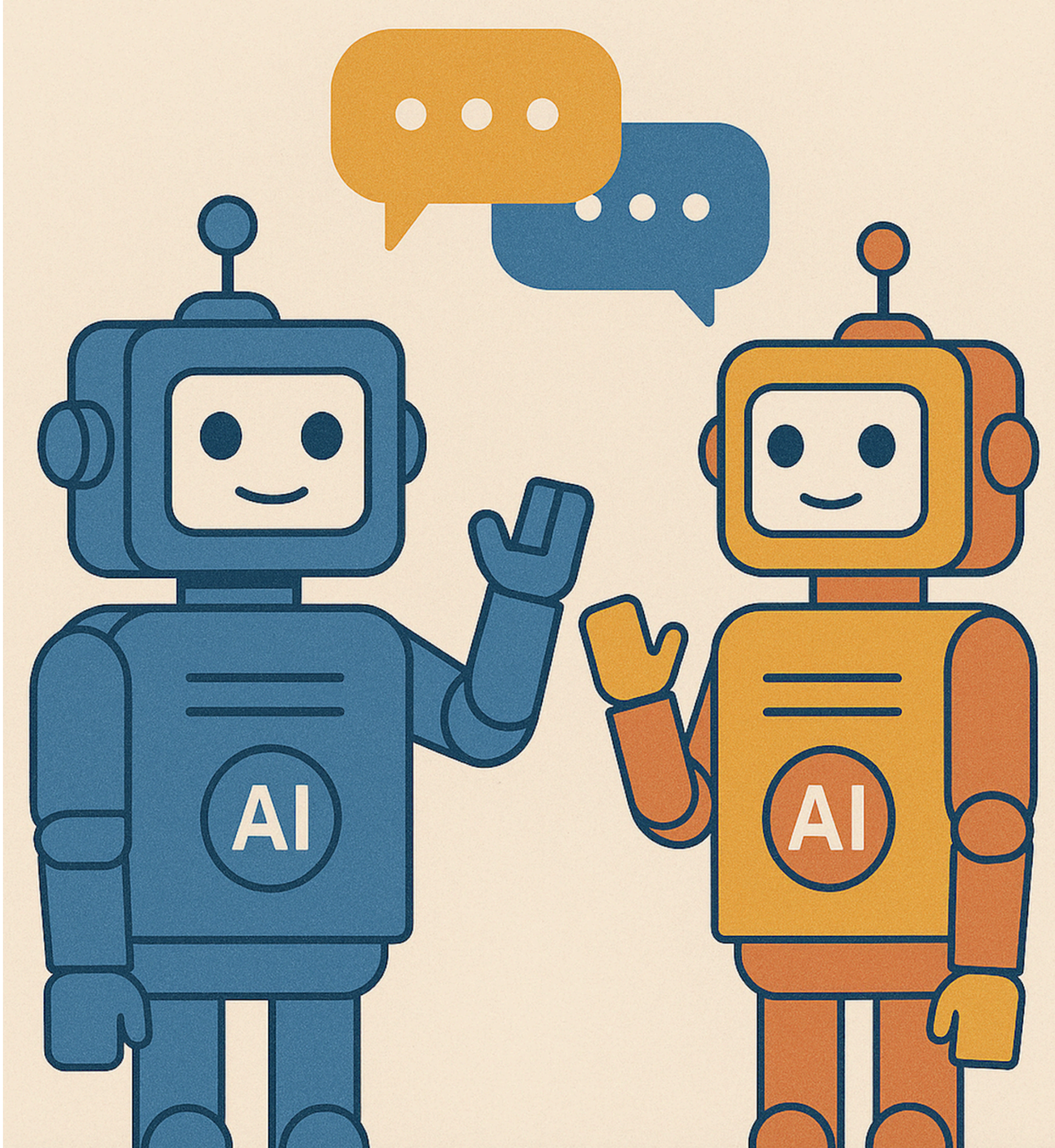
Learning Translations: Emergent Communication Pretraining for Cooperative Language Acquisition

by Dylan Cope and Peter McBurney
King's College London
IJCAI 2024, pages 40-48

SEMINAR KÜNSTLICHE INTELLIGENZ
MANSUR FROMM
16.07.2025

Agenda

1. Einführung & Motivation
2. Hintergrundkonzepte
3. Formale Definition von CLAP-Problem
4. Disentangling Competencies
5. Methoden zur Lösung von CLAP
6. Experimenteller Aufbau & Ergebnisse



