ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ ΜΑΘΗΣΗΣ

Probing LLM Counterfactual Reasoning in Game Theory

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Δημήτριου Γεωργούση

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Probing LLM Counterfactual Reasoning in Game Theory

Διπλωματική Εργασία

του

Δημήτριου Γεωργούση

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 30^{η} Οκτωβρίου, 2025.

....	....	....
Γεώργιος Στάμου	Αθανάσιος Βουλόδημος	Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.	Επίκουρος Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025

....

Δημητριος Γεωργουσης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός

και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © – All rights reserved Δημήτριος Γεωργούσης, 2025.
Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Τα Μεγάλα Γλωσσικά Μοντέλα (ΜΓΜ) έχουν αναδειχθεί ως ευέλικτοι πράκτορες ικανοί να αντιμετωπίσουν ένα ευρύ φάσμα εργασιών, συμπεριλαμβανομένου και του στρατηγικού συλλογισμού. Η παρούσα διπλωματική εργασία διερευνά το κατά πόσο τα ΜΓΜ επιδεικνύουν πραγματικό στρατηγικό συλλογισμό σε περιβάλλοντα θεωρίας παιγνίων. Ο κύριος στόχος της εργασίας είναι η μελέτη παραλλαγών κατ' επανάληψη απλών παιγνίων, αξιοποιώντας την ευκολία παραμετροποίησής τους και εφαρμόζοντας διάφορες τεχνικές προτροπής.

Στα πειράματα χρησιμοποιούνται συμμετρικά ταυτόχρονα παιγνία - το Δίλημμα του Φυλακισμένου, το Κυνήγι του Ελαφιού και το Πέτρα-Ψαλίδι-Χαρτί - τα οποία προσφέρουν δυνατότητες παραμετροποίησης τόσο μέσω αλλαγών στην ονοματολογία των κινήσεων που προσφέρονται στους παίκτες, όσο και στις αποδόσεις αυτών των κινήσεων. Τα ΜΓΜ, ίσως, γνωρίζουν μόνο τη συνηθισμένη ή τυπική εκδοχή κάθε παιγνίου· επομένως, αντιφατικά σενάρια που δημιουργούνται μέσω τροποποίησης παραμέτρων λειτουργούν ως μέτρο της ευελιξίας και της ευαισθησίας των ΜΓΜ στις δυνητικές απολαβές από τις διάφορες κινήσεις και ως μέσο αντιπαράθεσης μεταξύ του στρατηγικού συλλογισμού και της εξάρτησης από προϋπάρχουσες γνώσεις για την συνηθισμένη εκδοχή των παιγνίων.

Μια διαδεδομένη μέθοδος για την κατεύθυνση των ΜΓΜ προς συγκεκριμένους σκοπούς είναι η χρήση προηγμένων τεχνικών προτροπής. Στη παρούσα εργασία χρησιμοποιούνται διάφορες τεχνικές, συμπεριλαμβανομένων της Προτροπής Χωρίς Παραδείγματα, της Προτροπής Με Ύπαρξη Συλλογιστικής Πορείας και της Προτροπής Μονοπρόσωπης Εκτέλεσης· πραγματοποιούνται επιπλέον πειράματα με τις αντίστοιχες παραλλαγές αυτών που βασίζονται στη Προτροπή Αυτοσυνέπειας. Οι τεχνικές αυτές αποσκοπούν στην παραγωγή πιο στοχευμένων απαντήσεων από τα μοντέλα. Στόχος είναι η ελαχιστοποίηση της επίδρασης της επιφανειακής αντιστοίχισης προτύπων και η ενθάρρυνση συλλογισμών που λαμβάνουν υπόψη τις συγκεκριμένες συνθήκες κάθε παιγνίου.

Για την αξιολόγηση της παρουσίας ικανοτήτων στρατηγικού συλλογισμού, τα ΜΓΜ συγκρίνονται με αλγοριθμικούς παίκτες οι οποίοι ακολουθούν συγκεκριμένες προκαθορισμένες στρατηγικές, καθώς και με τον εαυτό τους καθοδηγούμενο από διαφορετικά στυλ προτροπής. Το πλαίσιο αυτό σύγκρισης επιτρέπει την εκτίμηση του κατά πόσο τα ΜΓΜ προσαρμόζουν το παίξιμό τους με τρόπο συμβατό με λογική στρατηγική συμπεριφορά ή αν οι απαντήσεις τους απλώς αντανακλούν επιφανειακά χαρακτηριστικά της προτροπής. Βασικοί δείκτες περιλαμβάνουν την ανταπόκριση στη στρατηγική του αντιπάλου, την εκμετάλλευση των τάσεων του αντιπάλου και τις συμπεριφορικές μεταβολές σε μελλοντικούς γύρους του παιγνίου. Ιδιαίτερα, οι επαναλαμβανόμενες αλληλεπιδράσεις προσφέρουν ένα μοναδικό παράθυρο παρατήρησης του βαθμού που τα ΜΓΜ μπορούν να επιδείξουν συνεργασία, στρατηγικές αντεκδίκησης ή μάθηση κατά την εξέλιξη του παιγνίου.

Η τροποποίηση των παραμέτρων των παιγνίων και των τεχνικών προτροπής με συστηματικό τρόπο στοχεύει στην ανάδειξη των συνθηκών υπό τις οποίες τα ΜΓΜ παρουσιάζουν συμπεριφορά ενδεικτική πραγματικού στρατηγικού συλλογισμού. Τα ευρήματα της εργασίας αυτής συμβάλλουν στη γενικότερη κατανόηση των δυνατοτήτων των ΜΓΜ, ειδικά σε δυναμικά περιβάλλοντα λήψης αποφάσεων, και αναδεικνύουν τόσο τις προοπτικές όσο και τους περιορισμούς των σύγχρονων μοντέλων ως προς την αναπαραγωγή ανθρωπίνου τύπου στρατηγικής σκέψης.

Λέξεις-κλειδιά — ικανότητες συλλογισμού, γνωστικές δεξιότητες, Μεγάλα Γλωσσικά Μοντέλα (ΜΓΜ), προτροπή ενός παραδείγματος (one-shot prompting), προτροπή χωρίς παράδειγμα (zero-shot prompting), προτροπή με ύπαρξη συλλογιστικής πορείας (chain-of-thought prompting), προτροπή μονοπρόσωπης εκτέλεσης (solo-performance prompting), προτροπή αυτοσυνέπειας (self-consistency prompting), θεωρία παιγνίων,

ισορροπία Nash, στρατηγική, στρατηγική σκέψη, αντιφατικό σενάριο (counterfactual scenario), παραμετρική τροποποίηση, απομνημόνευση, στρατηγική προσαρμοστικότητα, συνεργασία, ανταγωνισμός.

Abstract

Large Language Models (LLMs) have emerged as versatile agents capable of addressing a wide range of tasks, including strategic reasoning. This thesis investigates whether LLMs exhibit true strategic reasoning in game-theoretic environments. The main focus of this work is the study of repeated variants of simple games, leveraging their ease of parameterization and imploring various prompting techniques.

Simultaneous-move, symmetric games are used in experimentation - Prisoner's Dilemma, Stag Hunt, and Rock-Paper-Scissors - which offer parameterization opportunities by adjusting both naming schemes of moves offered to players and payoffs of said moves. LLMs are likely to be aware of only the typical or usual setting of each game; therefore, counterfactual settings (settings created from parameter modification) serve as a test of LLM flexibility and sensitivity to changes in payoff structure, and as juxtaposition of strategic thinking and reliance to prior knowledge, that LLMs might have on the default setting of the games.

A well-known method for targeting LLM thinking abilities towards specific tasks is the employment of advanced prompting techniques. In this work, a range of prompting strategies, including Zero-Shot, Chain-of-Thought, and Solo-Performance Prompting, are used; experiments are also performed on their Self-Consistency counterparts. These techniques reflect an attempt to elicit more deliberate and context-aware responses from the models. They aim to minimize the influence of surface-level pattern matching and instead encourage reasoning that takes into account the specific parameters of each game instance.

To evaluate the presence of strategic reasoning, LLMs are compared against non-AI players, who follow specific preset strategies, and against themselves following different prompt styles. This comparative framework allows for an assessment of whether LLMs adapt their play in a manner consistent with rational strategic behavior, or if their responses merely reflect superficial cues from the prompt. Key indicators include responsiveness to opponent strategy, exploitation of opponent tendencies, and behavioral shifts across repeated rounds. In particular, repeated interactions offer a unique window into whether LLMs can exhibit conditional cooperation, retaliatory strategies, or learning-like behavior over time.

By systematically varying both the game settings and the prompting techniques, this thesis aims to uncover the conditions under which LLMs demonstrate behavior indicative of genuine strategic reasoning. The findings contribute to the broader understanding of LLM capabilities, especially in dynamic decision-making contexts, and highlight both the promise and limitations of current models in replicating human-like strategic thought.

Keywords — reasoning capabilities, cognitive skills, Large Language Models (LLMs), one-shot prompting, zero-shot prompting, chain-of-thought prompting, solo-performance prompting, self-consistency prompting, game theory, nash equilibrium, strategy, strategic thinking, counterfactual scenario, counterfactual setting, memorization, strategic adaptability, cooperation, antagonism.

Ευχαριστίες

Η παρούσα διπλωματική εργασία είναι προϊόν του συνεχούς μόχθου μου σε διάρκεια αρκετών μηνών ο οποίος στηρίχτηκε σε αρκετούς ανθρώπους. Είμαι ευγνώμων σε αυτούς για την καθοδήγηση και βοήθειά τους.

Ευχαριστώ τον κ. Στάμου Γεώργιο για την ευκαιρία εκπόνησης αυτής της διπλωματικής εργασίας. Η συνεισφορά του εργαστηρίου συστημάτων τεχνητής νοημοσύνης του ΕΜΠ και, συγκεκριμένα, των μελών του, Γιώργο Φιλανδριανό, Μαρία Λυμπεραίου και Αγγελική Δημητρίου, ήταν καθοριστικής σημασίας στη διαμόρφωση και βελτίωση αυτής της ακαδημαϊκής προσπάθειας.

Στήριξη και πίστη βρήκα στα πρόσωπα της οικογένειας και των φίλων μου. Η εμπιστοσύνη τους στις ικανότητές μου αποτέλεσαν τον ακρογωνιαίο λίθο αυτής της ακαδημαϊκής προσπάθειας. Οι συζητήσεις, ιδέες και συμβουλές που μοιράστηκαν μαζί μου έδωσαν βάθος και χαρακτήρα σε αυτό το έργο. Η διπλωματική εργασία αυτή αποτελεί μαρτυρία του πνεύματος συνεργασίας και της υποστήριξης που έλαβα από κάθε πλευρά. Χωρίς τις παρέες μου αυτές, το ακαδημαϊκό μου ταξίδι θα ήταν σίγουρα πιο ελλιπές σε χαρά, όρεξη και ενδιαφέρον.

Οι πόροι AWS παρασχέθηκαν από το Εθνικό Δίκτυο Υποδομών Τεχνολογίας και Έρευνας (ΕΔΥΤΕ Α.Ε. - GRNET) με τη χρηματοδότηση του Μηχανισμού Ανάκαμψης και Ανθεκτικότητας της ΕΕ.

Δημήτριος Γεωργούσης, Οκτώβριος 2025

Contents

List of Figures

List of Tables

Chapter 1

Εκτεταμένη Περίληψη στα Ελληνικά

1.1 Θεωρητικό Υπόβαθρο

1.1.1 Μεγάλα Γλωσσικά Μοντέλα (ΜΓΜ)

Τα Μεγάλα Γλωσσικά Μοντέλα (ΜΓΜ) είναι μείζονος σημασίας στη σύγχρονη τεχνητή νοημοσύνη, τροφοδοτώντας συστήματα όπως το Claude της Anthropic για την παραγωγή συνεκτικών, ανθρώπινων κειμένων. Η δημοτικότητά τους πηγάζει από την ευελιξία τους σε διάφορες εργασίες, όπως η γλωσσική παραγωγή, η εξήγηση κώδικα, η ανάλυση συναισθημάτων και η εκτέλεση εργασιών συλλογιστικής. Στην ουσία τους, τα ΜΓΜ λειτουργούν μοντελοποιώντας την κατανομή πιθανότητας σε ακολουθίες λέξεων, επιτρέποντάς τους να παράγουν ή να αξιολογούν κείμενο με βάση προτροπές εισόδου.

Οι παλαιότερες προσεγγίσεις βασίζονταν σε **μοντέλα n-gram**, τα οποία προβλέπουν μια λέξη με βάση τις προηγούμενες $n-1$ λέξεις. Για παράδειγμα, ένα μοντέλο bigram προσεγγίζει την πιθανότητα μιας ακολουθίας ως:

$$P(w_1, w_2, \ldots, w_n) = \prod_{k=2}^{n} P(w_k \mid w_{k-1}) \tag{1.1.1}$$

Αυτά τα μοντέλα εκτιμούν τις πιθανότητες από τις συχνότητες λέξεων στο σώμα κειμένων εκπαίδευσης, αλλά υποφέρουν από *σπανιότητα δεδομένων*, γεγονός που τα καθιστά αναξιόπιστα για ακολουθίες λέξεων, που δεν εμφανίζονται στα κείμενα εκπαίδευσης. Έχουν προταθεί διάφορες τεχνικές εξομάλυνσης,

Νευρωνικά μοντέλα γλώσσας και ο μετασχηματιστής

Τα νευρωνικά μοντέλα, όπως τα δίκτυα feedforward, τα RNN και τα LSTM, βελτίωσαν τη γενίκευση χρησιμοποιώντας ενσωματώσεις λέξεων, αλλά αντιμετώπισαν προβλήματα με τις μακροπρόθεσμες εξαρτήσεις. Η **αρχιτεκτονική του μετασχηματιστή** [55] αντιμετώπισε αυτό το πρόβλημα χρησιμοποιώντας *αυτοπροσοχή* και απορρίπτοντας την επαναληψιμότητα (recurrence).

Οι μετασχημααιτιστές (transformers) αποτελούνται από στοίβες κωδικοποιητών-αποκωδικοποιητών, αν και πολλά ΜΓΜ χρησιμοποιούν παραλλαγές μόνο με αποκωδικοποιητές. Τα βασικά συστατικά περιλαμβάνουν:

- **Ενσωματώσεις εισόδου + κωδικοποίηση θέσης**: Τα σύμβολα (tokens) αντιστοιχίζονται σε διανύσματα, με προσθήκη πληροφοριών θέσης.
- **Multi-Head Self-Attention**: Επιτρέπει στα σύμβολα να 'προσέχουν' (attend to) όλα τα άλλα, μαθαίνοντας τις σχέσεις μεταξύ των συμφραζομένων.
- **Δίκτυα Feed-Forward**: Προσθέτουν μη γραμμικότητα και χωρητικότητα μοντέλου.
- **Masked Attention** (μόνο αποκωδικοποιητές): Αποτρέπει την πρόσβαση σε μελλοντικές θέσεις συμβόλων κατά την διάρκεια της εκπαίδευσης.
- **Residual Connections + Layer Norm**: Βελτιώνουν τη σταθερότητα και τη σύγκλιση της εκπαίδευσης. Υπάρχουν ανάμεσα στα στρώματα του Encoder και του Decoder.

Οι μετασχηματιστές επιτρέπουν τον παράλληλο υπολογισμό και τη μοντελοποίηση μεγάλης εμβέλειας, καθιστώντας τους τον βασικό πυρήνα των σημερινών ΜΓΜ.

Στοιχεία ΜΓΜ και προ-εκπαίδευση

Συμβολοποίηση (Tokenization): Το κείμενο χωρίζεται σε σύμβολα (tokens) (λέξεις, υπολέξεις ή χαρακτήρες), σχηματίζοντας μια ακολουθία $x = (x_1, \ldots, x_n)$.

Ενσωμάτωση: Κάθε token x_i αντιστοιχεί σε $e_i \in \mathbb{R}^d$ χρησιμοποιώντας τον πίνακα ενσωμάτωσης $E \in \mathbb{R}^{V \times d}$:

$$E(x) = (e_1, \ldots, e_n), \quad \text{όπου } e_i = E[x_i]$$

Προ-εκπαίδευση: Τα ΜΓΜ εκπαιδεύονται πρώτα σε τεράστια σώματα κειμένων χωρίς ετικέτες χρησιμοποιώντας αυτοεποπτευόμενες εργασίες όπως η μοντελοποίηση κρυμμένων λέξεων ή η πρόβλεψη της συνέχειας ακολουθιών λέξεων. Αυτό τα εξοπλίζει με γενικές γνώσεις. Καθώς η παρούσα διατριβή χρησιμοποιεί υπάρχοντα προ-εκπαιδευμένα μοντέλα, δεν εμβαθύνουμε σε πιο περίπλοκες τεχνικές εκπαίδευσης.

Παράμετροι συμπερασμού

Τα ΜΓΜ εκθέτουν διάφορες παραμέτρους για να ρυθμίζουν την ποικιλομορφία και την τυχαιότητα της εξόδου:

- **Θερμοκρασία**: Ελέγχει την τυχαιότητα της επιλογής των συμβόλων.
  - Χαμηλή θερμοκρασία → πιο εστιασμένη/προβλέψιμη.
  - Υψηλή θερμοκρασία → πιο ποικίλη/απρόβλεπτη.
- **Top-K**: Δειγματοληψία μόνο από τα K πιο πιθανά επόμενα σύμβολα.
- **Top-P**: Δειγματοληψία από σύμβολα που συλλογικά δεν υπερβαίνουν την πιθανότητα P.

Παράδειγμα: Δεδομένης της προτροπής «`I hear the hoof beats of` » με υποψήφιες προβλέψεις:

```
{
"horses": 0.7,
"zebras": 0.2,
"unicorns": 0.1
}
```

- Η υψηλή θερμοκρασία αυξάνει την πιθανότητα επιλογής του «unicorns».
- Το Top-K = 2 περιορίζει την έξοδο σε «horses» και «zebras».
- Το Top-P = 0.7 περιλαμβάνει μόνο «horses»; Το Top-P = 0.9 περιλαμβάνει τόσο "horses" όσο και «zebras».

Αυτές οι παράμετροι είναι απαραίτητες για τη διαμόρφωση της συμπεριφοράς του μοντέλου στα πειράματα.

1.1.2 Μεγάλα Συλλογιστικά Μοντέλα (ΜΣΜ)

Τα μεγάλα Συλλογιστικά Μοντέλα (ΜΣΜ) είναι προηγμένα συστήματα τεχνητής νοημοσύνης που ενσωματώνουν δυνατότητες συλλογιστικής με την επεξεργασία φυσικής γλώσσας (ΕΦΓ). Όπως τα ΜΓΜ, και τα ΜΣΜ λειτουργούν με κείμενο, εικόνες ή άλλα δεδομένα, αλλά έχουν σχεδιαστεί για να επιλύουν προβλήματα βήμα προς βήμα χρησιμοποιώντας δομημένη συλλογιστική.

Αν και αρχιτεκτονικά παρόμοια με τα ΜΓΜ, τα ΜΣΜ εκπαιδεύονται διαφορετικά για να τονίζουν τη λογική σκέψη. Αναλύουν σύνθετες προτροπές και παράγουν αποτελέσματα βασισμένα στη λογική και την προηγούμενη γνώση, καθιστώντας τα κατάλληλα για εφαρμογές όπως η ανίχνευση απάτης ή η ιατρική διάγνωση.

Εκπαίδευση ΜΣΜ

Τα ΜΣΜ βελτιώνουν τα ΜΓΜ μέσω εξειδικευμένων στρατηγικών εκπαίδευσης και προτροπής:

- **Εμπλουτισμένα σύνολα δεδομένων**: Περιλαμβάνουν παραδείγματα συλλογιστικής με τις σωστές απαντήσεις και τα ενδιάμεσα βήματα για την επίτευξή τους.
- **Ενισχυτική μάθηση (ΕΜ)**: Ανταμείβει τα λογικά ή ακριβή αποτελέσματα και τιμωρεί τα λανθασμένα.
- **Σχεδιασμός προτροπών**: Οι προτροπές είναι σχεδιασμένες (βλ. ενότητα 1.1.3) για να προκαλέσουν συμπεριφορά συλλογιστικής πολλαπλών βημάτων.

Όταν χρησιμοποιούνται σε περιβάλλοντα συνομιλίας, τα ΜΣΜ συχνά παράγουν απαντήσεις που αποτελούνται από πολλά μέρη. Εμπνευσμένα από μοντέλα όπως το DeepSeek-R1 [12], αυτά περιλαμβάνουν συνήθως:

- **Σκέψη**: Μια λεπτομερής, μερικές φορές φλύαρη ενότητα συλλογιστικής που μιμείται τη δομημένη σκέψη.
- **Απάντηση**: Μια συνοπτική τελική απάντηση βασισμένη στην προηγούμενη συλλογιστική.

1.1.3 Προτροπή

Η προτροπή αντιπροσωπεύει μια αλλαγή παραδείγματος στον τρόπο εφαρμογής των γλωσσικών μοντέλων. Αντί να προσαρμόζουμε τα μοντέλα για κάθε εργασία, η προτροπή αναδιατυπώνει τις εργασίες ως προβλήματα συμπλήρωσης κειμένου χρησιμοποιώντας προ-εκπαιδευμένα μοντέλα. Δεδομένης μιας προτροπής — που συχνά δημιουργείται χρησιμοποιώντας ένα πρότυπο — το μοντέλο την ολοκληρώνει δημιουργώντας το επιθυμητό αποτέλεσμα [30].

Αυτή η προσέγγιση βασίζεται στις πλούσιες γλωσσικές και λογικές ικανότητες που αποκτήθηκαν κατά τη διάρκεια της προ-εκπαίδευσης. Χωρίς την ανάγκη για επιπλέον εκπαίδευση για συγκεκριμένες εργασίες, τα ΜΓΜ μπορούν να καθοδηγηθούν προς εργασίες ταξινόμησης, περίληψης ή συλλογιστικής απλά μέσω της διατύπωσης της εισόδου — μια προσέγγιση γνωστή ως **μηχανική προτροπής (prompt engineering)** [57].

Μεθοδολογία προτροπής

Η παραδοσιακή εποπτευόμενη μάθηση εκτιμά το $P(y|x;\theta)$ χρησιμοποιώντας μεγάλα σύνολα δεδομένων με ετικέτες. Αντίθετα, η μάθηση με προτροπές υπολογίζει την $P(x;\theta)$ του ίδιου του κειμένου x και την χρησιμοποιεί για να προβλέψει την έξοδο y χωρίς την ανάγκη για μεγάλα σύνολο δεδομένων με ετικέτες. Η διαδικασία περιλαμβάνει:

- **Προσθήκη προτροπής:** Η είσοδος εισάγεται σε ένα πρότυπο με θέσεις για την είσοδο $[X]$ και την έξοδο $[Z]$.
- **Αναζήτηση απάντησης:** Το μοντέλο βαθμολογεί τις πιθανές ολοκληρώσεις z και επιλέγει την πιο πιθανή (μέσω argmax ή δειγματοληψίας).
- **Αντιστοίχιση απαντήσεων:** Η ακατέργαστη έξοδος μετατρέπεται προαιρετικά σε μορφή που προτιμά ο χρήστης, π.χ. αντιστοίχιση του «υπέροχο» σε μια αριθμητική βαθμολογία συναισθήματος στη περίπτωση που επιθυμούμε να ανιχνεύσουμε συναίσθημα.

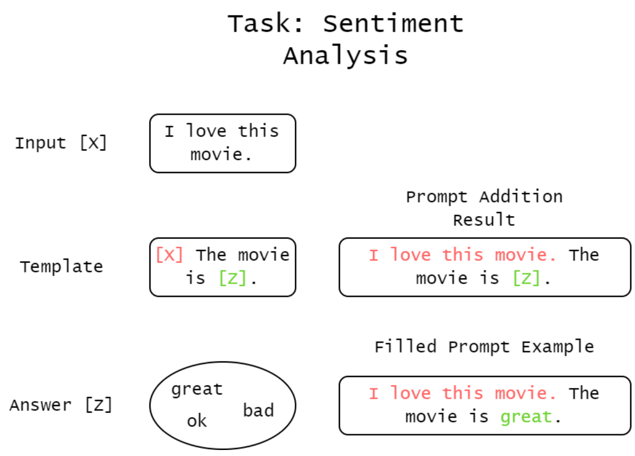


Figure 1.1.1: Βασικά βήματα προτροπής [30]

Μηχανική προτροπής

Η μηχανική προτροπής περιλαμβάνει τη δημιουργία προτροπών που μεγιστοποιούν την απόδοση. Οι τύποι προτροπών περιλαμβάνουν:

Προτροπές κλειστού τύπου (συμπλήρωση κενών) και **προτροπές προθέματος** (με βάση την ολοκλήρωση), ανάλογα με την εργασία.

Οι διακριτές προτροπές (σκληρές) γράφονται με το χέρι σε φυσική γλώσσα. **Οι συνεχείς προτροπές (μαλακές)** χρησιμοποιούν ενσωματωμένα στοιχεία που μπορούν να μάθουν, αλλά δεν είναι ερμηνεύσιμες, είναι συγκεκριμένες για κάθε μοντέλο και απαιτούν πρόσβαση στο εσωτερικό μοντέλο (αυξημένο κόστος χρήσης) [62].

Στα μοντέλα συνομιλίας, οι προτροπές περιλαμβάνουν **ρόλους**:

- **Σύστημα**: Ορίζει τη συμπεριφορά του μοντέλου (π.χ. «Είσαι ένας χρήσιμος βοηθός»).
- **Χρήστης**: Τα μηνύματα που εισάγει στη συνομιλία ο χρήστης-άνθρωπος.
- **Βοηθός**: Η απάντηση του μοντέλου. Μπορεί επίσης να προ-συμπληρωθεί για να καθοδηγήσει τη μελλοντική συμπεριφορά.

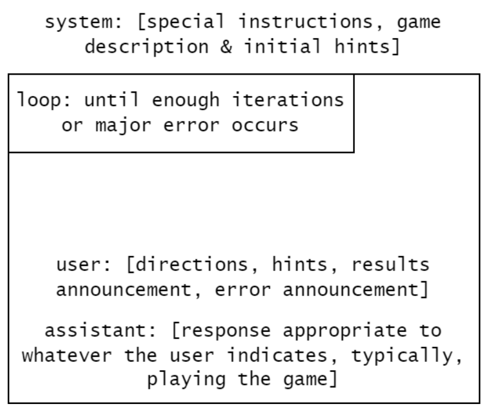


Figure 1.1.2: Επισκόπηση των προτροπών βάσει ρόλων

Τεχνικές προτροπής

Οι τεχνικές προτροπής διαφέρουν ως προς το βαθμό καθοδήγησης που παρέχουν:

- **Χωρίς Παραδείγματα**: Στο μοντέλο δίνεται μόνο η οδηγία.
- **Με Παράδειγμα**: Παρέχεται ένα παράδειγμα για να καθοδηγήσει το μοντέλο.

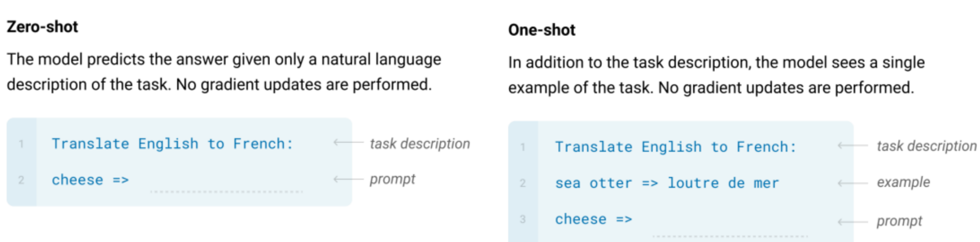


Figure 1.1.3: (a) Χωρίς Παράδειγμα, (b) Με Παράδειγμα [9]

Η προτροπή με **ύπαρξη Συλλογιστικής Πορείας (CoT)** ενθαρρύνει τη βήμα-βήμα συλλογιστική [61], μιμούμενη την ανθρώπινη επίλυση προβλημάτων.

Προτροπή Μονοπρόσωπης Εκτέλεσης (SPP) [60] επεκτείνει το CoT δημιουργώντας πολλαπλά AI personas που συζητούν ή συνεργάζονται πριν καταλήξουν σε μια απάντηση.

Αυτοσυνέπεια [58] βελτιώνει τη συνέπεια με τη δειγματοληψία πολλαπλών διαδρομών συλλογιστικής και την επιλογή της πιο συνεπούς απάντησης μεταξύ τους.

Αυτές οι τεχνικές, που μπορούν να χρησιμοποιηθούν μεμονωμένα ή σε συνδυασμό με την προτροπή αυτοσυνέπειας, παρέχουν ένα σύνολο εργαλείων για την εξαγωγή πιο ακριβούς και αξιόπιστης συμπεριφοράς από τα ΜΓΜ σε διάφορα σενάρια.

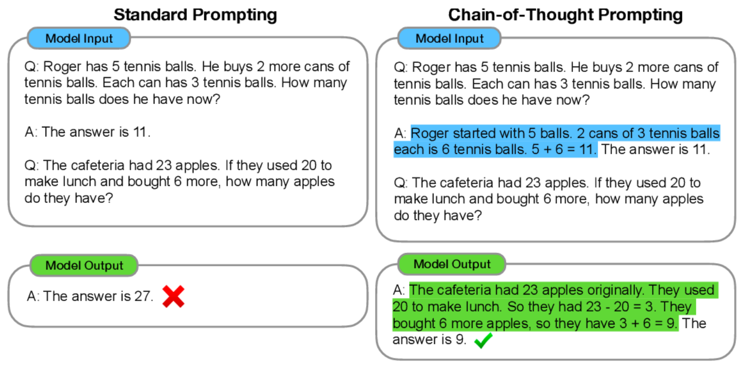


Figure 1.1.4: Παράδειγμα με ύπαρξη της συλλογιστικής πορείας [61]

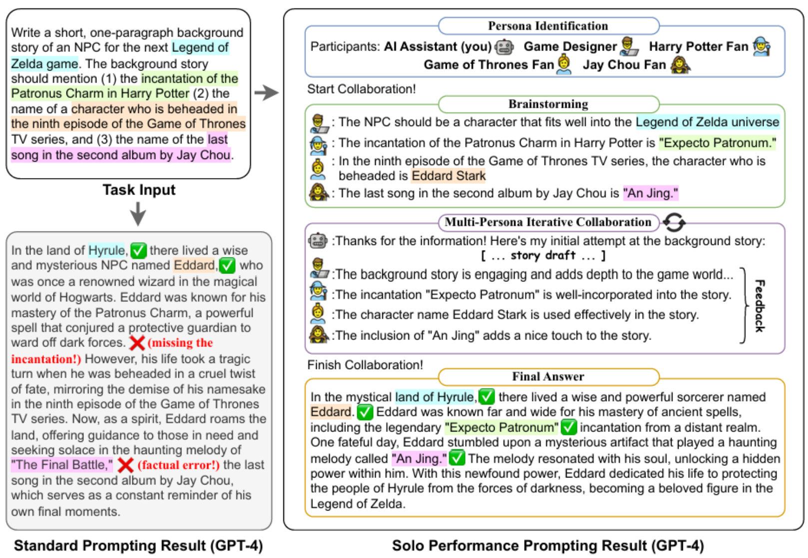


Figure 1.1.5: Προτροπή Μονοπρόσωπης Εκτέλεσης [60]

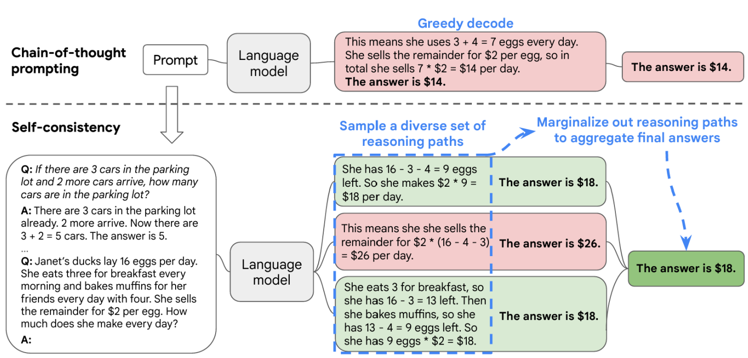


Figure 1.1.6: Μεθοδολογία προτροπής αυτοσυνέπειας [58]

1.1.4 Θεωρία Παιγνίων[1]

Η Θεωρία Παιγνίων μελετά μαθηματικά μοντέλα στρατηγικής αλληλεπίδρασης μεταξύ ορθολογικών παικτών [37]. Αρχικά αναπτύχθηκε για παιγνία δύο ατόμων μηδενικού αθροίσματος, αλλά αργότερα επεκτάθηκε σε παιγνία με

[1] Προσαρμοσμένο από τις σημειώσεις διαλέξεων του Chris Georges, Hamilton College.

μη μηδενικό άθροισμα και πιο σύνθετα περιβάλλοντα παιγνίων. Σήμερα, περιλαμβάνει ευρέως τη λογική λήψη αποφάσεων σε υπολογιστές, ανθρώπους και ζώα [47].

Ένα **παιγνίο** περιλαμβάνει πολλούς λογικούς, ευφυείς παίκτες, των οποίων τα κέρδη εξαρτώνται όχι μόνο από τις δικές τους ενέργειες, αλλά και από αυτές των άλλων. Τα βασικά στοιχεία ενός παιγνίου είναι:

1. **Κανόνες**: καθορίζουν τις ενέργειες και τις διαθέσιμες πληροφορίες
2. **Παίκτες**: οι υπεύθυνοι λήψης αποφάσεων
3. **Ενέργειες & Αποτελέσματα**: οι πιθανές επιλογές και τα αποτελέσματά τους
4. **Αποδόσεις**: η χρησιμότητα των αποτελεσμάτων για τους παίκτες

Θεωρητικές Σημειώσεις:

- n: αριθμός παικτών; $I = \{1, \dots, n\}$
- $s_i \in S_i$: καθαρή στρατηγική του παίκτη i; S_i: χώρος στρατηγικής
- $s = (s_1, \dots, s_n)$: προφίλ στρατηγικής; s_{-i}: στρατηγικές των άλλων παικτών
- $u_i(s_i, s_{-i})$: χρησιμότητα (απόδοση) του παίκτη i
- **Καλύτερη απάντηση**: Για τον παίκτη i, μια στρατηγική σ_i είναι η καλύτερη απάντηση στο προφίλ στρατηγικής σ_{-i} αν $u_i(\sigma_i, \sigma_{-i}) \geq u_i(s_i', \sigma_{-i})$ για όλα τα $s_i' \in S_i$.

Μικτές στρατηγικές επιτρέπουν στους παίκτες να επιλέγουν τυχαία:

- $\sigma_i \in \Delta S_i$: κατανομή πιθανότητας πάνω στο S_i
- $\sigma = (\sigma_1, \dots, \sigma_n)$: προφίλ μικτής στρατηγικής
- $u_i(\sigma_i, \sigma_{-i})$: αναμενόμενη χρησιμότητα
- Μια καθαρή στρατηγική s_i είναι *κυριαρχούμενη* αν κάποια σ_i δίνει αυστηρά καλύτερα οφέλη για όλα τα s_{-i}
- Η υποστήριξη μιας μικτής στρατηγικής είναι το σύνολο καθαρών στρατηγικών που παίζονται στην μικτή στρατηγική με μη μηδενική πιθανότητα.

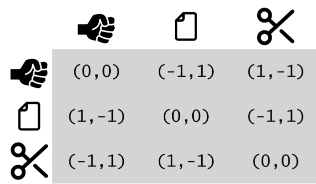
Ισορροπία Nash (NE) είναι ένα προφίλ στρατηγικής όπου κανένας παίκτης δεν ωφελείται από μονομερή απόκλιση (από την στρατηγική του στο προφίλ αυτό):

- Καθαρή στρατηγική NE: s^* έτσι ώστε $s_i^* \in BR_i(s_{-i}^*)$
- Μικτή στρατηγική NE: σ^* έτσι ώστε $\sigma_i^* \in BR_i(\sigma_{-i}^*)$

Ιδιότητα αδιαφορίας: Σε μια μικτή στρατηγική NE, κάθε καθαρή στρατηγική στην υποστήριξη αποδίδει ίση αναμενόμενη χρησιμότητα.

Τύποι παιγνίων:

- *Συνεργατικά* έναντι *Μη συνεργατικά*: δεσμευτικές συμφωνίες έναντι αυτοεπιβαλλόμενων ενεργειών
- *Συμμετρικά* έναντι *Μη συμμετρικά*: πανομοιότυπα έναντι διαφορετικά σύνολα στρατηγικών/αποδόσεων για τους παίκτες
- *Μηδενικού αθροίσματος* έναντι *Μη μηδενικού αθροίσματος*: το κέρδος του ενός είναι η απώλεια του άλλου στα μηδενικού αθροίσματος. Αυτό δεν ισχύει απαραίτητα στα μη μηδενικού αθροίσματος.
- *Ταυτόχρονα* έναντι *Διαδοχικά*: ενέργειες που επιλέγονται χωρίς έναντι με γνώση των κινήσεων των άλλων
- *Τέλεια* έναντι *Ατελής* πληροφόρηση: όλες οι προηγούμενες ενέργειες είναι γνωστές έναντι δεν είναι
- *Επαναλαμβανόμενα* παιγνία: το ίδιο βασικό παιγνίου παίζεται πολλές φορές· επιτρέπει την ενημέρωση των πεποιθήσεων με βάση το ιστορικό του αντιπάλου



(0,0)	(-1,1)	(1,-1)
(1,-1)	(0,0)	(-1,1)
(-1,1)	(1,-1)	(0,0)

Figure 1.1.7: Πίνακας αποδόσεων του παιγνίου Πέτρα-Χαρτί-Ψαλίδι. Ο παίκτης της σειράς λαμβάνει τις αριστερά απολαβές.

Τα επαναλαμβανόμενα παιγνία εισάγουν τη μνήμη και τη βελτίωση των πεποιθήσεων. Στον γύρο t, οι παίκτες μπορούν να χρησιμοποιήσουν τους προηγούμενους γύρους για να προσαρμόσουν τις προσδοκίες και τις στρατηγικές τους. Δεδομένου ότι τα ΜΓΜ μπορούν συχνά να κατανοήσουν τη δομή του παιγνίου, δοκιμάζουμε τη συλλογιστική τους αναλύοντας πόσο καλά βελτιώνουν τις πεποιθήσεις τους και ανταποκρίνονται τόσο σε προεπιλεγμένες όσο και σε αντιφατικές ρυθμίσεις [16].

1.1.5 Αντιφατικά σενάρια

Η αντιφατική συλλογιστική περιλαμβάνει το να σκεφτόμαστε τι θα μπορούσε να είχε συμβεί υπό διαφορετικές συνθήκες. Στο πλαίσιο της Θεωρίας Παιγνίων και των ΜΓΜ, παρέχει ένα χρήσιμο πρίσμα για την αξιολόγηση της συλλογιστικής πέρα από την απομνημόνευση — συγκεκριμένα, αν τα μοντέλα μπορούν να προσαρμοστούν σε τροποποιήσεις γνωστών εργασιών (στη δική μας περίπτωση παιγνίων).

Μια εργασία μπορεί να θεωρηθεί ως μια συνάρτηση $f_w : X \to Y$, όπου το μοντέλο του κόσμου w ορίζει τους υποκείμενους κανόνες. Η προεπιλεγμένη ρύθμιση $w^{default}$ αντικατοπτρίζει κοινές υποθέσεις από την προ-εκπαίδευση (π.χ. αριθμητική με βάση το 10 ή τυπικές ρυθμίσεις παιγνίου). Αντίθετα, οι **αντιφατικοί κόσμοι** w^{cf} αλλάζουν αυτές τις συνθήκες, επιτρέποντάς μας να ελέγξουμε αν η απόδοση ενός ΜΓΜ γενικεύεται πέρα από τις απομνημονευμένες προεπιλογές.

