore be replicated in more ecological settings.

3.6 Conclusion

In this chapter, we modelled sequential human behaviour captured in the Embodied Communication Game with deep neural networks. Here, participants establish a communication system from scratch to solve a collaborative task. We demonstrate that neural networks can learn the human behaviours associated with the creation of a new communication system. Manipulation of network types shows that bidirectional processing of sequential actions better explains human behaviour than unidirectional processing, hereby providing additional arguments for the existence of a planning mechanism for sequential signal production in humans. No relationship was found between Personal Need for Structure and participant-specific computational parameters, but our results suggest that humans combine random and uncertainty-directed exploration strategies to develop optimal behaviour in the ECG. Future research should attempt to extrapolate our results to communicative settings with complex linguistic signal exchange (e.g., between chatbots and humans). Additionally, experiments on the emergence of a more complex human–AI language will deepen the understanding of the relationship between natural and artificial biases that play a role during the emergence of communicative systems.