Einführung & Grundlagen

Multi-Agenten-Systeme: Agenten lernen Kommunikation & Kooperation

Emergent Communication

- Agenten entwickeln eigene Kommunikationssprache ohne Vorgaben
- Diskrete Nachrichten (ähnlich Wörter) mit konventionsbasierten Bedeutungen

Herausforderung

- Protokolle sind spezialisiert & oft nicht kompatibel zwischen Communities
- **Fehlende Generalisierung** bei Zusammenarbeit mit neuen Partnern

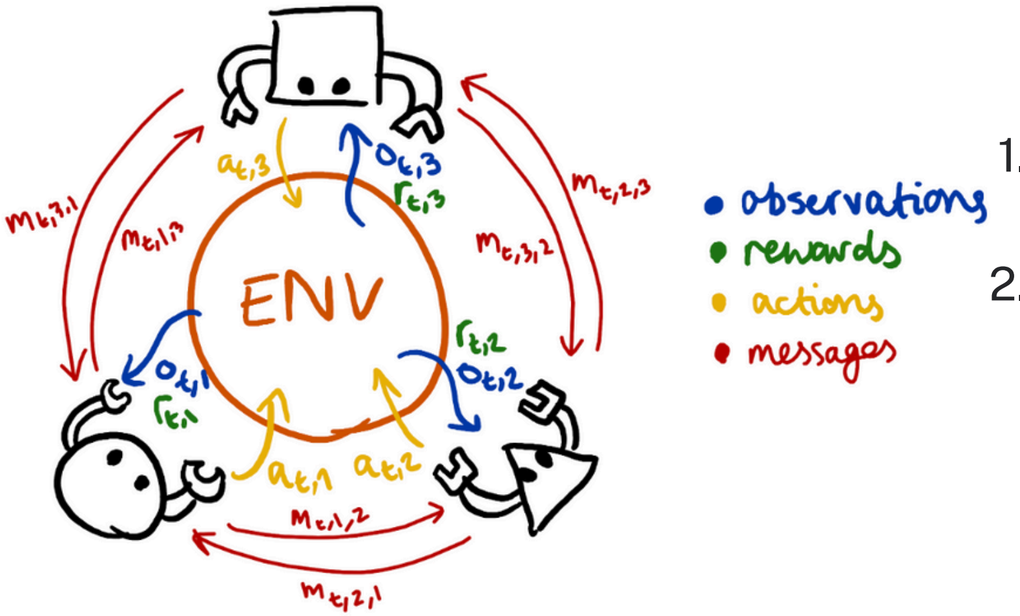
Zero-Shot-Coordination

- ZSC: Koordination ohne Vorwissen – meist unrealistisch
- Gemeinsame Sprache/Konventionen erleichtern Zusammenarbeit

Cooperative Language Acquisition (CLAP) & Lösungsansätze

Cooperative Language Acquisition Problem (CLAP)

- Neuer Agent („Joiner“) lernt Sprache & Verhalten einer Ziel-Community aus Beobachtungsdaten
- Ziel: nahtlose Kooperation ohne vorheriges Live-Training



Methoden zur Lösung

1. Behaviour Cloning (BC): Imitation aus Beobachtungsdaten
2. Emergent Communication Pretraining and Translation Learning (ECTL):
   - Self-Play-Training + Übersetzung zwischen Protokollen
   - Bessere Generalisierung & Robustheit

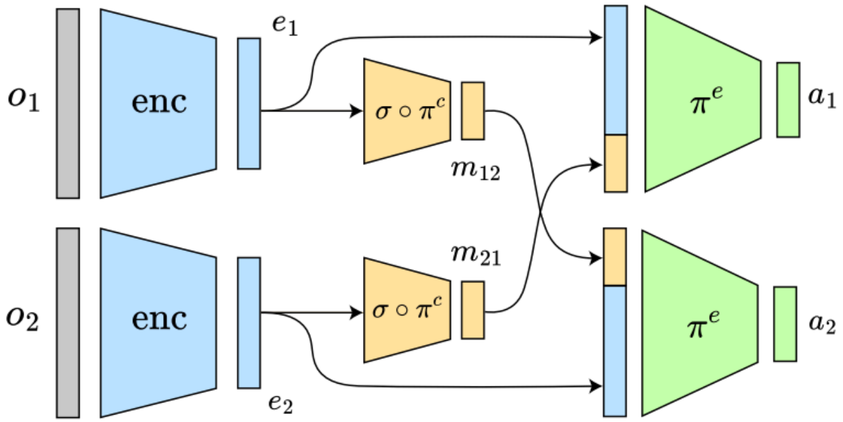
Decentralised Partially Observable Markov Decision Processes (Dec-POMDPs)