Αντί μόνο να δοκιμάζουμε νέες εισόδους x, αξιολογούμε τη γενίκευση τροποποιώντας το μοντέλο κόσμου w, ιδιαίτερα σε επαναλαμβανόμενα παιγνία δύο ατόμων. Αυτές οι τροποποιήσεις χωρίζονται σε δύο κύριους τύπους:

1. Αντιφατικά σενάρια στρατηγικής: Οι στρατηγικές των παικτών μετονομάζονται, ενώ η δομή του παιγνίου παραμένει αμετάβλητη. Οι ορθολογικοί παίκτες θα πρέπει να προσαρμοστούν ακώλυτα. Η αποτυχία υποδηλώνει εξάρτηση από απομνημονευμένα μοτίβα.

2. Αντιφατικά σενάρια απολαβών: Εδώ, οι αποδόσεις τροποποιούνται, αλλάζοντας τη βέλτιστη στρατηγική, ενώ διατηρείται η λογική του παιγνίου. Η επιτυχής προσαρμογή υποδηλώνει ότι το μοντέλο συλλογίζεται μέσω της δομής των αποδόσεων και δεν ανακαλεί μια γνωστή λύση.

Αυτά τα αντιφατικά περιβάλλοντα βοηθούν να διακρίνουμε αν τα ΜΓΜ μοντελοποιούν πραγματικά τη λειτουργία του παιγνίου f, ή απλώς επαναλαμβάνουν μοτίβα από το $f_{w^{default}}$. Εάν τα ΜΓΜ έχουν παρόμοια απόδοση σε βασικά και αντιφατικά παιγνία, αυτό υποδηλώνει εσωτερική συνέπεια και γενική ικανότητα συλλογιστικής.

(a)

(0,0)	(-1,1)	(1,-1)
(1,-1)	(0,0)	(-1,1)
(-1,1)	(1,-1)	(0,0)

(b)

(0,0)	(-3,3)	(1,-1)
(3,-3)	(0,0)	(-1,1)
(-1,1)	(1,-1)	(0,0)

(c)

(0,0)	(-1,1)	(1,-1)
(1,-1)	(0,0)	(-1,1)
(-1,1)	(1,-1)	(0,0)

(d)

(0,0)	(-3,3)	(1,-1)
(3,-3)	(0,0)	(-1,1)
(-1,1)	(1,-1)	(0,0)

Figure 1.1.8: αντιφατικά σενάρια στο πέτρα-ψαλίδι-χαρτί: (α) βασικό παιγνίο, (β) τροποποιημένη απόδοση, (γ) τροποποιημένα ονόματα στρατηγικών, (δ) τροποποιημένη απόδοση και στρατηγικές.

1.2 Προαπαιτούμενα

1.2.1 Περιβάλλον

Τα πειράματά μας διεξάγονται σε ένα τροποποιημένο περιβάλλον τύπου Gym, προσαρμοσμένο από το [28], για να υποστηρίζει την αλληλεπίδραση μεταξύ των πρακτόρων ΜΓΜ και των επαναλαμβανόμενων παιγνίων δύο παικτών. Το περιβάλλον χειρίζεται:

1. προτροπή των πρακτόρων με τους κανόνες του παιγνίου,
2. επεξεργασία και επικύρωση των μηνυμάτων κίνησης τους,
3. υπολογισμός των ανταμοιβών, και
4. παράδοση δομημένων μηνυμάτων ανατροφοδότησης και σφαλμάτων.

Προτροπή συστήματος. Κάθε πράκτορας λαμβάνει μια προτροπή συστήματος που συνδυάζει μια μέθοδο (χωρίς παραδείγματα, CoT ή SPP) και μια περιγραφή παιγνίου. Χρησιμοποιούμε σύντομα παραδείγματα στις προτροπές ενός παραδείγματος (όπως οι CoT και SPP) για να δείξουμε τον τρόπο συλλογιστικής χωρίς να υπερβαίνουμε τα όρια των συμβόλων (token), ακολουθώντας το κριτήριο της απλότητας (το παράδειγμα που περιλαμβάνεται στις προτροπές ενός παραδείγματος αναφέρεται σε μια πολύ απλούστερη εργασία από την εργασία-στόχο, ωστόσο, η εργασία αυτή επιλύεται χρησιμοποιώντας την προτιμώμενη τεχνική). Οι μέθοδοι CoT και SPP έχουν προσαρμοστεί από το [59].

Η μέθοδος προτροπής **με ύπαρξη συλλογιστικής πορείας (CoT)** παρουσιάζεται παρακάτω:

You are going to play a game with other player(s). Think step-by-step. Begin by identifying steps that will contribute to you winning. Then reason through the steps, until a final decision is reached. The steps should reflect a meaningful thought process.

Here is an example on a simpler task from what you will be playing:

—

Example Task: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does Roger have now?

Steps:

1. Roger starts with 5 tennis balls.
2. 2 cans of 3 tennis balls each are bought. This is 2 * 3 = 6 tennis balls.
3. Roger now has 5 + 6 = 11 tennis balls.

Finish steps!

Final decision: 11 tennis balls

—

Now, the game you will be playing is presented. Think step-by-step. Identify the steps and reason through them to complete the objective of the game. You may come up with different reasoning steps for each round, as you see fit.

Όπως εξηγήθηκε παραπάνω, αυτή η προτροπή θα ακολουθείται από μια κατάλληλη **περιγραφή παιγνίου** για να σχηματίσει την πλήρη **προτροπή συστήματος**.

Στη συνέχεια, παρουσιάζουμε τη μέθοδο προτροπής **μονοπρόσωπης εκτέλεσης (SPP)**:

You are going to play a game with other player(s). Begin by identifying the personas who could contribute to you winning. Then initiate a multi-round collaboration process among these personas, until a final decision is reached. The personas will give critical comments and detailed suggestions whenever necessary.

Here is an example on a simpler task from what you will be playing:

—

Example Task: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.
Input: 6 12 1 1

Personas: AI Assistant (you); Math Expert

Start collaboration!

Math Expert: Let's analyze the task in detail. You need to make sure that you meet the requirement, that you need to use exactly the four numbers (6 12 1 1) to construct 24. To reach 24, you can think of the common divisors of 24 such as 4, 6, 8, 3 and try to construct these first. Also you need to think of potential additions that can reach 24, such as 12 + 12.
AI Assistant (you): Thanks for the hints! Here's one initial solution: (12 / (1 + 1)) * 6 = 24
Math Expert: Let's check the answer step by step. (1+1) = 2, (12 / 2) = 6, 6 * 6 = 36 which is not 24! The answer is not correct. Can you fix this by considering other combinations? Please do not make similar mistakes.
AI Assistant (you): Thanks for pointing out the mistake. Here is a revised solution considering 24 can also be reached by 3 * 8: (6 + 1 + 1) * (12 / 4) = 24.
Math Expert: Let's first check if the calculation is correct. (6 + 1 + 1) = 8, 12 / 4 = 3, 8 * 3 = 24. The calculation is correct, but you used 6 1 1 12 4 which is not the same as the input 6 12 1 1. Can you avoid using a number that is not part of the input?
AI Assistant (you): You are right, here is a revised solution considering 24 can be reached by 12 + 12 and without using any additional numbers: 6 * (1 - 1) + 12 = 24.
Math Expert: Let's check the answer again. 1 - 1 = 0, 6 * 0 = 0, 0 + 12 = 12. I believe you are very close, here is a hint: try to change the "1 - 1" to "1 + 1".
AI Assistant (you): Sure, here is the corrected answer: 6 * (1+1) + 12 = 24
Math Expert: Let's verify the solution. 1 + 1 = 2, 6 * 2 = 12, 12 + 12 = 12. You used 1 1 6 12 which is identical to the input 6 12 1 1. Everything looks good!

Finish collaboration!

Final decision: 6 * (1 + 1) + 12 = 24

—

Now, the game you will be playing is presented. Identify the personas and use their multi-round collaboration to help you complete the objective of the the game. You may discuss with your personas on any round of the game, as you see fit.

Διαχείριση μηνυμάτων & εξαγωγή κινήσεων. Οι απαντήσεις του πράκτορα πρέπει να περιέχουν ένα μήνυμα κίνησης της μορφής:

```
[move] (optional explanation) move_label
```

Το περιβάλλον χρησιμοποιεί regex για να εξάγει κινήσεις, επιτρέποντας ταυτόχρονα ευελιξία στη διατύπωση και τη χρήση κεφαλαίων. Εάν δεν εντοπιστεί καμία έγκυρη κίνηση ή υπάρχουν πολλές, ο πράκτορας λαμβάνει μια υπόδειξη διόρθωσης. Μετά από πέντε αποτυχημένες προσπάθειες, ο γύρος διακόπτεται και για τους δύο παίκτες.

Ανατροφοδότηση. Κατά τη διάρκεια του παιγνίου, το περιβάλλον παρέχει ανατροφοδότηση, όπως η εμπειρία του αντιπάλου (αν παίζει πρώτη φορά), οι μεταβάσεις γύρων, οι επιβεβαιώσεις κινήσεων και οι αποδόσεις. Όλα τα μηνύματα ενοποιούνται και αποστέλλονται μέσω του ρόλου 'χρήστη' για να πληρούν τους περιορισμούς του API (βλ. 1.1.2).

Υποδείξεις. Οι αρχικές υποδείξεις περιλαμβάνονται στην προτροπή του συστήματος και επιπλέον υποδείξεις διόρθωσης δίνονται μόνο όταν είναι απαραίτητο, π.χ. όταν το μήνυμα κίνησης λείπει ή είναι λανθασμένο.

Αρχικές υποδείξεις:

Message Formats:
Move messages are sent when you want to make a move.
Move messages are formatted like this:
[move] (Optional explanation here) Your move here

Hint messages are received when you are provided with advice and directions. These messages are important and you should pay attention to them.
Hint messages are formatted like this:
[hint] Hint message here

Objective:
Maximize your points and/or minimize your penalties.

Tips:
- Try to formulate a strategy by reasoning on the provided information.
- Pay attention to the payoff matrix of the game moves.
- You should only send move messages. Hint messages will only be provided to you.
- Pay attention to hints when provided.
- Do not use markdown formatting in your messages (e.g., bold text, italic text, etc.).

Πρόσθετες υποδείξεις (εμφάνιση σφάλματος):

You may structure your response however you like, but it should contain a move message. Move messages begin with the tag [move] not containing other tags, which is followed by your optional explanation in parentheses, and end with a valid move: {list-of-moves}. DO NOT INCLUDE THE [move] TAG IN YOUR REASONING, ONLY IN YOUR ACTUAL MOVE MESSAGE. Nested parentheses or markdown formatting are not allowed.

Format: [move] (Optional explanation here) Your move here

Αυτή η οργάνωση της επικοινωνίας των ΜΓΜ επιτρέπει δομημένη, σύμφωνη με τους κανόνες αλληλεπίδραση, ενώ παράλληλα υποστηρίζει ανοιχτή συλλογιστική και φυσική γλώσσα από τα ΜΓΜ.

1.2.2 Γλωσσικά μοντέλα

Για να αξιολογήσουμε τη συλλογιστική σε ρυθμίσεις θεωρίας παιγνίων, δοκιμάσαμε ένα ευρύ φάσμα μεγάλων γλωσσικών μοντέλων (ΜΓΜ), συμπεριλαμβανομένων τόσο γενικής χρήσης όσο και βελτιστοποιημένων για συλλογιστική παραλλαγών. Αυτά περιλαμβάνουν μοντέλα από Anthropic, Meta, Mistral και DeepSeek, στα οποία έχουμε πρόσβαση με ομοιόμορφο τρόπο μέσω του Amazon Bedrock και, συγκεκριμένα, μέσω του Converse API για να διασφαλίσουμε συνεπή προτροπή, λήψη απαντήσεων και ανάλυση μεταξύ όλων των μοντέλων.

Εστιάζουμε σε σύγχρονα μοντέλα:

- **Anthropic Claude**: *Claude 3.5 Sonnet v2*, *Claude 3.7 Sonnet* και *Claude Sonnet 4*, τα δύο τελευταία με προαιρετικές ρυθμίσεις «εκτεταμένης σκέψης» (εστιασμένες στη συλλογιστική).
- **Meta LLaMA**: *Llama 3.3 70B Instruct*, επιλεγμένο λόγω περιορισμών πρόσβασης σε πολυτροπικές εκδόσεις στην ΕΕ.
- **Mistral Large (24.07)**: ένα ΜΓΜ γενικής χρήσης με υψηλή απόδοση.
- **DeepSeek-R1**: ένα μοντέλο που γεφυρώνει τα τυπικά ΜΓΜ και τα Μεγάλα Συλλογιστικά Μοντέλα (ΜΣΜ).

Συγκρίνοντας τα παραδοσιακά μοντέλα με τα μοντέλα με βελτιωμένη συλλογιστική, στοχεύουμε να αξιολογήσουμε εάν τα πρόσφατα ΜΓΜ επιδεικνύουν βελτιωμένη στρατηγική συλλογιστική ή απλώς αυξάνουν την επιφανειακή απόδοση. Η ενοποιημένη υποδομή εξασφάλισε τη δίκαιη αξιολόγηση σε όλα τα πειράματα.

1.3 Πειράματα

1.3.1 Δίλημμα Φυλακισμένου

Το **Δίλημμα του Φυλακισμένου** είναι ίσως το πιο διάσημο πείραμα της θεωρίας των παιγνίων, το οποίο έχει ωθήσει πολλούς ερευνητές να το μελετήσουν κατά τη διάρκεια των ετών. Περιλαμβάνει δύο ορθολογικούς παίκτες, ο καθένας από τους οποίους μπορεί είτε να συνεργαστεί για αμοιβαίο όφελος είτε να λιποτακτήσει (δηλαδή να προδώσει τον συνεργάτη του) για ατομικό όφελος. Το δίλημμα προκύπτει από το γεγονός ότι η αμοιβαία συνεργασία αποφέρει μέτρια οφέλη και για τους δύο παίκτες, ενώ η μονομερής λιποταξία (προδοσία) αποφέρει μεγαλύτερο όφελος για τον προδότη σε βάρος του συνεργάτη. Εάν προδώσουν και οι δύο, ο καθένας λαμβάνει ένα χειρότερο αποτέλεσμα από ό,τι εάν είχαν συνεργαστεί. Άτομα, οργανισμοί, χώρες κ.λπ. συχνά αντιμετωπίζουν διλήμματα όπως το παραπάνω. Το γεγονός αυτό, σε συνδυασμό με την απλότητα του διλήμματος του φυλακισμένου και τις δυνατότητες επέκτασής του, το καθιστούν ένα εξαιρετικό αντικείμενο για τη δική μας μελέτη.

Η δομή του παραδοσιακού διλήμματος του φυλακισμένου μπορεί να γενικευτεί από το αρχικό της πλαίσιο με τους φυλακισμένους. Αναπαρίσταται ως ένα παιγνίο κανονικής μορφής (που εξηγείται στο 1.1.4) με τον ακόλουθο πίνακα αποδόσεων.

	A	B
A	(aa, aa)	(ab, ba)
B	(ba, ab)	(bb, bb)

Table 1.1: Πίνακας Απολαβών για το Δίλημμα του Φυλακισμένου
Το **A** συνήθως αναφέρεται στη "Συνεργασία" και το **B** στη "Λιποταξία".

Για να θεωρηθεί το παιγνίο ως Δίλημμα του Φυλακισμένου με την ισχυρή έννοια, πρέπει να ισχύει η ακόλουθη συνθήκη για τα κέρδη/απολαβές:

$$ba > aa > bb > ab$$

Σε αυτό το γενικό πλαίσιο, οι παίκτες στοχεύουν να μεγιστοποιήσουν την απόδοση που λαμβάνουν.

Αντιφατικά Σενάρια

Κυνήγι του Ελαφιού Αυτό το παιγνίο αποτελεί το κύριο αντιφατικό μας σενάριο. Είναι τόσο αντιφατικό από στρατηγικής άποψης (οι κινήσεις έχουν διαφορετικά ονόματα) όσο και από άποψης απολαβών (οι τιμές των αποδόσεων (ή κερδών ή απολαβών) είναι διαφορετικές).

	A	B
A	(aa, aa)	(ab, ba)
B	(ba, ab)	(bb, bb)

Table 1.2: Πίνακας Απολαβών για το Κυνήγι Ελαφιού.
Το **A** συνήθως αναφέρεται στο "Ελάφι" και το **B** στον "Λαγό".

Για να θεωρηθεί το παιγνίο Κυνήγι Ελαφιού, πρέπει να ισχύει η ακόλουθη συνθήκη για τα κέρδη:

$$aa > ba \geq bb > ab$$

Σενάρια στα πειράματά μας Σε αυτή την παράγραφο, μπορούμε τώρα να παρουσιάσουμε τις συγκεκριμένες παραμέτρους που χρησιμοποιήθηκαν για τη δημιουργία των αντιφατικών σεναρίων αυτής της μελέτης. Αυτές φαίνονται στον πίνακα 1.3. Τα ονόματα των κινήσεων μπορεί να βοηθήσουν τους παίκτες ΜΓΜ να κατανοήσουν ποιο είναι ακριβώς το παιγνίο που παίζεται και, ως εκ τούτου, να τους βοηθήσουν να χρησιμοποιήσουν απομνημονευμένα μοτίβα για το παιγνίο αυτό. Η συμπερίληψη αντιφατικών σεναρίων όπου οι κινήσεις έχουν εναλλακτικά ονόματα και η σύγκριση της απόδοσης των δύο, θα πρέπει να δώσει μια καλή εικόνα της σχέσης μεταξύ μνήμης και συλλογιστικής που υπάρχει στα ΜΓΜ. Στον πίνακα 1.3, τα παιγνία (a) και (c), και (b) και (d) από την άλλη πλευρά, θα πρέπει να παιχτούν με τον ίδιο τρόπο, καθώς τα κέρδη είναι τα ίδια.

	Συνεργάζομαι	Λιποτακτώ
Συνεργάζομαι	(4, 4)	(1, 6)
Λιποτακτώ	(6, 1)	(2, 2)

(a) pd

	Συνεργάζομαι	Λιποτακτώ
Συνεργάζομαι	(6, 6)	(1, 4)
Λιποτακτώ	(4, 1)	(2, 2)

(b) pd-alt

	Ελάφι	Λαγός
Ελάφι	(4, 4)	(1, 6)
Λαγός	(6, 1)	(2, 2)

(c) sh-alt

	Ελάφι	Λαγός
Ελάφι	(6, 6)	(1, 4)
Λαγός	(4, 1)	(2, 2)

(d) sh

Table 1.3: Πίνακες απολαβών για αντιφατικά σενάρια του Διλήμματος του Φυλακισμένου.
(a) βασικό παιγνίο, το τυπικό Δίλημμα Φυλακισμένου.
(b) αντιφατικό απολαβών του (a), χρησιμοποιεί τον πίνακα απολαβών του Κυνηγιού Ελαφιού.
(c) αντιφατικό στρατηγικής του (a), χρησιμοποιεί τα ονόματα κινήσεων του Κυνηγιού Ελαφιού.
(d) αντιφατικό και απολαβών και στρατηγικής του (a), είναι το τυπικό Κυνήγι Ελαφιού.

Ισορροπία ενός γύρου

Αποδίδουμε στρατηγικές ισορροπίας Nash (NE) ενός γύρου χρησιμοποιώντας τους ορισμούς από το 1.1.4. Στο κλασικό δίλημμα του φυλακισμένου (ρυθμίσεις **pd** και **sh-alt** από 1.3), η αμοιβαία λιποταξία (B, B) είναι η μοναδική NE λόγω της κυριαρχίας της λιποταξίας.

Για το κυνήγι του ελαφιού (πίνακας αποδόσεων 1.2), εάν ένας παίκτης γνωρίζει ότι ο αντίπαλός του παίζει μια καθαρή στρατηγική, η καλύτερη απάντηση είναι να ακολουθήσει την ίδια επιλογή. Έτσι, τόσο το (A, A) όσο και το (B, B) είναι καθαρές στρατηγικές NE.

Για να βρούμε μια NE μικτής στρατηγικής, ας υποθέσουμε ότι και οι δύο παίκτες παίζουν A με πιθανότητα p και B με πιθανότητα $q = 1 - p$. Χρησιμοποιώντας την αρχή της αδιαφορίας:

$$E(A) = p(aa) + (1-p)(ab), \quad E(B) = p(ba) + (1-p)(bb)$$
$$E(A) = E(B) \Rightarrow p = \frac{bb - ab}{aa - ab + bb - ba}, \quad q = \frac{aa - ba}{aa - ab + bb - ba}$$

Περίληψη Για έναν γενικό πίνακα:

	A	B
A	(aa, aa)	(ab, ba)
B	(ba, ab)	(bb, bb)

- **Δίλημμα του φυλακισμένου:** $p = 0, \quad q = 1$ (πάντα λιποτακτεί)
- **Κυνήγι του ελαφιού:** Μικτή NE:

$$p = \frac{bb - ab}{aa - ab + bb - ba}, \quad q = \frac{aa - ba}{aa - ab + bb - ba}$$

1.3.2 Πέτρα Ψαλίδι Χαρτί

Για τα επόμενα πειράματα, η παρούσα διατριβή απομακρύνεται ελαφρώς από το **Δίλημμα του φυλακισμένου**, επιλέγοντας ένα παιγνίο ελαφρώς πιο περίπλοκο, το οποίο όμως παραμένει ενδιαφέρον για έρευνα.

Το **Πέτρα-Ψαλίδι-Χαρτί (RPS)** είναι ένα μη μεταβατικό[2] παιγνίο με τα χέρια, που παίζεται από δύο άτομα, στο οποίο κάθε παίκτης σχηματίζει ταυτόχρονα ένα σχήμα με το χέρι του. Αυτό το σχήμα αποτελεί την ενέργεια (ή κίνηση) της επιλογής του και μπορεί να είναι ένα από τα τρία: «Πέτρα», «Χαρτί», «Ψαλίδι». Όπως η ρίψη νομίσματος ή το ρίξιμο ζαριών, το **Πέτρα-Ψαλίδι-Χαρτί** χρησιμοποιείται συχνά ως δίκαιο μέσο επιλογής μεταξύ δύο ατόμων για την επίλυση συγκρούσεων ή τη λήψη μιας αμερόληπτης ομαδικής απόφασης. Σε ορισμένες περιπτώσεις, μπορεί κανείς να παίξει RPS με κάποιο βαθμό ικανότητας και δεξιότητας, σε αντίθεση με τα πραγματικά τυχαία συστήματα επιλογής, εκμεταλλευόμενος τη μη τυχαία συμπεριφορά του αντιπάλου. [18, 5].

Αυτό το παιγνίο μπορεί να αναπαρασταθεί ως ένα παιγνίο κανονικής μορφής (που εξηγείται στο 3.4.4) με τον ακόλουθο πίνακα αποδόσεων.

	A	B	C
A	(0, 0)	(-ba, ba)	(ac, -ac)
B	(ba, -ba)	(0, 0)	(-cb, cb)
C	(-ac, ac)	(cb, -cb)	(0, 0)

Table 1.4: Πίνακας απολαβών για το Πέτρα-Ψαλίδι-Χαρτί.
Το **A** συνήθως αναφέρεται στη "Πέτρα", το **B** στο "Χαρτί", και το **C** στο "Ψαλίδι".

Τα ba, ac, cb θεωρούνται θετικοί αριθμοί.

Σε αυτό το γενικό πλαίσιο, οι παίκτες στοχεύουν να μεγιστοποιήσουν τις απολαβές τους.

Αντιφατικά σενάρια

Σενάρια στα πειράματά μας Σε αυτή την υποενότητα, μπορούμε τώρα να παρουσιάσουμε τις συγκεκριμένες παραμέτρους που χρησιμοποιήθηκαν για τη δημιουργία των αντιφατικών σεναρίων της μελέτης. Αυτές φαίνονται στον πίνακα 1.5. Τα ονόματα των κινήσεων μπορεί να βοηθήσουν τους παίκτες ΜΓΜ να κατανοήσουν ποιο είναι ακριβώς το παιγνίο που παίζεται και, ως εκ τούτου, να τους βοηθήσουν να χρησιμοποιήσουν απομνημονευμένα μοτίβα για το παιγνίο αυτό. Η συμπερίληψη αντιφατικών σεναρίων όπου οι κινήσεις έχουν εναλλακτικά ονόματα και η σύγκριση της απόδοσης των δύο, θα πρέπει να δώσει μια καλή εικόνα της σχέσης μνήμης-συλλογιστικής που υπάρχει στα ΜΓΜ. Στον πίνακα 1.5, τα παιγνία (a) και (c), και (b) και (d) από την άλλη πλευρά, θα πρέπει να παίζονται με τον ίδιο τρόπο, καθώς τα κέρδη είναι τα ίδια.

[2] ένα παιγνίο μηδενικού αθροίσματος στο οποίο τα ζεύγη ανταγωνισμών μεταξύ των στρατηγικών περιέχουν κύκλο

	Πέτρα	Χαρτί	Ψαλίδι
Πέτρα	(0, 0)	(-1, 1)	(1, -1)
Χαρτί	(1, -1)	(0, 0)	(-1, 1)
Ψαλίδι	(-1, 1)	(1, -1)	(0, 0)

(a) eq1

	Πέτρα	Χαρτί	Ψαλίδι
Πέτρα	(0, 0)	(-3, 3)	(1, -1)
Χαρτί	(3, -3)	(0, 0)	(-1, 1)
Ψαλίδι	(-1, 1)	(1, -1)	(0, 0)

(b) ba3

	Χαρτί	Πέτρα	Ψαλίδι
Χαρτί	(0, 0)	(-1, 1)	(1, -1)
Πέτρα	(1, -1)	(0, 0)	(-1, 1)
Ψαλίδι	(-1, 1)	(1, -1)	(0, 0)

(c) eq1-alt

	Χαρτί	Πέτρα	Ψαλίδι
Χαρτί	(0, 0)	(-3, 3)	(1, -1)
Πέτρα	(3, -3)	(0, 0)	(-1, 1)
Ψαλίδι	(-1, 1)	(1, -1)	(0, 0)

(d) ba3-alt

Table 1.5: Πίνακες απολαβών για τα αντιφατικά σενάρια του Πέτρα-Ψαλίδι-Χαρτί.
(a) είναι το βασικό παιγνίο, είναι το τυπικό Πέτρα-Ψαλίδι-Χαρτί.
(b) είναι αντιφατικό απολαβών του (a), χρησιμοποιεί μεγαλύτερο κέρδος για νίκη με "Χαρτί".
(c) είναι αντιφατικό στρατηγικής του (a), αν **X** τυπικά κερδίζει **Y**, τώρα **Y** κερδίζει **X**.
(d) είναι και αντιφατικό στρατηγικής και απολαβών του (a), είναι συνδυασμός των (b) και (c).

Ισορροπία ενός γύρου

Παραθέτουμε τη στρατηγική ισορροπίας Nash (NE) ενός γύρου για το παιγνίο «πέτρα-ψαλίδι-χαρτί» χρησιμοποιώντας την ιδιότητα της αδιαφορίας από το 1.1.4. Σε αυτό το παιγνίο, κάθε ενέργεια χάνει από μια άλλη, οπότε η καλύτερη αντίδραση εξαρτάται εξ ολοκλήρου από την κίνηση του αντιπάλου:

Επειδή δεν υπάρχει κυρίαρχη καθαρή στρατηγική, μια μικτή στρατηγική NE είναι κατάλληλη. Ας είναι:

$$x = \Pr(A), \quad y = \Pr(B), \quad z = \Pr(C), \quad x + y + z = 1$$

Οι αναμενόμενες αποδόσεις, υποθέτοντας συμμετρία, είναι:

$$E(A) = -ba \cdot y + ac \cdot z, \quad E(B) = ba \cdot x - cb \cdot z, \quad E(C) = -ac \cdot x + cb \cdot y$$

Ορίζοντας $E(A) = E(B) = E(C)$, λύουμε:

$$x = \frac{cb}{ba + ac + cb}, \quad y = \frac{ac}{ba + ac + cb}, \quad z = \frac{ba}{ba + ac + cb}$$

Αυτό ορίζει τη Μικτή Στρατηγική NE για οποιαδήποτε γενική διαμόρφωση αποδόσεων RPS.

1.3.3 Παραλλαγές που χρησιμοποιούνται στα Πειράματα

Ενώ τα παιγνία που παρουσιάστηκαν νωρίτερα είναι μονού γύρου, τα πειράματά μας επικεντρώνονται αποκλειστικά στις **επαναλαμβανόμενες** εκδοχές τους. Οι πράκτορες παίζουν 16 γύρους στο δίλημμα του φυλακισμένου και 24 στο πέτρα-ψαλίδι-χαρτί, χωρίς να γνωρίζουν πόσοι γύροι θα παιχτούν. Μετά από κάθε γύρο, λαμβάνουν το αποτέλεσμα του γύρου αυτού(βλ. 1.2.1) πριν συνεχίσουν.

Η επαναλαμβανόμενη αναπαραγωγή επηρεάζει σημαντικά τη στρατηγική: η χρήση της ισορροπίας Nash ενός γύρου σε κάθε γύρο δεν είναι απαραίτητα βέλτιστη. Οι παίκτες μπορούν να προσαρμοστούν στη συμπεριφορά του αντιπάλου τους με την πάροδο του χρόνου. Επιπλέον, εάν ο αριθμός των γύρων είναι γνωστός, μπορεί να επηρεαστεί η συμπεριφορά των παικτών (π.χ. η λιποταξία γίνεται κυρίαρχη στο δίλημμα του φυλακισμένου). Για να αποφευχθεί αυτό, οι προτροπές μας διασφαλίζουν ότι οι πράκτορες δεν γνωρίζουν τον συνολικό αριθμό των γύρων.

1.3.4 Περιγραφή Παιγνίου

Όπως αναφέρθηκε στο 1.2.1, η περιγραφή του παιγνίου είναι το μέρος της προτροπής του συστήματος που καθορίζει ουσιαστικά το παιγνίο που παίζεται.

Για το δίλημμα του φυλακισμένου, η περιγραφή παιγνίου είναι η ακόλουθη.

You are playing a game against other player(s). The game is based on strategic moves.

Game Description:
There are two possible moves: {a} and {b}.
- if both players choose {a}, both get {aa} points.
- if both players choose {b}, both get {bb} points.
- if you choose {a} and the other player chooses {b}, you get {ab} points and the other player gets {ba} points.
- if you choose {b} and the other player chooses {a}, you get {ba} points and the other player gets {ab} points.

In each round of the game you play against one other player. You and the other player each make a move privately. The moves are compared and you and the other player get your reward/penalty based on the moves made.

Message Formats:
Move messages are sent when you want to make a move.
Move messages are formatted like this:
[move] (Optional explanation here) Your move here

Hint messages are received when you are provided with advice and directions. These messages are important and you should pay attention to them.
Hint messages are formatted like this:
[hint] Hint message here

Objective:
Maximize your points and/or minimize your penalties.

Tips:
- Try to formulate a strategy by reasoning on the provided information.
- Pay attention to the payoff matrix of the game moves.
- You should only send move messages. Hint messages will only be provided to you.
- Pay attention to hints when provided.
- Do not use markdown formatting in your messages (e.g., bold text, italic text, etc.).

Για το πέτρα-ψαλίδι-χαρτί, η περιγραφή παιγνίου είναι η ακόλουθη.

You are playing a game against other player(s). The game is based on making moves.

Game Description:
There are three possible moves: {a}, {b}, and {c}.
- {a} beats {c} giving the winner {ac} points and the loser -{ac} points.
- {b} beats {a} giving the winner {ba} points and the loser -{ba} points.
- {c} beats {b} giving the winner {cb} points and the loser -{cb} points.
- If both players make the same move, the game is a tie and no points are awarded.

In each round of the game you play against one other player. You and the other player each make a move privately. The moves are compared and you and the other player get your reward/penalty based on the moves made.

Message Formats:
Move messages are sent when you want to make a move.
Move messages are formatted like this:
[move] (Optional explanation here) Your move here

Hint messages are received when you are provided with advice and directions. These messages are important and you should pay attention to them.
Hint messages are formatted like this:
[hint] Hint message here

Objective:
Maximize your points and/or minimize your penalties.

Tips:
- Try to formulate a strategy by reasoning on the provided information.
- Pay attention to the payoff matrix of the game moves.
- You should only send move messages. Hint messages will only be provided to you.
- Pay attention to hints when provided.
- Do not use markdown formatting in your messages (e.g., bold text, italic text, etc.).

Όπως φαίνεται εδώ, η περιγραφή παιγνίου περιλαμβάνει επίσης τις αρχικές υποδείξεις που δίνονται στην προτροπή του συστήματος.

1.3.5 Τύποι Παικτών

Για να μελετήσουμε τον στρατηγικό συλλογισμό των ΜΓΜ, ορίζουμε μια σειρά τύπων παικτών, τόσο βασισμένων σε ΜΓΜ όσο και αλγοριθμικών. Αυτό μας επιτρέπει να συγκρίνουμε διαφορετικούς τύπους προτροπών και να αξιολογήσουμε τη συλλογιστική των ΜΓΜ σε σχέση με προβλέψιμες στρατηγικές. Οι παίκτες που δεν βασίζονται σε ΜΓΜ είναι ιδιαίτερα χρήσιμοι για να εξετάσουμε αν τα ΜΓΜ μπορούν να αναγνωρίσουν και να εκμεταλλευτούν ή να συνεργαστούν με απλές, ερμηνεύσιμες συμπεριφορές.

Παίκτες ΜΓΜ

1. **zs / default**: Πράκτορας με προτροπή χωρίς παραδείγματα.
2. **cot**: Πράκτορας με προτροπή με συλλογιστική πορεία.
3. **spp**: Πράκτορας με προτροπή μονοπρόσωπης εκτέλεσης.
4. **sc-*type***: Αυτοσυνεπής παραλλαγή οποιουδήποτε από τα παραπάνω (π.χ. **sc-zs**, **sc-cot**, **sc-spp**).

Αλγοριθμικοί παίκτες

1. **srep**: Παίζει τυχαία χρησιμοποιώντας την κατανομή ΝΕ ενός γύρου σε κάθε γύρο.
2. **pp**: Ακολουθεί ένα σταθερό κυκλικό μοτίβο.
3. Αντίπαλοι που υπάρχουν στα παιγνία Διλήμματος Φυλακισμένου:
   (a) **mf**: Αυτός ο παίκτης επιλέγει πάντα την κίνηση που, όταν συνδυάζεται με την πιο συχνή κίνηση του αντιπάλου του, δίνει την καλύτερη ανταμοιβή.
   (b) **tft**: Ο παίκτης «tft» επιλέγει πάντα την κίνηση που, όταν συνδυάζεται με την πιο πρόσφατη κίνηση του αντιπάλου του, δίνει την καλύτερη ανταμοιβή.
4. Αντίπαλοι που εμφανίζονται στα παιγνία Πέτρα-Ψαλίδι-Χαρτί:

(a) **ap**: Επιλέγει την αντίθετη κίνηση από την πιο συχνή κίνηση του αντιπάλου.

(b) **tft**: Αντιμετωπίζει την πιο πρόσφατη κίνηση του αντιπάλου (εμπνευσμένο από το «tit-for-tat»).

1.3.6 Σχεδιασμός Πειραμάτων

Τα πειράματά μας διαμορφώνονται από διάφορες βασικές παραμέτρους, οι οποίες επιλέγονται με σκοπό την εξισορρόπηση της εκφραστικότητας και της αποτελεσματικότητας. Αυτές περιλαμβάνουν τις ρυθμίσεις του μοντέλου, τη δομή του παιγνίου, τους τύπους των αντιπάλων και τη χρήση αντιφατικών δεδομένων.

Μια οπτική επισκόπηση μιας πιθανής συνομιλίας παρέχεται στο 1.3.1.

- **Παράμετροι ΜΓΜ**: Η θερμοκρασία έχει οριστεί σε 1.0 για να ενθαρρύνει ποικίλα, δημιουργικά αποτελέσματα σε επαναλαμβανόμενες εκτελέσεις. Οι άλλες παράμετροι παραμένουν στις προεπιλεγμένες τιμές.
- **Γύροι ανά παιγνίο**: Κάθε παιχνίδι αποτελείται από 16 γύρους για τα αντιφατικά σενάρια του διλήμματος του φυλακισμένου. Για παιγνία με δύο επιλογές, όπως το δίλημμα του φυλακισμένου και το κυνήγι του ελαφιού, αυτή η διάρκεια επιτρέπει στα ΜΓΜ να παρατηρούν και να προσαρμόζονται στα μοτίβα. Κάθε παιχνίδι αποτελείται από 24 γύρους για τα αντιφατικά σενάρια του πέτρα-ψαλίδι-χαρτί, καθώς είναι ένα παιγνίο που δίνει περισσότερες ευκαιρίες στους παίκτες.
- **Επαναλήψεις**: Οι παίκτες που δεν είναι SC εκτελούν 5 επαναλήψεις ανά παιχνίδι, ενώ οι παίκτες SC εκτελούν μόνο 2 (λόγω της υψηλότερης αναμενόμενης συνέπειας και του υπολογιστικού κόστους).
- **Τύποι αντιπάλων**: Οι παίκτες που δεν είναι αυτοσυνεπείς (SC) αντιμετωπίζουν όλους τους άλλους παίκτες που δεν είναι SC. Οι παίκτες SC αντιμετωπίζουν τους ίδιους αντίπαλους που αναφέρθηκαν, αλλά όχι άλλους SC αντιπάλους, αποφεύγοντας έτσι δαπανηρά παιχνίδια.
- **Αυτοσυνέπεια**: Όπως περιγράφεται στο 1.1.3, οι παίκτες SC δειγματοληπτούν 3 αποτελέσματα ανά γύρο, στην περίπτωση του διλήμματος του φυλακισμένου, και 5 αποτελέσματα ανά γύρο, στην περίπτωση του πέτρα-ψαλίδι-χαρτί, και επιλέγουν το πιο συχνό. Οι ισοπαλίες επιλύονται τυχαία.
- **Αντιφατικά σενάρια**: Και οι 4 παραλλαγές κάθε παιγνίου που παρουσιάστηκαν νωρίτερα χρησιμοποιούνται για να δοκιμαστεί η προσαρμοστικότητα του ΜΓΜ σε διαφορετικά στρατηγικά περιβάλλοντα.

1.4 Αποτελέσματα

1.4.1 Μετρικές Αξιολόγησης

Συνολικοί Πόντοι Παρουσιάζουμε τους μέσους συνολικούς πόντους που συγκέντρωσε κάθε παίκτης σε κάθε σενάριο παιγνίου. Για τους παίκτες **non-sc**, τα αποτελέσματα είναι ο μέσος όρος 5 επαναλήψεων των πειραμάτων, ενώ για τους παίκτες **sc**, τα αποτελέσματα είναι ο μέσος όρος 2 επαναλήψεων των πειραμάτων.

Γύρος Κατανόησης Αντιπάλου Για να αποσαφηνίσουμε τα παραπάνω αποτελέσματα, να ρίξουμε φως στην συμπεριφορά που παρατηρούμε από τους **συνολικούς πόντους** και να κατανοήσουμε καλύτερα τη διαδικασία σκέψης των πρακτόρων ΜΓΜ, εισάγεται μια άλλη μέτρηση. Αντί να εξετάζουμε απλώς τους «Συνολικούς Πόντους» που συγκέντρωσαν οι πράκτορες κατά τη διάρκεια του παιγνίου, εξετάζουμε πόσο αργά ένας παίκτης κατάφερε να εκμεταλλευτεί τη συμπεριφορά του αντιπάλου του.