- Mehrere Agenten handeln gemeinsam unter teilweiser Beobachtung
- Jeder Agent sieht nur einen Teil der Umgebung
- Ziel: gemeinsame Belohnung maximieren
- Formal definiert durch $M = (S, A, T, r, \Omega, O)$:
  - *States S*: Umgebungszustände
  - *Actions* $A = \Pi_i A_i$:Aktionen je Agent
  - *Transition Function* $T : S \times A \times S \to [0, 1]$: Zustandsübergänge
  - *Reward Function* $r : S \times A \times S \to \mathbb{R}$: Team-Belohnung
  - *Observations* $\Omega = \{\Omega_i\}$, *Observation Function* $O : S \to \Pi_i \Omega_i$:Teilbeobachtungen
- Policies π_i entscheiden Aktionen basierend auf Beobachtungen
- Erfahrung als Verlaufssequenzen (Trajektorien) von Beobachtungen, Aktionen und Belohnungen

Emergent Communication in Multi-Agent Systems

- Agenten lernen eigene Kommunikationssprache (emergent communication)
- Aktionen:
  - Environment Actions A_i^e
  - Communicative Actions A_i^c(Nachrichten)
- Nachrichten über kostenfreie **cheap-talk channels**
- Policies:
  - Kommunikationspolicy $\pi_i^c : \Omega_i \to A_i^c$
  - Umweltpolicy $\pi_i^e : \Omega_i \times \Sigma^{n_s} \to A_i^e$
- Kommunikation erkennbar durch positive signalling & positive listening

CLAP



1

CLAP

Nahtlose Integration eines neuen Agenten (Joiner) in ein bestehendes Team

2

Ziel-Community

Team von Agenten (z.B. Agent A & B), die effektiv kommunizieren und kooperieren

3

Joiner ersetzt ein Teammitglied und muss sofort kommunizieren & zusammenarbeiten

4

Relevanz

Schnelle Integration ohne Trial-and-Error, wichtig für autonome Fahrzeuge, Roboterteams etc.

1

Ziel

- Joiner-Agent soll einen Agenten πk aus Team $\prod = \{\pi_i\}$ ersetzen
- Joiner soll maximale Team-Leistung erreichen

$$\pi_{\mathcal{J}} \in \arg\max_{\pi'} R(\{\pi'\} \cup \Pi^{-k}) \quad (1)$$

2

Zero-Shot Evaluation

- Joiner muss sofort beim ersten Einsatz funktionieren
- Kein Lernen durch Interaktion möglich (stört Teamkoordination)
- Lernen erfolgt ausschließlich vor dem Einsatz mit vorhandenen Daten

3

Verfügbare Daten für den Joiner

Datensatz D_k mit Interaktionen des zu ersetzenden Agenten π_k:

- Sender-Beobachtung
- Gesendete Nachricht
- Empfänger-Beobachtung
- Aktion des Empfängers

4

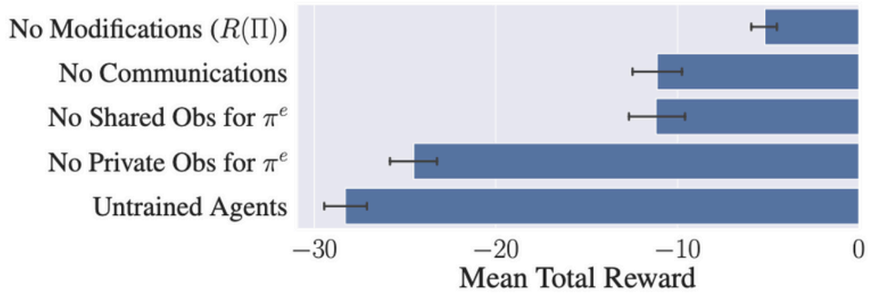
Bedeutung des Kommunikationsprotokolls

- Nachrichten sind anfangs bedeutungslos und variabel
- Nur Aktionen beobachten reicht nicht
- Joiner muss Kommunikationssprache lernen, um sinnvoll zu senden und zu verstehen
- Schlüssel für erfolgreiche Team-Koordination

Environment-Level vs. Communicative Competencies

- **Environment-level competencies:** Fähigkeiten zur Interaktion mit der Umwelt (Bewegen, Objekte manipulieren, Navigation)
- **Communicative competencies:** Fähigkeiten zum Austausch und Verständnis von Nachrichten
- Beide sind für Team-Erfolg essenziell: Umwelt-Fähigkeiten ermöglichen Handlungen, Kommunikation ermöglicht Koordination
- Trennung wichtig, da Umwelt-Fähigkeiten oft unabhängig von Kommunikation gelernt werden können
- Kommunikation basiert auf geteilten, nicht direkt beobachtbaren Protokollen

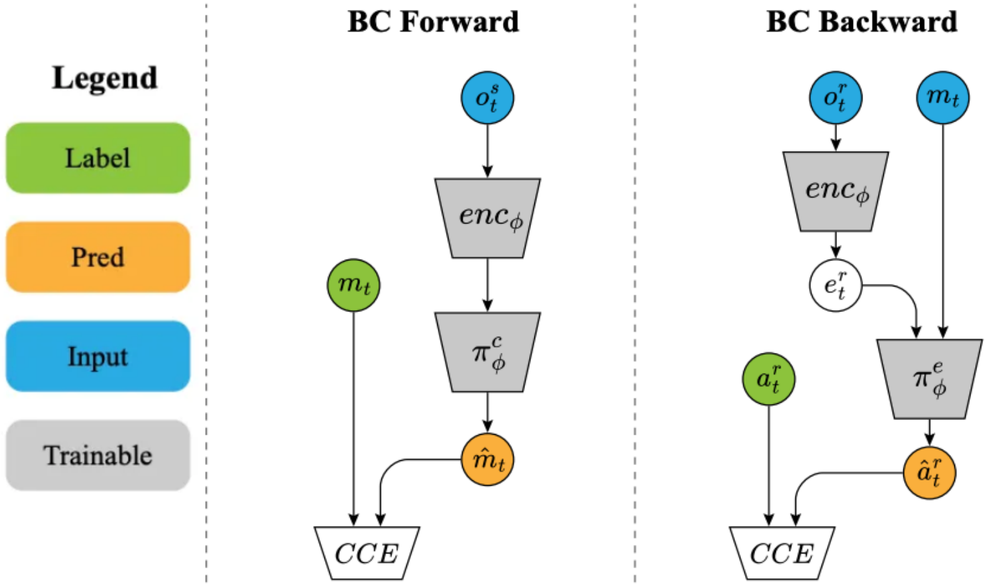
Ablationsmethodik (Informationen gezielt entfernen)



(a) Ablated agents in the target community.

- Beobachtungen aufgeteilt in
  - Private (lokale) Beobachtungen l_t^i
  - Geteilte (globale) Beobachtungen g_t
- **Kommunikation blockieren:** Leistungsabfall R → R′ zeigt Bedeutung der Kommunikation
- **Private Info aus environment-level Policy entfernen:**
  - Eingabe wird zu $o'_t = (0, g_t)$
  - Policy: πi′(ot,Mt)= $\pi'_i(o_t, M_t) = (\pi_i^e(o'_t, M_t), \pi_i^c(o_t))$
  - Sinkt Leistung R′ → R″, sind private Infos wichtig für Umweltfähigkeiten
- Keine Veränderung der Kommunikationspolicy
- Keine Leistungseinbuße? Agent hängt allein von Kommunikation oder globalen Infos ab