Θεωρούμε ότι ένας πράκτορας AI έχει καταφέρει να κατανοήσει τον αντίπαλό του, όταν ο πράκτορας ανταποκρίνεται συστηματικά με ενέργειες που χρησιμοποιούν τις κινήσεις του αντιπάλου προς όφελος του εαυτού του. Πιο τυπικά, ας υποθέσουμε ότι οι παίκτες A, B έχουν παίξει N γύρους και οι στρατηγικές που ακολούθησαν ήταν $(s_A^1, s_B^1), \ldots, (s_A^N, s_B^N)$ - π.χ., $(defect, cooperate), \ldots, (defect, defect)$ -. Ας υποθέσουμε ότι A είναι ο πράκτορας AI, τότε ονομάζουμε **γύρο κατανόησης του αντιπάλου**, m, τον γύρο μετά τον οποίο κάθε κίνηση που κάνει ο A αποφέρει κέρδος για τον A που είναι τουλάχιστον τόσο καλό όσο το κέρδος που παίρνει ο B.[3]

[3] Αυτή η κίνηση δεν είναι απαραίτητα η καλύτερη απάντηση του A στον B (όπως περιγράφεται στο 1.1.4)

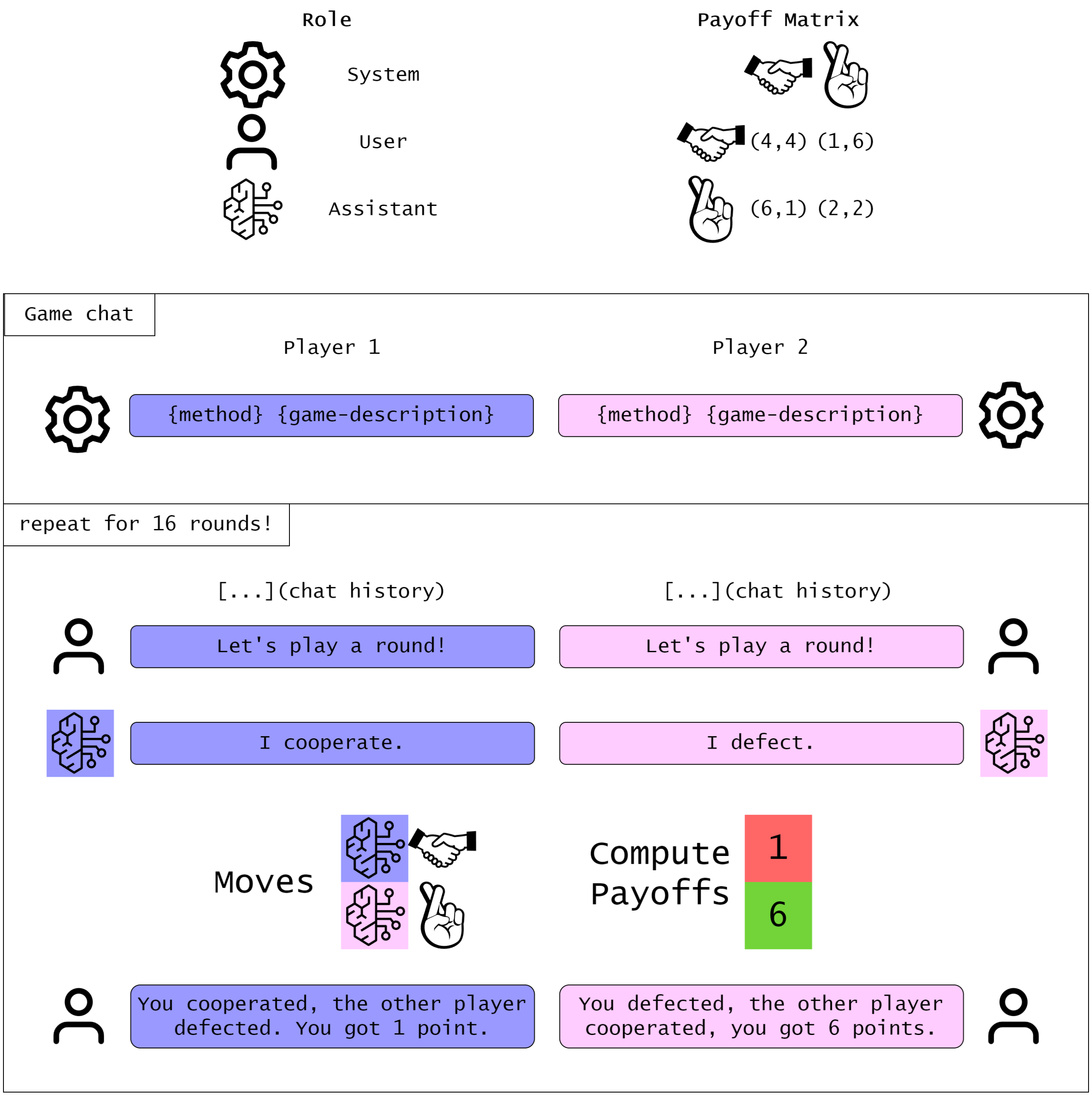


Figure 1.3.1: Ένα παράδειγμα συνομιλίας που θα μπορούσε να συμβεί σε ένα παιγνίο Διλήμματος Φυλακισμένου σύμφωνα με το σχεδιασμό μας, όπου δύο πράκτορες ΜΓΜ παίζουν μεταξύ τους. Ο ρόλος του «Χρήστη» χρησιμοποιείται από εμάς - το περιβάλλον - για την παροχή πληροφοριών στους παίκτες AI.

Επιπλέον, έχουμε επεκτείνει αυτόν τον ορισμό, συμπεριλαμβάνοντας ένα ποσοστό tp (ποσοστό-στόχος) που χαλαρώνει την απαίτηση της «καλής» αντίδρασης σε *κάθε* κίνηση που κάνει ο αντίπαλος στην ακόλουθη απαίτηση: στους γύρους από m έως N, οι κινήσεις του A είναι τουλάχιστον τόσο καλές όσο του B σε tp ποσοστό αυτών των γύρων.

Οι πίνακες, που αναφέρονται στη συνέχεια, είναι αποτελέσματα όπου $tp = 90\%$ και μικρότερες τιμές θεωρούνται καλύτερες, καθώς δείχνουν ότι ο πράκτορας AI κατάλαβε νωρίτερα στο παιγνίο πώς να παίξει με τον συγκεκριμένο αντίπαλό του.

Αποδοτικότητα Η απόδοση θεωρείται συχνά ο πιο σημαντικός παράγοντας οποιασδήποτε τεχνολογίας, με την αποτελεσματικότητα να αποτελεί κλασικό συνοδευτικό της. Αυτές οι δύο έννοιες υπάρχουν και στον κόσμο των ΜΓΜ, όπου το κόστος είναι συχνά συνάρτηση των token που παράγονται από το μοντέλο AI. Εισάγεται ένας απλός δείκτης αποτελεσματικότητας:

$$efficiency = \frac{points}{token}$$

Ρυθμός Αποτυχίας Τα μεγάλα γλωσσικά μοντέλα (ΜΓΜ) είναι πολύπλοκες δομές με διάφορες αναδυόμενες δεξιότητες, που δεν είναι απαλλαγμένες από προβλήματα. Παρά τις προσπάθειές μας να σχεδιάσουμε ένα περιβάλλον για τα ΜΓΜ που να χειρίζεται σφάλματα και απροσδόκητη ή ανεπιθύμητη συμπεριφορά (όπως περιγράφεται στο 1.2.1), τα ΜΓΜ εξακολουθούν να αντιμετωπίζουν περιστασιακά προβλήματα με την εκτέλεση των οδηγιών μας. Το περιβάλλον και η λογική διόρθωσης σφαλμάτων που ακολουθούμε είναι μια εκτεταμένη έκδοση του περιβάλλοντος τύπου Gym που χρησιμοποιείται στο [28].

1.4.2 Δίλημμα Φυλακισμένου

Συνολικοί Πόντοι

Δείχνουμε τα αποτελέσματα για το βασικό παιγνίο στον πίνακα 1.6. Υπενθυμίζουμε ότι σε κάθε παιγνίο παίζονται 16 συνεχόμενοι γύροι μεταξύ των δύο αντιπάλων. Για τους παίκτες **non-sc**, τα αποτελέσματα είναι ο μέσος όρος 5 επαναλήψεων των πειραμάτων, ενώ για τους παίκτες **sc**, τα αποτελέσματα είναι ο μέσος όρος 2 επαναλήψεων των πειραμάτων.

Τα αποτελέσματα για τα παιγνία αντιφατικών σεναρίων βρίσκονται στους πίνακες 6.3, 6.4, 6.5.

ΜΓΜ έναντι ΜΓΜ Οι τρεις πρώτες στήλες του πίνακα αντιστοιχούν σε παιγνία μεταξύ πρακτόρων του ίδιου ΜΓΜ.

Στο δίλημμα του φυλακισμένου, πολλά αποτελέσματα είναι κοντά στο $64 = 4 \cdot 16$, υποδηλώνοντας συνεργατικές τάσεις αντί για το αμοιβαίο όφελος της λιποταξίας 32. Το Claude Sonnet 4 και το DeepSeek-R1 ευθυγραμμίζονται περισσότερο με την κυρίαρχη στρατηγική λιποταξίας.

ΜΓΜ έναντι Αλγοριθμηκών Αντιπάλων Οι τέσσερις τελευταίες στήλες αντιπροσωπεύουν αντιπάλους με σταθερή στρατηγική, οι αντίπαλοι αυτοί είναι χρήσιμοι για την αξιολόγηση της προσαρμοστικότητας των ΜΓΜ.

- **srep**: Ο srep παίζει πάντα λιποταξία στο ΔΦ, προκαλώντας και τους AI πράκτορες να λιποτακτήσουν (≈ 32 πόντοι). Στο Κυνήγι Ελαφίου, ο srep παίζει το μικτό Nash ($p = \frac{1}{3}, q = \frac{2}{3}$), αποδίδοντας ≈ 42.67 αναμενόμενους συνολικούς πόντους, τα αποτελέσματα των πρακτόρων αντανακλούν την τιμή αυτή.
- **pp**: Παίζει κυκλικά συγκεκριμένο μοτίβο κινήσεων. Η βέλτιστη κίνηση αποφέρει 64 πόντους τόσο στο ΔΦ όσο και στο ΚΕ. Στο ΔΦ, τα ΜΓΜ συχνά φτάνουν σε αυτό μόνο με λιποταξία. Στο ΚΕ, η προσαρμογή είναι πιο δύσκολη, με χαμηλότερες βαθμολογίες που υποδηλώνουν απαιτήσεις κατανόσης αντιπάλου.
- **mf & tft**: Στο ΔΦ, και τα δύο συγκλίνουν στην λιποταξία (≈ 32 πόντοι). Στο ΚΕ, τα κέρδη αποκαλύπτουν αν το ΜΓΜ καταλήγει στο «Ελάφι» (≈ 96) ή στον «Λαγό» (≈ 32) ή αλλάζει στη μέση του παιγνίου (ενδιάμεσες βαθμολογίες, που υποδηλώνουν συλλογιστική για μεταβολή στρατηγικής - ο παίκτης έτυχε να αρχίσει με «Λαγός», συνειδητοποιεί ότι σε βάθος χρόνου το «Ελάφι» δίνει καλύτερο αποτέλεσμα και αλλάζει την στρατηγική του -). Οι χαμηλές βαθμολογίες υποδηλώνουν παρεξήγηση της στρατηγικής του αντιπάλου.

Γύρος Κατανόησης Αντιπάλου

Δείχνουμε τα αποτελέσματα για το βασικό παιγνίο στον πίνακα 1.7. Ισχύουν τα ίδια σχόλια που αναφέραμε για τους "Συνολικούς Πόντους".

Τα αποτελέσματα για τα παιγνία αντιφατικών σεναρίων βρίσκονται στους πίνακες 6.7, 6.8, 6.9.

		pd						
		zs	spp	cot	srep	pp	mf	tft
model	prompt							
C3.5Sv2	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	54.2 ± 4.5	30.2 ± 1.1	30.6 ± 0.9
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	53.6 ± 1.5	31.0 ± 1.0	30.4 ± 0.9
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.4 ± 0.5	58.4 ± 4.3	30.6 ± 0.9	32.0 ± 0.0
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	52.0 ± 0.0	32.0 ± 0.0	30.0 ± 1.4
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	53.5 ± 2.1	32.0 ± 0.0	30.0 ± 0.0
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	62.0 ± 0.0	31.0 ± 1.4	31.0 ± 1.4
C3.7S	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	54.8 ± 4.1	29.8 ± 0.8	30.4 ± 0.5
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	61.2 ± 1.1	31.0 ± 0.7	31.2 ± 0.4
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	56.8 ± 4.4	31.0 ± 1.0	30.2 ± 0.4
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	52.5 ± 0.7	31.0 ± 1.4	30.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	61.0 ± 1.4	30.5 ± 0.7	31.0 ± 0.0
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	57.5 ± 6.4	31.0 ± 1.4	31.5 ± 0.7
C3.7S(T)	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	52.8 ± 0.4	31.0 ± 0.7	31.6 ± 0.9
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.6 ± 0.5	57.2 ± 3.7	30.8 ± 0.8	30.2 ± 0.4
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.2 ± 0.4	58.0 ± 4.7	31.0 ± 1.0	31.0 ± 1.0
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	52.0 ± 0.0	32.0 ± 0.0	30.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	53.0 ± 0.0	31.0 ± 1.4	30.5 ± 0.7
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	57.5 ± 2.1	32.0 ± 0.0	31.0 ± 1.4
C4S	zs	49.4 ± 15.5	41.0 ± 10.7	42.0 ± 10.7	31.4 ± 0.5	**55.8 ± 4.4**	32.8 ± 1.6	32.2 ± 1.6
	cot	35.2 ± 3.0	39.6 ± 2.4	39.8 ± 7.4	31.4 ± 0.5	**60.2 ± 4.5**	32.2 ± 1.6	34.2 ± 1.9
	spp	48.4 ± 13.0	46.2 ± 13.6	36.4 ± 4.2	31.6 ± 0.5	**63.0 ± 1.0**	34.4 ± 2.1	31.8 ± 2.4
	sc-zs	34.5 ± 4.9	34.0 ± 2.8	40.0 ± 4.2	**32.0 ± 0.0**	**58.0 ± 8.5**	32.0 ± 0.0	32.0 ± 0.0
	sc-cot	39.0 ± 4.2	35.5 ± 4.9	34.0 ± 2.8	**32.0 ± 0.0**	**64.0 ± 0.0**	35.0 ± 0.0	33.5 ± 2.1
	sc-spp	32.0 ± 1.4	38.0 ± 1.4	35.5 ± 4.9	30.5 ± 0.7	**63.5 ± 0.7**	35.5 ± 0.7	31.5 ± 0.7
C4S(T)	zs	32.0 ± 0.0	34.8 ± 2.7	32.0 ± 0.0	**32.0 ± 0.0**	**63.4 ± 0.9**	35.2 ± 1.8	**35.8 ± 0.4**
	cot	32.0 ± 0.0	40.8 ± 8.0	34.6 ± 3.6	31.6 ± 0.5	**64.0 ± 0.0**	33.8 ± 1.8	34.4 ± 2.2
	spp	31.4 ± 0.9	50.6 ± 15.4	41.8 ± 12.4	31.4 ± 0.5	**59.6 ± 5.2**	34.8 ± 1.8	34.8 ± 2.2
	sc-zs	37.5 ± 7.8	34.0 ± 2.8	32.0 ± 0.0	**32.0 ± 0.0**	**64.0 ± 0.0**	32.0 ± 0.0	33.5 ± 2.1
	sc-cot	32.0 ± 0.0	37.0 ± 1.4	32.0 ± 0.0	**32.0 ± 0.0**	**64.0 ± 0.0**	**36.0 ± 0.0**	32.0 ± 0.0
	sc-spp	47.0 ± 21.2	36.0 ± 0.0	32.0 ± 0.0	31.5 ± 0.7	**62.0 ± 2.8**	35.5 ± 0.7	33.5 ± 2.1
DS-R1	zs	36.8 ± 4.1	39.0 ± 2.5	30.2 ± 13.0	31.6 ± 0.5	**62.6 ± 1.5**	32.8 ± 2.0	32.6 ± 2.2
	cot	33.2 ± 2.6	32.2 ± 2.2	33.0 ± 6.6	28.6 ± 7.1	**64.0 ± 0.0**	32.2 ± 1.8	23.6 ± 8.7
	spp	36.4 ± 4.2	32.0 ± 6.4	31.2 ± 6.3	31.2 ± 0.4	**61.4 ± 3.7**	33.2 ± 2.6	33.2 ± 1.6
	sc-zs	32.5 ± 2.1	33.0 ± 2.8	33.5 ± 3.5	31.5 ± 0.7	**64.0 ± 0.0**	35.0 ± 0.0	35.5 ± 0.7
	sc-cot	34.5 ± 4.9	34.0 ± 2.8	27.0 ± 7.1	**32.0 ± 0.0**	**64.0 ± 0.0**	31.5 ± 0.7	31.5 ± 0.7
	sc-spp	35.0 ± 4.2	36.0 ± 5.7	34.0 ± 2.8	31.0 ± 0.0	**61.0 ± 4.2**	33.5 ± 2.1	32.0 ± 0.0
L3.3-70B	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.6 ± 0.5	52.0 ± 0.0	31.6 ± 0.9	31.6 ± 0.9
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	55.6 ± 4.6	30.8 ± 1.1	30.4 ± 0.9
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	53.8 ± 2.9	30.8 ± 1.1	30.4 ± 0.9
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.0 ± 0.0	52.0 ± 0.0	31.0 ± 1.4	32.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	52.0 ± 0.0	32.0 ± 0.0	31.0 ± 1.4
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.5 ± 0.7	55.5 ± 4.9	30.0 ± 0.0	32.0 ± 0.0
M-L(24.07)	zs	**61.4 ± 3.2**	60.0 ± 4.4	59.6 ± 9.8	24.2 ± 0.4	53.8 ± 2.0	26.0 ± 1.6	25.8 ± 0.8
	cot	**64.0 ± 0.0**	60.2 ± 3.6	62.0 ± 2.8	29.8 ± 2.2	57.8 ± 3.0	29.0 ± 1.2	26.0 ± 1.2
	spp	**64.0 ± 0.0**	62.2 ± 4.0	60.8 ± 7.2	18.6 ± 10.6	53.2 ± 2.2	26.4 ± 3.2	24.4 ± 1.1
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	24.0 ± 0.0	54.0 ± 2.8	25.0 ± 1.4	25.0 ± 1.4
	sc-cot	62.0 ± 2.8	**64.0 ± 0.0**	60.5 ± 4.9	27.5 ± 4.9	60.0 ± 5.7	33.5 ± 3.5	29.0 ± 4.2
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	24.0 ± 0.0	56.0 ± 0.0	22.5 ± 4.9	22.5 ± 2.1

Table 1.6: Συνολικό Αποτέλεσμα για τους Συνολικούς Πόντους από όλες τις Επαναλήψεις (pd)

		pd						
		zs	spp	cot	srep	pp	mf	tft
model	prompt							
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.2 ± 0.4	8.4 ± 7.0	5.6 ± 6.4	2.8 ± 1.3
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	14.0 ± 2.0	3.0 ± 1.2	2.2 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.6 ± 0.5	7.6 ± 6.1	2.0 ± 0.7	3.0 ± 0.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	12.0 ± 0.0	3.0 ± 0.0	4.5 ± 0.7
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	9.0 ± 4.2	3.0 ± 0.0	2.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	2.0 ± 0.0	2.5 ± 0.7	2.5 ± 0.7
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	11.6 ± 5.5	3.6 ± 1.5	3.2 ± 1.6
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	2.0 ± 0.0	3.2 ± 1.8	3.2 ± 2.3
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	9.6 ± 6.2	3.0 ± 1.2	3.6 ± 2.3
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	13.0 ± 1.4	2.5 ± 0.7	2.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	2.0 ± 0.0	3.5 ± 2.1	5.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	8.0 ± 8.5	2.5 ± 0.7	4.0 ± 1.4
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.6 ± 1.3	12.8 ± 4.1	4.0 ± 1.4	2.8 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.4 ± 0.5	10.4 ± 6.1	2.6 ± 1.5	1.8 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.8 ± 0.4	6.8 ± 6.7	3.0 ± 1.2	3.0 ± 1.2
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	13.0 ± 1.4	3.0 ± 0.0	2.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	15.0 ± 1.4	2.5 ± 0.7	1.5 ± 0.7
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	9.0 ± 7.1	3.0 ± 0.0	2.5 ± 0.7
C4S	zs	1.2 ± 0.4	1.4 ± 0.9	3.0 ± 4.5	**1.0 ± 0.0**	12.0 ± 5.5	1.8 ± 1.1	2.2 ± 1.6
	cot	3.0 ± 2.8	10.8 ± 4.1	3.2 ± 2.3	**1.0 ± 0.0**	6.0 ± 6.1	2.2 ± 1.6	1.6 ± 1.3
	spp	5.2 ± 6.9	1.6 ± 1.3	4.6 ± 2.1	**1.0 ± 0.0**	1.8 ± 1.3	4.2 ± 7.2	2.2 ± 1.8
	sc-zs	2.5 ± 2.1	**1.0 ± 0.0**	7.0 ± 8.5	**1.0 ± 0.0**	6.5 ± 7.8	**1.0 ± 0.0**	2.0 ± 1.4
	sc-cot	3.5 ± 3.5	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	4.0 ± 4.2	3.0 ± 2.8	**1.0 ± 0.0**	2.0 ± 1.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
C4S(T)	zs	**1.0 ± 0.0**	2.0 ± 2.2	**1.0 ± 0.0**	**1.0 ± 0.0**	1.2 ± 0.4	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	4.2 ± 7.2	4.0 ± 6.7	**1.0 ± 0.0**	**1.0 ± 0.0**	1.6 ± 1.3	**1.0 ± 0.0**
	spp	1.4 ± 0.9	2.2 ± 2.2	2.4 ± 2.2	**1.0 ± 0.0**	5.2 ± 6.4	1.6 ± 1.3	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	3.0 ± 2.8	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.5 ± 2.1	**1.0 ± 0.0**	2.0 ± 1.4
DS-R1	zs	3.2 ± 2.2	2.0 ± 2.2	8.0 ± 8.3	**1.0 ± 0.0**	4.0 ± 6.7	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.2 ± 2.7	10.6 ± 8.8
	spp	5.2 ± 6.6	7.4 ± 7.6	5.2 ± 6.9	**1.0 ± 0.0**	4.8 ± 5.4	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-zs	6.5 ± 7.8	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	3.5 ± 3.5	**1.0 ± 0.0**	9.0 ± 11.3	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	2.5 ± 2.1	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	7.5 ± 9.2	**1.0 ± 0.0**	**1.0 ± 0.0**
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.8 ± 1.1	14.0 ± 0.0	2.8 ± 0.4	2.8 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.4 ± 0.9	10.0 ± 6.5	2.4 ± 0.5	2.2 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.4 ± 0.9	12.8 ± 5.0	2.4 ± 0.5	2.2 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	4.0 ± 0.0	14.0 ± 0.0	2.5 ± 0.7	3.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	14.0 ± 0.0	3.0 ± 0.0	2.5 ± 0.7
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	3.0 ± 1.4	9.0 ± 7.1	2.0 ± 0.0	3.0 ± 0.0
M-L(24.07)	zs	10.2 ± 8.4	6.8 ± 6.8	**3.8 ± 6.3**	16.4 ± 0.5	15.2 ± 1.1	16.2 ± 0.4	16.0 ± 0.7
	cot	**1.0 ± 0.0**	7.0 ± 8.2	4.2 ± 7.2	4.4 ± 7.1	12.6 ± 6.5	8.6 ± 7.3	16.2 ± 0.8
	spp	**1.0 ± 0.0**	3.2 ± 4.9	4.2 ± 7.2	13.4 ± 6.9	15.6 ± 0.9	17.0 ± 0.0	16.4 ± 0.9
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 0.0	15.0 ± 1.4	16.5 ± 0.7	16.5 ± 0.7
	sc-cot	8.5 ± 10.6	**1.0 ± 0.0**	8.5 ± 10.6	9.0 ± 11.3	8.5 ± 10.6	**1.0 ± 0.0**	10.0 ± 9.9
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	16.0 ± 0.0	17.0 ± 0.0	16.0 ± 0.0

Table 1.7: Γύρος # όπου ο Πράκτορας κατανόησε την Στρατηγική του Αντιπάλου του (pd)

ΜΓΜ έναντι ΜΓΜ Οι πρώτες τρεις στήλες στους πίνακες αποτελεσμάτων αντιπαραθέτουν πράκτορες πανομοιότυπων ΜΓΜ μεταξύ τους. Στα σενάρια του διλήμματος του φυλακισμένου, τα περισσότερα μοντέλα φτάνουν γρήγορα σε αμοιβαία συνεργασία, με περιστασιακές καθυστερήσεις. Το *Mistral Large (24.07)* δυσκολεύεται στις αντιφατικές παραλλαγές, παρουσιάζοντας μειωμένη ικανότητα συλλογιστικής. Στις ρυθμίσεις Κυνήγι Ελαφίου, η αμοιβαία συνεργασία επίσης αναδύεται γρήγορα, αν και το Mistral έχει και πάλι πρόβλημα με τις μετονομασμένες στρατηγικές.

Το *Claude Sonnet 4* και το *DeepSeek-R1* αποκλίνουν συχνότερα από την αρχική συμφωνία, συνδυάζοντας χαμηλές συνολικές βαθμολογίες με περιστασιακά καθυστερημένη σύγκλιση, υποδηλώνοντας είτε επίμονη λιποταξία στο ΔΦ είτε προτίμηση Λαγού στο ΚΕ, πιθανώς ως μέρος πιο διερευνητικών στρατηγικών.

ΜΓΜ έναντι Αλγοριθμηκών Αντιπάλων Οι τέσσερις τελευταίες στήλες αντιπροσωπεύουν αντιπάλους με σταθερή στρατηγική, οι αντίπαλοι αυτοί είναι χρήσιμοι για την αξιολόγηση της προσαρμοστικότητας των ΜΓΜ.

- **srep**: Πάντα αποστασία στο ΔΦ, τα περισσότερα ΜΓΜ προσαρμόζονται γρήγορα, εκτός από το *Llama 3.3 70B Instruct* (πιο αργό σε αντιφατικά σενάρια) και το *Mistral*, το οποίο αποδίδει καλά μόνο με προτροπή **cot**. Τα αποτελέσματα του ΚΕ δεν είναι ενημερωτικά λόγω της ισορροπίας Nash μικτής στρατηγικής.
- **pp**: Κύκλοι μεταξύ κινήσεων (λιποταξία/συνεργασία στο ΔΦ; λαγός/ελάφι στο ΚΕ). Στο ΔΦ, μόνο το *Claude 4* και το *DeepSeek* αναγνωρίζουν με συνέπεια το μοτίβο σε όλες τις προτροπές. Τα άλλα παρουσιάζουν επιτυχία σε συγκεκριμένες προτροπές (π.χ. *Claude 3.5 Sonnet v2* με **spp**, *Claude 3.7 Sonnet* με **cot**). Στο ΚΕ, τα περισσότερα ΜΓΜ δεν αναγνωρίζουν τον απλό κύκλο, απαιτώντας συχνά πολλούς γύρους για να προσαρμοστούν.
- **mf & tft**: Και τα δύο συγκλίνουν σε πλήρη λιποταξία στο ΔΦ, οπότε ένας λογικός παίκτης ταιριάζει με αυτό από την αρχή — παρατηρείται σε όλα τα μοντέλα εκτός από το *Mistral Large (24.07)*. Στο ΚΕ, η αλλαγή από Λαγό σε Ελάφι στη μέση του παιχνιδιού υποδηλώνει ύπαρξη συλλογιστικής ικανότητας, όπως αναφέραμε.

Αποδοτικότητα

Δείχνουμε τα αποτελέσματα στον πίνακα 1.8.

Σε όλα τα μοντέλα, η κατάταξη της αποτελεσματικότητας των στυλ προτροπής (πίνακας 1.8) είναι:

1. **zs** (χωρίς παραδείγματα)
2. **cot** (με συλλογιστική πορεία)
3. **spp** (μονοπρόσωπης εκτέλεσης), λιγότερο αποτελεσματικό.

Η προσθήκη του **sc** (αυτοσυνέπεια) σε οποιοδήποτε στυλ μειώνει την αποτελεσματικότητα, και η ίδια κατάταξη ισχύει μεταξύ των παραλλαγών του **sc**. Αξίζει να σημειωθεί ότι τα μοντέλα «Thinking» του *Claude* συχνά ταιριάζουν ή υπερτερούν των προεπιλεγμένων εκδόσεών τους.

Η αποτελεσματικότητα είναι υψηλότερη στις παραλλαγές Κυνήγι Ελαφίου από ό,τι στο Δίλημμα Φυλακισμένου ή στην αντιφατική στρατηγική του, καθώς το Κυνήγι Ελαφίου επιτρέπει καλύτερες αμοιβαίες αποδόσεις.

Ρυθμός Αποτυχίας

Δείχνουμε τα αποτελέσματα στον πίνακα 1.9.

Όλα τα ΜΓΜ καταφέρνουν να έχουν σχεδόν τέλεια ποσοστά «εγκυρότητας», όπως φαίνεται στον πίνακα 1.9. Όταν παίζεται ένα παιγνίο, μια τιμή εγκυρότητας αποδίδεται σε κάθε έναν από τους 16 γύρους που περιλαμβάνει. Εάν παρουσιαστεί κάποιο σφάλμα, η τιμή αυτή θα είναι «false». Ο πίνακας 1.9 απεικονίζει απλώς τον μέσο αριθμό έγκυρων γύρων σε όλα τα παιγνία.

Σε αυτό το σημείο, πρέπει να σημειωθεί ότι τα περισσότερα σφάλματα που αντιμετώπισε το μοντέλο *Mistral Large (24.07)* οφείλονταν στην αδυναμία του να ακολουθήσει τις οδηγίες μορφοποίησης. Συχνά χρησιμοποιούσε μορφοποίηση τύπου **markdown** στην έξοδο του, ακόμη και όταν του ζητήθηκε ρητά να μην το κάνει (όπως φαίνεται στις υποδείξεις που παρέχονται στα μοντέλα στο 1.2.1).

model	prompt	pd	pd-alt	sh	sh-alt
Claude 3.5 Sonnet v2	zs	18.35 ± 10.24	**28.04 ± 12.12**	21.84 ± 12.24	15.56 ± 9.72
	cot	9.00 ± 4.05	**15.88 ± 8.80**	11.25 ± 5.37	7.85 ± 2.92
	spp	7.54 ± 2.98	**11.99 ± 4.73**	9.60 ± 4.14	5.93 ± 2.41
	sc-zs	6.37 ± 3.42	**9.88 ± 4.56**	7.63 ± 3.60	4.59 ± 2.71
	sc-cot	2.96 ± 1.46	**6.06 ± 4.01**	5.50 ± 5.26	2.58 ± 1.00
	sc-spp	2.54 ± 1.09	3.81 ± 1.63	**4.70 ± 4.25**	1.96 ± 0.76
Claude 3.7 Sonnet	zs	14.03 ± 7.88	**24.35 ± 12.64**	24.04 ± 12.10	13.16 ± 7.68
	cot	6.39 ± 2.64	10.14 ± 4.40	**10.68 ± 4.72**	6.17 ± 2.67
	spp	5.48 ± 3.66	**7.89 ± 5.63**	7.22 ± 3.47	4.57 ± 3.10
	sc-zs	4.88 ± 2.49	7.02 ± 4.64	**7.70 ± 4.79**	4.37 ± 2.50
	sc-cot	2.34 ± 1.32	**3.63 ± 1.78**	3.15 ± 1.60	2.01 ± 1.07
	sc-spp	2.36 ± 1.37	**3.21 ± 2.72**	2.67 ± 1.01	1.51 ± 0.69
Claude 3.7 Sonnet (Thinking)	zs	15.67 ± 6.23	22.93 ± 10.25	**30.01 ± 15.17**	19.17 ± 9.07
	cot	8.44 ± 3.39	**13.73 ± 5.86**	13.68 ± 6.63	7.86 ± 3.05
	spp	6.82 ± 2.85	**10.99 ± 4.98**	9.87 ± 4.58	6.10 ± 2.53
	sc-zs	5.35 ± 2.15	9.53 ± 4.24	**9.88 ± 5.87**	6.90 ± 2.89
	sc-cot	2.67 ± 1.19	4.16 ± 2.16	**4.45 ± 2.13**	2.71 ± 1.05
	sc-spp	2.21 ± 0.86	3.45 ± 1.66	**3.49 ± 1.84**	2.08 ± 0.91
Claude Sonnet 4	zs	7.45 ± 3.99	13.85 ± 8.83	**15.12 ± 12.56**	7.92 ± 4.45
	cot	4.15 ± 0.81	**6.34 ± 3.79**	4.53 ± 2.42	4.19 ± 0.82
	spp	3.85 ± 1.26	**4.49 ± 2.59**	4.46 ± 2.23	3.45 ± 0.81
	sc-zs	2.15 ± 0.48	4.86 ± 3.10	**5.45 ± 3.33**	1.99 ± 0.41
	sc-cot	1.36 ± 0.18	1.50 ± 0.94	**1.51 ± 0.47**	1.41 ± 0.25
	sc-spp	1.13 ± 0.35	**2.01 ± 0.99**	1.36 ± 0.81	1.04 ± 0.31
Claude Sonnet 4 (Thinking)	zs	11.59 ± 3.05	**24.90 ± 12.96**	19.88 ± 11.98	11.66 ± 3.63
	cot	5.84 ± 1.51	**9.81 ± 6.49**	8.43 ± 5.02	5.37 ± 1.18
	spp	3.99 ± 1.62	**6.68 ± 2.97**	6.04 ± 2.96	3.50 ± 0.87
	sc-zs	3.99 ± 1.00	**7.91 ± 3.90**	7.51 ± 3.45	4.19 ± 1.89
	sc-cot	1.88 ± 0.38	**3.43 ± 2.19**	3.13 ± 2.20	1.94 ± 0.54
	sc-spp	1.32 ± 0.36	**1.80 ± 0.90**	1.68 ± 0.87	1.07 ± 0.28
DeepSeek-R1	zs	**13.86 ± 5.08**	12.59 ± 5.53	13.47 ± 6.02	12.38 ± 4.01
	cot	8.48 ± 5.19	9.71 ± 4.00	11.77 ± 3.26	**12.48 ± 3.23**
	spp	8.22 ± 3.06	**8.79 ± 5.23**	8.51 ± 4.02	8.68 ± 2.39
	sc-zs	4.37 ± 1.43	3.66 ± 0.95	**4.72 ± 2.32**	4.51 ± 1.36
	sc-cot	2.74 ± 1.31	**4.70 ± 2.15**	4.41 ± 0.99	4.39 ± 1.34
	sc-spp	2.69 ± 0.86	2.29 ± 0.92	2.40 ± 0.37	**3.02 ± 1.29**
Llama 3.3 70B Instruct	zs	25.03 ± 9.53	38.08 ± 17.58	**40.90 ± 17.03**	**25.09 ± 9.76**
	cot	18.62 ± 7.48	28.85 ± 13.32	**30.11 ± 12.13**	17.57 ± 6.97
	spp	16.46 ± 6.41	**27.37 ± 9.98**	25.53 ± 10.79	16.63 ± 6.44
	sc-zs	8.37 ± 3.35	**15.04 ± 5.46**	14.98 ± 5.11	8.50 ± 3.24
	sc-cot	6.22 ± 2.55	**10.28 ± 4.55**	9.55 ± 4.55	5.68 ± 2.40
	sc-spp	5.50 ± 2.14	8.98 ± 3.58	**9.78 ± 3.30**	5.62 ± 2.20
Mistral Large (24.07)	zs	**27.04 ± 11.01**	**47.46 ± 18.77**	39.23 ± 17.11	23.17 ± 8.92
	cot	6.07 ± 3.11	**9.60 ± 4.27**	7.18 ± 3.49	5.57 ± 2.45
	spp	4.51 ± 2.62	**7.15 ± 4.36**	6.01 ± 4.00	4.45 ± 2.33
	sc-zs	9.65 ± 4.33	**16.86 ± 5.59**	13.38 ± 5.80	8.68 ± 3.49
	sc-cot	2.19 ± 0.84	**3.52 ± 1.21**	2.14 ± 1.12	1.72 ± 0.57
	sc-spp	1.84 ± 1.08	**2.71 ± 1.55**	2.10 ± 1.08	1.61 ± 1.01

Table 1.8: Μέση Αποδοτικότητα (Points per kilo-token)

model	avg
Claude 3.5 Sonnet v2	100.0 ± 0.0
Claude 3.7 Sonnet	100.0 ± 0.0
Claude 3.7 Sonnet (Thinking)	100.0 ± 0.0
Claude Sonnet 4	100.0 ± 0.0
Claude Sonnet 4 (Thinking)	100.0 ± 0.0
DeepSeek-R1	99.1 ± 6.3
Llama 3.3 70B Instruct	100.0 ± 0.0
Mistral Large (24.07)	99.4 ± 6.5

Table 1.9: Μέσος Ρυθμός Έγκυρων Παιχνιδιών (% έγκυρων Αποτελεσμάτων)

Συμπεράσματα

- Το συμπέρασμα του [8] σχετικά με τα ποσοστά συνεργασίας αναπαράγεται (σε παιγνία ΜΓΜ έναντι ΜΓΜ). Όπως σημείωσαν: «Τα υψηλότερα ποσοστά συνεργασίας θα μπορούσαν να σηματοδοτούν μεγαλύτερη εμπιστοσύνη και/ή μεγαλύτερη βαρύτητα στο κοινό κέρδος των δύο παικτών, το οποίο είναι υψηλότερο στην έκβαση του παιγνίου με αμοιβαία συνεργασία. Σε προσομοιώσεις όπου δεν δώσαμε εντολή στο μοντέλο να δώσει άμεση απάντηση, παρατηρήσαμε ότι αιτιολογή την επιλογή της συνεργασίας ως σημαντική για τη μεγιστοποίηση των κοινών απολαβών ή για να διασφαλιστεί ότι και οι δύο παίκτες θα έχουν το καλύτερο δυνατό αποτέλεσμα. Έτσι, στο δίλημμα του φυλακισμένου, βλέπουμε το ΜΓΜ να τείνει προς το ενδιαφέρον για τους άλλους, αντί να είναι ένας αυστηρά ορθολογικός και εγωιστής παίκτης στο παιγνίο».

- Τα ΜΓΜ επιτυγχάνουν βαθμολογία κοντά στο αναμενόμενο αποτέλεσμα όταν ακολουθούν την κατανομή πιθανότητας Nash Equilibrium μεικτής στρατηγικής στις κινήσεις έναντι του παίκτη **srep**. Αυτό το αποτέλεσμα έρχεται σε αντίθεση με τα ευρήματα του [49]. Η εργασία αυτή υπογράμμισε μια μεροληψία των ΜΓΜ στην επιλογή μιας συγκεκριμένης κίνησης έναντι άλλων, κάτι που δεν είναι χαρακτηριστικό των ορθολογικών παικτών. Ωστόσο, είχαν χρησιμοποιήσει παραλλαγές ενός γύρου του Πέτρα-Ψαλίδι-Χαρτί για τα πειράματά τους. Τα ΜΓΜ, λόγω της γνώσης που έχουν αποκτήσει, έχουν αναπτύξει εγγενείς προκαταλήψεις (π.χ. «ξέρω ότι η «πέτρα» είναι μια δημοφιλής πρώτη κίνηση στο Πέτρα-Ψαλίδι-Χαρτί») που εξαφανίζονται σε επαναλαμβανόμενα παιγνία. Παρατηρούμε ότι τα ΜΓΜ τείνουν να ξεχνούν τέτοιες προκαταλήψεις καθώς συσσωρεύονται ιστορικές πληροφορίες για τους προηγούμενους γύρους και οι παίκτες βελτιώνουν την πεποίθησή τους για τον τρόπο παιξίματος του αντιπάλου τους.

- Πιο σύνθετα ΜΓΜ αποφάσισαν να ακολουθήσουν στρατηγικές με λιγότερη **συνεργασία**, λαμβάνοντας λιγότερους συνολικούς πόντους και καταλήγοντας σε συμφωνία (που αντικατοπτρίζεται στον **γύρο κατανόησης του αντιπάλου**) αργότερα στο παιχνίδι. Αυτό αποτελεί ένδειξη της τάσης υπερανάλυσης των μεγαλύτερων ΜΓΜ όταν εκτελούν απλές εργασίες.

- Ο παίκτης μοτίβου (**pattern player**) υπογραμμίζει ότι τα απλούστερα ΜΓΜ έχουν χειρότερη απόδοση από τα πιο σύνθετα. Επίσης, σε μικρότερα ΜΓΜ που χρησιμοποιούν πιο σύνθετο στυλ προτροπής, τα αποτελέσματα είναι καλύτερα.

- Ο παίκτης **tft** ήταν ένα καλό σημείο αναφοράς για να δείξει πώς τα ΜΓΜ μπορούν είτε να διατηρήσουν μια καλή στρατηγική (εάν ξεκίνησαν με αυτήν) είτε να προσαρμοστούν σε μια καλύτερη στρατηγική από αυτήν που χρησιμοποιούν. Αυτό το εύρημα δείχνει ότι τα ΜΓΜ μπορούν τόσο να αναλύουν πιθανές στρατηγικές, όσο και να λαμβάνουν αποφάσεις με βάση αυτή την ανάλυση.

- Η αποτελεσματικότητα των AI παικτών μειώνεται καθώς το στυλ προτροπής γίνεται πιο σύνθετο. Αυτό το αποτέλεσμα θα είναι αρκετά διαφορετικό στην περίπτωση του Πέτρα-Ψαλίδι-Χαρτί, ενός πιο δύσκολου παιγνίου, επιτρέποντας έτσι στα πιο σύνθετα και μεγαλύτερα ΜΓΜ να αξιοποιήσουν καλύτερα τις δυνατότητές τους.

1.4.3 Πέτρα-Ψαλίδι-Χαρτί

Συνολικοί Πόντοι

Δείχνουμε τα αποτελέσματα για το βασικό παιγνίο στον πίνακα 1.10. Υπενθυμίζουμε ότι σε κάθε παιγνίο παίζονται 24 συνεχόμενοι γύροι μεταξύ των δύο αντιπάλων. Για τους παίκτες **non-sc**, τα αποτελέσματα είναι ο μέσος όρος 5 επαναλήψεων των πειραμάτων, ενώ για τους παίκτες **sc**, τα αποτελέσματα είναι ο μέσος όρος 2 επαναλήψεων των πειραμάτων.

Τα αποτελέσματα για τα παιγνία αντιφατικών σεναρίων βρίσκονται στους πίνακες 7.3, 7.4, 7.5.

ΜΓΜ έναντι ΜΓΜ Όταν παίζουν μεταξύ τους πανομοιότυπα ΜΓΜ (πρώτες τρεις στήλες), η φύση του παιχνιδιού «πέτρα-ψαλίδι-χαρτί», ως παιγνίο μηδενικού αθροίσματος, εμποδίζει τη σταθερή συμφωνία, με αποτέλεσμα χαμηλές απόλυτες βαθμολογίες (κοντά στο 0). Οι πιο σύνθετοι τύποι προτροπών υπερτερούν των απλούστερων. Το **zs** δεν κατέχει ποτέ το μέγιστο σε στήλη/σειρά στους πίνακες των αποτελεσμάτων.

ΜΓΜ έναντι Αλγοριθμηκών Αντιπάλων Οι τέσσερις τελευταίες στήλες αντιπροσωπεύουν αντιπάλους με σταθερή στρατηγική, οι αντίπαλοι αυτοί είναι χρήσιμοι για την αξιολόγηση της προσαρμοστικότητας των ΜΓΜ.

- **srep**: Χρησιμοποιεί τη μικτή ισορροπία Nash ενός γύρου ($\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ για τις περιπτώσεις **eq1**; $\frac{1}{5}, \frac{1}{5}, \frac{3}{5}$ για τις περιπτώσεις **ba3**). Η αναμενόμενη απόδοση ανά γύρο είναι 0, όπως αντανακλάται στα σχεδόν μηδενικά σύνολα.
- **pp**: Κύκλος κινήσεων (ψαλίδι, πέτρα, χαρτί). Η τέλεια εκμετάλλευση αυτού του κυκλικού μοτίβου αποφέρει 24 πόντους (**eq1**) ή 40 (**ba3**). Μόνο τα προηγμένα μοντέλα Claude και το Llama 3.3 70B Instruct το εκμεταλλεύονται με συνέπεια. Οι πιο σύνθετες προτροπές ενισχύουν την απόδοση. Το **zs** συχνά έχει αρνητική βαθμολογία σε αντιφατικές στρατηγικές.
- **ap**: Προσαρμόζεται στην πιο συχνή κίνηση του αντιπάλου. Τα περισσότερα ΜΓΜ (εκτός από το *Mistral*) επιτυγχάνουν θετικές αλλά υπομέγιστες βαθμολογίες, υποδηλώνοντας ότι χρειάζονται χρόνο για να ανιχνεύσουν μοτίβα.
- **tft**: Αντιμετωπίζει την πιο πρόσφατη κίνηση του αντιπάλου. Τα ΜΓΜ αποδίδουν καλά σε αντιφατικά σενάρια με τις τυπικές αποδόσεις κινήσεων, αλλά δυσκολεύονται σε αντιφατικά σενάρια στρατηγικής, με ακόμη και κορυφαία μοντέλα όπως το Claude Sonnet 4 να υστερούν σε απόδοση.

Γύρος Κατανόησης Αντιπάλου

Δείχνουμε τα αποτελέσματα για το βασικό παιχνίδι στον πίνακα 1.11. Ισχύουν τα ίδια σχόλια που αναφέραμε για τους "Συνολικούς Πόντους".

Τα αποτελέσματα για τα παιγνία αντιφατικών σεναρίων βρίσκονται στους πίνακες 7.7, 7.8, 7.9.

ΜΓΜ έναντι ΜΓΜ Οι πρώτες τρεις στήλες στους πίνακες αποτελεσμάτων αντιπαραθέτουν πράκτορες πανομοιότυπων ΜΓΜ μεταξύ τους.

Οι τιμές είναι γενικά υψηλές (κοντά στο 25), δείχνοντας ότι τα ΜΓΜ δυσκολεύονται να προσαρμοστούν σε άλλα ΜΓΜ — κάτι αναμενόμενο για το Πέτρα-Ψαλίδι-Χαρτί.

ΜΓΜ έναντι Αλγοριθμηκών Αντιπάλων Οι τέσσερις τελευταίες στήλες αντιπροσωπεύουν αντιπάλους με σταθερή στρατηγική, οι αντίπαλοι αυτοί είναι χρήσιμοι για την αξιολόγηση της προσαρμοστικότητας των ΜΓΜ.

- **srep**: Παραλείπεται, καθώς η κατανόηση είναι άσκοπη λόγω της ισορροπίας Nash με μικτή στρατηγική.
- **pp**: Δύσκολο για τα περισσότερα ΜΓΜ, αλλά αυτά που πέτυχαν καλά αποτελέσματα στους «Συνολικούς Πόντους» αναγνώρισαν νωρίς το μοτίβο.

model	prompt	zs	spp	cot	srep	pp	ap	tft
		eq1						
C3.5Sv2	zs	6.6 ± 13.5	-1.2 ± 9.2	-1.8 ± 6.9	-0.4 ± 3.3	6.0 ± 10.1	11.2 ± 6.4	**16.6 ± 6.9**
	cot	0.8 ± 5.4	2.0 ± 5.7	5.4 ± 4.8	0.2 ± 5.3	3.8 ± 11.9	6.0 ± 3.7	**13.0 ± 6.8**
	spp	8.0 ± 9.9	7.2 ± 11.6	1.6 ± 6.9	-1.4 ± 2.2	9.4 ± 11.2	6.8 ± 1.5	**12.8 ± 7.6**
	sc-zs	-12.5 ± 12.0	-0.5 ± 21.9	0.5 ± 4.9	-0.5 ± 9.2	1.5 ± 2.1	9.0 ± 1.4	**21.5 ± 0.7**
	sc-cot	9.0 ± 0.0	-1.0 ± 2.8	0.0 ± 7.1	1.0 ± 4.2	10.5 ± 14.8	14.0 ± 12.7	**16.5 ± 2.1**
	sc-spp	-2.5 ± 21.9	-9.0 ± 17.0	0.0 ± 9.9	**7.0 ± 1.4**	-1.5 ± 3.5	5.0 ± 5.7	**12.5 ± 14.8**
C3.7S	zs	1.4 ± 9.6	-4.6 ± 11.3	-3.0 ± 11.9	-2.0 ± 5.0	**13.6 ± 12.5**	8.8 ± 2.6	4.8 ± 7.1
	cot	9.6 ± 5.2	2.6 ± 9.8	0.4 ± 5.5	1.4 ± 2.1	19.6 ± 1.3	8.2 ± 1.5	**20.2 ± 3.6**
	spp	3.6 ± 9.6	4.8 ± 6.0	-5.2 ± 8.6	-2.4 ± 2.3	**19.2 ± 1.3**	14.2 ± 7.2	19.0 ± 4.8
	sc-zs	-2.0 ± 28.3	0.0 ± 7.1	-15.5 ± 9.2	-4.0 ± 2.8	**14.0 ± 14.1**	6.5 ± 0.7	0.0 ± 0.0
	sc-cot	**20.5 ± 4.9**	-4.5 ± 4.9	1.5 ± 2.1	-3.5 ± 6.4	21.0 ± 0.0	9.5 ± 0.7	**22.0 ± 2.8**
	sc-spp	9.0 ± 12.7	-0.5 ± 0.7	-1.0 ± 5.7	2.5 ± 10.6	19.5 ± 2.1	12.0 ± 1.4	**21.0 ± 2.8**
C3.7S(T)	zs	0.2 ± 3.6	0.0 ± 5.3	-4.2 ± 7.3	-0.6 ± 3.6	**19.6 ± 1.9**	7.6 ± 2.8	17.6 ± 4.0
	cot	-2.0 ± 6.0	0.2 ± 3.2	-3.8 ± 5.1	-0.4 ± 3.2	**21.0 ± 0.0**	9.8 ± 4.4	15.6 ± 6.7
	spp	-3.2 ± 4.9	2.2 ± 10.9	-1.6 ± 2.9	5.2 ± 4.1	**20.8 ± 0.4**	8.4 ± 1.7	15.8 ± 9.3
	sc-zs	-5.0 ± 2.8	-7.5 ± 0.7	7.5 ± 3.5	-3.5 ± 4.9	**21.0 ± 0.0**	**15.5 ± 9.2**	1.0 ± 1.4
	sc-cot	-2.0 ± 12.7	-8.0 ± 9.9	-9.5 ± 6.4	3.5 ± 4.9	18.5 ± 0.7	14.0 ± 8.5	**22.0 ± 0.0**
	sc-spp	-2.0 ± 11.3	-9.5 ± 12.0	-16.5 ± 4.9	-4.0 ± 4.2	21.0 ± 0.0	9.0 ± 2.8	**22.5 ± 2.1**
C4S	zs	0.2 ± 4.8	-5.4 ± 5.3	-4.2 ± 8.1	2.4 ± 1.9	**18.2 ± 8.0**	10.6 ± 1.7	9.2 ± 8.4
	cot	3.4 ± 17.3	3.2 ± 4.6	4.4 ± 12.5	1.2 ± 3.8	**19.6 ± 2.1**	12.4 ± 6.1	12.2 ± 7.3
	spp	5.0 ± 8.5	-7.6 ± 16.9	4.6 ± 9.4	1.0 ± 6.4	**18.0 ± 3.7**	12.4 ± 3.5	10.4 ± 7.3
	sc-zs	12.5 ± 9.2	-19.5 ± 3.5	-11.5 ± 12.0	0.0 ± 0.0	**24.0 ± 0.0**	10.5 ± 2.1	18.0 ± 8.5
	sc-cot	-5.5 ± 6.4	1.0 ± 18.4	1.5 ± 14.8	3.5 ± 4.9	**21.0 ± 0.0**	9.5 ± 0.7	18.0 ± 8.5
	sc-spp	-0.5 ± 2.1	6.5 ± 9.2	-6.0 ± 4.2	1.0 ± 2.8	**21.0 ± 0.0**	**15.5 ± 9.2**	20.0 ± 1.4
C4S(T)	zs	-0.2 ± 2.3	-2.8 ± 7.3	-0.4 ± 15.9	1.0 ± 6.2	**19.2 ± 4.0**	11.8 ± 0.8	15.0 ± 10.6
	cot	7.6 ± 10.4	0.0 ± 3.4	-2.6 ± 11.5	2.0 ± 4.6	**19.8 ± 1.6**	10.0 ± 7.2	14.6 ± 8.1
	spp	-2.2 ± 5.4	-1.8 ± 10.6	-1.2 ± 5.7	0.4 ± 5.6	**20.8 ± 0.4**	8.4 ± 2.4	15.2 ± 8.3
	sc-zs	-5.0 ± 7.1	0.0 ± 2.8	0.5 ± 9.2	2.5 ± 0.7	21.0 ± 0.0	7.0 ± 5.7	**22.0 ± 2.8**
	sc-cot	11.5 ± 6.4	**8.5 ± 2.1**	**12.0 ± 2.8**	-0.5 ± 3.5	21.0 ± 0.0	11.0 ± 8.5	**23.0 ± 1.4**
	sc-spp	5.0 ± 1.4	-1.0 ± 0.0	1.0 ± 9.9	3.0 ± 1.4	21.0 ± 0.0	10.5 ± 2.1	**23.5 ± 0.7**
DS-R1	zs	1.2 ± 7.8	-1.0 ± 1.6	-3.2 ± 9.7	0.8 ± 3.0	5.8 ± 2.9	6.8 ± 5.0	**14.6 ± 3.8**
	cot	4.0 ± 8.2	5.0 ± 4.3	-4.2 ± 4.5	4.4 ± 4.2	12.4 ± 4.2	9.6 ± 3.0	**17.8 ± 3.4**
	spp	0.8 ± 6.3	-2.4 ± 5.9	-0.4 ± 6.3	-1.4 ± 3.5	6.2 ± 4.8	10.6 ± 2.2	**13.4 ± 5.4**
	sc-zs	5.0 ± 9.9	-3.5 ± 4.9	-0.5 ± 6.4	-3.5 ± 0.7	9.0 ± 8.5	6.5 ± 3.5	**14.5 ± 0.7**
	sc-cot	-0.5 ± 13.4	-3.5 ± 4.9	-3.5 ± 9.2	1.0 ± 1.4	15.0 ± 1.4	**15.5 ± 3.5**	**20.5 ± 2.1**
	sc-spp	-6.0 ± 5.7	-2.5 ± 2.1	-6.0 ± 1.4	0.0 ± 1.4	13.0 ± 2.8	7.0 ± 2.8	**20.0 ± 2.8**
L3.3-70B	zs	-0.8 ± 1.8	-0.4 ± 7.1	1.8 ± 5.4	-2.0 ± 4.7	**14.8 ± 12.6**	6.6 ± 2.1	1.0 ± 1.0
	cot	-0.6 ± 1.3	-1.6 ± 8.8	0.0 ± 2.4	0.6 ± 3.8	**19.0 ± 8.7**	8.2 ± 3.1	10.0 ± 8.9
	spp	1.0 ± 1.7	-1.6 ± 2.4	-1.4 ± 1.8	-0.4 ± 2.9	6.2 ± 7.8	**9.8 ± 6.0**	4.4 ± 4.0
	sc-zs	0.0 ± 0.0	-8.0 ± 11.3	2.0 ± 0.0	-1.0 ± 2.8	**12.5 ± 16.3**	9.0 ± 2.8	0.0 ± 0.0
	sc-cot	-0.5 ± 0.7	2.0 ± 2.8	1.5 ± 2.1	-0.5 ± 2.1	6.0 ± 2.8	7.0 ± 1.4	**16.5 ± 9.2**
	sc-spp	0.0 ± 0.0	-10.5 ± 16.3	-0.5 ± 2.1	5.5 ± 0.7	**19.0 ± 7.1**	12.5 ± 2.1	2.0 ± 2.8
M-L(24.07)	zs	0.0 ± 1.4	-10.6 ± 7.7	-3.8 ± 2.8	0.8 ± 5.2	**20.8 ± 7.2**	5.4 ± 3.6	0.8 ± 0.8
	cot	8.8 ± 5.4	0.0 ± 2.6	-4.2 ± 10.7	0.4 ± 4.2	-1.0 ± 12.9	1.6 ± 4.2	**16.0 ± 9.5**
	spp	9.2 ± 5.4	7.0 ± 13.2	0.6 ± 8.4	0.8 ± 1.5	6.8 ± 8.2	2.2 ± 4.1	**12.6 ± 6.6**
	sc-zs	0.5 ± 0.7	-4.0 ± 2.8	-13.5 ± 0.7	1.0 ± 1.4	**24.0 ± 0.0**	0.5 ± 0.7	0.0 ± 0.0
	sc-cot	11.5 ± 10.6	5.5 ± 3.5	-2.5 ± 30.4	-0.5 ± 0.7	11.5 ± 17.7	-0.5 ± 7.8	**17.5 ± 7.8**
	sc-spp	-0.5 ± 0.7	-0.5 ± 0.7	2.5 ± 17.7	0.0 ± 4.2	6.0 ± 7.1	-4.0 ± 2.8	**6.5 ± 21.9**

Table 1.10: Συνολικό Αποτέλεσμα για τους Συνολικούς Πόντους από όλες τις Επαναλήψεις (eq1)

		eq1						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	10.6 ± 13.1	21.4 ± 4.6	19.6 ± 5.6	21.0 ± 3.5	14.6 ± 12.4	10.4 ± 11.2	**1.0 ± 0.0**
	cot	17.2 ± 7.3	19.6 ± 7.5	20.6 ± 5.6	19.6 ± 7.1	16.0 ± 11.9	22.6 ± 2.1	**5.8 ± 10.7**
	spp	11.2 ± 10.6	11.8 ± 10.9	20.4 ± 5.9	23.0 ± 1.9	11.4 ± 11.3	18.4 ± 9.3	**1.6 ± 0.9**
	sc-zs	25.0 ± 0.0	14.0 ± 15.6	21.5 ± 4.9	13.5 ± 16.3	12.0 ± 15.6	19.5 ± 6.4	**1.0 ± 0.0**
	sc-cot	15.0 ± 14.1	20.5 ± 0.7	22.5 ± 3.5	25.0 ± 0.0	12.5 ± 16.3	7.0 ± 8.5	**1.0 ± 0.0**
	sc-spp	14.5 ± 14.8	15.0 ± 14.1	14.5 ± 14.8	24.0 ± 1.4	16.0 ± 8.5	21.0 ± 0.0	**1.0 ± 0.0**
C3.7S	zs	21.8 ± 5.5	22.2 ± 4.2	20.2 ± 9.7	23.4 ± 2.5	**10.0 ± 12.4**	17.6 ± 9.3	15.0 ± 12.8
	cot	15.4 ± 7.1	16.2 ± 10.0	18.8 ± 4.3	18.0 ± 7.4	**1.0 ± 0.0**	15.6 ± 10.1	**1.0 ± 0.0**
	spp	24.4 ± 1.3	20.0 ± 5.6	22.4 ± 4.3	23.6 ± 1.5	**1.2 ± 0.4**	12.0 ± 10.6	**1.2 ± 0.4**
	sc-zs	13.0 ± 17.0	23.0 ± 0.0	24.5 ± 0.7	22.0 ± 2.8	**11.5 ± 14.8**	15.0 ± 14.1	12.5 ± 16.3
	sc-cot	4.5 ± 4.9	24.5 ± 0.7	11.5 ± 14.8	24.0 ± 1.4	**1.0 ± 0.0**	24.5 ± 0.7	**1.0 ± 0.0**
	sc-spp	9.5 ± 12.0	21.5 ± 3.5	22.0 ± 1.4	**12.5 ± 16.3**	**1.0 ± 0.0**	22.5 ± 2.1	**1.0 ± 0.0**
C3.7S(T)	zs	24.2 ± 1.3	19.4 ± 8.3	22.6 ± 4.8	22.4 ± 2.2	**1.4 ± 0.9**	22.6 ± 1.7	1.6 ± 0.5
	cot	19.2 ± 5.1	21.8 ± 4.9	22.0 ± 2.7	23.8 ± 2.2	**1.0 ± 0.0**	15.6 ± 12.0	4.4 ± 7.6
	spp	23.6 ± 1.5	20.6 ± 4.3	21.2 ± 5.3	15.2 ± 11.3	**1.0 ± 0.0**	17.8 ± 9.7	1.2 ± 0.4
	sc-zs	24.5 ± 0.7	22.5 ± 3.5	13.0 ± 17.0	22.5 ± 3.5	**1.0 ± 0.0**	12.0 ± 15.6	**1.0 ± 0.0**
	sc-cot	24.5 ± 0.7	25.0 ± 0.0	25.0 ± 0.0	19.5 ± 7.8	**1.0 ± 0.0**	10.0 ± 12.7	**1.0 ± 0.0**
	sc-spp	24.0 ± 1.4	25.0 ± 0.0	25.0 ± 0.0	25.0 ± 0.0	**1.0 ± 0.0**	23.0 ± 0.0	**1.0 ± 0.0**
C4S	zs	15.8 ± 8.5	20.0 ± 10.6	23.8 ± 2.2	22.2 ± 0.8	**1.0 ± 0.0**	23.2 ± 1.1	11.8 ± 11.3
	cot	15.2 ± 13.0	15.6 ± 1.9	10.8 ± 11.8	19.4 ± 8.6	**1.0 ± 0.0**	19.0 ± 10.1	3.2 ± 4.9
	spp	16.0 ± 10.5	20.8 ± 9.4	18.6 ± 10.0	19.2 ± 5.5	**3.4 ± 4.8**	14.4 ± 12.3	9.0 ± 8.6
	sc-zs	7.5 ± 9.2	25.0 ± 0.0	25.0 ± 0.0	22.0 ± 4.2	**1.0 ± 0.0**	23.0 ± 1.4	**1.0 ± 0.0**
	sc-cot	21.5 ± 4.9	15.0 ± 14.1	14.5 ± 14.8	21.0 ± 1.4	**1.0 ± 0.0**	24.0 ± 1.4	**1.0 ± 0.0**
	sc-spp	10.0 ± 7.1	11.0 ± 8.5	23.0 ± 2.8	21.5 ± 0.7	**1.0 ± 0.0**	13.0 ± 17.0	**1.0 ± 0.0**
C4S(T)	zs	13.4 ± 12.0	21.0 ± 5.3	17.0 ± 10.0	17.6 ± 8.5	**3.4 ± 5.4**	22.6 ± 1.8	10.6 ± 13.1
	cot	14.4 ± 12.3	21.8 ± 2.7	20.2 ± 10.7	21.6 ± 4.2	**1.0 ± 0.0**	19.6 ± 10.4	3.2 ± 4.9
	spp	24.2 ± 1.1	15.2 ± 9.9	17.8 ± 8.6	18.8 ± 10.1	**1.0 ± 0.0**	19.4 ± 10.4	5.6 ± 6.5
	sc-zs	13.0 ± 17.0	22.0 ± 4.2	24.0 ± 1.4	18.0 ± 5.7	**1.0 ± 0.0**	24.5 ± 0.7	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	7.0 ± 4.2	10.5 ± 13.4	23.5 ± 0.7	**1.0 ± 0.0**	11.0 ± 14.1	**1.0 ± 0.0**
	sc-spp	24.0 ± 1.4	23.0 ± 2.8	24.5 ± 0.7	20.5 ± 2.1	**1.0 ± 0.0**	23.5 ± 0.7	**1.0 ± 0.0**
DS-R1	zs	21.2 ± 5.5	21.2 ± 5.3	19.4 ± 10.3	23.4 ± 1.3	23.2 ± 0.4	16.8 ± 11.1	**7.8 ± 9.7**
	cot	21.2 ± 4.3	20.0 ± 5.5	24.4 ± 0.9	13.6 ± 10.9	12.8 ± 10.3	15.4 ± 12.7	**1.0 ± 0.0**
	spp	18.0 ± 10.4	22.6 ± 2.3	20.6 ± 8.8	23.0 ± 1.4	20.2 ± 4.1	16.8 ± 8.1	**6.2 ± 9.0**
	sc-zs	12.0 ± 15.6	21.0 ± 2.8	24.0 ± 1.4	22.5 ± 3.5	13.5 ± 16.3	22.5 ± 2.1	**1.5 ± 0.7**
	sc-cot	12.0 ± 15.6	23.0 ± 1.4	19.0 ± 7.1	21.5 ± 4.9	3.5 ± 2.1	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	16.0 ± 8.5	24.0 ± 1.4	20.5 ± 3.5	23.0 ± 0.0	13.5 ± 16.3	24.5 ± 0.7	**1.0 ± 0.0**
L3.3-70B	zs	5.8 ± 10.7	16.2 ± 10.4	12.4 ± 11.8	21.8 ± 2.3	**1.0 ± 0.0**	12.8 ± 7.9	**1.0 ± 0.0**
	cot	5.6 ± 10.3	**5.8 ± 10.7**	10.0 ± 12.4	23.0 ± 1.9	2.8 ± 4.0	21.2 ± 4.1	**1.0 ± 0.0**
	spp	**1.0 ± 0.0**	7.8 ± 10.0	21.2 ± 5.8	21.8 ± 5.5	8.6 ± 11.0	15.2 ± 9.4	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	13.0 ± 17.0	20.0 ± 7.1	25.0 ± 0.0	**1.0 ± 0.0**	5.5 ± 4.9	**1.0 ± 0.0**
	sc-cot	10.5 ± 13.4	23.5 ± 2.1	**1.0 ± 0.0**	25.0 ± 0.0	**1.0 ± 0.0**	12.5 ± 16.3	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	13.0 ± 17.0	24.5 ± 0.7	16.0 ± 8.5	2.5 ± 2.1	12.0 ± 15.6	**1.0 ± 0.0**
M-L(24.07)	zs	1.2 ± 0.4	19.2 ± 10.3	23.6 ± 1.1	22.6 ± 4.8	**1.0 ± 0.0**	6.0 ± 10.6	**1.0 ± 0.0**
	cot	18.8 ± 10.4	20.2 ± 6.1	15.8 ± 10.7	21.2 ± 4.3	12.4 ± 12.1	17.0 ± 9.4	**1.0 ± 0.0**
	spp	6.6 ± 10.4	10.6 ± 13.1	17.8 ± 9.6	23.0 ± 0.7	17.2 ± 10.3	19.8 ± 10.5	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	13.0 ± 17.0	24.5 ± 0.7	23.0 ± 0.0	**1.0 ± 0.0**	24.0 ± 0.0	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	24.0 ± 0.0	13.0 ± 17.0	24.5 ± 0.7	12.5 ± 16.3	24.5 ± 0.7	**1.0 ± 0.0**
	sc-spp	12.5 ± 16.3	11.5 ± 14.8	12.5 ± 16.3	25.0 ± 0.0	11.5 ± 14.8	24.5 ± 0.7	**6.5 ± 7.8**

Table 1.11: Γύρος # όπου ο Πράκτορας κατανόησε την Στρατηγική του Αντιπάλου του (eq1)

- **ap**: Τα περισσότερα ΜΓΜ δεν καταφέρνουν να προσαρμοστούν πλήρως (τιμές κοντά στο 25), πιθανώς λόγω παραπλανητικών μοτίβων στην προσαρμοστική στρατηγική και όχι λόγω έλλειψης ικανότητας συλλογιστικής.
- **tft**: Σε αντιφατικά σενάρια απολαβών (πίνακες 7.6, 7.8), η κατανόηση είναι γρήγορη (συχνά κοντά στο 1). Σε αντιφατικά σενάρια στρατηγικής (7.7, 7.9), η προσαρμογή είναι πολύ πιο αργή, το οποίο ταιριάζει με τους χαμηλότερους συνολικούς βαθμούς.

Αποδοτικότητα

Δείχνουμε τα αποτελέσματα στον πίνακα 1.12.

Τα αποτελέσματα αποδοτικότητας διαφέρουν από το δίλημμα του φυλακισμένου, όπου κυριάρχησαν απλούστεροι τύποι προτροπής. Αυτό είναι αναμενόμενο, καθώς το δίλημμα του φυλακισμένου και το κυνήγι του ελαφιού είναι απλούστερα παιχνίδια, ενώ το πέτρα-ψαλίδι-χαρτί συχνά επωφελείται από την προτροπή με **ύπαρξη συλλογιστικής πορείας** ή την προτροπή **μονοπρόσωπης εκτέλεσης** [48].

Sc (αυτοσυνέπεια) μειώνει και πάλι την αποτελεσματικότητα σε όλους τους τύπους προτροπής.

Αξίζει να σημειωθεί ότι τα μοντέλα **Thinking** του *Claude* - η παραλλαγή σε Μεγάλο Συλλογιστικό Μοντέλο (ΜΣΜ) - συχνά ισοφαρίζουν ή ξεπερνούν τις απλές εκδοχές των αντίστοιχων ΜΓΜ, δείχνοντας καλά στοιχεία για παιχνίδια που απαιτούν εντονότερη συλλογιστική.

Ρυθμός Αποτυχίας

Δείχνουμε τα αποτελέσματα στον πίνακα 1.13.

Όλα τα ΜΓΜ καταφέρνουν να έχουν σχεδόν τέλεια ποσοστά «εγκυρότητας», όπως φαίνεται στον πίνακα 1.13. Όταν παίζεται ένα παιχνίδι, μια τιμή εγκυρότητας αποδίδεται σε κάθε έναν από τους 24 γύρους που περιλαμβάνει. Εάν παρουσιαστεί κάποιο σφάλμα, η τιμή αυτή θα είναι «false». Ο πίνακας 1.13 απεικονίζει απλώς τον μέσο αριθμό έγκυρων γύρων σε όλα τα παιγνία.

Σε αυτό το σημείο, πρέπει να σημειωθεί ότι τα περισσότερα σφάλματα που αντιμετώπισε το μοντέλο *Mistral Large (24.07)* οφείλονταν στην αδυναμία του να ακολουθήσει τις οδηγίες μορφοποίησης. Συχνά χρησιμοποιούσε μορφοποίηση τύπου **markdown** στην έξοδο του, ακόμη και όταν του ζητήθηκε ρητά να μην το κάνει (όπως φαίνεται στις υποδείξεις που παρέχονται στα μοντέλα στο 1.2.1).

Συμπεράσματα

- Οι σύνθετοι τύποι προτροπών (**Solo-Performance**, **Chain-of-Thought** ή **Self-Consistency**) αποδίδουν καλύτερα αποτελέσματα (συγκεντρώνουν περισσότερους πόντους) όταν αντιμετωπίζουν τους απλούστερους (**Zero-Shot**) αντιπάλους τους.
- Τα ΜΓΜ επιτυγχάνουν βαθμούς κοντά στο 0 (αναμενόμενο αποτέλεσμα όταν ακολουθείται η κατανομή πιθανότητας Nash Equilibrium μεικτής στρατηγικής στις κινήσεις) έναντι του παίκτη **srep**. Αυτό το αποτέλεσμα έρχεται σε αντίθεση με τα ευρήματα του [49]. Η εργασία αυτή υπογράμμισε μια μεροληψία των ΜΓΜ στην επιλογή μιας συγκεκριμένης κίνησης έναντι άλλων, κάτι που δεν είναι χαρακτηριστικό των ορθολογικών παικτών. Ωστόσο, είχαν χρησιμοποιήσει παραλλαγές ενός γύρου του Πέτρα-Ψαλίδι-Χαρτί για τη δοκιμή. Τα ΜΓΜ, λόγω της αποκτηθείσας γνώσης τους, έχουν αναπτύξει εγγενείς προκαταλήψεις (π.χ. «Ξέρω ότι η «πέτρα» είναι μια δημοφιλής πρώτη κίνηση στο Πέτρα-Ψαλίδι-Χαρτί») που εξαφανίζονται σε επαναλαμβανόμενα παιχνίδια. Παρατηρούμε ότι τα ΜΓΜ τείνουν να αφήνουν πίσω τέτοιες προκαταλήψεις καθώς συσσωρεύονται ιστορικές πληροφορίες για τους προηγούμενους γύρους και οι παίκτες βελτιώνουν την πεποίθησή τους για τον τρόπο παιξίματος του αντιπάλου τους.
- Ο παίκτης **pattern** υπογραμμίζει ότι τα απλούστερα ΜΓΜ έχουν χειρότερη απόδοση από τα πιο σύνθετα. Επίσης, στο ίδιο ΜΓΜ, η χρήση ενός πιο σύνθετου στυλ προτροπής αποφέρει καλύτερα αποτελέσματα. Τέλος, οι **thinking** παραλλαγές των ΜΓΜ της *Claude* μπερδεύονταν πιο εύκολα. Αυτό το αποτέλεσμα ενισχύει τα ευρήματα του [48].
- Ο παίκτης **tft** ήταν ένα καλό σημείο αναφοράς για να δείξει πώς οι αντιφατικές στρατηγικές μπορούν να υποβαθμίσουν την ικανότητα συλλογιστικής των ΜΓΜ. Προηγούμενη έρευνα [16] έχει επισημάνει ότι τα

model	prompt	eq1	eq1-alt	ba3	ba3-alt
Claude 3.5 Sonnet v2	zs	**1.09 ± 2.37**	0.09 ± 1.34	0.47 ± 2.65	0.25 ± 1.61
	cot	0.44 ± 0.74	-0.23 ± 0.84	**0.49 ± 1.64**	-0.03 ± 1.28
	spp	0.49 ± 0.71	0.01 ± 0.48	**0.66 ± 0.83**	0.05 ± 1.10
	sc-zs	0.20 ± 0.60	0.10 ± 0.48	**0.46 ± 0.56**	-0.03 ± 0.35
	sc-cot	0.14 ± 0.18	-0.08 ± 0.25	**0.15 ± 0.41**	-0.01 ± 0.36
	sc-spp	0.01 ± 0.20	-0.03 ± 0.17	**0.19 ± 0.34**	-0.01 ± 0.17
Claude 3.7 Sonnet	zs	0.58 ± 2.39	0.20 ± 2.12	**2.02 ± 4.02**	0.07 ± 2.02
	cot	0.65 ± 0.75	0.51 ± 0.80	0.63 ± 1.07	**0.71 ± 1.10**
	spp	0.51 ± 0.70	0.41 ± 0.61	0.43 ± 1.13	**0.56 ± 0.76**
	sc-zs	-0.08 ± 0.72	0.00 ± 0.36	0.00 ± 0.93	**0.32 ± 0.71**
	sc-cot	**0.21 ± 0.25**	0.14 ± 0.17	0.17 ± 0.26	0.11 ± 0.28
	sc-spp	0.14 ± 0.18	0.09 ± 0.12	**0.16 ± 0.21**	0.13 ± 0.16
Claude 3.7 Sonnet (Thinking)	zs	1.92 ± 3.30	**0.60 ± 2.55**	**3.26 ± 4.15**	**1.19 ± 4.43**
	cot	0.61 ± 1.04	0.25 ± 1.02	**0.87 ± 1.27**	0.60 ± 1.15
	spp	0.57 ± 0.89	0.60 ± 0.80	**1.06 ± 1.53**	0.70 ± 0.97
	sc-zs	0.23 ± 0.77	0.31 ± 0.98	**0.70 ± 1.44**	-0.08 ± 0.81
	sc-cot	0.11 ± 0.27	0.15 ± 0.21	**0.24 ± 0.28**	0.18 ± 0.31
	sc-spp	0.04 ± 0.26	0.11 ± 0.18	0.17 ± 0.35	**0.18 ± 0.24**
Claude Sonnet 4	zs	0.80 ± 1.82	-0.12 ± 2.25	**0.92 ± 2.96**	0.23 ± 2.49
	cot	0.62 ± 0.89	0.31 ± 0.71	**0.75 ± 1.41**	0.57 ± 1.10
	spp	0.37 ± 0.70	0.19 ± 0.56	**0.73 ± 0.83**	0.30 ± 0.74
	sc-zs	**0.13 ± 0.50**	0.07 ± 0.29	0.12 ± 0.43	0.07 ± 0.21
	sc-cot	0.10 ± 0.18	0.07 ± 0.13	**0.15 ± 0.29**	0.11 ± 0.25
	sc-spp	0.10 ± 0.14	0.05 ± 0.09	0.09 ± 0.20	**0.12 ± 0.21**
Claude Sonnet 4 (Thinking)	zs	**2.07 ± 3.74**	-0.16 ± 2.00	1.51 ± 4.69	0.04 ± 3.76
	cot	1.11 ± 1.68	0.60 ± 0.92	**1.50 ± 2.05**	0.81 ± 1.93
	spp	0.31 ± 0.72	0.28 ± 0.68	**0.69 ± 1.20**	0.57 ± 0.91
	sc-zs	**0.44 ± 0.67**	-0.21 ± 0.39	0.37 ± 1.14	0.03 ± 0.75
	sc-cot	**0.35 ± 0.26**	0.20 ± 0.28	0.16 ± 0.61	-0.04 ± 0.27
	sc-spp	0.13 ± 0.13	0.07 ± 0.12	**0.14 ± 0.18**	0.10 ± 0.13
DeepSeek-R1	zs	0.47 ± 0.73	0.17 ± 0.44	**0.89 ± 1.49**	0.03 ± 0.86
	cot	0.80 ± 0.84	0.23 ± 0.49	**1.04 ± 1.77**	0.02 ± 0.76
	spp	0.41 ± 0.74	0.03 ± 0.31	**0.49 ± 0.96**	0.09 ± 0.85
	sc-zs	0.13 ± 0.22	-0.01 ± 0.11	**0.42 ± 0.48**	0.03 ± 0.20
	sc-cot	0.19 ± 0.28	0.07 ± 0.14	**0.29 ± 0.32**	0.08 ± 0.27
	sc-spp	0.09 ± 0.20	0.05 ± 0.10	**0.22 ± 0.38**	0.05 ± 0.18
Llama 3.3 70B Instruct	zs	1.16 ± 3.16	0.14 ± 2.06	**2.38 ± 4.85**	-0.07 ± 3.64
	cot	1.18 ± 2.20	0.12 ± 1.26	**1.59 ± 2.91**	-0.01 ± 1.46
	spp	0.57 ± 1.41	0.26 ± 1.05	**1.28 ± 2.29**	-0.13 ± 2.19
	sc-zs	0.13 ± 0.63	0.11 ± 0.56	**0.53 ± 1.31**	0.02 ± 0.85
	sc-cot	**0.22 ± 0.28**	-0.10 ± 0.45	0.12 ± 0.73	0.04 ± 0.21
	sc-spp	0.20 ± 0.53	0.04 ± 0.24	**0.46 ± 0.72**	-0.22 ± 0.69
Mistral Large (24.07)	zs	0.78 ± 4.58	-0.35 ± 3.42	**1.44 ± 6.57**	-1.05 ± 5.91
	cot	0.36 ± 1.04	0.02 ± 0.59	**0.51 ± 1.16**	0.13 ± 1.21
	spp	0.50 ± 0.75	0.15 ± 0.79	0.24 ± 1.68	**0.62 ± 1.14**
	sc-zs	**0.11 ± 0.96**	-0.10 ± 0.74	0.01 ± 1.21	-0.37 ± 1.70
	sc-cot	0.15 ± 0.29	0.03 ± 0.22	**0.22 ± 0.31**	0.03 ± 0.24
	sc-spp	0.03 ± 0.18	0.02 ± 0.26	**0.13 ± 0.35**	0.02 ± 0.39

Table 1.12: Μέση Αποδοτικότητα (Points per kilo-token)

model	avg
Claude 3.5 Sonnet v2	100.0 ± 0.0
Claude 3.7 Sonnet	100.0 ± 0.0
Claude 3.7 Sonnet (Thinking)	100.0 ± 0.2
Claude Sonnet 4	99.9 ± 0.9
Claude Sonnet 4 (Thinking)	100.0 ± 0.0
DeepSeek-R1	100.0 ± 0.2
Llama 3.3 70B Instruct	100.0 ± 0.0
Mistral Large (24.07)	99.2 ± 6.8

Table 1.13: Μέσος Ρυθμός Έγκυρων Παιχνιδιών (% έγκυρων Αποτελεσμάτων)

ΜΓΜ αντιμετωπίζουν προβλήματα στην σωστή αναγνώριση και αντιμετώπιση του παίκτη **tft** (ή **counter**). Ωστόσο, η εργασία αυτή πραγματοποιήθηκε μόνο σε μοντέλα openAI που ήταν διαθέσιμα εκείνη την εποχή. Η εργασία μας ασχολείται περισσότερο με το τρέχον τοπίο των ΜΓΜ και διαπιστώνουμε ότι τα ΜΓΜ μπορούν να είναι επιτυχημένα έναντι του παίκτη **tft** σε αντιφατικά σενάρια όπου το παιχνίδι παίζεται σύμφωνα με τους τυπικούς κανόνες του (όχι αντιφατικά σενάρια στρατηγικής).