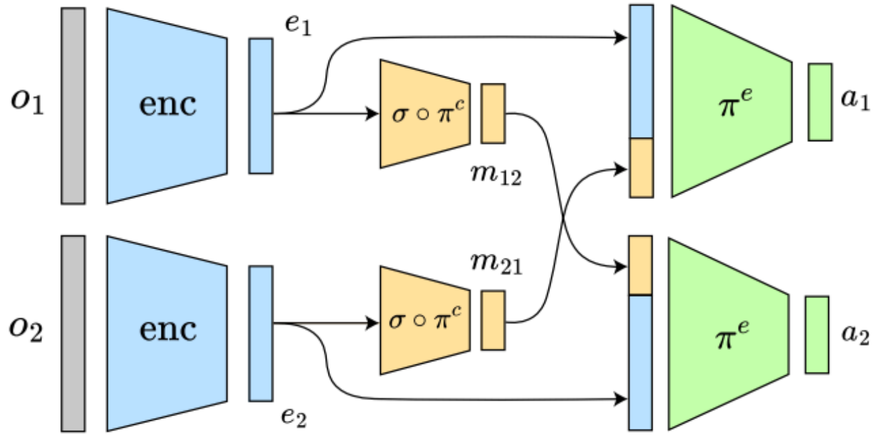
Behavior Cloning (BC)



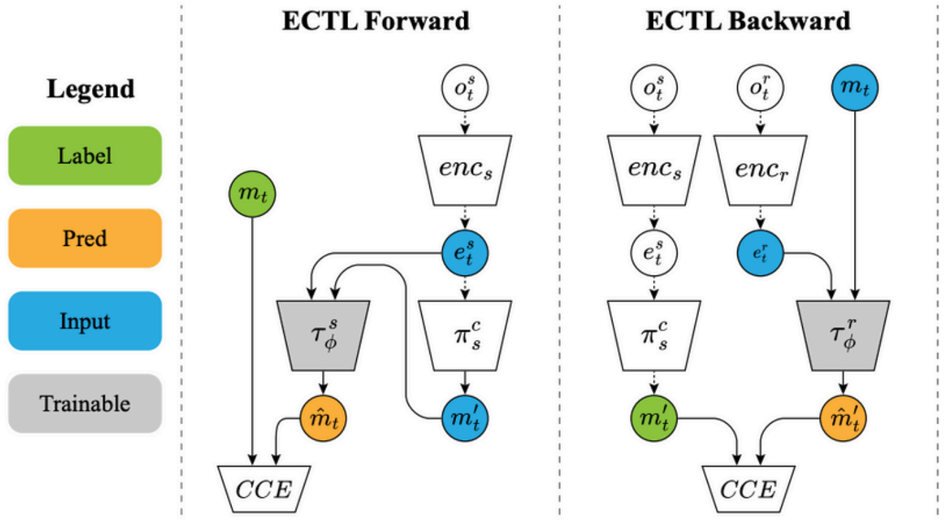
- Klassische Imitationslerntechnik: Joiner imitiert Ziel-Agenten
- Nutzt Datensatz mit Sender-Beobachtung, Nachricht, Empfänger-Beobachtung, Empfänger-Aktion
- Dataset in zwei Teile:
  - Signalling: Vorhersage gesendeter Nachrichten
  - Listening: Vorhersage empfangener Aktionen
- Überwachtes Lernen mit **categorical cross-entropy (CCE)** als Verlust
- Architektur: Kodieren → Nachrichten vorhersagen (Forward) → Aktionen vorhersagen (Backward)
- Einschränkungen: Brüchigkeit, kumulierende Fehler, schlechte Generalisierung bei unbekannten Situationen

Emergent Communication Pretraining and Translation Learning (ECTL)

- **Schritt 1**: Self-Play Pretraining einer neuen Agenten-Community mit emergent communication
  - Architektur: Observation Encoder, Communication Head $\sigma o \pi^c$, Action Head π^e
- **Schritt 2**: Translation Learning
  - Zwei Übersetzungsfunktionen:
    - Signalling: τ_ϕ^s (pretrained → Zielprotokoll)
    - Listening: τ_ψ^r (Zielprotokoll → pretrained)
  - Training mit Ziel-Community-Daten
- Vorteile:
  - Robustheit gegen unbekannte Situationen
  - Daten-Effizienz durch Fokus auf Protokoll-Mapping
  - Generalisierung auf verschiedene Zielteams
- Joiner-Policy: Kombination von pretrained Kommunikation + Translation

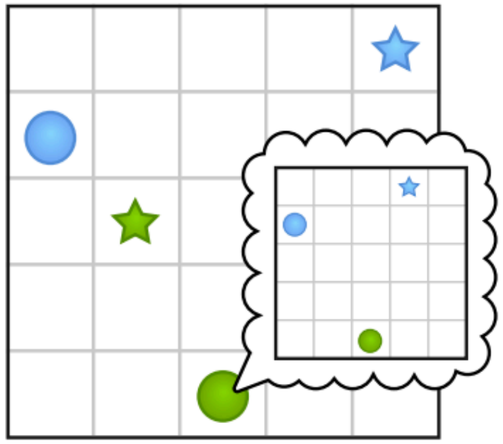


(a) Joint agents architecture (two agents) with communication.

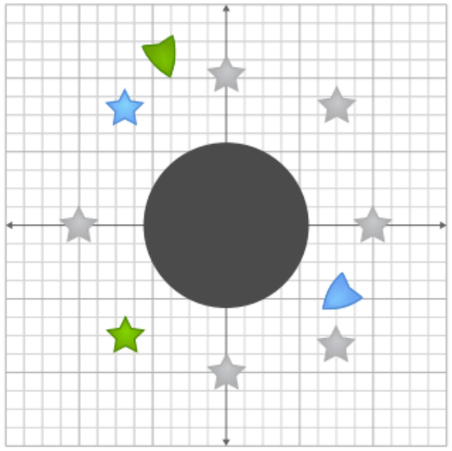


Experimenteller Aufbau

- Zwei Testumgebungen für BC & ECTL:
  - **Gridworld**: Diskretes 5x5 Raster, Agenten kommunizieren Ziele, sehen nur grobe Vermutungen
  - **Driving Game**: Kontinuierlicher Raum, Fahrzeuge steuern zu 8 Zielen, „Grube“ mit hohen Strafen
- Gridworld erzwingt Kommunikation, Driving Game erfordert zusätzlich Risiko-Vermeidung



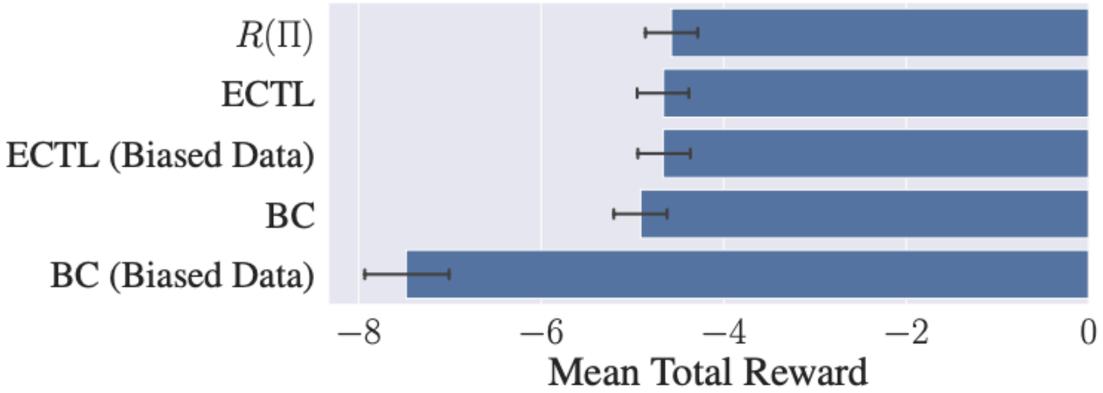
(b) Illustration of the gridworld environment (Section 5.1).



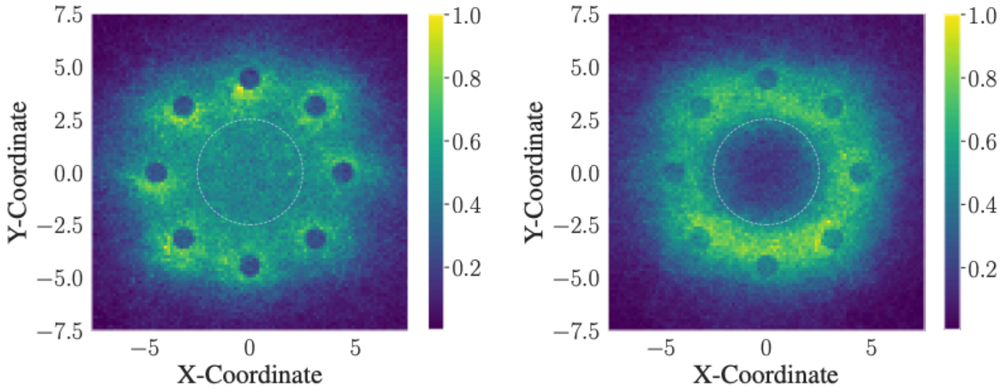
(c) Illustration of the driving environment (Section 5.1).