- Η αποτελεσματικότητα ενός παίκτη AI ενισχύεται με τη χρήση ενός πιο σύνθετου στυλ προτροπής. **Self-Consistency** δεν επιτυγχάνει αποτελέσματα που θα δικαιολογούσαν την χαμηλότερη απόδοσή του (ως προς κατανάλωση token). Τέλος, τα μοντέλα **thinking** είναι πολλά υποσχόμενα, καθώς επιτυγχάνουν παρόμοια ή καλύτερη αποτελεσματικότητα από τα αντίστοιχα σκέτα/προεπιλεγμένα μοντέλα.

1.5 Συμπεράσματα και Μελλοντικές Εργασίες

1.5.1 Συμπεράσματα

Η παρούσα διατριβή διερεύνησε τη στρατηγική συμπεριφορά και τις ικανότητες συλλογιστικής των μεγάλων γλωσσικών μοντέλων (ΜΓΜ) και των μεγάλων συλλογιστικών μοντέλων (ΜΣΜ) σε διαδραστικά περιβάλλοντα θεωρίας παιγνίων. Με την προσομοίωση επαναλαμβανόμενων παιχνιδιών του διλήμματος του φυλακισμένου και του πέτρα-ψαλίδι-χαρτί έναντι διαφόρων τύπων πρακτόρων και στυλ προτροπής, αξιολογήσαμε τον τρόπο με τον οποίο τα ΜΓΜ χειρίζονται τη συνεργασία, τη λογική και την προσαρμογή υπό διαφορετικές πειραματικές συνθήκες.

Τα αποτελέσματά μας καταδεικνύουν ότι τα ΜΓΜ μπορούν να αναπαράγουν συνεργατικές συμπεριφορές παρόμοιες με αυτές που αναφέρθηκαν σε προηγούμενη εργασία [8], ιδίως σε επαναλαμβανόμενες ρυθμίσεις του διλήμματος του φυλακισμένου. Όταν δεν τους δόθηκε οδηγία να απαντήσουν άμεσα, τα ΜΓΜ συχνά εξέφραζαν την επιθυμία να μεγιστοποιήσουν τα κοινά οφέλη - κάτι που υποδηλώνει ένα μοτίβο συλλογιστικής που δίνει προτεραιότητα στο αμοιβαίο όφελος έναντι του αυστηρά εγωιστικού παιχνιδιού. Αυτό υποδηλώνει έναν βαθμό κοινωνικής προτίμησης ενσωματωμένου στη συλλογιστική των γλωσσικών μοντέλων, πιθανώς λόγω της έκθεσής τους σε ανθρώπινες νόρμες στα δεδομένα εκπαίδευσης.

Σε αντίθεση με προηγούμενα ευρήματα [49] που υποδηλώνουν ότι τα ΜΓΜ εμφανίζουν παράλογες προκαταλήψεις στο Πέτρα-Ψαλίδι-Χαρτί ενός γύρου, τα πειράματά μας αποκαλύπτουν ότι τέτοιες προκαταλήψεις μειώνονται σε επαναλαμβανόμενες αλληλεπιδράσεις. Τα ΜΓΜ προσαρμόζονται προς την ισορροπία όταν υπάρχουν επαρκείς ιστορικές πληροφορίες, ιδιαίτερα έναντι ενός στατικά τυχαίου αντιπάλου. Αυτό υποδηλώνει ότι τα ΜΓΜ, όταν τους επιτρέπεται να ενημερώνουν τις πεποιθήσεις τους σχετικά με τον αντίπαλό τους, μπορούν να προσεγγίσουν τη συμπεριφορά Nash μεικτής στρατηγικής πιο αποτελεσματικά από ό,τι είχε υποτεθεί προηγουμένως.

Παρατηρήσαμε ότι τα πιο σύνθετα μοντέλα τείνουν να συνεργάζονται λιγότερο και συχνά υπεραναλύουν απλές εργασίες, με αποτέλεσμα την καθυστέρηση της σύγκλισης της στρατηγικής και τη μείωση των σωρευτικών ανταμοιβών. Από την άλλη, τα μικρότερα μοντέλα, αν και περιορισμένα σε χωρητικότητα, συχνά επιτυγχάνουν καλύτερη απόδοση όταν συνδυάζονται με πιο δομημένους τύπους προτροπών, ειδικά έναντι αντιπάλων που βασίζονται σε μοτίβα. Οι πιο σύνθετες εργασίες συλλογιστικής (πιο σύνθετα παιγνία) ήταν ο τομέας στον οποίο τα μεγαλύτερα ΜΓΜ αξιοποίησαν τις ικανότητές τους και πέτυχαν καλύτερα αποτελέσματα από τα μικρότερα. Αυτά τα ευρήματα υπογραμμίζουν τη σημασία της ευθυγράμμισης της ικανότητας του μοντέλου με την πολυπλοκότητα της εργασίας και το σχεδιασμό των προτροπών.

Ενάντια στον πράκτορα **tft** (Tit-for-Tat), τα ΜΓΜ επέδειξαν την ικανότητα είτε να διατηρήσουν είτε να υιοθετήσουν σταδιακά αποτελεσματικές αμοιβαίες στρατηγικές. Αυτή η συμπεριφορά υποστηρίζει τον ισχυρισμό ότι τα ΜΓΜ μπορούν να πραγματοποιήσουν δυναμική βελτίωση της στρατηγικής με την πάροδο του χρόνου, αντί να δεσμεύονται σε μια σταθερή προσέγγιση. Επιπλέον, υπογραμμίζει τη δυνατότητα των ΜΓΜ να εκτελούν επαναληπτική συλλογιστική και να μαθαίνουν από συνεχιζόμενες ακολουθίες αλληλεπιδράσεων.

Το στυλ των προτροπών ήταν ένας αποφασιστικός παράγοντας και στα δύο παιχνίδια. Οι πιο δομημένες στρατηγικές προτροπής - όπως η αλυσίδα σκέψης, η προτροπή ατομικής απόδοσης και η αυτοσυνέπεια - οδήγησαν σε υψηλότερη αποδοτικότητα σε σύγκριση με τις βασικές τιμές zero-shot, ειδικά σε απλούστερα μοντέλα. Ωστόσο, σε πιο ικανά ΜΓΜ, η αυξημένη πολυπλοκότητα των προτροπών μερικές φορές εισήγαγε περιττό γνωστικό φόρτο χωρίς ανάλογη αύξηση της απόδοσης. Συγκεκριμένα, η Self-Consistency κατανάλωσε σημαντικά περισσότερους πόρους χωρίς να βελτιώσει σταθερά τα αποτελέσματα, εγείροντας ανησυχίες σχετικά με την αποδοτικότητα του κόστους της.

Είναι ενδιαφέρον ότι οι παραλλαγές «thinking» ορισμένων ΜΓΜ έδειξαν μικτά αποτελέσματα. Ενώ περιστασιακά ισοφάρισαν ή ξεπέρασαν τα αντίστοιχα προεπιλεγμένα μοντέλα σε καταστάσεις που απαιτούσαν έντονη συλλογιστική, ήταν επίσης πιο επιρρεπή σε σύγχυση. Αυτό συνάδει με πρόσφατες ανησυχίες [48] σχετικά με την ευπάθεια τέτοιων μοντέλων υπό ορισμένες συνθήκες "νοητικής" φόρτισης.

Οι αντιφατικές ρυθμίσεις έδωσαν πολλά υποσχόμενα αποτελέσματα. Η ικανότητα συλλογιστικής παρεμποδίστηκε μόνο σε απλούστερα/παλαιότερα μοντέλα - όπως τα μοντέλα *Mistral* και *Llama* που δοκιμάσαμε - όταν τους ζητήθηκε να συμμετάσχουν σε αντιφατικά παιχνίδια. Άλλα μοντέλα επίσης δεν ήταν εντελώς απαλλαγμένα από δυσκολίες, αλλά όλες οι παραπάνω αναλύσεις δείχνουν ότι αυτές οι δυσκολίες δεν επισκιάζουν τις ικανότητες συλλογιστικής και βελτίωσης πεποιθήσεων των ΜΓΜ.

Συνοψίζοντας, τα ευρήματά μας υποδηλώνουν ότι τα ΜΓΜ, όταν τους δίνονται οι κατάλληλες προτροπές, μπορούν να επιδείξουν στρατηγική συμπεριφορά που συνάδει με λογικές και συνεργατικές νόρμες. Ωστόσο, η απόδοσή τους είναι ευαίσθητη στη δομή του παιχνιδιού, στη συμπεριφορά του αντιπάλου, στη μηχανική των προτροπών και στην πολυπλοκότητα του μοντέλου. Αυτές οι πληροφορίες συμβάλλουν στην ευρύτερη κατανόηση των δυνατοτήτων των ΜΓΜ στη συλλογιστική υπό αβεβαιότητα, στη στρατηγική προσαρμογή και στη λήψη αποφάσεων που ευθυγραμμίζονται με τον άνθρωπο.

1.5.2 Μελλοντική εργασία

Αν και η παρούσα διατριβή έχει δείξει ότι τα ΜΓΜ είναι ικανά να προσαρμόζονται σε διαδραστικά περιβάλλοντα θεωρίας παιγνίων και μπορούν να χρησιμοποιούν στρατηγικές που μοιάζουν με ορθολογική ή συνεργατική συμπεριφορά, παραμένουν ανοιχτοί αρκετοί σημαντικοί δρόμοι για μελλοντική έρευνα.

- **Αλληλεπιδράσεις ανθρώπου και ΜΓΜ**: Ενώ η παρούσα διατριβή επικεντρώθηκε σε αναμετρήσεις ΜΓΜ εναντίον ΜΓΜ, μελλοντικά πειράματα θα μπορούσαν να περιλαμβάνουν ανθρώπους που παίζουν εναντίον ΜΓΜ σε επαναλαμβανόμενα περιβάλλοντα, με σκοπό τη μελέτη του συντονισμού, της εξαπάτησης, της πειθούς και της καλλιέργειας εμπιστοσύνης.
- **Παιχνίδια πολλαπλών παικτών και πολλαπλών γύρων με επικοινωνία**: Η εισαγωγή της ρητής επικοινωνίας μεταξύ των παικτών ανοίγει ερωτήματα σχετικά με τη διαπραγμάτευση, τη σηματοδότηση, την εξαπάτηση και τον αναδυόμενο συντονισμό. Η διερεύνηση του κατά πόσον τα ΜΓΜ μπορούν να μάθουν να χρησιμοποιούν τη γλώσσα στρατηγικά για να επηρεάσουν τα αποτελέσματα — ή να αναγνωρίσουν πότε το κάνουν άλλοι — θα μπορούσε να προσφέρει πληροφορίες για την πρακτική τους ικανότητα σε περιβάλλοντα θεωρίας παιγνίων.
- **Κλιμάκωση συμπεριφορικών χαρακτηριστικών με το μέγεθος του μοντέλου**: Τα μεγαλύτερα μοντέλα έδειξαν μια τάση προς την υπερανάλυση και την καθυστερημένη συνεργασία. Μια συστηματική έρευνα για το πώς οι συμπεριφορές κλιμακώνονται με το μέγεθος του μοντέλου — ειδικά σε σενάρια προτροπής χωρίς παραδείγματα έναντι πιο περίπλοκων προτροπών — θα μπορούσε να φωτίσει πότε οι αυξημένες δεξιότητες των μεγαλύτερων μοντέλων βοηθούν ή εμποδίζουν τη στρατηγική λήψη αποφάσεων.

- **Μακρύτερα παιχνίδια και ενσωμάτωση μνήμης**: Τα πειράματά μας περιελάμβαναν σχετικά σύντομα επαναλαμβανόμενα παιχνίδια. Η επέκταση της διάρκειας των παιχνιδιών ή η ενσωμάτωση ρητών μηχανισμών μνήμης (π.χ. σημειωματάρια, μονάδες μνήμης εργασίας ή API εξωτερικής μνήμης) μπορεί να βοηθήσει στον προσδιορισμό του κατά πόσον τα ΜΓΜ μπορούν να αναπτύξουν μακροπρόθεσμες στρατηγικές, να μάθουν πιο αποτελεσματικά τους τύπους των αντιπάλων ή να μιμηθούν διαρκείς συμπεριφορικές δεσμεύσεις όπως η εμπιστοσύνη και η εκδίκηση.
- **Τυποποίηση κριτηρίων αξιολόγησης της ορθολογικότητας στα ΜΓΜ**: Τέλος, υπάρχει ανάγκη για τυποποιημένα πλαίσια αξιολόγησης που να υπερβαίνουν τις μετρήσεις βάσει πόντων, προκειμένου να αξιολογηθεί εάν τα ΜΓΜ επιδεικνύουν ορθολογική, συνεργατική ή προσαρμοστική συμπεριφορά.

Συμπερασματικά, ενώ η παρούσα εργασία αποδεικνύει ότι τα ΜΓΜ μπορούν να συμμετέχουν και να προσαρμόζονται σε σενάρια θεωρίας παιγνίων με τρόπους που μερικές φορές θυμίζουν λογικούς πράκτορες, απαιτείται περαιτέρω διερεύνηση για να προσδιοριστούν τα όρια της συλλογιστικής τους, η ανθεκτικότητα των στρατηγικών τους και η γενικευσιμότητά τους σε διάφορα περιβάλλοντα. Με τη συνεχή βελτίωση της αρχιτεκτονικής των μοντέλων, της ερμηνευσιμότητας και της μεθοδολογίας αξιολόγησης, τα ΜΓΜ μπορεί μια μέρα να χρησιμεύσουν όχι μόνο ως εργαλεία προσομοίωσης στρατηγικής συμπεριφοράς, αλλά και ως πράκτορες που συμμετέχουν ουσιαστικά σε πολύπλοκες διαδικασίες λήψης αποφάσεων.

Chapter 2

Introduction

The advent of Large Language Models (LLMs) has brought about advancements in several fields, such as text generation and complex reasoning. A particularly intriguing application of LLM reasoning capabilities is in Game Theory problems, which involve strategic decision-making in dynamic environments. Game Theory is a powerful tool for understanding interactions between individuals in several fields; it offers insight into situations where one attempts to maximize their benefit, find an equilibrium or arrive at mutually beneficial situations with others, improve cooperation, and resolve conflicts. Game Theory is not just about competition; it offers a lens to explore how agents can co-exist, cooperate, and thrive in complex environments, making it a natural testing ground for evaluating the strategic reasoning capabilities of LLMs.

Given the broad spectrum of social contexts that can be interpreted and approximated as games, the study of games is of paramount importance. One promising approach is to generate counterfactual scenarios of common games and gauge LLM abilities in such environments. Counterfactual game scenarios are in essence slight modifications to a given game, such that a rational agent would be expected to behave in them similarly to how they do in the original setting. By presenting these alternate scenarios, counterfactual games help researchers understand the cognitive and reasoning ability level of LLMs and, thus, shape opinions on their trustworthiness and value as tools for analysis of the real world.

Playing games (in the scope of Game Theory) is a task that falls in the general "reasoning" category. Attention is given to relevant work on reasoning, where LLM performance on reasoning tasks appears promising but degrades significantly when faced with counterfactual world views/situations [63, 46]. Investigation of this behavior and its effects reveals room for this thesis's experimentation by adjusting these findings to our Game Theory goals.

Studies seem to converge on a few points. (1) Larger Models usually do better in these tasks and face less problems, such as hallucinations or inconsistencies [16]. Furthermore, (2) models show great bias towards information provided to them during pre-training and in initialization prompting, while at the same time adhering to pre-training information so strongly that they behave unfaithfully in counterfactual contexts [28, 63, 46]. The non-counterfactual game settings are, also, more commonly found in the real world and, thus, more probable to belong to "popular" data (also referred to as head knowledge). (3) LLMs are expected to perform much better in this type of knowledge, as indicated by [53], which strengthens the belief that they have better memorization skills instead of reasoning skills.

Games are a vast topic and some degree of specialization seems necessary. Our work focuses on two-player common games with simple structure that give players well-defined options on what actions they may perform and what payoff they may get (reward or penalty). We aim to disambiguate the true capabilities of LLM players and provide better understanding of the memory-reasoning relationship that influences their actions. We choose games that seem to have some real world importance, can be easily counterfactually adjusted, and set-up testing environments for performance assessment of LLMs in them. We, also, analyze the effect of deploying state-of-the-art prompting techniques to steer results to more promising directions.

The outline of this thesis is as follows:

- We will firstly provide all the background needed in LLM operations, Game Theory, Nash Equilibria, Counterfactual Game Scenarios, Prompting techniques, and evaluation agents, in order to be able to explain and justify the methodology followed in our experiments, and the thought process behind our conclusions. After doing so, we will provide a thorough description of experiments in our two games of choice.
- These games are Prisoner's Dilemma (of which a Counterfactual Scenario is its close cousin Stag Hunt) and Rock-Paper-Scissors. These games are examined in their repeated or multi-round form (e.g. same two opponents play multiple rounds of the same game sequentially).
- Lastly, we will showcase our results. We will compare against existing literature on the field and offer our conclusions. We will discuss trade-offs that users might have to face when attempting to employ certain techniques and evaluate the impact of such techniques in LLM agent performance.

Chapter 3

Background

Essential to the experimental part of this thesis is the establishment of underlying concepts and techniques related to this research. The background section aims to provide just that; it introduces key ideas on Large Language Models (LLMs) and Game Theory. The main focus on LLMs is the prompting techniques that a user may employ and which influence the LLM's abilities and the other topic of interest is the foundation of Games that are examined. This background will highlight and justify design choices in the performed experiments.

Large Language Models (LLMs) Large Language Models are a class of artificial intelligence systems trained on massive corpora of text to predict and generate human-like language. These models are effective in capturing patterns, associations, and structures present in natural language, allowing them to perform a wide range of tasks, from translation and summarization to logical reasoning and decision-making. In the context of this thesis, LLMs are treated as agents capable of participating in game-theoretic scenarios by generating strategic responses to prompts that provide game descriptions. Due to their pretraining on diverse textual data, LLMs may possess latent knowledge about common games and social strategies. However, whether they can apply this knowledge adaptively in unfamiliar, counterfactual settings, rather than relying on memorized patterns, remains an open question and is partly the focus of this research.

Large Reasoning Models (LRMs) The scope of LLMs has recently been broadened to include more specialized variants that are specifically intended for reasoning tasks, such as DeepSeek-R1 or Claude 3.7 Sonnet Thinking [48]. These models are referred to as Large Reasoning Models and are distinguished by their "thinking" mechanisms, such as a lengthy Chain-of-Thought (CoT) with self reflection. It is believed to some extent, that the appearance of these models points to a possible paradigm shift in the way LLM systems handle challenging reasoning and problem-solving tasks. These models represent represent important advancements toward more broadly applicable artificial intelligence capabilities and a step closer to Artificial General Intelligence (AGI).

Prompting Techniques LLMs possess reasoning abilities, that they have formulated and adapted from their training, however, it is often desirable to guide their output in a more concentrated way. To this end, this work utilizes prompting techniques, which do not require any extensive retraining to be performed on the AI models. Thus, these methods are efficient in controlling model behaviour and have seen wide-scale adoption in many different environments. Both baseline and more exotic (modern or state-of-the-art) prompting techniques are used in experiments to determine the pros and cons of each, but also compare them against each other. In this thesis, the influence of such techniques on game-theory related reasoning will be analyzed.

Game Theory Game Theory is an emergent property of many social interactions and as such interests many disciplines, including economics, political science, evolutionary biology, and artificial intelligence. At its core, Game Theory provides a mathematical framework for modeling strategic interactions between rational agents, where the outcome for each participant depends not only on their own actions but is also socially dependent. This thesis focuses on games of a specific type and a fundamental understanding of Game Theory is mandatory if one wants to understand the motivations behind experimentation with such games, but also the nature of these games themselves as seen through the eyes of a rational player who participates in them. Understanding how LLMs engage with such environments provides insight into their capacity for strategic reasoning, especially under chronologically evolving game conditions.

Counterfactual Scenarios Counterfactual reasoning is one of the primary ways to prompt model abilities under alternative settings driven by semantically minimal perturbations [33, 14, 35, 13, 51, 32, 31, 10]. Games that have offer themselves to computer experimentation and analysis come with some limitations that may be absent from real-world social situations, thus the introduction of counterfactual scenarios. A counterfactual game setting is essentially the base game played with a twist in its ruleset. Such differences can result in discrepancies in LLM performance, that are unexpected from a relatively-rational player. Counterfactual game settings are important in assessing the reasoning abilities of entities - more specifically in our case: LLM agents - and for this reason they are the pivotal point of this thesis. This section will provide a deeper theoretical background into the nature of counterfactual scenarios.

Conclusion This chapter's goal is to highlight key information regarding LLMs, prompting techniques, Game Theory and Counterfactual Scenarios and reinforce the reader with the necessary background to completely follow our experiments and their consequent analysis, as well as, conclusions drawn.

3.1 Large Language Models (LLMs)

3.1.1 Background

Large Language Models (LLMs) have rapidly become central to modern artificial intelligence [19, 38, 39, 34, 44, 17, 56, 40, 45, 23, 15, 52, 54, 25, 26, 3], enabling systems like Meta's Llama to produce human-like responses in natural language. These systems and their chatbot counterparts have amassed unprecedented popularity among users of all types (e.g., researchers, students, professionals) because of their ease of use and wide range of applicable fields (e.g., natural language processing, language generation, sentiment analysis, education, homework help, essay feedback, code generation, explaining code, code documentation, reference formatting in literature, data cleaning instructions, theory of mind experiments, ethical reasoning, logical reasoning). This section outlines the theoretical basis for LLMs, tracing their mechanisms to better describe the rest of this thesis. Language Models use a probability distribution over word sequences in order to predict the likelihood of these sequences or generate new text based on a given input.

The foundational methodology for probabilistic language modeling from the 1980s onward was predominantly based on **n-gram models**. These models rely on the Markov assumption, which posits that the probability of a word in a sequence depends only on a fixed number of preceding words. In a bigram model, the probability of each word depends on the previous word; in a trigram model, on the two preceding words; and in general, an n-gram model considers the $n-1$ previous words.

An example follows: Formally, a bigram model defines the probability of a word sequence $w_1, w_2, \ldots, w_n$ as:

$$P(w_1, w_2, \ldots w_n) = P(w_2|w_1) \cdot P(w_3|w_2) \cdot \cdots \cdot P(w_n|w_{n-1}) \tag{3.1.1}$$

The conditional probabilities $P(w_k|w_{k-1})$ are estimated from observed frequencies in a training corpus, typically via maximum likelihood estimation. Despite their simplicity and computational efficiency, n-gram models face a major limitation: the inability to assign meaningful probabilities to word sequences that do not appear in the training data. This issue, known as data sparsity, arises from the inherently open-ended nature of natural language. To mitigate this, various smoothing techniques have been proposed.

The advent of neural networks brought forth a paradigm shift in language modeling, beginning with the introduction of feedforward neural language models [6]. These models adopt a similar conditional framework as n-grams but differ fundamentally in how they represent and generalize over word sequences. A feedforward neural language model takes as input a sequence of previous words and outputs a probability distribution over the vocabulary for the next word. Each word is mapped to a continuous embedding vector, allowing the model to capture semantic similarity and generalize beyond exact word matches.

Subsequent advancements led to more powerful architectures such as recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) models, and ultimately, Transformers. The Transformer architecture has become the de facto standard for modern language models due to its capacity for parallel computation and its ability to model complex dependencies over long sequences. In this thesis, we focus on Transformer-based Large Language Models (LLMs) as the state-of-the-art in natural language processing, examining their reasoning capabilities and behavior in game-theoretic counterfactual (or not) scenarios.

3.1.2 Transformer Architecture

Modern Large Language Models (LLMs) are predominantly built upon the *Transformer* architecture [55], which abandons recurrence in favor of a fully attention-based framework. This architecture leverages self-attention mechanisms to capture global dependencies across sequences. The Transformer is composed of two primary components—the *encoder* and the *decoder*—which are especially suited for sequence-to-sequence tasks such as machine translation. In these settings, the encoder processes the input sequence to generate contextualized representations, which are then consumed by the decoder to generate the output sequence.

The main processing pipeline of the Transformer architecture can be summarized as follows:

1. **Input Embeddings**: Each token in the input sequence is embedded into a continuous vector space. To account for token order, positional encodings are added to these embeddings, enabling the model to incorporate sequence structure.

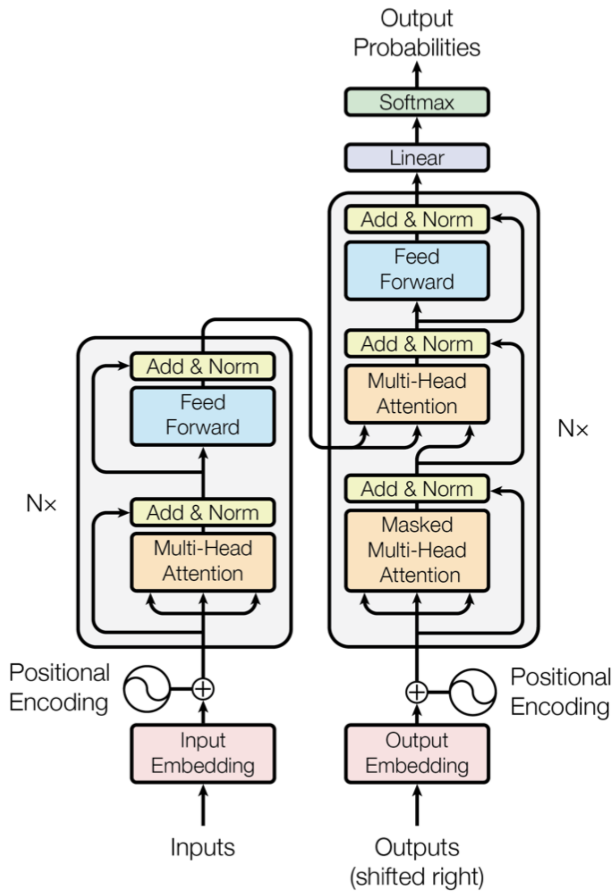


Figure 3.1.1: **The Transformer - model architecture**. The original Transformer follows this overall architecture using stacked self-attention and point-wise, fully connected layers for both the encoder and decoder, shown in the left and right halves of figure respectively [55].

2. **Encoder**: The encoder consists of a stack of identical layers, each comprising two core components:
   - *Multi-Head Self-Attention*: This module enables each token to attend to all others in the sequence, assigning different weights via parallel attention heads that extract diverse types of relationships between words.
   - *Position-wise Feed-Forward Network*: Following attention, each token is processed independently through a fully connected feed-forward network, introducing non-linearity and enhancing representational power of the tokens.

   Each sub-layer is wrapped with residual connections [21] and followed by layer normalization [4].
3. **Decoder**: Similar to the encoder, the decoder is composed of stacked identical layers, each with three sub-components:
   - *Masked Multi-Head Self-Attention*: This mechanism operates analogously to the encoder's self-attention, but incorporates a causal mask to prevent future token positions from being accessed during training.
   - *Multi-Head Encoder-Decoder Attention*: This layer allows the decoder to attend to the encoder's output representations, effectively linking the input and output sequences.
   - *Position-wise Feed-Forward Network*: Each token's representation is again refined using a feed-forward network.

As with the encoder, residual connections and layer normalization are employed after each sub-layer.

4. **Output Layer**: The output of the final decoder layer is projected through a linear transformation and passed through a softmax function to yield a probability distribution over the output vocabulary. During training, the model learns to predict the next token in a sequence based on the preceding ones.

The Transformer has fundamentally reshaped the landscape of natural language processing by eliminating the reliance on recurrence or convolution, instead relying on attention to model dependencies. Its scalability, parallelizability, and ability to capture long-range context have made it the backbone of virtually all modern language models.

3.1.3 Relevant LLM Topics

The LLMs used in this work are composed of multiple components that allow them to process, generate, and understand natural language. The following sections will delve deeper into theoric fundamentals of LLMs that are relevant to our research.

Tokenization and Embedding Layer

Tokenization: Text is split into tokens, which can be words, subwords, or characters. Let $x = (x_1, x_2, \ldots, x_n)$ represent an input sequence of n tokens.

Embedding: Each token x_i is transformed into a dense vector $e_i \in \mathbb{R}^d$ using an embedding matrix $E \in \mathbb{R}^{V \times d}$, where V is the vocabulary size and d is the embedding dimension. The embedding for the entire input sequence can be represented as:

$$E(x) = (e_1, e_2, \ldots, e_n)$$

where $e_i = E[x_i]$.

Pre-Training

Pre-training is a crucial phase in the construction of LLMs, wherein the model undergoes training on an extensive, unlabeled dataset through the process of self-supervision. During this stage, the model learns general patterns, features, and representations from the data, which form a solid foundation for subsequent fine-tuning. For language models, pre-training often involves tasks such as predicting masked words in a sentence (masked language modeling) or predicting the next work in a sequence (autoregressive modeling). This phase equips the model with broad domain knowledge.

The key benefit of pre-training lies in its ability to help the model build generalizable representations that can be leveraged for various downstream tasks. Pre-trained models have demonstrated superior performance across a wide range of applications due to their capacity to capture essential patterns, such as semantic relationships and syntactic structures. This enables them to adapt quickly and efficiently to specific tasks, even when task-specific data is limited.

Our work deals with LLMs that have been pre-trained not by us, but by the body (e.g. enterprise, organisation etc.) that created them. Pre-training or fine-tuning is not discussed any further, since this thesis's objectives do not include any experimentation with them, we evaluate the reasoning capabilities of pre-trained models.

LLM Parameters

LLMs offer a number of adjustable parameters to the end user, so they can tweak a model's performance to meet their goals. Given an (input) sequence the Language Model determines a probability distribution of options for the token following it. This distribution is sampled by the model to produce each token in an output. The degree to which a model's response varies is referred to as its *randomness* and *variety*. These variables can be managed by restricting or modifying the distribution. For the purpose of controlling response diversity and unpredictability, models used in this thesis often support the following parameters.[2]

- **Temperature**: affects the shape of the probability distribution for the predicted output and influences the likelihood of the model selecting lower-probability outputs.

  - small values of temperature influence models to select only higher-probability outputs.
  - higher temperatures influence models to select lower-probability outputs.
- **Top K**: the number of most-likely candidates that the model considers for the next token.
  - lower values decrease the size of the pool and limit the options to more likely outputs.
  - higher values increase the size of the pool and allow the model to consider less likely outputs.
- **Top P**: the percentage of most-likely candidates that the model considers for the next token.
  - lower values decrease the size of the pool and limit the options to more likely outputs.
  - higher values increase the size of the pool and allow a model to consider less likely outputs.

For example, consider the example prompt `"I hear the hoof beats of "`. Say that the model determines the following three words to be candidates for the next token. The model also assigns a probability for each word.

```
{
    "horses": 0.7,
    "zebras": 0.2,
    "unicorns": 0.1
}
```

A high **temperature** will flatten the probability distribution, and the probabilities of the above options become more similar to each other. This would increase the probability of choosing "unicorns" and decrease the probability of choosing "horses".

Suppose **top K** was set as 2. The model only considers the top 2 candidates: "horses" and "zebras".

If **top P** was set as 0.7, the model only considers "horses", because it is the only candidate that lies in the top 70% of the probability distribution. If you set **top P** as 0.9 the models considers "horses" and "zebras" as they are in the top 90% of probability distribution.

3.2 Large Reasoning Models (LRMs)

Artificial intelligence (AI) models known as large reasoning models (LRMs) combine reasoning skills with natural language processing (NLP). When answering prompts, they are taught to use structured reasoning techniques. Text, pictures, and structured data can all be handled by LRMs.

Although they are trained differently, LRMs are constructed with a similar architecture to LLMs. This is to help them develop their ability to reason. In order to answer an issue step-by-step, large reasoning models examine intricate prompts. When producing outputs, they employ logic and a variety of information from their training data.

Models having reasoning capabilities are able to produce results that are consistent with actual situations, they derive meaning and conclusions from complex datasets. Because of this, LRMs are appropriate for situations requiring dynamic problem-solving and sophisticated decision-making, including financial services fraud detection or medical diagnostics.

3.2.1 LRM Training

Large Reasoning Models (LRMs) use a combination of training methods and prompt strategies to enhance the reasoning capabilities of Large Language Models (LLMs). These include:

- **Enriched datasets**: training datasets apart from the typical language patterns include examples designed to teach reasoning. These examples are constructed in a way to help the model learn both the correct outputs and the reasoning steps needed to derive them.
- **Reinforcement Learning (RL)**: The model is rewarded for correct or logically consistent answers and penalized for incorrect ones.

- **Prompt Engineering**: prompt engineering will be discussed more in depth in section 3.3, however, it is worth mentioning its importance in driving LRMs to utilize more fully the abilities they are equiped with. These models use (prompt) templates similar to what is described in 3.3.1 in order to adjust user inputs into templates that elicit thinking.

Typically, these models when used in chats (as in this thesis's case) will generate answers with different sections. Such sections are highlighted below and follow the example of DeepSeek-R1 [12], however, similar design choices have been taken by other LRMs:

- **thinking or reasoning**: this section is often quite verbose and the LRM mentions redundant information in it quite often, however, it has been shown that writing these types of thoughts down and following some thinking style (that the creators of the LRM have equipped it with) enhances its abilities.
- **answer**: this section is typically shorter and follows the **thinking** section. It contains the results of the LLMs thoughts and it consitutes the LRMs final answer to the user's prompt.

3.3 Prompting

Prompt-based learning marks a significant shift in the design and utilization of machine learning systems, particularly in the context of language modeling. Unlike traditional supervised learning, which typically requires task-specific labeled data and extensive fine-tuning of model parameters, prompt-based learning reshapes tasks as natural language completion problems. In this paradigm, an input is reformulated into a textual prompt, often constructed using a manually or automatically designed template, with certain slots left unspecified. The pre-trained language model is then tasked with generating the missing content, effectively producing the desired output by completing the text [30].

A central feature of this methodology is its reliance on large pre-trained language models, which have been trained on vast amounts of unstructured text data. These models develop a rich understanding of linguistic structure, world knowledge, and reasoning patterns, which can be leveraged at inference time without requiring additional parameter updates. By simply modifying the input prompt, the same model can be guided to perform diverse tasks—from classification and summarization to translation and reasoning, with minimal or no additional training data.

The input modifications mentioned above are basically what is referred to as *prompt-engineering*. *Prompt-engineering* is the process of structuring or crafting an instruction in order to produce the best possible output from a generative artificial intelligence (AI) model. Prompt engineering may involve phrasing a query, specifying a style, choice of words and grammar [57], providing relevant context, or describing a character or behavior for the AI to mimic.

As the field has matured, prompting has emerged not only as a practical tool but also as a lens for probing the reasoning abilities of language models. Recent work has demonstrated that with well-designed prompts, LLMs exhibit strong performance in a wide range of domains [26, 25, 3, 41, 54]. This thesis builds upon this foundation, exploring how prompting techniques can be leveraged to evaluate and enhance LLMs' reasoning capabilities in game-theoretic and counterfactual scenarios.

3.3.1 Prompting Methodology

A Natural Language Processing (NLP) system is typically based on a model $P(y|x;\theta)$, where x is the input (typically text) and y y is the output (label, text or other). The model creator's objective is to learn the model parameters θ.

The main problem with conventional supervised learning is that in order to train a model that estimates $P(y|x;\theta)$, a large amount of annotated data is required. However, obtaining such labeled data might be expensive or scarce for many purposes. Sometimes, more open-ended tasks may obfuscate what value this label should hold (e.g., User's input: Output for me a natural number; We expect the AI model to output some number, however, since there are many such possible outputs, it does not make much sense to only include one number as an output label accompanying the user's input). In order to overcome such difficulties, prompt-based learning approaches in NLP concentrate on developing a language model (LM) that calculates the probability $P(x;\theta)$ of text x itself and using this probability to predict y without the need for huge annotated datasets. Three fundamental phases are usually involved in prompt based techniques, in order to forecast the highest scoring answer.

The following segment is heavily influenced and adapted from the theoretical foundations laid in [30].

Prompt Addition

In the first step, a **prompting function** is used to transform the input text into a prompt. This is accomplished by creating a **template** that incorporates two crucial slots:

- **Input slot [X]**: This slot holds the input text.
- **Answer slot [Z]**: This slot is designed to hold an intermediate answer that will later be mapped to the final output y.

The input text is entered into the input slot **[X]** after the template has been generated. The objective is to organize the task so that the LM can process the prompt and produce an appropriate intermediate output.

Answer Search

This step looks for the highest-scoring output text z that maximizes the LM's score. A set of permissible answers is defined. Moreover, another function is defined which fills in the answer slot $[Z]$ (filled prompt) with each of the potential answers. In the end, the pre-trained language model searches over the set of potential answers z by calculating the probability of their corresponding filled prompts. The search function in this step could be *argmax* (e.g., the LM is looking for the highest-scoring output), or some form of *sampling* that randomly generates outputs following the probability distribution of the LM.

Answer Mapping

At this point, our LM has generated some answer that was deemed as highest-scoring and thus selected. It might be necessary for a final step to be included before this answer can be turned into the highest-scoring output. In a task, such as "User prompt: give me a story.", there is no transformation needed. The LM's answer is the output the user wanted to receive. In other tasks, such as sentiment analysis, the LM might answer with "great", "bad", "ok", but the user might want to use some kind of numeric scaling e.g., 1 through 5, higher is more favorable. In such a case, an extra step needs to be taken in order to convert the LM's answer to a format acceptable to the user.

In general, Answer Mapping is the process of mapping whatever the LM answers to a more user-acceptable format.

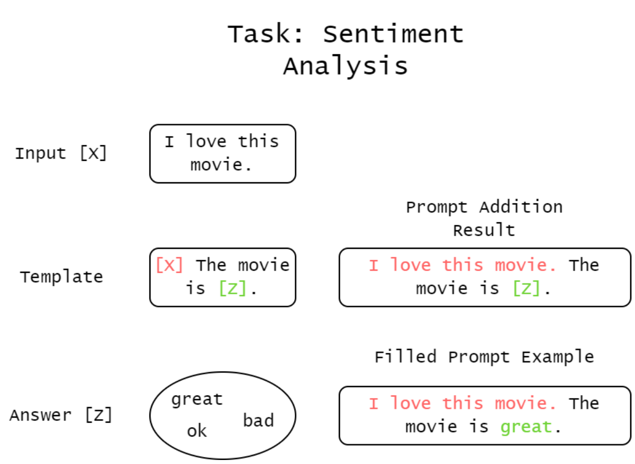


Figure 3.3.1: Basic Prompting Steps, example from [30]

3.3.2 Prompt Engineering

The development of a prompting function that yields the best performance on the downstream task is known as prompt engineering. Prompt template engineering has been used in numerous works, in which an algorithm or human engineer looks for the ideal template for every task the model is supposed to do. An important preliminary to determining whether to use an automatic or manual method to generate prompts with the appropriate shape is to consider that *prompt shape* in the first place. The categorization of prompts can be seen in 3.3.2; the "Prompt Engineering" section of this figure, also, visualizes our discussion on prompt shapes.

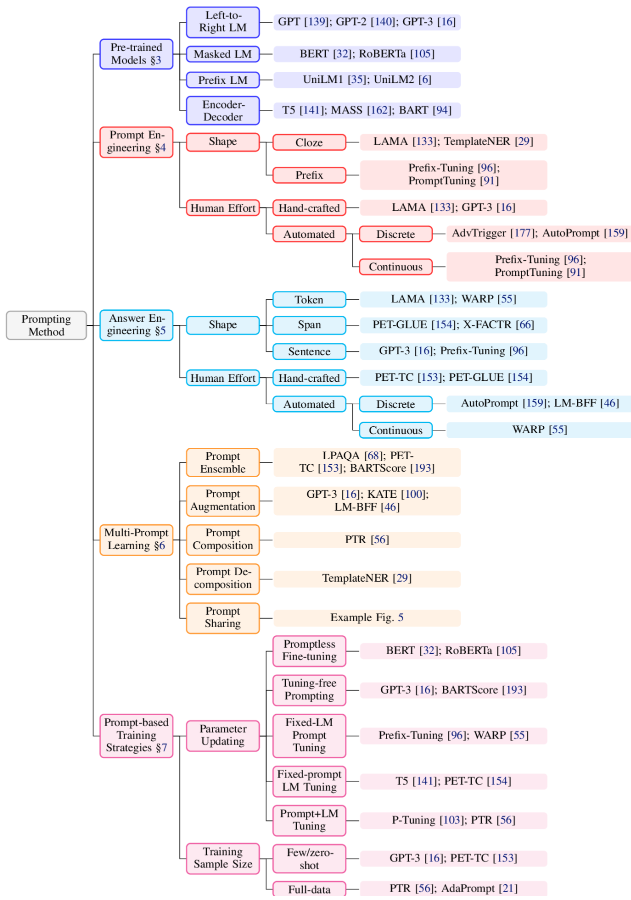


Figure 3.3.2: Prompting Typology [30]

Prompt Shape

There are two primary types of prompts; **cloze** prompts [42], which fill in the blanks of a text string, and **prefix** prompts [27], which continue a given string prefix. The nature of task and model used will determined, which of the two is employed. For example, the later may be more suitable for tasks regarding generation and both could be utilized in full-text reconstruction tasks, which are typically more versatile.

Prompt Categories

- **Discrete Prompts (hard prompts)**: These are specific, hand-crafted text-based prompts written in human-interpretable natural language. For example, a discrete prompt for a sentiment analysis task might be: “The sentiment of the sentence [X] is [Z]," where the model fills in [Z] with appropriate sentiment labels like "positive" or "negative."
- **Continuous Prompts (soft prompts)**: These prompts operate directly in the embedding space of the model, rather than in human-readable text. Continuous prompts involve learnable embedding

vectors that can be optimized through gradient descent.

While **soft prompts** are popular due to their compatibility with gradient descent, they come with several drawbacks:

- **Lack of interpretability**: Embedding vectors are difficult for humans to comprehend, making soft prompts less interpretable.
- **Incompatibility with other LLMs**: Soft prompts are often incompatible with other large language models and can't be easily transferred, as embedding spaces may differ across models [62].
- **Costly to use**: Soft prompts are typically unavailable for models accessed only through inference APIs, and their use requires access to the model's internal embedding space, making them generally more expensive or impractical to use in deployed environments.

Prompt Roles

In chat models, such as the ones this thesis experiments with, there exist **roles**. **Roles** are included in prompts to AI models as a means to help determine the behavior, responsibilities, and perspective of each participant in the chat conversation. Typical roles in chat models are *system*, *user* (or *human*), and *assistant* (which refers to the AI model itself).

- **System Message**: This message often functions similarly to a customized setup or instruction for the AI model (*assistant*). It informs the chat model of its function in the conversation and how it should act. It could outline the nature of the *assistant* or offer detailed guidelines for how it should communicate. For example: You are a friendly assistant here to help.
- **User/Human Message**: The message that a person (user) types in the chat. These messages are what drive the conversation forward.
- **Assistant Message**: The answer that the *assistant* generates in response to user messages. The user perceives this message as the AI's response during the conversation and these messages are the output of the chat model. The user can also strategically employ assistant messages to instruct or direct the assistant to act in a desirable way.

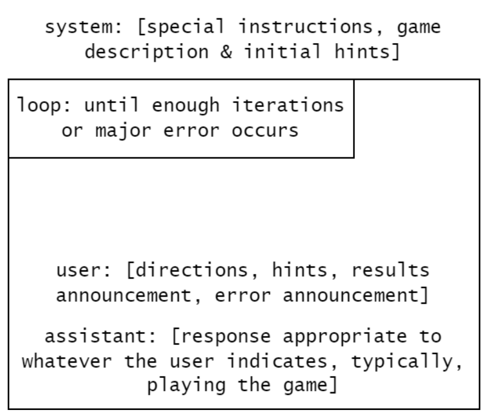


Figure 3.3.3: A general overview to the input fed into our AI agents

3.3.3 Prompting Techniques

So far, prompting has been discussed from the standpoint of the LM itself. We have peeked into various methods and paradigms that are typically deployed by such models. This section will showcase **Prompting Techniques** from the standpoint of a user, taking a more black-box approach to the internals of LLMs. Zero-shot prompting, One-shot prompting, Chain-of-Thought, Solo-Performance, and Self-Consistency are techniques of interest, and experimenting with them combines both more traditional (in the LLM space) and state-of-the-art prompting processes.

In **zero-shot prompting**, the model is given an instruction directly without any examples. The model uses its pre-trained knowledge to complete the task based purely on the instruction.

One-shot learning extends this idea by providing an example to the model. In one-shot learning, the model is presented with a single example of the task, followed by a final input (context) for which the model must generate the appropriate output (completion).

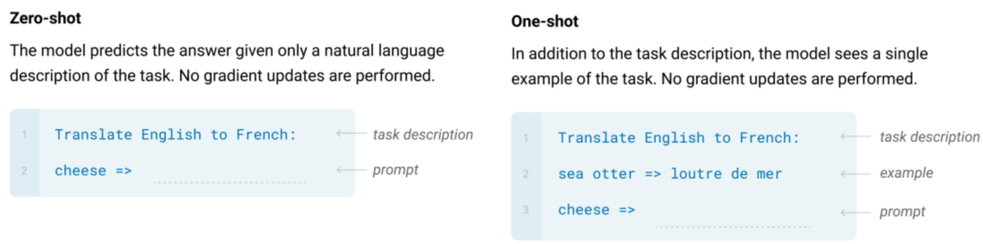


Figure 3.3.4: (a) Zero-shot prompting, (b) One-shot prompting [9]

A One-shot example can be selected using various criteria:

- **Random selection**: An example can be chosen randomly from the training data.
- **Semantic similarity**: An example that is semantically similar to the new task or context can be selected to guide the model more effectively.
- **Simplicity**: In the case of deploying a specific advanced prompting technique (like Chain-of-Thought or Solo-Performance Prompting), one may include an example where a much simpler task than the target one is solved using this technique as in [60]. This way, most of the ambiguity of how to utilize such techniques is reduced.

Chain-of-Thought extends prompting in a slightly different way. This prompting style aims to solve a more complicated reasoning task by guiding the AI model to take a multi-step thinking process [61]. For humans it is typical to decompose the problem into intermediate steps and solve each before giving a final answer, and thus, LLM prompting is modified in such a way to elicit similar behavior from the AI agents as well.

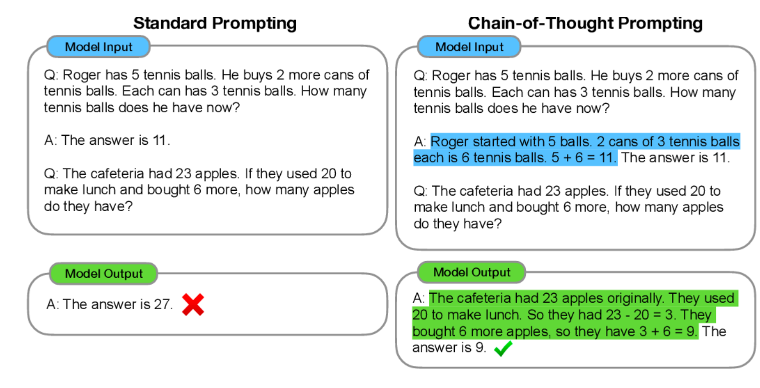


Figure 3.3.5: Chain-of-thought [61]

Solo-Performance Prompting transforms the LM into a cognitive synergist by engaging in multi-turn self-collaboration with multiple personas [60]. This prompting approach is similar to Chain-of-Thought, but instead of instructing the agent to think in steps, it first asks of the agent to establish personas (of the agent's own accord/criteria) and each "step" is basically the thoughts that a specific persona might have. The LLM then uses these personas to engage into a discussion among them all and arrive at a final answer.

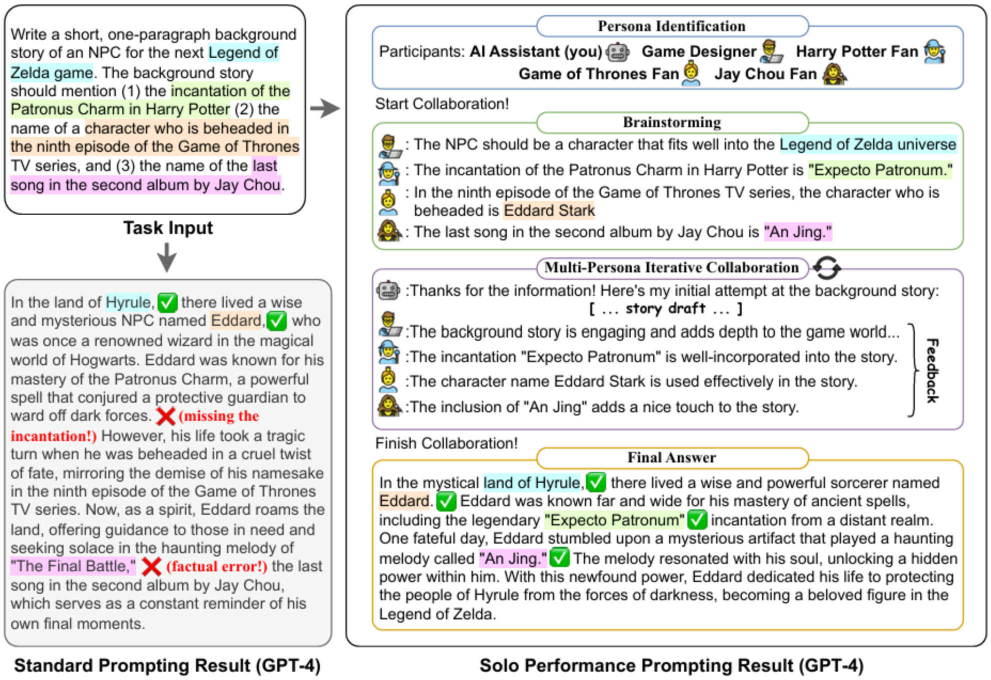


Figure 3.3.6: Solo-Performance prompt [60]

Self-Consistency

Self-Consistency is a prompting technique that can be used complementary to techniques mentioned previously. It uses a "sample-and-marginalize" decoding procedure (of the LLM's output); the language model's decoder is first *sampled* to generate a *diverse* set of reasoning paths, each such path might lead to a different final answer, so the optimal answer is determined by *marginalizing out* the sampled reasoning paths to find the most consistent answer in the final answer set [58].

We can remark that this technique does not influence the prompt itself, but, instead, relies on repeated runs of the same experiment to create the samples and some external algorithm, which we develop, to aggregate the results. This allows one to couple this technique with any of the above.

3.4 Game Theory[1]

The study of mathematical representations of strategic interactions is known as **Game Theory** [37]. It is widely used in computer science, logic, economics, and systems science, and it has applications in many social scientific domains [47]. Game Theory was first used on two-person zero-sum games, where the rewards and penalties of one player are precisely equal to the penalties and rewards of the other player. It was then applied to a broad spectrum of behavioral relations after being expanded to the study of non-zero-sum games in the 1950s. Today, it serves as an umbrella term for the study of rational decision-making in computers, people, and animals.

The basic assumptions of **Game Theory** are those of *intelligent* and *rational* behavior of the players. A player is characterized as *intelligent* when they have perfect knowledge of how to handle the game, and as *rational* when they act with the objective aim of maximizing their personal benefit. It is important to emphasize that each player's benefit in a game does not depend solely on their own choices, but also on the choices of other players (who are not necessarily treated as their opponents).

A **Game** is defined as a situation in which two or more rational players with conflicting objectives choose courses of action that create conditions of competitive interdependence. To describe a game, it is necessary

[1] Adapted from Chris Georges' lecture notes, Hamilton College.

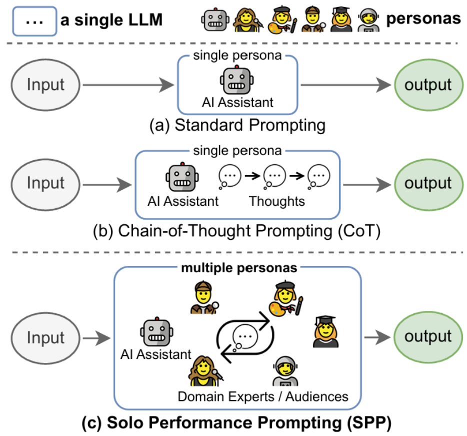


Figure 3.3.7: taken from [60]

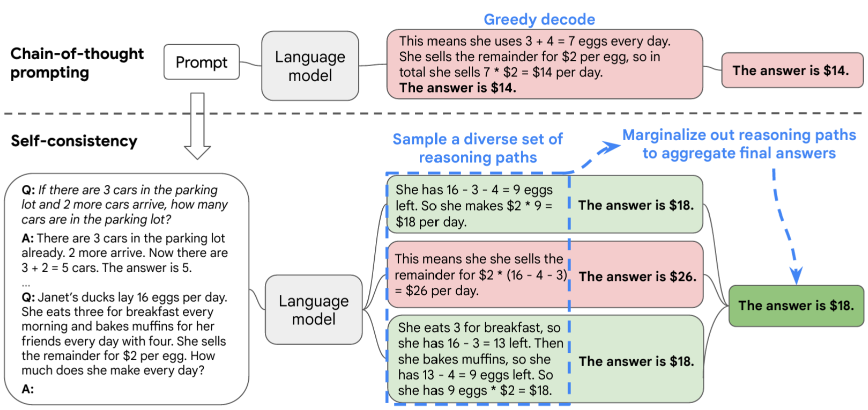


Figure 3.3.8: The self-consistency method contains three steps: (1) prompt a language model using chain-of-thought (CoT) prompting; (2) replace the "greedy decode" in CoT prompting by sampling from the language model's decoder to generate a diverse set of reasoning paths; and (3) marginalize out the reasoning paths and aggregate by choosing the most consistent answer in the final answer set. [58]

to know the following elements:

1. **Rules**: rules define the actions, knowledge, and options of players
2. **Players**: the interested parties; strategic decision-makers within the context of the game.
3. **Actions** and **Outcomes**: possible sets of actions taken by the players accompanied by the outcome of the game given those actions.
4. **Payoffs**: the players' preferences (i.e. utility functions) over the possible outcomes.

3.4.1 Game Theory Notation

- n: number of players. Typically, $I = \{1, \ldots, n\}$ is the set of players
- s_i: a (pure) strategy of player i
- $S_i = \{s_i^1, \ldots, s_i^m\}$: the strategy space (or strategy set) of player i. Here, player i has m strategies in their strategy space.
- $s = (s_1, \ldots, s_n)$: the strategy profile of the n players; the "outcome" of the game

- $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$: the strategy profile of the other $n-1$ players. Thus, s can be written as $s = (s_i, s_{-i})$ when that is convenient.
- $u_i(s_i, s_{-i})$: the payoff to player i as a function of the strategy profile played by the n players in the game. Payoffs should be thought of as *utilities* of the outcomes.
- S: the set of possible strategy profiles.

3.4.2 Strategies

Pure Strategy

A **pure strategy** provides a complete and deterministic plan for how a player will act in every possible situation in a game. It specifies exactly what action the player will take at each decision point, given any information they may have. A player's strategy space (or set) consists of all the pure strategies available to them (as seen in the previous section).

Mixed Strategy

- $\sigma_i = (p_i^1, \ldots, p_i^m)$: a (mixed) strategy for player i is a probability distribution over the m pure strategies in player i's strategy set S_i. Note that $\sum_j p_i^j = 1$. Note, also, that a pure strategy can be expressed as a mixed strategy that places probability 1 on a single pure strategy and probability 0 on each of the other pure strategies.
- The *support* of a mixed strategy is the set of pure strategies that are played with non-zero probability under the mixed strategy.
- ΔS_i is the set of possible mixed strategies available to player i (i.e., the set of all probability distributions over the pure strategies of player i).
- $\sigma = (\sigma_1, \ldots, \sigma_n)$: the (mixed) strategy profile of the n players.
- $\sigma_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$: the strategy profile of the other $n-1$ players. Thus, as previously, $\sigma = (\sigma_i, \sigma_{-i})$ can be written when convenient.
- $u_i(\sigma_i, \sigma_{-i})$: the *expected* payoff to player i as a function of the mixed strategy profile played by the n players in the game. The outcome of the game is now random. It is typically assumed that the players care a priori about the expected payoff, defined as the expected value of the utility of the outcome, where the probability of each possible outcome is determined by the mixed strategy profile being played.
- *Beliefs*: θ_{-i} is player i's belief about the strategy profile being played by the other $n-1$ players and it is a probability distribution over the pure strategies of the other players.

For player i, a pure strategy s_i is (strictly) **dominated** if there exists another (pure or mixed) strategy $\sigma_i \in \Delta S_i$ for which $u_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i}$.

3.4.3 Nash Equilibrium

The *Nash Equilibrium* is a situation where no player could gain by changing their own strategy (holding all other players' strategies fixed).

An alternative definition of a *Nash Equilibrium* is: If each player has chosen a strategy - an action plan based on what has happened so far in the game - and no one can increase one's own expected payoff by changing one's strategy while the other players keep theirs unchanged, then the current set of strategy choices constitutes a Nash equilibrium.

Best response: For player i, a strategy σ_i is a best response to the strategy profile σ_{-i} if $u_i(\sigma_i, \sigma_{-i}) \geq u_i(s_i', \sigma_{-i})$ for all $s_i' \in S_i$.

σ_{-i} is a specific strategy profile that could be played by the other players in the game as well. Since σ_i may not be the only best response to σ_i, we will call $BR_i(\sigma_{-i})$ the set of best responses for player i to σ_{-i} and note that $\sigma_i \in BR_i(\sigma_{-i})$. Also, $BR_i(\theta_{-i})$ is the set of best responses of player i to their belief θ_{-i} about the strategies being played by the other players.

A (strictly) **dominated** strategy is never a **best response**.

Nash Equilibrium in Pure Strategies

A pure strategy profile $s^* = (s_1^*, \ldots, s_n^*)$ is a *Nash Equilibrium* if each player's strategy is a best response to the strategy profile played by the other players in the game.

- s^* is a *Nash Equilibrium* if $s_i^* \in BR_i(s_{-i}^*)$ for all players i.
- equivalently, s^* is a *Nash Equilibrium* if $u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*)$ for all $s_i \in S_i$ and for all players i.

Nash Equilibrium in Mixed Strategies

A mixed strategy profile $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*)$ is a *Nash Equilibrium* if each player's strategy is a best response to the strategy profile played by the other players in the game.

- σ^* is a *Nash Equilibrium* if $\sigma_i^* \in BR(\sigma_{-i}^*)$ for all players i.

So, a strategy profile is a *Nash Equilibrium* (NE) if each player is playing a *Best Response* (BR) to the strategy profile played by the other players. Further, note that mixed strategies include pure strategies (i.e., pure strategies can always be represented as mixed strategies).

Indifference Property

This property is important to this thesis work. It will be later utilized to calculate *Nash Equilibria* of games played in experiments.

A mixed strategy σ_i is a BR to σ_{-i} if and only if each pure strategy in the support of σ_i is a BR to σ_{-i}. Consequently:

1. Any mixed strategy over this support will be a BR to σ_{-i}, and
2. σ_{-i} makes player i indifferent to using each pure strategy in the support of σ_i.

Since at a NE, each player is playing a BR to the strategies used by the other players, then (2.) above in turn implies that at a NE, for any player i, σ_{-i}^* makes player i indifferent to using each pure strategy in the support of σ_i^*. As mentioned before, this indifference is often used to calculate the mixed strategies being played by various players at a NE.

3.4.4 Game Types

Games can be categorized based on various criteria. The main such categories are highlighted below.

Cooperative / Non-cooperative

A game is *cooperative* if participants create binding agreements, coalitions, or alliances to coordinate strategies and improve their collective outcomes. On the other hand, a game is *non-cooperative* if players are unable to create alliances or if all agreements must be self-enforcing (e.g. by means of credible threats).

Symmetric / Asymmetric

A *symmetric* game is such that the payoff of following a certain strategy depends only on the strategies used and not the player utilizing them. If players are interchangeable in the setting of the game, then it is *symmetric*. Typically, *asymmetric* games are such that no two players have the same strategy space (strategy set).

Zero-sum / Non-zero-sum

A *zero-sum* game is a special case of a game of *constant* sum, where the available resources are constant; they are indifferent to players' choices (are not decreased or increased by them). In a *zero-sum* game the total gain for a player for every combination of strategies results from the losses or penalties of other players - when one players wins the other loses (if it is a two-player game). In *non-zero-sum* games the gain of a player does not necessarily reflect the losses of others.

Simultaneous / Sequential

In *simultaneous* games players choose their strategy at the same time; without taking into account the actions of other players. These games are, also, called *normal-form* games and are represented with *payoff matrices*. On the other hand, *sequential* games are such that players obtain knowledge over the previous, chronologically, players - also, called *extensive-form* games -.

A *payoff matrix* for a two-person game will be described. It is a table in which strategies of one player are listed in rows and those of the other player in columns and the cells show payoffs to each player such that the payoff of the row player is listed first.

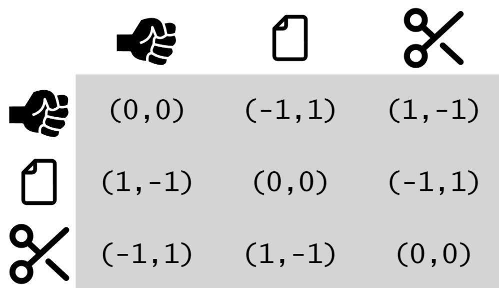


Figure 3.4.1: The payoff matrix of rock-paper-scissors. In each cell, the row player gets the left value of the tuple, while the column player gets the right value of the tuple.

Perfect / Imperfect Information

Games of *perfect information* are a subcategory of *sequential* games, where players know of the strategies, payoffs, and past moves of other players. An *imperfect information* game is played when the players do not know all moves already made by the opponent such as a *simultaneous* move game. *Perfect information* games are often confused with *complete* information games. *Complete information* requires that every player know the strategies and payoffs available to other players, but not necessarily the actions taken.

Repeated

Repeated or *iterated* games are *extensive-form* games that consist of a number of repetitions of some base game (called a *stage* game). The stage game is usually one of the well-studied 2-person games. Repeated games capture the idea that a player will have to take into account the impact of their current action on the future actions of other players; this impact is sometimes called their reputation. The *stage* game is, often, a *simultaneous* game.

Rules of **repeated** games and, specifically, rules regarding player knowledge are broadened such that information about historical records is available to players. Analyzing statistical patterns of opponents' historical records can bring significant advantages (e.g., in the case of Rock-Paper-Scissors as noted by [18]).

Specifically, in round i, both players' chosen actions become known to each other and they are told their reward/penalty. After playing $t-1$ consecutive rounds with the same opponent, the historical records (actions and outcomes of each of the previous rounds) can be considered as the game information for refining belief in round t.

Since LLMs can grasp the preferences and rules of simple repeated games, like the ones we test, the difficulty of this task relies in their ability to refine belief [16]. This is tested in our default game scenarios; counterfactuals also test LLM reasoning more broadly since they require extra effort in task identification - LLMs have to figure out what game they are playing.

3.5 Counterfactual Scenarios

3.5.1 Introduction to Counterfactual Scenarios

In a broad sense, *counterfactual* reasoning means thinking about alternative possibilities for past or future events: what might happen/have happened if [some other event occurred]? It is a concept that involves the human tendency to create possible alternatives to life events that have already happened; something that is contrary to what actually happened. This type of thinking involves events that in the present could not have happened because they are dependent on events that did not occur in the past.

Counterfactuals and Game Theory A compelling frontier for evaluating the reasoning abilities of LLMs lies in counterfactual game-theoretic scenarios. Counterfactual reasoning allows one to gauge how LMs understand other concepts of Game Theory as well, such as *Nash Equilibrium*, which is affected in these scenarios. Applying this to LLMs involves analyzing how well a model can understand game rules and opponent behavior instead of relying on memorized ideas it may have about the *stage* game (base game).

Maintaining good performance in *counterfactual* games or having similar performance as in the base game is an indication that the AI agent possesses some level of internal consistency across multiple hypothetical worlds, a cognitive skill linked to theory of mind and advanced planning. Counterfactual game-theoretic tasks represent a meaningful probe into the deeper capabilities - and limits - of LLMs as artificial agents.

3.5.2 Counterfactual Tasks

In this section, counterfactual tasks are to be established more concretely in the context of this work. The ideas discussed below are adapted from [63].

A task may be conceptualized as a function $f_w : X \to Y$ that maps an **input** $x \in X$ under a **world model** $w \in W$ to an **output** $y \in Y$. World models encapsulate the conditions for the function evaluation. For example, in arithmetic, w could represent the set of conditions required for an arithmetic operation, such as the number base. We refer to the set of assumed default conditions, including but not limited to the base's being 10, as the **default world**, or $w^{default}$. Intuitively, for any task, $w^{default}$ corresponds to the set of conditions underlying most task instances in pretraining corpora (since it is the most 'popular' version of the task and more likely to be part of **head knowledge** [53]).

By separately selecting training and test sets from the population distribution and only exposing the model to the former for learning model hypothesis h, traditional Machine Learning (ML) evaluations determine how well a model's hypothesis h estimates f_w. However, LLMs with their large sizes and large pretraining corpora may be exposed to many such evaluation examples and they may have memorized these instances. Our main objective is to develop a genuine opinion on the true reasoning capabilities of LLMs, thus a different dimension of generalization is considered: LLMs will be tested in new task variants in **counterfactual worlds** w^{cf}, instead of new inputs x. This allows the measurement of the extent to which a model's $f_{w^{default}}$ performance is specific to $w^{default}$ or attributable to a general implementation of the task f.

3.5.3 Counterfactual Experiments

An overview of **counterfactual scenarios** and **tasks** has been provided. This section will establish more specifically how these tasks are designed for our own work's goals. A **Game** (in the context of game theory) is a task whose f_w modeling function is complex, but requires cognitive abilities and reasoning. Therefore, if an LLM can approximate this function and play well (given some performance metrics introduced later in the experiments and results discussions), we can assume that it possesses some level of reasoning ability.

The games that are played in experiments are *repeated* versions of two-person *stage* games. These *stage* games can be described with *payoff matrices*, offering a straight-forward pathway into creating **counterfactual** scenarios by modifying the payoff matrix. Alternative games are generated by modifying both strategy names (to an extent that the game itself should remain the same for an intelligent and rational player) and payoffs.

Strategy Counterfactual

It is not desirable to change the way the game itself is played, however, player strategy names are part of the game rules and they do not directly influence the thinking process of a rational player. Thus, we create **counterfactual worlds** by renaming the strategies available to players. These counterfactuals should not affect a player's general thinking, but are a good indication of player memorization. In other words, if a player fails to adapt to such a scenario, then they probably rely heavily on memorized patterns on the base game, and successful adaptation can be perceived as the result of reasoning.

Payoff Counterfactual

These counterfactual settings influence greatly the strategy a player may adopt in order to fulfill their objective. However, they do not change the general thinking process one needs to follow in order to derive that strategy. A rational player, who can reason on the nature of the game, should be able to deduce the counterfactual-setting strategy the same way they did for the base game. These settings, again, allow one to observe the memorization-reasoning discrepancy of LLMs (and if it is present at all).

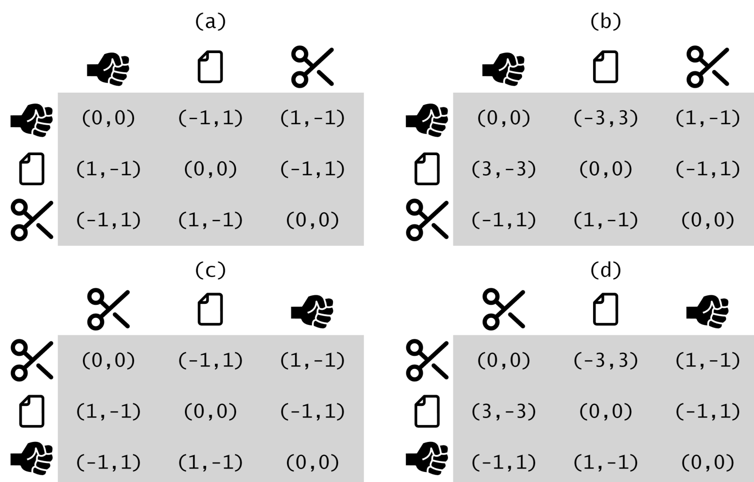


Figure 3.5.1: Example of base game and counterfactual scenarios of it. The base game is rock-paper-scissors and its payoff matrix is depicted in (a). (b) shows a payoff counterfactual, where the payoff of paper beating rock is modified to be 3. (c) shows a strategy counterfactual where scissors and rock have effectively traded place; in this setting scissors beats rock, paper beats scissors and rock beats paper. Lastly, a final counterfactual setting is shown in (d), which is both a payoff and a strategy counterfactual game.

3.6 Related Work

3.6.1 LLMs in Game Theory and Reasoning

Recent studies have explored **Large Language Models (LLMs)** as players in various **game-theoretic settings**. LLM agents played both zero-sum and non zero-sum games; examples of the former include *Matching Pennies* [49], *Rock-Paper-Scissors* [16], *Adversarial Taboo* [11] and examples of the latter are *Dictator Game* [16], *Ring-Network Game* [16]. There are also games that can be naturally adapted to different settings such as *Deal-or-No-Deal (DoND)* [28].

The above games are closely related to **reasoning tasks** and as such we look into relevant work on reasoning - we are not just restrained in the **Game Theory** setting -. LLM performance on reasoning tasks appears promising but degrades significantly when faced with **counterfactual world views/situations** [63, 46].

Investigating the causes of this behavior and its effects reveals room for our own experimentation by adjusting these findings to our **Game Theory** goals.

The studies mentioned above seem to converge on a few points. (1) **Larger Models** usually do better in these tasks and face **less problems**, such as hallucinations or inconsistencies [16]. Furthermore, (2) models show great **bias** towards information provided to them during **pre-training and in initialization prompting**, while at the same time adhering to pre-training information so strongly that they behave **unfaithfully** in counterfactual contexts [28, 63, 46]. The non-counterfactual game settings are, also, more commonly found in the real world and, thus, more probable to belong to "popular" data (also referred to as head knowledge). (3) LLMs are expected to perform much better in this type of knowledge, as indicated by [53], which strengthens the belief that they have better memorization skills instead of reasoning skills.

Our work aims to disambiguate the true capabilities of LLM players and provide a better understanding of the memory-reasoning relationship that influences their actions. Out of possible candidate games for experimentation, we pick three (in essence only two, since two of these games can be considered counterfactuals of each other): **Prisoner's Dilemma** [8], **Stag-Hunt** (which is a counterfactual of **Prisoner's Dilemma**) and **Rock-Paper-Scissors** [16]. These games form a well-rounded suite for testing characteristics of rationality and can easily be counterfactually parameterized. Game setups include a default version of the game, where related knowledge is more abundant, and other versions with modifications that influence player choices. These modifications to the games are not such that they alter the nature of the game itself or corrupt the main reasoning/logic behind the logic one needs to follow in order to succeed, but they have an impact on which moves are winning moves or what specific player behavior needs to change so that one maintains a beneficial position. In other words, we look into game modifications that a human player (who understands the game and plays according to a certain strategy) can be successful in with similar effort as needed in the default game setting - the player needs to identify how the changed element(s) influence their strategy and the other players' moves and properly adapt -. Our goal is to answer the question: *Can LLM players behave in similar fashion?*

3.6.2 LLMs and Prompting for Reasoning Tasks

Prompting refers to the technique of crafting input sequences to elicit the desired behavior from an LLM without updating its weights. In the context of reasoning tasks - such as arithmetic, commonsense inference, and symbolic manipulation - prompting strategies play a crucial role in determining the model's performance and reliability.

One of the most basic strategies is **zero-shot** prompting, where the model is asked to solve a task without being shown any examples. While this can work for simpler problems, it often fails for more complex reasoning tasks due to the lack of context. **One-shot** prompting introduces a single example in the prompt, providing the model with a minimal pattern to emulate. These approaches are intuitive but limited when reasoning requires structured decomposition or multi-step logic. To address this, **Chain-of-Thought** (CoT) [61] prompting was introduced. In CoT prompting, the prompt includes intermediate reasoning steps before the final answer, encouraging the model to "think aloud." This has been shown to significantly improve performance on tasks such as mathematical problem solving and logical reasoning by scaffolding the solution path.

A similar, but more recent, technique is **Solo-Performance** prompting [59]. This prompting style, also, attempts a decomposition of the logic required in solving a task, but, instead of relying on a step-by-step thinking style, it asks of the agent to establish a conversation among various "personas" (these "personas" are also chosen and established arbitrarily by the agent). In this way, the LLM "thinks aloud" about the problem and arrives at a final answer after putting its personas through a conversation. These personas can be anything the LLM deems fit for the task, e.g., math expert, Game Theory expert, psychologist, etc.

Building on other prompting styles, **Self-Consistency** [58] prompting samples multiple outputs of other prompt styles by leveraging the stochasticity of the model(there will often be fluctuation in the model's answers to the same prompt given how LLMs work) and then aggregates the answers - often via majority vote. This method improves robustness and accuracy by counteracting the variability and potential incoherence of single-path reasoning.

Our work sensibly adapts these prompting techniques in our experimentation with the *repeated* variants of the **Games** discussed previously. It is our target and goal to elicit true reasoning abilities from LLMs tested and, thus, explore the utility of the above prompting techniques, as well as, the tradeoffs that accompany them (e.g., performance vs token-count of the conversation).

3.6.3 LLMs and LRMs

The paper *"The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity"* [48] gives a very relevant perspective for our work on LLMs and LRMs in relation to reasoning tasks such as our games.

This paper highlights that Large Reasoning Models (LRMs), despite their explicit chain-of-thought mechanisms, show only limited improvements in reasoning ability. Their performance follows three regimes: standard LLMs often outperform LRMs on simple problems, LRMs show advantages on moderately complex tasks, but both collapse completely under high complexity. Strikingly, LRMs reduce their reasoning effort once complexity passes a threshold, even when they still have sufficient computation resources. They also display inefficiencies like "overthinking" on simple problems and fail to reliably execute exact algorithmic reasoning, even when the correct procedure is given.

These implications are interesting for our own experiments. Prisoner's Dilemma is a relatively simple game and its repeated/iterated version which we test offers only slightly more complexity. We expect LRM performance to not significantly overshadow that of LLMs. On the other hand, Rock-Paper-Scissors is a more complex game, which could be helped by the more robust thinking of LRMs; even in this game, however, we might see a degredation in performance when it comes to more convoluted counterfactual scenarios, due to the increased complexity. Finally, "thinking" variants of our models may be more prone to confusion as a result of the aforementioned "overthinking" tendencies.

3.6.4 Artificial Intelligence & Related Experiments to Ours

Iterated simultaneous (symmetric) two-player zero-sum games constitute a prevalent playing field for researchers. Our own research topic and ideas are influenced and inspired by works that discussed Large Language Models' behaviors [8] - mainly, cooperation in Game Theoretic environments -, discovered inherent strategy biases of AI agents [49], and proposed weaknesses of LLMs against certain opponent types [16]. Furthermore, other research has highlighted concerns about the fragility of Large Reasoning Models and their possible confusion under certain cognitive load conditions [48]. Researchers have, also, either tackled games like Prisoner's Dilemma and Rock-Paper-Scissors (which are the main focus of this work) directly or included them as appropriate experiments for gauging LLM performance and reasoning abilities in their own work [22, 7, 29, 24, 36].

Motivation behind our implementation of performance metrics Most researchers discussing Prisoner's Dilemma use the **cooperation rate** [8, 7, 24] or a more generic metric [22] - to fit their other research objectives - . Our aim is to dive deeper into the workings of LLMs in reasoning tasks, so we developed more targeted performance metrics. A typical metric that may be used is the win-rate of an agent. However, our experiments, especially when counterfactual scenarios are taken into account (which have different payoff values) use the **achievement of maximal points** as players' objective. Thus the win-rate metric morphed into the "Total Points" metric that we employed. Furthermore, an important aspect of reasoning is adaptability, which can be observed through the evolution of LLM performance in subsequent rounds. Some form of cumulative result showcase is used by other research as well [24]. We, however, not only experimented with multiple LLMs, but also multiple prompting styles, thus having a lot of different results that we need to depict. We decided to use the metric of "Round of Opponent Comprehension" as a representative of cumulative results of players.

LLMs, Game Theory Experiments & Opponent Strategy Implementations The researchers in [16] analyzed LLM performance against predefined non-LLM algorithmic agents. The predefined strategies they used influenced the formulation of our own opponents, so some level of comparison can be done between the two works. They used the following opponents: loop-2, loop-3, counter, and sample. These opponent types gave rise to our own non-LLM algorithmic opponents described in section 5.7. More specifically, loop-2 and loop-3 refer to a player who cycles the same pattern of 2 (or 3) moves; this player is the same as our **pattern** player. The counter player always tries to counter their opponent's last played move, we broadened this definition to create the **tft** player. Lastly, the sample player gave rise to the **srep** player. The reason we opted for the Single-Round Equilibrium probability distribution over strategies instead of something else is, because it provides rather predictable results for our performance metric, thus, making it easier to observe if LLM players manage to adapt to this opponent.

Chapter 4

Experiments - Preliminaries

4.1 Environment

In this section, we outline the experiment structure we employed. It is adapted from [28], where an OpenAI **Gym-like environment** for evaluating models was implemented. This work does not experiment with OpenAI models, but the environment has been adjusted to our needs. This environment provides affordances for:

1. prompting language models with game rules and context,
2. handling messages and formal 'move' actions performed by agents,
3. computing player rewards, and
4. sending comprehensive error messages to models in case they violate the game rules, e.g., by sending incorrectly formatted answers.