Leistungsvergleich BC vs. ECTL

- **Gridworld**:
  - BC & ECTL ähnlich bei umfassenden Daten
  - BC leidet bei verzerrten Trainingsdaten, ECTL bleibt robust
- **Driving Game**:
  - ECTL meidet Grube besser, zeigt stabilere Leistung bei limitierten Daten
  - BC stark beeinträchtigt durch Grube und verzerrte Daten



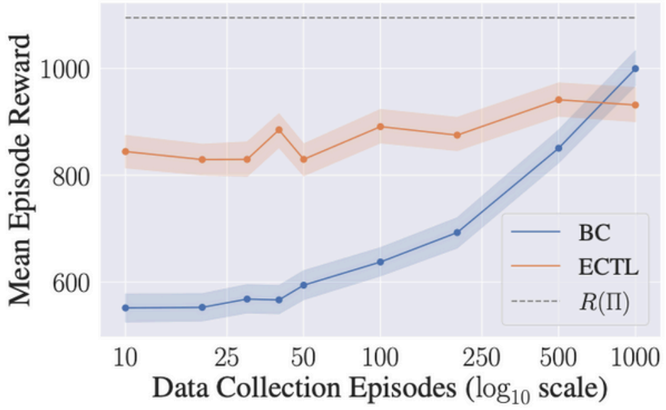
(b) Mean reward for different teams.



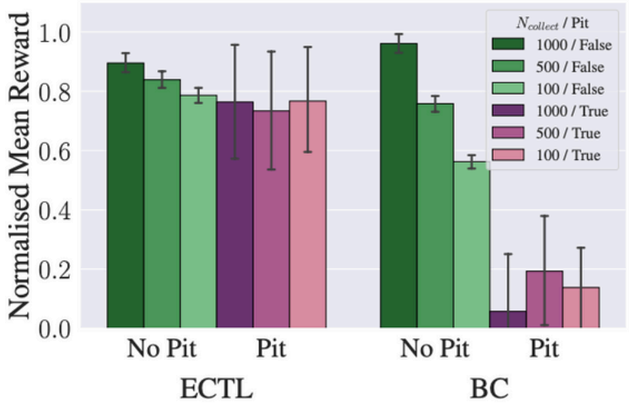
(a) (b)

Zentrale Erkenntnisse & Menschliche Kommunikation

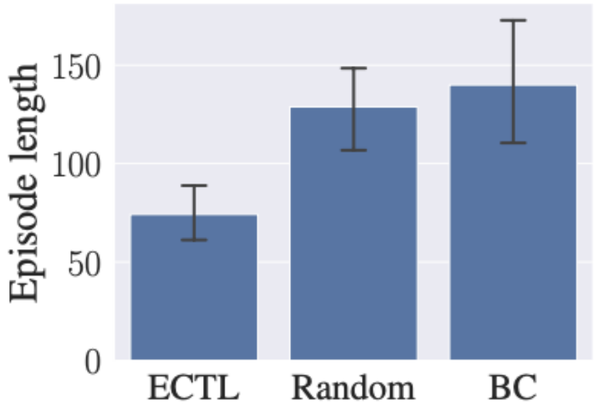
- **Robustheit von ECTL**: Gute Leistung trotz verzerrter/unvollständiger Daten
- **Menschliche Kommunikation**:
  - Menschliche Aktionen wurden in Nachrichten übersetzt
  - Nur ECTL konnte effektiv mit Menschen kommunizieren
  - Kürzere Episodenlängen bei ECTL → effizientere Aufgabenerfüllung



(a) The mean CLAP-Replace performance on the Driving Game (No Pit) for ECTL and BC, against the number data collection episodes ($N_{collect}$) from the target community.



(b) Mean reward for each method normalised by the original team mean rewards. Shows the impact of the pit on CLAP performance, highlighting BC's relative fragility.



(d)

Vielen Dank für Ihre Aufmerksamkeit!

- Fragen?
- Diskussion & Feedback willkommen

Quellen:

- Learning Translations: Emergent Communication Pretraining for Cooperative Language Acquisition
- https://dylancope.com/