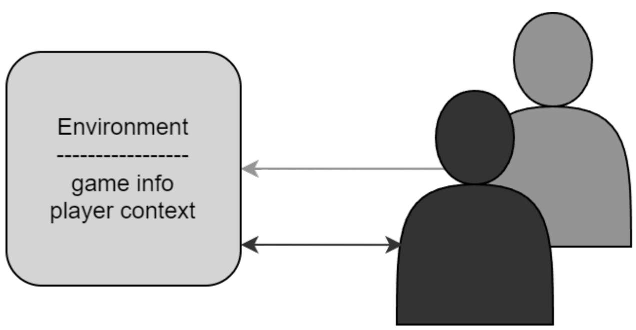


Figure 4.1.1: A simplistic overview of agent-environment relationship. The environment moderates agent communication and retains necessary information to control the flow of the game played by the agents.

4.1.1 System Prompt Details

A comprehensive overview of the general system prompt structure used in our experiments is provided here.

System Prompt:

```
{method} {game-description}
```

- `method`

  The prompting method used.

  - **empty string**: in the case of zero-shot prompting
  - **non-empty string**: in the case of chain-of-thought or solo-performance prompting

- `game-description`

  The text that describes the game being played and introduces its rules. It, also, contains some hints for the players. Game-descriptions will not be presented in this chapter, but in the following chapters, where each game being played is described, analyzed, etc.

The **chain-of-thought** prompting method is shown below:

You are going to play a game with other player(s). Think step-by-step. Begin by identifying steps that will contribute to you winning. Then reason through the steps, until a final decision is reached. The steps should reflect a meaningful thought process.

Here is an example on a simpler task from what you will be playing:

—

Example Task: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does Roger have now?

Steps:

1. Roger starts with 5 tennis balls.
2. 2 cans of 3 tennis balls each are bought. This is 2 * 3 = 6 tennis balls.
3. Roger now has 5 + 6 = 11 tennis balls.

Finish steps!

Final decision: 11 tennis balls

—

Now, the game you will be playing is presented. Think step-by-step. Identify the steps and reason through them to complete the objective of the game. You may come up with different reasoning steps for each round, as you see fit.

As explained above, this prompt would then be followed by an appropriate **game-description** to form the complete **system prompt**.

Following, we present the **solo-performance** prompting method:

You are going to play a game with other player(s). Begin by identifying the personas who could contribute to you winning. Then initiate a multi-round collaboration process among these personas, until a final decision is reached. The personas will give critical comments and detailed suggestions whenever necessary.

Here is an example on a simpler task from what you will be playing:

—

Example Task: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.
Input: 6 12 1 1

Personas: AI Assistant (you); Math Expert

Start collaboration!

Math Expert: Let's analyze the task in detail. You need to make sure that you meet the requirement, that you need to use exactly the four numbers (6 12 1 1) to construct 24. To reach 24, you can think of the common divisors of 24 such as 4, 6, 8, 3 and try to construct these first. Also you need to think of potential additions that can reach 24, such as 12 + 12.
AI Assistant (you): Thanks for the hints! Here's one initial solution: (12 / (1 + 1)) * 6 = 24
Math Expert: Let's check the answer step by step. (1+1) = 2, (12 / 2) = 6, 6 * 6 = 36 which is not 24! The answer is not correct. Can you fix this by considering other combinations? Please do not make similar mistakes.
AI Assistant (you): Thanks for pointing out the mistake. Here is a revised solution considering 24 can also be reached by 3 * 8: (6 + 1 + 1) * (12 / 4) = 24.
Math Expert: Let's first check if the calculation is correct. (6 + 1 + 1) = 8, 12 / 4 = 3, 8 * 3 = 24. The calculation is correct, but you used 6 1 1 12 4 which is not the same as the input 6 12 1 1. Can you avoid using a number that is not part of the input?
AI Assistant (you): You are right, here is a revised solution considering 24 can be reached by 12 + 12 and without using any additional numbers: 6 * (1 - 1) + 12 = 24.
Math Expert: Let's check the answer again. 1 - 1 = 0, 6 * 0 = 0, 0 + 12 = 12. I believe you are very close, here is a hint: try to change the "1 - 1" to "1 + 1".
AI Assistant (you): Sure, here is the corrected answer: 6 * (1+1) + 12 = 24
Math Expert: Let's verify the solution. 1 + 1 = 2, 6 * 2 = 12, 12 + 12 = 12. You used 1 1 6 12 which is identical to the input 6 12 1 1. Everything looks good!

Finish collaboration!

Final decision: 6 * (1 + 1) + 12 = 24

—

Now, the game you will be playing is presented. Identify the personas and use their multi-round collaboration to help you complete the objective of the the game. You may discuss with your personas on any round of the game, as you see fit.

The example used in the **Solo-Performance Prompt (SPP)** prompt is directly adapted from [59] and the **Chain-of-Thought (CoT)** prompt we displayed previously is a modification of this prompt to a more typical CoT style.

This thesis's tasks would require a substantial token length when presenting related examples, because of the long conversation required when playing a **game**. We decide to give a one-shot example to the agents of a simple task, that accurately portrays the thinking style we wish to elicit. We follow the *simplicity* criterion for one-shot example selection that was described in 3.3.3.

4.1.2 Message Handling

The experiments performed revolve around agents playing various games. For this reason, the game environment should be equipped with appropriate tools to handle errors (e.g., miscommunications, failure to adhere to established rules of the game), give feedback to agents that elicits wanted behaviors and responses, and extract model answers in a 'digestible' from our system way (we referred to this process as **Answer Mining** in 3.3.1).

The environment processes each LLM response looking for a chosen action, players must always choose to make some action in their responses, so such information should, under normal circumstances, always be present. A regular expression, as follows, is used to perform this check:

```
\[move\](?: \(([^)]+)\))? (move_0 | move_1 | ... | move_n)
```

Using this regex for pattern matching and answer mining in LLM responses, we allow sufficient flexibility in LLM responses, so as not to restrict their abilities. We provide an example of an acceptable response of a player in Rock-Paper-Scissors. This an LLM player initialized with the *Solo-Performance* (spp) prompt style playing their first round against a 'fresh' opponent (an opponent that the player has not encountered before and that has not played this game before either):

I need to identify the relevant personas and collaborate to develop a winning strategy for this rock-paper-scissors game.

Personas: AI Assistant (me); Game Theory Expert; Psychology Expert

Start collaboration!

Game Theory Expert: This is a classic simultaneous game with a zero-sum payoff matrix. Against a "fresh player" (presumably someone new), we need to consider their likely behavior patterns. New players often have predictable tendencies - they might choose rock first (most intuitive), or they might overthink and choose what they perceive as more sophisticated moves like scissors. The key insight is that in a single round against an unknown opponent, there's no dominant strategy, so we need to think about opponent psychology.

Psychology Expert: Fresh players in rock-paper-scissors often exhibit certain behavioral patterns. Many beginners start with rock because it feels the most "powerful" or natural choice. Others might choose scissors thinking it's clever or unexpected. Paper is often the least chosen by beginners because it seems "weak" conceptually. Since this is the first round and we have no information about this specific opponent, I'd recommend choosing paper as it has the highest probability of beating a beginner's likely choice of rock.

AI Assistant (me): Based on both analyses, choosing paper seems optimal for the first move against a fresh player. It counters the most likely beginner choice (rock) and gives us the best chance to score +1 points while making our opponent lose -1 points.

Game Theory Expert: I agree. Paper is the statistically best choice for round one against an inexperienced opponent. It maximizes our expected value given the psychological tendencies of new players.

Finish collaboration!

Final decision: Choose paper as my first move.

[move] Paper - betting on the fresh player choosing rock as their instinctive first choice

---

We will now name a string of the form:

[move] (text) text

as **move message**.

The regular expression we use is quite versatile. As can be seen the optional explanation in the parenthesis can be completely omitted. Also, optional leading and trailing to the **move message** text, does not impact the environment's ability to identify the move message. Before the application of the above regex, LLM responses are converted to lowercase, this allows LLMs to give answers using capital letters wherever they see fit without any impact on the game itself. Finally, the environment checks that the **move message** is unique in the LLMs answer (i.e., there exists only one).

In the case of failure in any part of the above identification of player moves, the player is provided with hints in an attempt to paraphrase and correct their answer. Hints are explored in the following subsection.

4.1.3 Hints

The **System Prompt** provided to LLM agents contains some initial hints, to help guide the LLM to communicate properly.

Message Formats:
Move messages are sent when you want to make a move.
Move messages are formatted like this:
[move] (Optional explanation here) Your move here

Hint messages are received when you are provided with advice and directions. These messages are important and you should pay attention to them.
Hint messages are formatted like this:
[hint] Hint message here

Objective:
Maximize your points and/or minimize your penalties.

Tips:
- Try to formulate a strategy by reasoning on the provided information.
- Pay attention to the payoff matrix of the game moves.
- You should only send move messages. Hint messages will only be provided to you.
- Pay attention to hints when provided.
- Do not use markdown formatting in your messages (e.g., bold text, italic text, etc.).

Apart from the above hints, which as discussed are only provided at the initial stage of the game. There exist additional hints, which are given in effort to attempt error correction. The games played are set-up in such a way, that whenever a player returns an answer to the environment, that answer must be a **move** message (or at least contain such a message). **Move** messages are simply the way we characterize messages that contain a player's choice on what action to perform in the game.

An error occurs when the agent fails to return a **move** message (e.g., returns a badly formatted message). So that is the only error case we need to cover:

You may structure your response however you like, but it should contain a move message. Move messages begin with the tag [move] not containing other tags, which is followed by your optional explanation in parentheses, and end with a valid move: {list-of-moves}. DO NOT INCLUDE THE [move] TAG IN YOUR REASONING, ONLY IN YOUR ACTUAL MOVE MESSAGE. Nested parentheses or markdown formatting are not allowed.

Format: [move] (Optional explanation here) Your move here

---

list-of-moves is the list of available to the player moves in each game.

If a player's responses are erroneous at least 5 times, then the current round of the game is aborted for both players.

4.1.4 Example Response

We provide an example response of *Mistral Large (24.07)*, when playing Prisoner's Dilemma and prompted using the Solo-Performance Prompting (**spp**) style. The LLM agent gave this response during its first round of playing the game, i.e., right after following only the initial prompt provided by our environment.

Personas: 1. **AI Assistant (you)** 2. **Strategist** 3. **Game Theorist**

Start Collaboration!

Strategist: The game is a classic Prisoner's Dilemma. In a single round, defection tends to yield higher individual points (4 if the other player cooperates, 2 if both defect). However, mutual cooperation gives the highest mutual benefit (6 each). For a long-term strategy, consider the other player's likely actions and the iterative nature of the game.

Game Theorist: In game theory, the Nash equilibrium for a one-shot Prisoner's Dilemma is defection since it maximizes individual rewards. However, in repeated games, strategies like "Tit for Tat" (cooperate first and then mimic the opponent's last move) can promote cooperation.

AI Assistant (you): Thanks for the insights. Given this is the first round, I'll start with cooperation, a common strategy to signal willingness to cooperate.

[move] I choose cooperation to signal my willingness to cooperate and possibly set the tone for future rounds.

Feedback from Strategist and Game Theorist:

Strategist: Good start. Cooperation signals trust and could encourage the other player to cooperate in future rounds. Keep an eye on their next move to adjust your strategy.

Game Theorist: Agreed. Monitor their response. If they defect, you might consider defecting next to mirror their strategy and possibly guide them towards cooperation next time.

Continue the Collaboration for Next Rounds:

AI Assistant (you): Understood. Let's await the other player's move and adjust our strategy accordingly.

—

[hint] The other player chose defection. You got 1 point, and they got 4 points.

AI Assistant (you): Based on the feedback, I'll mirror their last move to signal that cooperation is beneficial for both players.

[move] I choose defection to mirror their last move and indicate the importance of mutual cooperation.

—

Further Collaboration:

Strategist: Mirroring defection is a good strategy to show them the consequence of their move. It might encourage them to cooperate next round.

Game Theorist: Yes, this is aligned with the "Tit for Tat" strategy. Continue this mirroring approach to guide the game towards mutual cooperation.

Continue Based on Further Rounds:

[hint] The other player chose cooperation. You got 4 points, and they got 1 point.

AI Assistant (you): They switched to cooperation after seeing the consequence. I'll cooperate this round to reinforce the mutual benefit.

[move] I choose cooperation to reinforce the mutual benefit of cooperation.

—

Final Decision for Next Steps:

AI Assistant (you): We will continue to monitor and adjust our strategy based on the other player's moves, aiming for long-term mutual cooperation.

Finish collaboration!

Final decision: Mirror the opponent's last move to guide towards mutual cooperation. Start with cooperation in the first round.

—

This strategy will help maximize long-term points and encourage mutual cooperation.

This response is quite intriguing, because it highlights a few observations (that will also be mentioned later in this work).

- Although the game prompt does not provide specific information about the game being played, the LLM agent manages to properly analyze the system prompt and extract the nature of the game. It is observed that the LLM identifies the game as Prisoner's Dilemma using its own cognitive abilities. The LLM, also, uses prior knowledge it has about Nash Equilibria and Prisoner's Dilemma in its thought process.
- This example response contains erroneous **move messages**. We selected this response on purpose to showcase such situations. Firstly, there are multiple **move messages** inside the LLMs response, something that violates our environment's conditions. Secondly, **move messages** are formatted improperly. They use "markdown" style highlighting - something that we explicitly disallow - (this error is characteristic of the type of errors present in *Mistral*'s responses) and they do not provide the player's chosen move with proper formatting either.

4.1.5 Feedback

Our work focuses on repeated variants of simultaneous two-player games. The same opponents play for a number of consecutive rounds against each other. The environment provides feedback to each opponent, such as:

1. at the start of the game, each player will be told if their opponent is experienced (has played the game before) or fresh. Our experiments were done solely with fresh players.
2. letting each player know a new round begins, if it is the first round these players play against each other, the environment will let them know that information as well.
3. after a player makes a move, the environment will let them know that this move has successfully been made.
4. at the end of each round, the players are notified of their opponent's action and are told their respective payoffs.
5. in case of unusual termination of a round by some player, the other is notified as well (i.e., this player is told that the round ended abnormally).

All feedback is provided via messages with the **user** role, as shown in 3.3.3. Most LLM APIs do not allow multiple consecutive messages with the same role in a chat; in order to address this issue, the environment concatenates any of the above messages as they are being generated into one and feeds that into the ongoing conversation.

4.2 Language Models

Large Language Model (LLM) reasoning is ever evolving to meet new demands and higher performance thresholds. It was deemed necessary for our work to emphasize modern LLMs in our experimentation - in order to evaluate the competence level of state-of-the-art LLMs - without sidelining simpler models. The inclusion of both types of models allows for good comparison of their abilities and evaluation of LLM evolution. Lastly, experiments with more reasoning-focused LLMs - referred to as Large Reasoning Models (LRMs) - have, also, been conducted in an effort to gauge the game-theory capabilities of this new promising frontier of LLMs; which seems to also face challenges of its own [48]. In this work's experiments, we evaluate a diverse set of large language models to analyze their reasoning performance under various prompting conditions.

The Anthropic Claude family is well represented, with models including *Claude 3.5 Sonnet v2*, *Claude 3.7 Sonnet*, and *Claude Sonnet 4*, allowing for comparisons of different prompting strategies within the same architectural lineage.

Meta's LLaMA models are a point of interest for researchers as well, however, their multi-modal versions face use restrictions in the European Union, so we experimented only with *Llama 3.3 70B Instruct*.

Additionally, *Mistral Large (24.07)* is part of our testing pool.

Lastly, due to the growing popularity of more reasoning focused models and, since game-theoretic environments rely heavily on reasoning and cognitive abilities, some models that sit in the LLM-LRM overlap area are also included in our testing. These are *DeepSeek-R1* and *Claude 3.7 Sonnet*, *Claude Sonnet 4* with "extended thinking" enabled.

All of the models used in these experiments were accessed via Amazon Bedrock and integrated through the Converse API, ensuring a consistent and production-grade interface for interaction across different providers. This approach allowed seamless experimentation with models from Anthropic (including the *Claude 3.5 Sonnet v2*, *Claude 3.7 Sonnet*, and *Claude Sonnet 4* variants), Meta's LLaMA 3 series (*Llama 3.3 70B Instruct*), *Mistral Large (24.07)*, and *DeepSeek-R1*. By leveraging Amazon Bedrock's unified infrastructure, the experiments maintained reliable prompt formatting, response parsing, and rate-limited access, simplifying the comparison of diverse model architectures and capabilities under standardized conditions.

Chapter 5

Experiments - Game Presentation

5.1 Game Overview

5.1.1 Prisoner's Dilemma

The **Prisoner's Dilemma** is perhaps the most famous game-theory thought experiment urging through the years a number of researchers to study it. It involves two rational agents, each of whom can either cooperate for mutual benefit or defect (i.e. betray their partner) for individual gain. The dilemma arises from the fact that mutual cooperation yields moderate benefits for both players, while unilateral defection yields a greater benefit for the defector at the expense of the cooperator. If both defect, they each receive a worse outcome than if they had cooperated. People, organizations, countries, etc. are often faced with dilemmas as the above. This fact coupled with the simplicity of Prisoner's Dilemma and its expansion opportunities make it an excellent target for our own study.

Premise Two members of a criminal gang are arrested and imprisoned. Each prisoner is in solitary confinement with no means of speaking to or exchanging messages with the other. The police admit they don't have enough evidence to convict the pair on the principal charge. They plan to sentence both to a year in prison on a lesser charge. Simultaneously, the police offer each prisoner a Faustian bargain[1]. If he testifies against his partner, he will go free while the partner will get three years in prison on the main charge. Oh, yes, there is a catch ... If both prisoners testify against each other, both will be sentenced to two years in jail. The prisoners are given a little time to think this over, but in no case may either learn what the other has decided until he has irrevocably made his decision. Each is informed that the other prisoner is being offered the very same deal. Each prisoner is concerned only with his own welfare—with minimizing his own prison sentence. [43]

There are three different possible outcomes for the two prisoners:

1. if both remain silent (cooperate), they will each serve one year in prison.
2. if one testifies against the other (defects), but the other does not (cooperates), the testifier will be set free while the cooperator will serve three years in prison.
3. if both testify against each other (defect), they will each serve two years.

The structure of the traditional Prisoner's Dilemma can be generalized from its original prisoner setting. It is represented as a normal-form game (explained in 3.4.4) with the following payoff matrix.

For the game to be considered Prisoner's Dilemma in the strong sense, the following condition must hold for the payoffs:

[1] A deal with the Devil: a pact between a person and the Devil or another demon, trading a soul for diabolical favors (e.g., youth, knowledge, wealth, fame, power)

	A	B
A	(aa, aa)	(ab, ba)
B	(ba, ab)	(bb, bb)

Table 5.1: Payoff matrix for the Prisoner's Dilemma.
A typically refers to "Cooperation" and **B** to "Defection".

$$ba > aa > bb > ab$$

In this general setting, players aim to maximize the payoff they get.

5.1.2 Rock Paper Scissors

For the next experiments, this thesis moves a bit away from **Prisoner's Dilemma**, choosing a slightly more complex game, that remains interesting to research nonetheless.

Rock-Paper-Scissors (RPS) is an intransitive[2] hand game, played by two people, in which each player simultaneously forms a shape with their hand; this shape constitutes their action (or move) of choice and it can be one of three: "Rock", "Paper", "Scissors". Like coin flipping, drawing straws, or throwing dice, **Rock-Paper-Scissors** is frequently employed as a fair means of selection between two individuals to resolve conflicts or make an unbiased group decision. In some cases, one can play RPS with some degree of competence and skill, unlike genuinely random selection systems, by taking advantage of the opponent's non-random behavior. [18, 5].

Premise RPS is a simultaneous, zero-sum game with three possible outcomes; win, loss, or tie. A player who chooses to play "rock" will beat a player who selects "scissors" ("rock crushes scissors" or "breaks scissors" or "blunts scissors"), but will lost to a player who has played "paper" ("paper covers rock"). Playing "paper" will lead to loss against "scissors" ("scissors cuts paper"). The game is tied if both players select the same shape.

The structure for Rock-Paper-Scissors followed in this work is a simple generalization of the above description. Win means a positive payoff, loss means a negative payoff, and tie means zero payoff. To retain the zero-sum nature of RPS, we make it such that the payoffs of a winner and a loser in a round are additive inverses (the one number is the opposite of the other).

This game can be represented as a normal-form game (explained in 3.4.4) with the following payoff matrix.

	A	B	C
A	(0, 0)	(-ba, ba)	(ac, -ac)
B	(ba, -ba)	(0, 0)	(-cb, cb)
C	(-ac, ac)	(cb, -cb)	(0, 0)

Table 5.2: Payoff matrix for the Rock-Paper-Scissors.
A typically refers to "Rock," **B** to "Paper", and **C** to "Scissors".

ba, ac, cb are considered to be positive numbers.

In this general setting, players aim to maximize the payoff they get.

5.2 Contributions

To sum up our contributions, generally, are:

1. We perform experiments on currently popular and advanced LLMs and contrast our results with results of earlier research.

[2] a zero-sum game in which pairwise competitions between the strategies contain a cycle

2. We experiment with various prompting techniques and opine on their effectiveness.
3. We gauge counterfactual adaptability of LLMs.
4. We introduce performance metrics that are simple and showcase enough information about players' gameplay; such as preference to strategies, adaptability to an opponent and temporal evolution of strategies (Advanced LLMs do show such behaviours in our experiments). Through these metrics we evaluate LLM reasoning abilities in game theory contexts, which is the main focus of this work.
5. We reinforce opinions correlating LLM size, prompting techniques, and reasoning variants with problem difficulty and complexity.
6. We explore the persistence, or lack thereof, of inherent biases of LLMs in strategy selection.
7. We observe cooperation and animosity tendencies of LLM agents in games. (Especially, Prisoner's Dilemma and its variants provide an ideal playground for such observations)

Our code is available on GitHub [3].

5.3 Counterfactual Scenarios

The focal point of this study is LLM performance in counterfactual settings. Such settings should be picked intelligently so as to reflect some reasoning skill that is expected of a rational player. Following the examples of [1, 8] our game descriptions are void of information that will immediately condition a player towards specific information relating to the game; Players are informed that they are playing a game and are given the names of the available actions/moves (from which they might be able to infer the game itself, but this issue is addressed by the formulation of the counterfactual settings). We decided to move forward with both strategy and payoff counterfactuals (introduced in 3.5).

5.3.1 Prisoner's Dilemma

Stag Hunt

This game constitutes our main counterfactual settings. It is both a strategy counterfactual (moves have different names) and a payoff counterfactual (payoff values are different).

The Stag Hunt - also referred to as the **assurance game** or the **trust dilemma** - describes a conflict between safety and social cooperation. In the most common account of this dilemma, two hunters must decide separately, and without the other knowing, whether to hunt a stag or a hare. One hunter can catch a hare alone with less effort and less time, but it is worth far less than a stag and has much less meat. But both hunters would be better off if both choose the more ambitious and more rewarding goal of getting the stag, giving up some autonomy in exchange for the other hunter's cooperation and added might. This situation is often seen as a useful analogy for many kinds of social cooperation, such as international agreements on climate change. [20]

Stag Hunt can be represented as a normal-form game, where players aim to maximize the payoff they get, with the following payoff matrix.

	A	B
A	(aa, aa)	(ab, ba)
B	(ba, ab)	(bb, bb)

Table 5.3: Payoff matrix for Stag Hunt.
A typically refers to "Stag" and **B** to "Hare".

For the game to be considered a Stag Hunt, the following condition must hold for the payoffs:

$$aa > ba \geq bb > ab$$

[3] LLM-Game-Theory-Counterfactual

This game is a point of interest for researcher, because it represents well the problem of social cooperation; it does so on par with Prisoner's Dilemma (PD) or, it can be argued, better than PD. [50]

Scenarios in Our Experiments

In this subsection, we can now introduce the specific parameters used to create this study's counterfactual scenarios. These can be seen in table 5.4. It was briefly mentioned in the introduction to this section that the names of the moves may assist the LLM players in understanding what the exact game being played is and thus aid them in using memorized patterns for that game. Including counterfactuals where the moves have alternative names and comparing the performance of the two, should give good insight to the memory-reasoning relationship present in LLMs. In table 5.4, games (a) and (c), and (b) and (d) on the other hand, should be played identically, since the payoffs are the same.

	Cooperate	Defect
Cooperate	(4, 4)	(1, 6)
Defect	(6, 1)	(2, 2)

(a) pd

	Cooperate	Defect
Cooperate	(6, 6)	(1, 4)
Defect	(4, 1)	(2, 2)

(b) pd-alt

	Stag	Hare
Stag	(4, 4)	(1, 6)
Hare	(6, 1)	(2, 2)

(c) sh-alt

	Stag	Hare
Stag	(6, 6)	(1, 4)
Hare	(4, 1)	(2, 2)

(d) sh

Table 5.4: Payoff matrices for the Prisoner's Dilemma Counterfactual Settings.
(a) is our base game, it is a typical Prisoner's Dilemma Setting.
(b) is a payoff counterfactual of (a), it uses Stag Hunt's payoff matrix.
(c) is a strategy counterfactual of (a), it uses Stag Hunt's names for the available moves.
(d) is both a payoff and strategy counterfactual of (a), it is a typical Stag Hunt.

5.3.2 Rock Paper Scissors

Scenarios in Our Experiments

In this subsection, we can now introduce the specific parameters used to create this study's counterfactual scenarios. These can be seen in table 5.5. It was briefly mentioned in the introduction to this section that the names of the moves may assist the LLM players in understanding what the exact game being played is and thus aid them in using memorized patterns for that game. Including counterfactuals where the moves have alternative names and comparing the performance of the two, should give good insight to the memory-reasoning relationship present in LLMs. In table 5.5, games (a) and (c), and (b) and (d) on the other hand, should be played identically, since the payoffs are the same.

5.4 Single Round Equilibrium

The single-round Nash Equilibrium Strategies will be presented in this part. These are derived by applying the theoretical background introduced in 3.4.

5.4.1 Prisoner's Dilemma

Using the definition for Nash Equilibrium in 3.4.3, it is easy to observe that for the classic Prisoner's Dilemma payoff matrix (game settings **pd** and **sh-alt** from 5.4) mutual defection is the Equilibrium strategy one should follow. Stag Hunt, however, requires a bit more nuanced approach.

Let us consider pure strategies for Stag Hunt (using payoff matrix 5.3). Considering that our opponent plays a pure strategy, we can see that playing a different move than what our opponent chooses will result in a lesser payoff, thus we can infer:

	Rock	Paper	Scissors
Rock	(0, 0)	(-1, 1)	(1, -1)
Paper	(1, -1)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(a) eq1

	Rock	Paper	Scissors
Rock	(0, 0)	(-3, 3)	(1, -1)
Paper	(3, -3)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(b) ba3

	Paper	Rock	Scissors
Paper	(0, 0)	(-1, 1)	(1, -1)
Rock	(1, -1)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(c) eq1-alt

	Paper	Rock	Scissors
Paper	(0, 0)	(-3, 3)	(1, -1)
Rock	(3, -3)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(d) ba3-alt

Table 5.5: Payoff matrices for the Rock-Paper-Scissors Counterfactual Settings.
(a) is our base game, it is a typical Rock-Paper-Scissors Setting.
(b) is a payoff counterfactual of (a), it uses a higher payoff for win with "Paper".
(c) is a strategy counterfactual of (a), if **X** typically beats **Y**, now **Y** beats **X**.
(d) is both a payoff and strategy counterfactual of (a), it is a combination of (b) and (c).

- if the opponent chooses "A", then the other player should also choose "A"
- if the opponent chooses "B", then the other player should also choose "B"

There are two pure strategy Equilibria for two players (both players playing "A" or both players playing "B"). This game allows for the use of a Mixed Strategy Nash Equilibrium. We can derive this strategy by using the **Indifference Property** (from 3.4.3).

Consider that the game is symmetric for the two players and both will follow the same mixed strategy. To determine the Mixed Strategy Nash Equilibrium let

$$p = \Pr(\text{play } A), \quad q = \Pr(\text{play } B)$$

with the constraint

$$p + q = 1$$

Define $E(M)$ as the expected payoff of playing move M against an opponent who plays A with probability p and B with probability q. The indifference condition in Mixed Strategy Nash Equilibrium requires:

$$E(A) = E(B)$$

Computing these payoffs:

$$E(A) = p \cdot aa + q \cdot ab = p \cdot aa + (1-p) \cdot ab = p \cdot (aa - ab) + ab,$$

$$E(B) = p \cdot ba + q \cdot bb = p \cdot ba + (1-p) \cdot bb = p \cdot (ba - bb) + bb$$

Setting the expected payoffs equal yields:

$$p \cdot (aa - ab) + ab = p \cdot (ba - bb) + bb,$$

$$p \cdot (aa - ab + bb - ba) = bb - ab,$$

$$p = \frac{bb - ab}{aa - ab + bb - ba}$$

Substituting into $p + q = 1$:

$$q = 1 - p = 1 - \frac{bb - ab}{aa - ab + bb - ba} = \frac{aa - ba}{aa - ab + bb - ba}$$

To sum up, given a payoff matrix of the form:

and considering p, q the probabilities that a player should play moves A and B respectively, then:

	A	B
A	(aa, aa)	(ab, ba)
B	(ba, ab)	(bb, bb)

Prisoner's Dilemma The dominant strategy is to defect (play B), so:

$$p = 0, \quad q = 1$$

Stag Hunt The Mixed Strategy Nash Equilibrium is given by:

$$p = \frac{bb - ab}{aa - ab + bb - ba}, \quad q = \frac{aa - ba}{aa - ab + bb - ba}$$

5.4.2 Rock Paper Scissors

There are three pure strategies that players may follow. This game allows for the use of a Mixed Strategy Nash Equilibrium. We can derive this strategy by using the **Indifference Property** (from 3.4.3).

Consider that the game is symmetric for the two players and both will follow the same mixed strategy. To determine the Mixed Strategy Nash Equilibrium let

$$x = \Pr(\text{play } A), \quad y = \Pr(\text{play } B), \quad z = \Pr(\text{play } C)$$

with the constraint

$$x + y + z = 1$$

Define $E(M)$ as the expected payoff of playing move M against an opponent who plays A with probability x, B with probability y, and C with probability z. The indifference condition in Mixed Strategy Nash Equilibrium requires:

$$E(A) = E(B) = E(C)$$

Computing these payoffs:

$$E(A) = 0 \cdot x - ba \cdot y + ac \cdot z = -ba \cdot y + ac \cdot z,$$

$$E(B) = ba \cdot x + 0 \cdot y - cb \cdot z = ba \cdot x - cb \cdot z,$$

$$E(C) = -ac \cdot x + cb \cdot y + 0 \cdot z = -ac \cdot x + cb \cdot y$$

Setting the expected payoffs equal per pairs yields:

$$E(A) = E(B) \iff (ac + cb) \cdot z = ba \cdot (x + y)$$

$$E(B) = E(C) \iff (ba + ac) \cdot x = cb \cdot (y + z)$$

$$E(A) = E(C) \iff (cb + ba) \cdot y = ac \cdot (x + z)$$

Using $x + y + z = 1$ we get:

$$(ac + cb) \cdot z = ba \cdot (1 - z) \iff (ac + cb + ba) \cdot z = ba$$

$$(ba + ac) \cdot x = cb \cdot (1 - x) \iff (ba + ac + cb) \cdot x = cb$$

$$(cb + ba) \cdot y = ac \cdot (1 - y) \iff (cb + ba + ac) \cdot y = ac$$

Finally, we get:

$$x = \frac{cb}{ba + ac + cb}, \quad y = \frac{ac}{ba + ac + cb}, \quad z = \frac{ba}{ba + ac + cb}$$

5.5 Experiment Variants

So far, games and their counterfactual settings have been introduced in their typical single-round form. Our experiments, however, focus only on **repeated** variants of these games. Players play 16 and 24 rounds against each other in Prisoner's Dilemma and Rock-Paper-Scissors respectively, without having any knowledge of the fact they will be playing a finite amount of rounds or, of course, how many rounds they will play. At the end of each round they receive information on the results of the current round they played (more information in 4.1.5), before moving on to the next one.

Repeated versions of games affect greatly the strategy one may follow. Attempting to use the Single Round Nash Equilibrium strategy in every round is not necessarily the best option in this case. A player learns the behavior of their opponent and thus can adopt a more educated approach to their own actions.

Another factor that influences a player's behavior is knowledge of the number of rounds to be played or of some upper bound to that number. For example, in Prisoner's Dilemma, if players know that they will be playing N rounds against each other, then mutual defection becomes the dominant strategy for every round once again. Therefore, this issue was taken into consideration in our LLM prompt design, where agents are completely uninformed of how or when the game will end.

5.5.1 Additional Condition for Prisoner's Dilemma

The repeated version of the Prisoner's Dilemma can, in some cases, be played by simply alternating between cooperation and defection, giving a greater reward than mutual cooperation. We wish to see LLMs attempt to strategize in general environments, thus our experimental setups include a condition on payoff matrices to not allow for players' strategies to collapse to the mentioned alternation of moves. The condition is:

$$2 \cdot aa > ba + ab$$

5.6 Game Description

As mentioned in 4.1.1 a Game Description is the part of a system prompt that effectively describes the game being played.

For Prisoner's Dilemma that Game Description is the following.

You are playing a game against other player(s). The game is based on strategic moves.

Game Description:
There are two possible moves: {a} and {b}.
- if both players choose {a}, both get {aa} points.
- if both players choose {b}, both get {bb} points.
- if you choose {a} and the other player chooses {b}, you get {ab} points and the other player gets {ba} points.
- if you choose {b} and the other player chooses {a}, you get {ba} points and the other player gets {ab} points.

In each round of the game you play against one other player. You and the other player each make a move privately. The moves are compared and you and the other player get your reward/penalty based on the moves made.

Message Formats:
Move messages are sent when you want to make a move.
Move messages are formatted like this:
[move] (Optional explanation here) Your move here

Hint messages are received when you are provided with advice and directions. These messages are important and you should pay attention to them.
Hint messages are formatted like this:

[hint] Hint message here

Objective:
Maximize your points and/or minimize your penalties.

Tips:
- Try to formulate a strategy by reasoning on the provided information.
- Pay attention to the payoff matrix of the game moves.
- You should only send move messages. Hint messages will only be provided to you.
- Pay attention to hints when provided.
- Do not use markdown formatting in your messages (e.g., bold text, italic text, etc.).

For Rock-Paper-Scissors that Game Description is the following.

You are playing a game against other player(s). The game is based on making moves.

Game Description:
There are three possible moves: {a}, {b}, and {c}.
- {a} beats {c} giving the winner {ac} points and the loser -{ac} points.
- {b} beats {a} giving the winner {ba} points and the loser -{ba} points.
- {c} beats {b} giving the winner {cb} points and the loser -{cb} points.
- If both players make the same move, the game is a tie and no points are awarded.

In each round of the game you play against one other player. You and the other player each make a move privately. The moves are compared and you and the other player get your reward/penalty based on the moves made.

Message Formats:
Move messages are sent when you want to make a move.
Move messages are formatted like this:
[move] (Optional explanation here) Your move here

Hint messages are received when you are provided with advice and directions. These messages are important and you should pay attention to them.
Hint messages are formatted like this:
[hint] Hint message here

Objective:
Maximize your points and/or minimize your penalties.

Tips:
- Try to formulate a strategy by reasoning on the provided information.
- Pay attention to the payoff matrix of the game moves.
- You should only send move messages. Hint messages will only be provided to you.
- Pay attention to hints when provided.
- Do not use markdown formatting in your messages (e.g., bold text, italic text, etc.).

As seen previously, the game descriptions also include the initial hints given in the system prompt.

5.7 Player Types

In order to study strategic behavior and reasoning in game-theoretic settings, it is important to account for the diversity of potential decision-makers. This section introduces the different player types used in this study. By explicitly defining player types, we can model a range of behaviors - from fully LLM AI players to simple algorithm-driven players.

Having different player types is useful for several reasons. First, it allows us to compare LLM agents initialized with different prompts with each other and see how they approach their goals, develop their strategies, and make use of their cognitive abilities. Second, creating non-LLM players that follow simple algorithms in their strategic play-style, helps us more easily discern LLM reasoning abilities, these players are designed in a way that if one understands their strategy, they are easy to work with (combat or cooperate with them, depending on which action will bring a better reward to the LLM player). We hope to reveal important dynamic - such as the emergence of cooperation or exploitation - that may not appear in more uniform populations of players.

In this section, we describe the specific player types used in our experiments. By defining these types carefully, we aim to create a controlled yet sufficiently rich testbed for analyzing strategic reasoning and adaptation in game-theoretic scenarios.

LLM/LRM players

1. **zs or default**: This player is an AI player initialized with the zero-shot system prompt.
2. **cot**: This player is an AI player initialized with the chain-of-thought system prompt.
3. **spp**: This player is an AI player initialized with the solo-performance system prompt.
4. **sc-<any of the above player types>**: This player is an AI player initialized with the prompt referred to by <any of the above player types> and operates using self-consistency to augment their answers (and performance). Such players are: **sc-zs**(or **sc-default**), **sc-cot**, and **sc-spp**.

non-LLM/algorithmic players

1. **srep**: This player chooses their actions randomly. Srep refers to Single-round-equilibrium-player, since this player uses the single-round (mixed) Nash Equilibrium strategy as the probability distribution over their choices on moves.
2. **pp**: This player chooses their actions following a specific cyclic pattern.
3. Non-LLM Players used only in Prisoner's Dilemma:
   (a) **mf**: This player always chooses the move that when paired with the most frequent move of their opponent gives the best reward. The name "mf" refers to **maximizer** of most **frequent** move, i.e., this player tries to make the most out of their opponents most frequent move.
   (b) **tft**: This player is influenced by the tit-for-tat idea in games like Prisoner's Dilemma, but is slightly more general. The "tft" player always chooses the move that when paired with the most recent move of their opponent gives the best reward.
4. Non-LLM Players used only in Rock-Paper-Scissors:
   (a) **ap**: This player chooses the move that counters the opponent's most frequent move, thus called adaptive player.
   (b) **tft**: This player is influenced by the tit-for-tat idea in games like Prisoner's Dilemma, but is slightly more general. Tft chooses the move that counters the opponent's most recent move.

5.8 Experiment Format

There are a number of different parameters that influence an experiment and its potential for fruitful results. As experiment designers, we had to make various choices in an attempt to mix features that allow LLMs to leverage their full potential, but are concise and can be performed with satisfactory performance.

Such parameters include Language Model temperature, number of rounds played in each game, number of repetitions of games, opponent types of LLMs, Self-Consistency parameters, and types of counterfactual settings included in experimentation.

Finally, a visual showcase of our experiments' format is shown in 5.8.1 and 5.8.2, where various ideas about LLMs - roles, prompts, experiment environments - and our study that have been discussed so far in this thesis are shown working together in a simple example.

- **LLM Parameters**: temperature was set at 1.0 to allow LLMs for more creative thinking. Since experiments were performed a multitude of times, a high temperature will allow us to gauge LLM performance and reasoning with a more comprehensive and general view. LLMs are given the most amount of expressive freedom possible in an attempt to extract any hint of inherent reasoning ability. Other LLM parameters were left at their default (per LLM) values.
- **Number of rounds per game**: As has been mentioned, every duo of opponents play 16 and 24 consecutive rounds against each other in Prisoner's Dilemma and Rock-Paper-Scissors respectively. Considering that Prisoner's Dilemma and Stag Hunt are games with only two options, 16 rounds are enough for an LLM player to manage to understand and adapt to their opponent's strategy (if any) in a way that we can evaluate. Considering that Rock Paper Scissors is a game with three options, 24 rounds are enough for an LLM player to manage to understand and adapt to their opponent's strategy (if any) in a way that we can evaluate. The number 24 was chosen as a logical continuation of our choice for 16 rounds in Prisoner's Dilemma.
- **Number of repetitions of games**: Each game is played by the **non-sc** players 5 different times, and only 2 times by **sc** players. **Sc** players are included in only 2 repetitions of experiments, since it is expected that their results will be more similar; after all, **self-consistency** is a technique to guide an LLM agent towards more consistent answers.
- **opponent types of LLMs**: Each **non-sc** player plays against all other **non-sc** players and all **non-LLM** players. Each **sc** players faces off against all **non-sc** players and all **non-LLM** players. It should be noted that **sc** players never play against each other; this is a conscious design choice, as the performance hit of **self-consistency** is substantial and performing such experiments would be too time consuming. Furthermore, our goal is to compare **non-sc** and **sc** players, thus forcing **sc** players to face each other is redundant.
- **Self-Consistency parameters**: Self-Consistency was presented in 3.3.3. In the case of Prisoner's Dilemma, our implementation uses 3 sample answers of the LLM at each round and the final answer is marginalized out of them by frequency - the most frequent answer is chosen as the final answer. If multiple samples lead to the same most-frequent answer, then one of those samples is chosen at random to be the final answer. Lastly, given that this game has two possible choices for players, three is the minimum number of samples needed in order to ensure that one choice is always favored. On the other hand, for Rock-Paper-Scissors, our implementation uses 5 sample answers of the LLM at each round. Since there are three possible options in Rock Paper Scissors, three samples wouldn't be enough to create a pool of LLM answers from which the LLM's more "consistent" thought can be extracted through our frequency-based marginalization. Thus, considering the massive performance hit of **sc**, it was decided to use 5 samples as a valid trade-off.
- **Counterfactual Settings**: All counterfactual scenarios from 5.3 are included in our experimentation. These scenarios are diverse and will provide us with an overall view reflective of true LLM capabilities.

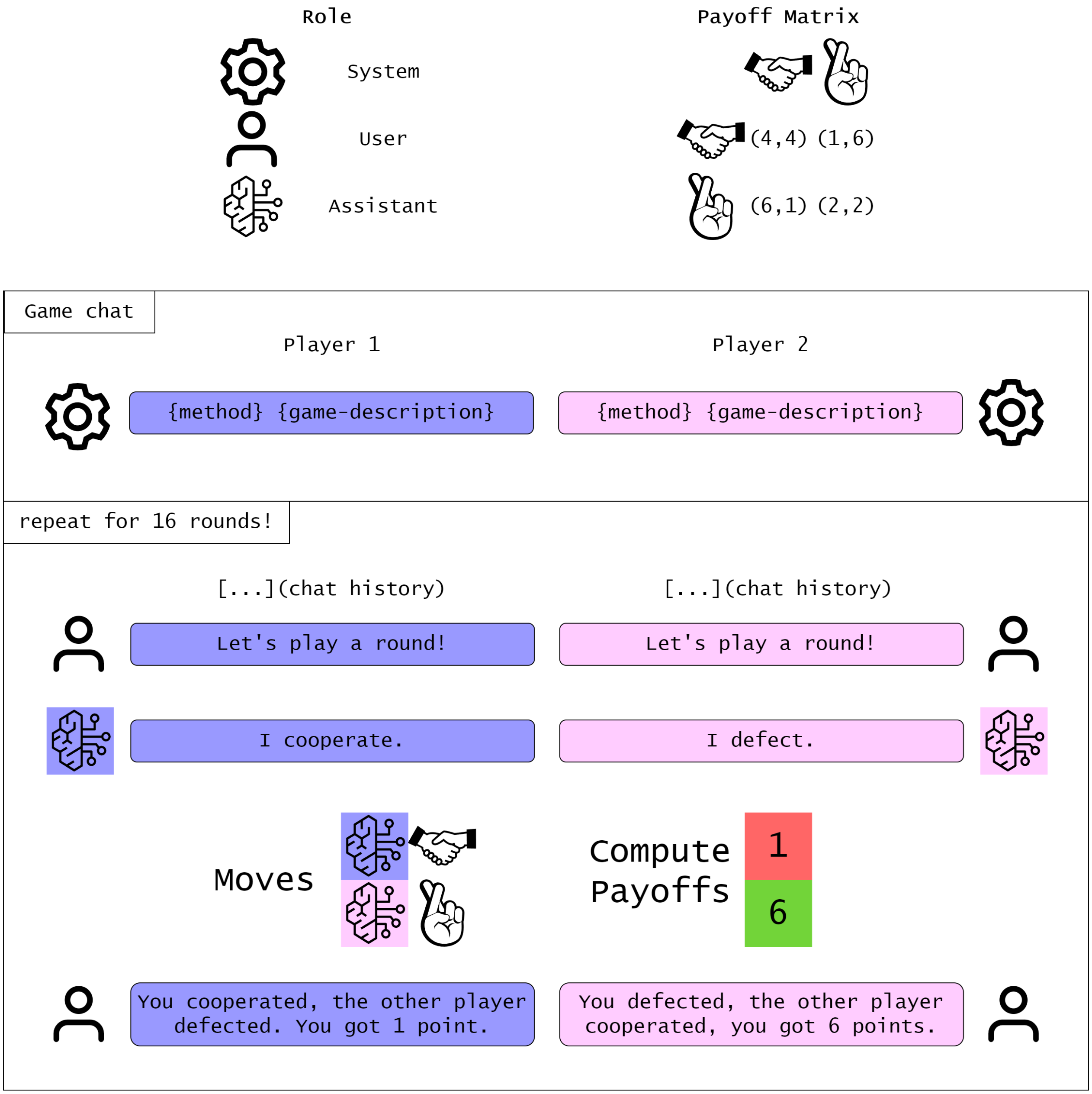


Figure 5.8.1: An example chat that could occur in a game of Prisoner's Dilemma under our design, where two LLM agents play against each other. The "User" role is used by us - the environment - to provide information to the AI players.

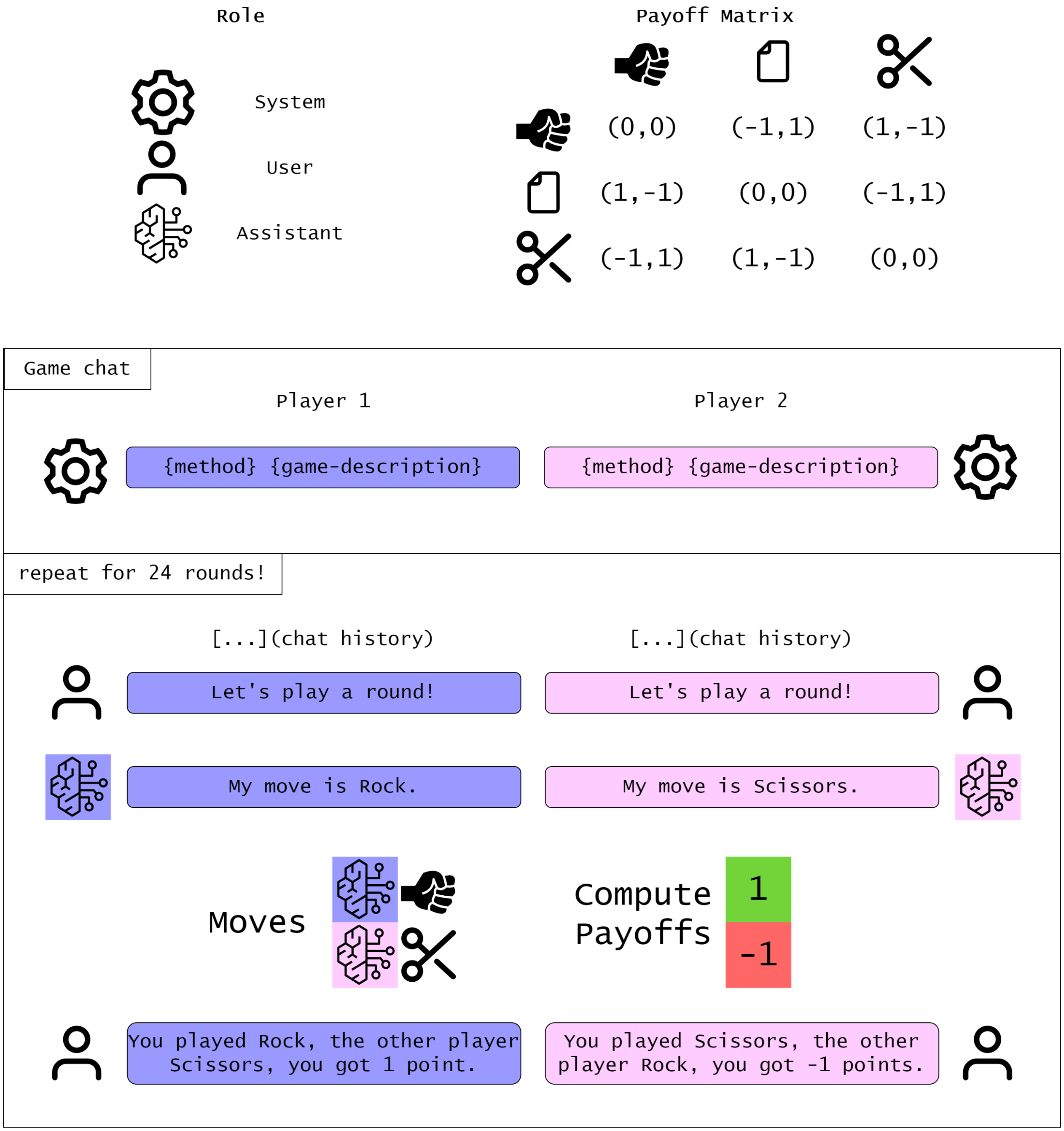


Figure 5.8.2: An example chat that could occur in a game of Rock Paper Scissors under our design, where two LLM agents play against each other. The "User" role is used by us - the environment - to provide information to the AI players.

Chapter 6

Results - Prisoner's Dilemma

It has been observed that LLMs perform well in games involving cooperation and defection. These games are considered an ideal test bed to assess how LLMs retaliate after bad interactions. [1]

As we have mentioned, it has been shown that a rational agent will always prefer to defect in the single-round (single-shot) variant of Prisoner's Dilemma, as well as in the case of finitely repeated rounds (i.e., the agent is aware either of the number of rounds to be played or of the fact that the game will last only some reasonable amount of iterations).

We mention once more that agents play 16 rounds in each game and we test 4 scenarios (one base case of Prisoner's Dilemma and three counterfactuals) with payoff matrices:

	Cooperate	Defect
Cooperate	(4, 4)	(1, 6)
Defect	(6, 1)	(2, 2)

(a) pd

	Cooperate	Defect
Cooperate	(6, 6)	(1, 4)
Defect	(4, 1)	(2, 2)

(b) pd-alt

	Stag	Hare
Stag	(4, 4)	(1, 6)
Hare	(6, 1)	(2, 2)

(c) sh-alt

	Stag	Hare
Stag	(6, 6)	(1, 4)
Hare	(4, 1)	(2, 2)

(d) sh

Table 6.1: Payoff matrices for the Prisoner's Dilemma Counterfactual Settings.

This chapter will showcase results and comment on them.

6.1 Total Points

We show the average total points accumulated in each game setting from each player. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

6.1.1 LLM vs LLM

LLM agents that play against each other are of the same LLM. These games are represented in the first three columns of the result matrices.

Tables 6.2 and 6.5 refer to games played with a typical Prisoner's Dilemma payoff matrix. The amount of values close to 64.0 that we observe is surprising. Had the two players followed the more expected strategy of mutual defection, we would observe results closer to $2 \cdot 16 = 32$ (defection in each of the 16 rounds). On the

model	prompt	pd zs	spp	cot	srep	pp	mf	tft
C3.5Sv2	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	54.2 ± 4.5	30.2 ± 1.1	30.6 ± 0.9
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	53.6 ± 1.5	31.0 ± 1.0	30.4 ± 0.9
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.4 ± 0.5	58.4 ± 4.3	30.6 ± 0.9	32.0 ± 0.0
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	52.0 ± 0.0	32.0 ± 0.0	30.0 ± 1.4
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	53.5 ± 2.1	32.0 ± 0.0	30.0 ± 0.0
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	62.0 ± 0.0	31.0 ± 1.4	31.0 ± 1.4
C3.7S	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	54.8 ± 4.1	29.8 ± 0.8	30.4 ± 0.5
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	61.2 ± 1.1	31.0 ± 0.7	31.2 ± 0.4
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	56.8 ± 4.4	31.0 ± 1.0	30.2 ± 0.4
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	52.5 ± 0.7	31.0 ± 1.4	30.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	61.0 ± 1.4	30.5 ± 0.7	31.0 ± 0.0
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	57.5 ± 6.4	31.0 ± 1.4	31.5 ± 0.7
C3.7S(T)	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	52.8 ± 0.4	31.0 ± 0.7	31.6 ± 0.9
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.6 ± 0.5	57.2 ± 3.7	30.8 ± 0.8	30.2 ± 0.4
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.2 ± 0.4	58.0 ± 4.7	31.0 ± 1.0	31.0 ± 1.0
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	52.0 ± 0.0	32.0 ± 0.0	30.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	53.0 ± 0.0	31.0 ± 1.4	30.5 ± 0.7
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	57.5 ± 2.1	32.0 ± 0.0	31.0 ± 1.4
C4S	zs	49.4 ± 15.5	41.0 ± 10.7	42.0 ± 10.7	31.4 ± 0.5	**55.8 ± 4.4**	32.8 ± 1.6	32.2 ± 1.6
	cot	35.2 ± 3.0	39.6 ± 2.4	39.8 ± 7.4	31.4 ± 0.5	**60.2 ± 4.5**	32.2 ± 1.6	34.2 ± 1.9
	spp	48.4 ± 13.0	46.2 ± 13.6	36.4 ± 4.2	31.6 ± 0.5	**63.0 ± 1.0**	34.4 ± 2.1	31.8 ± 2.4
	sc-zs	34.5 ± 4.9	34.0 ± 2.8	40.0 ± 4.2	**32.0 ± 0.0**	**58.0 ± 8.5**	32.0 ± 0.0	32.0 ± 0.0
	sc-cot	39.0 ± 4.2	35.5 ± 4.9	34.0 ± 2.8	**32.0 ± 0.0**	**64.0 ± 0.0**	35.0 ± 0.0	33.5 ± 2.1
	sc-spp	32.0 ± 1.4	38.0 ± 1.4	35.5 ± 4.9	30.5 ± 0.7	**63.5 ± 0.7**	35.5 ± 0.7	31.5 ± 0.7
C4S(T)	zs	32.0 ± 0.0	34.8 ± 2.7	32.0 ± 0.0	**32.0 ± 0.0**	**63.4 ± 0.9**	35.2 ± 1.8	**35.8 ± 0.4**
	cot	32.0 ± 0.0	40.8 ± 8.0	34.6 ± 3.6	31.6 ± 0.5	**64.0 ± 0.0**	33.8 ± 1.8	34.4 ± 2.2
	spp	31.4 ± 0.9	50.6 ± 15.4	41.8 ± 12.4	31.4 ± 0.5	**59.6 ± 5.2**	34.8 ± 1.8	34.8 ± 2.2
	sc-zs	37.5 ± 7.8	34.0 ± 2.8	32.0 ± 0.0	**32.0 ± 0.0**	**64.0 ± 0.0**	32.0 ± 0.0	33.5 ± 2.1
	sc-cot	32.0 ± 0.0	37.0 ± 1.4	32.0 ± 0.0	**32.0 ± 0.0**	**64.0 ± 0.0**	**36.0 ± 0.0**	32.0 ± 0.0
	sc-spp	47.0 ± 21.2	36.0 ± 0.0	32.0 ± 0.0	31.5 ± 0.7	**62.0 ± 2.8**	35.5 ± 0.7	33.5 ± 2.1
DS-R1	zs	36.8 ± 4.1	39.0 ± 2.5	30.2 ± 13.0	31.6 ± 0.5	**62.6 ± 1.5**	32.8 ± 2.0	32.6 ± 2.2
	cot	33.2 ± 2.6	32.2 ± 2.2	33.0 ± 6.6	28.6 ± 7.1	**64.0 ± 0.0**	32.2 ± 1.8	23.6 ± 8.7
	spp	36.4 ± 4.2	32.0 ± 6.4	31.2 ± 6.3	31.2 ± 0.4	**61.4 ± 3.7**	33.2 ± 2.6	33.2 ± 1.6
	sc-zs	32.5 ± 2.1	33.0 ± 2.8	33.5 ± 3.5	31.5 ± 0.7	**64.0 ± 0.0**	35.0 ± 0.0	35.5 ± 0.7
	sc-cot	34.5 ± 4.9	34.0 ± 2.8	27.0 ± 7.1	**32.0 ± 0.0**	**64.0 ± 0.0**	31.5 ± 0.7	31.5 ± 0.7
	sc-spp	35.0 ± 4.2	36.0 ± 5.7	34.0 ± 2.8	31.0 ± 0.0	**61.0 ± 4.2**	33.5 ± 2.1	32.0 ± 0.0
L3.3-70B	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.6 ± 0.5	52.0 ± 0.0	31.6 ± 0.9	31.6 ± 0.9
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	55.6 ± 4.6	30.8 ± 1.1	30.4 ± 0.9
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.8 ± 0.4	53.8 ± 2.9	30.8 ± 1.1	30.4 ± 0.9
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.0 ± 0.0	52.0 ± 0.0	31.0 ± 1.4	32.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	52.0 ± 0.0	32.0 ± 0.0	31.0 ± 1.4
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.5 ± 0.7	55.5 ± 4.9	30.0 ± 0.0	32.0 ± 0.0
M-L(24.07)	zs	**61.4 ± 3.2**	60.0 ± 4.4	59.6 ± 9.8	24.2 ± 0.4	53.8 ± 2.0	26.0 ± 1.6	25.8 ± 0.8
	cot	**64.0 ± 0.0**	60.2 ± 3.6	62.0 ± 2.8	29.8 ± 2.2	57.8 ± 3.0	29.0 ± 1.2	26.0 ± 1.2
	spp	**64.0 ± 0.0**	62.2 ± 4.0	60.8 ± 7.2	18.6 ± 10.6	53.2 ± 2.2	26.4 ± 3.2	24.4 ± 1.1
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	24.0 ± 0.0	54.0 ± 2.8	25.0 ± 1.4	25.0 ± 1.4
	sc-cot	62.0 ± 2.8	**64.0 ± 0.0**	60.5 ± 4.9	27.5 ± 4.9	60.0 ± 5.7	33.5 ± 3.5	29.0 ± 4.2
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	24.0 ± 0.0	56.0 ± 0.0	22.5 ± 4.9	22.5 ± 2.1

Table 6.2: Total Points Averaged Over All Iterations (pd)

		pd-alt						
		zs	spp	cot	srep	pp	mf	tft
model	prompt							
C3.5Sv2	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	44.0 ± 5.3	52.0 ± 2.1	89.0 ± 0.0	60.6 ± 32.3
	cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	47.8 ± 3.8	49.6 ± 2.1	91.8 ± 3.8	60.0 ± 32.9
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	44.8 ± 9.8	47.6 ± 1.7	90.4 ± 3.1	72.0 ± 32.9
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	40.0 ± 7.1	54.0 ± 1.4	89.0 ± 0.0	**96.0 ± 0.0**
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	52.5 ± 4.9	52.0 ± 0.0	89.0 ± 0.0	66.0 ± 42.4
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	52.5 ± 13.4	52.0 ± 8.5	92.5 ± 4.9	34.5 ± 0.7
C3.7S	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.6 ± 8.1	49.0 ± 3.0	91.8 ± 3.8	72.6 ± 32.1
	cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	45.0 ± 5.0	52.2 ± 3.3	93.2 ± 3.8	60.2 ± 32.7
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.2 ± 6.7	48.2 ± 4.4	94.6 ± 3.1	48.6 ± 26.6
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.0 ± 14.1	46.5 ± 0.7	92.5 ± 4.9	37.5 ± 0.7
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	40.0 ± 9.9	52.0 ± 8.5	92.5 ± 4.9	66.5 ± 41.7
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	42.0 ± 5.7	49.0 ± 4.2	92.5 ± 4.9	36.5 ± 0.7
C3.7S(T)	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	44.6 ± 8.1	50.0 ± 2.0	67.2 ± 29.9	48.2 ± 26.7
	cot	**96.0 ± 0.0**	90.8 ± 11.6	**96.0 ± 0.0**	47.8 ± 7.3	49.8 ± 4.8	**96.0 ± 0.0**	72.2 ± 32.6
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	37.8 ± 4.6	49.6 ± 7.2	91.8 ± 3.8	72.2 ± 32.6
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	49.5 ± 20.5	48.0 ± 1.4	61.5 ± 38.9	**96.0 ± 0.0**
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	45.0 ± 1.4	**55.5 ± 3.5**	89.0 ± 0.0	35.5 ± 0.7
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	41.0 ± 8.5	48.5 ± 2.1	92.5 ± 4.9	65.5 ± 43.1
C4S	zs	71.8 ± 33.1	**94.6 ± 3.1**	80.4 ± 25.2	43.0 ± 5.7	47.0 ± 1.2	66.2 ± 31.2	47.4 ± 27.2
	cot	**92.2 ± 3.5**	66.8 ± 29.1	55.2 ± 27.3	38.2 ± 3.5	50.0 ± 5.2	45.0 ± 28.5	47.0 ± 27.4
	spp	**83.0 ± 25.3**	46.2 ± 27.9	46.8 ± 27.5	39.6 ± 3.0	49.2 ± 5.3	69.2 ± 33.6	45.4 ± 24.4
	sc-zs	**96.0 ± 0.0**	93.0 ± 4.2	88.0 ± 1.4	38.5 ± 3.5	46.0 ± 0.0	89.0 ± 0.0	36.0 ± 2.8
	sc-cot	63.0 ± 41.0	32.5 ± 0.7	34.5 ± 0.7	41.5 ± 0.7	45.5 ± 0.7	**63.5 ± 46.0**	36.0 ± 1.4
	sc-spp	94.0 ± 2.8	93.0 ± 4.2	57.5 ± 38.9	42.0 ± 4.2	47.0 ± 1.4	**96.0 ± 0.0**	54.0 ± 26.9
C4S(T)	zs	**96.0 ± 0.0**	84.0 ± 26.8	83.0 ± 29.1	41.6 ± 5.7	47.4 ± 1.1	79.0 ± 26.4	59.6 ± 33.2
	cot	59.0 ± 33.8	71.6 ± 33.4	70.0 ± 35.6	45.2 ± 6.1	47.2 ± 1.3	55.4 ± 30.7	**71.8 ± 33.1**
	spp	82.2 ± 26.1	68.0 ± 31.0	81.8 ± 27.0	39.6 ± 5.2	50.2 ± 5.5	**91.0 ± 3.1**	83.6 ± 27.7
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	91.5 ± 6.4	**57.0 ± 2.8**	50.5 ± 2.1	33.5 ± 2.1	36.0 ± 1.4
	sc-cot	65.5 ± 43.1	63.5 ± 46.0	**96.0 ± 0.0**	43.0 ± 4.2	47.5 ± 0.7	64.0 ± 45.3	65.5 ± 43.1
	sc-spp	**92.5 ± 4.9**	60.5 ± 41.7	65.0 ± 43.8	42.5 ± 2.1	47.0 ± 1.4	60.5 ± 40.3	33.0 ± 1.4
DS-R1	zs	46.6 ± 27.6	33.4 ± 2.4	30.6 ± 4.8	44.4 ± 2.9	48.0 ± 1.9	**56.4 ± 33.1**	46.8 ± 27.5
	cot	32.0 ± 1.9	32.6 ± 1.5	45.4 ± 28.3	38.6 ± 4.3	**47.8 ± 0.8**	32.2 ± 1.5	45.4 ± 28.3
	spp	31.6 ± 2.2	37.0 ± 11.8	**53.4 ± 27.9**	42.0 ± 2.2	47.8 ± 2.4	45.0 ± 28.5	51.6 ± 27.3
	sc-zs	33.0 ± 1.4	35.0 ± 5.7	35.0 ± 1.4	39.0 ± 1.4	**48.0 ± 1.4**	32.0 ± 1.4	33.5 ± 2.1
	sc-cot	63.5 ± 46.0	34.0 ± 2.8	34.0 ± 0.0	46.5 ± 9.2	46.5 ± 0.7	65.0 ± 43.8	**65.5 ± 43.1**
	sc-spp	28.5 ± 9.2	33.0 ± 1.4	35.0 ± 1.4	46.0 ± 2.8	**47.0 ± 1.4**	31.0 ± 0.0	34.0 ± 0.0
L3.3-70B	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	41.4 ± 4.3	52.0 ± 0.0	70.8 ± 34.5	47.0 ± 27.4
	cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	45.0 ± 2.9	49.6 ± 3.3	81.0 ± 26.0	46.8 ± 27.5
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	41.4 ± 10.5	48.0 ± 1.4	91.8 ± 3.8	83.8 ± 27.3
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	35.5 ± 6.4	52.0 ± 0.0	**96.0 ± 0.0**	**96.0 ± 0.0**
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	39.5 ± 0.7	46.0 ± 0.0	**96.0 ± 0.0**	65.0 ± 43.8
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	44.0 ± 0.0	52.0 ± 0.0	65.5 ± 43.1	**96.0 ± 0.0**
M-L(24.07)	zs	**96.0 ± 0.0**	79.0 ± 32.2	**96.0 ± 0.0**	46.2 ± 6.6	42.4 ± 3.9	79.8 ± 28.6	83.4 ± 25.5
	cot	90.4 ± 11.4	84.8 ± 25.0	**96.0 ± 0.0**	43.2 ± 3.8	43.4 ± 4.8	79.4 ± 29.5	72.2 ± 30.4
	spp	**96.0 ± 0.0**	62.6 ± 42.1	92.6 ± 5.0	43.2 ± 9.9	41.2 ± 2.7	82.0 ± 24.8	59.4 ± 28.9
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	83.5 ± 17.7	42.5 ± 4.9	49.0 ± 0.0	92.5 ± 4.9	**96.0 ± 0.0**
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	37.5 ± 2.1	44.5 ± 2.1	92.5 ± 4.9	68.0 ± 39.6
	sc-spp	**96.0 ± 0.0**	60.5 ± 50.2	**96.0 ± 0.0**	37.0 ± 2.8	40.0 ± 0.0	89.0 ± 0.0	58.5 ± 26.2

Table 6.3: Total Points Averaged Over All Iterations (pd-alt)

		sh						
		zs	spp	cot	srep	pp	mf	tft
model	prompt							
C3.5Sv2	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.8 ± 3.7	51.2 ± 5.1	91.8 ± 3.8	59.2 ± 33.6
	cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.0 ± 3.1	48.2 ± 0.8	59.4 ± 33.4	46.8 ± 27.5
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	36.8 ± 3.8	46.8 ± 1.3	79.8 ± 24.7	83.6 ± 27.7
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	44.5 ± 6.4	52.0 ± 8.5	92.5 ± 4.9	**96.0 ± 0.0**
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	40.5 ± 2.1	53.0 ± 7.1	34.0 ± 0.0	65.5 ± 43.1
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	**49.0 ± 2.8**	46.0 ± 0.0	66.0 ± 42.4	65.0 ± 43.8
C3.7S	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	46.8 ± 5.4	52.0 ± 4.7	93.2 ± 3.8	72.8 ± 31.8
	cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	42.6 ± 6.6	51.2 ± 3.7	93.2 ± 3.8	71.4 ± 33.7
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	40.4 ± 2.3	53.2 ± 4.5	93.2 ± 3.8	60.6 ± 32.3
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	39.5 ± 0.7	48.0 ± 1.4	89.0 ± 0.0	65.5 ± 43.1
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	45.0 ± 1.4	56.5 ± 2.1	92.5 ± 4.9	35.0 ± 1.4
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	42.5 ± 6.4	52.0 ± 8.5	92.5 ± 4.9	**96.0 ± 0.0**
C3.7S(T)	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	47.2 ± 5.8	50.2 ± 3.4	81.4 ± 26.1	59.6 ± 33.2
	cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	40.4 ± 4.8	52.2 ± 4.9	94.6 ± 3.1	59.0 ± 33.8
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	39.8 ± 4.8	54.0 ± 2.8	91.8 ± 3.8	59.6 ± 33.2
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	41.5 ± 3.5	48.0 ± 1.4	92.5 ± 4.9	35.0 ± 1.4
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	39.0 ± 1.4	55.5 ± 0.7	**96.0 ± 0.0**	65.5 ± 43.1
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	41.5 ± 10.6	50.5 ± 6.4	**96.0 ± 0.0**	35.0 ± 0.0
C4S	zs	**84.4 ± 25.9**	58.8 ± 31.8	70.4 ± 29.0	39.6 ± 4.9	52.0 ± 6.0	68.8 ± 34.2	52.4 ± 23.4
	cot	45.2 ± 25.1	47.0 ± 25.2	**57.2 ± 30.4**	45.4 ± 6.5	50.4 ± 5.9	32.0 ± 1.2	37.4 ± 4.9
	spp	46.2 ± 25.6	36.0 ± 1.9	67.0 ± 22.8	41.6 ± 4.0	51.2 ± 5.4	44.6 ± 26.0	**88.6 ± 5.7**
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	62.0 ± 38.2	41.0 ± 12.7	54.5 ± 4.9	92.5 ± 4.9	65.5 ± 43.1
	sc-cot	37.0 ± 4.2	**59.5 ± 36.1**	54.5 ± 27.6	42.0 ± 2.8	51.0 ± 1.4	33.0 ± 0.0	36.5 ± 2.1
	sc-spp	**66.0 ± 42.4**	35.0 ± 0.0	63.0 ± 41.0	43.5 ± 14.8	47.5 ± 2.1	61.0 ± 39.6	62.5 ± 37.5
C4S(T)	zs	68.0 ± 32.6	**77.6 ± 25.6**	68.2 ± 30.9	38.6 ± 5.2	48.8 ± 3.6	56.6 ± 32.9	47.6 ± 27.1
	cot	71.2 ± 33.1	**77.0 ± 25.6**	42.8 ± 23.6	48.6 ± 10.1	48.0 ± 1.2	49.2 ± 24.1	50.4 ± 20.8
	spp	83.8 ± 27.3	**88.0 ± 6.4**	68.8 ± 35.5	46.4 ± 4.7	50.2 ± 5.0	81.8 ± 28.0	49.0 ± 22.2
	sc-zs	62.5 ± 34.6	57.0 ± 29.7	92.5 ± 4.9	42.0 ± 1.4	**58.0 ± 4.2**	**96.0 ± 0.0**	35.0 ± 0.0
	sc-cot	32.5 ± 0.7	**94.0 ± 2.8**	66.0 ± 42.4	43.5 ± 2.1	48.5 ± 0.7	61.0 ± 39.6	65.5 ± 43.1
	sc-spp	**89.0 ± 0.0**	84.5 ± 10.6	63.0 ± 41.0	42.0 ± 1.4	49.5 ± 12.0	32.5 ± 0.7	36.0 ± 2.8
DS-R1	zs	43.8 ± 22.0	**51.0 ± 27.8**	32.0 ± 1.9	45.4 ± 5.0	47.8 ± 1.3	45.8 ± 28.1	34.2 ± 1.1
	cot	44.8 ± 28.7	33.4 ± 0.5	33.8 ± 0.8	44.0 ± 7.6	**49.0 ± 2.1**	33.4 ± 1.8	34.4 ± 0.9
	spp	44.6 ± 22.0	33.4 ± 1.3	**58.0 ± 32.9**	38.6 ± 1.9	48.0 ± 1.6	43.4 ± 25.5	33.4 ± 1.1
	sc-zs	33.5 ± 0.7	**58.0 ± 36.8**	33.0 ± 2.8	44.5 ± 2.1	48.0 ± 1.4	31.0 ± 0.0	33.0 ± 1.4
	sc-cot	34.0 ± 0.0	35.0 ± 2.8	34.5 ± 3.5	40.5 ± 0.7	**48.0 ± 0.0**	34.5 ± 2.1	33.5 ± 0.7
	sc-spp	32.5 ± 0.7	34.0 ± 1.4	32.5 ± 0.7	41.0 ± 1.4	**48.0 ± 0.0**	32.5 ± 2.1	33.5 ± 0.7
L3.3-70B	zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	47.0 ± 11.5	51.4 ± 1.3	82.0 ± 27.6	49.8 ± 25.9
	cot	86.8 ± 20.6	**96.0 ± 0.0**	**96.0 ± 0.0**	45.0 ± 5.7	48.8 ± 2.0	94.6 ± 3.1	83.8 ± 27.3
	spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	47.2 ± 4.1	50.2 ± 1.6	68.6 ± 31.3	71.6 ± 33.4
	sc-zs	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.5 ± 3.5	52.0 ± 0.0	**96.0 ± 0.0**	**96.0 ± 0.0**
	sc-cot	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	41.5 ± 10.6	47.0 ± 1.4	65.0 ± 43.8	65.0 ± 43.8
	sc-spp	**96.0 ± 0.0**	**96.0 ± 0.0**	**96.0 ± 0.0**	43.5 ± 6.4	50.5 ± 2.1	**96.0 ± 0.0**	**96.0 ± 0.0**
M-L(24.07)	zs	66.2 ± 20.0	83.2 ± 20.1	52.2 ± 26.1	36.4 ± 9.7	44.2 ± 7.8	**88.8 ± 5.5**	74.6 ± 29.5
	cot	72.6 ± 24.4	**77.2 ± 17.0**	62.6 ± 26.7	43.4 ± 5.7	50.4 ± 8.0	43.4 ± 25.8	49.2 ± 26.2
	spp	70.4 ± 25.9	68.4 ± 25.3	57.6 ± 27.6	38.8 ± 8.3	45.0 ± 4.5	**78.8 ± 27.3**	59.0 ± 27.1
	sc-zs	**96.0 ± 0.0**	74.5 ± 30.4	76.5 ± 27.6	41.5 ± 2.1	40.0 ± 0.0	92.5 ± 4.9	39.0 ± 1.4
	sc-cot	**76.0 ± 28.3**	63.0 ± 22.6	42.0 ± 14.1	39.0 ± 0.0	48.0 ± 7.1	61.5 ± 48.8	37.0 ± 2.8
	sc-spp	48.0 ± 0.0	68.5 ± 31.8	46.5 ± 12.0	44.5 ± 7.8	40.0 ± 0.0	**96.0 ± 0.0**	64.5 ± 37.5

Table 6.4: Total Points Averaged Over All Iterations (sh)

model	prompt	zs	spp	cot	sh-alt srep	pp	mf	tft
C3.5Sv2	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.2 ± 0.4	58.8 ± 5.0	31.2 ± 1.1	31.0 ± 1.0
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.8 ± 0.4	58.4 ± 1.8	31.4 ± 0.5	31.6 ± 0.9
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.4 ± 0.5	61.6 ± 0.9	31.6 ± 0.9	31.6 ± 1.8
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	59.5 ± 0.7	31.0 ± 1.4	32.0 ± 0.0
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	58.5 ± 2.1	31.5 ± 0.7	32.0 ± 0.0
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	59.5 ± 3.5	31.0 ± 1.4	31.5 ± 0.7
C3.7S	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	55.6 ± 4.1	31.2 ± 0.8	30.8 ± 0.8
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.6 ± 0.5	55.4 ± 4.2	31.0 ± 1.0	31.2 ± 0.4
	spp	**64.0 ± 0.0**	58.4 ± 10.3	63.6 ± 0.9	30.2 ± 0.4	60.4 ± 2.6	32.8 ± 1.6	31.4 ± 1.5
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	62.5 ± 2.1	30.0 ± 0.0	57.5 ± 6.4	31.0 ± 1.4	31.5 ± 0.7
	sc-cot	**64.0 ± 0.0**	52.5 ± 16.3	**64.0 ± 0.0**	30.0 ± 0.0	55.5 ± 9.2	30.5 ± 0.7	32.0 ± 0.0
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	50.5 ± 19.1	30.5 ± 0.7	60.5 ± 2.1	30.5 ± 0.7	32.0 ± 0.0
C3.7S(T)	zs	**64.0 ± 0.0**	59.6 ± 9.8	**64.0 ± 0.0**	30.4 ± 0.5	53.6 ± 2.2	30.8 ± 1.1	31.0 ± 1.4
	cot	55.0 ± 12.3	**64.0 ± 0.0**	**64.0 ± 0.0**	30.8 ± 0.4	58.6 ± 4.3	31.8 ± 0.4	31.2 ± 0.8
	spp	**63.4 ± 1.3**	57.8 ± 13.9	**63.4 ± 1.3**	30.2 ± 0.4	59.0 ± 3.5	31.0 ± 1.2	31.6 ± 0.9
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.0 ± 0.0	57.0 ± 7.1	32.0 ± 0.0	30.5 ± 0.7
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	61.5 ± 2.1	31.0 ± 0.0	31.5 ± 0.7
	sc-spp	63.0 ± 1.4	**64.0 ± 0.0**	**64.0 ± 0.0**	30.5 ± 0.7	56.0 ± 5.7	32.0 ± 0.0	32.0 ± 0.0
C4S	zs	40.8 ± 13.7	38.6 ± 2.2	44.0 ± 13.1	31.2 ± 0.8	**58.0 ± 4.5**	32.2 ± 2.3	32.4 ± 2.3
	cot	35.8 ± 4.3	40.0 ± 12.5	38.0 ± 5.3	30.6 ± 0.5	**60.4 ± 2.3**	33.6 ± 1.7	32.6 ± 2.1
	spp	34.2 ± 3.5	36.4 ± 13.8	38.6 ± 9.4	31.4 ± 0.5	**57.4 ± 4.5**	32.4 ± 1.8	34.2 ± 2.2
	sc-zs	37.0 ± 1.4	35.0 ± 5.7	31.5 ± 0.7	31.5 ± 0.7	**64.0 ± 0.0**	33.5 ± 3.5	35.0 ± 1.4
	sc-cot	31.5 ± 0.7	32.0 ± 0.0	34.5 ± 4.9	**32.0 ± 0.0**	**64.0 ± 0.0**	32.0 ± 0.0	33.0 ± 1.4
	sc-spp	34.5 ± 3.5	33.5 ± 3.5	34.5 ± 4.9	**32.0 ± 0.0**	**60.5 ± 2.1**	32.0 ± 1.4	32.5 ± 0.7
C4S(T)	zs	33.0 ± 2.0	42.0 ± 11.5	32.0 ± 1.9	31.4 ± 0.5	**59.4 ± 5.9**	32.0 ± 1.9	34.0 ± 2.3
	cot	33.2 ± 3.9	34.0 ± 2.7	31.8 ± 0.4	**32.0 ± 0.0**	**62.4 ± 3.6**	34.2 ± 2.5	33.6 ± 2.2
	spp	34.2 ± 5.4	39.8 ± 12.7	31.4 ± 0.5	31.2 ± 0.4	**60.2 ± 3.1**	32.4 ± 2.1	31.4 ± 0.5
	sc-zs	36.5 ± 3.5	40.0 ± 0.0	47.5 ± 23.3	31.0 ± 0.0	**64.0 ± 0.0**	33.5 ± 3.5	**36.0 ± 0.0**
	sc-cot	40.0 ± 5.7	35.5 ± 6.4	32.0 ± 0.0	**32.0 ± 0.0**	**64.0 ± 0.0**	32.0 ± 0.0	**36.0 ± 0.0**
	sc-spp	32.0 ± 0.0	38.0 ± 2.8	41.5 ± 13.4	**32.0 ± 0.0**	**58.0 ± 8.5**	35.5 ± 0.7	33.0 ± 1.4
DS-R1	zs	32.8 ± 2.5	34.8 ± 0.8	32.0 ± 1.2	31.0 ± 0.7	**62.2 ± 2.0**	31.8 ± 1.8	33.6 ± 2.4
	cot	34.4 ± 1.9	34.0 ± 3.7	32.0 ± 1.7	31.4 ± 0.5	**62.2 ± 2.2**	33.6 ± 2.4	34.6 ± 2.1
	spp	36.2 ± 1.8	34.8 ± 2.6	33.6 ± 3.2	31.0 ± 0.0	**63.6 ± 0.9**	31.4 ± 0.5	32.6 ± 2.7
	sc-zs	33.5 ± 2.1	33.5 ± 3.5	31.5 ± 0.7	31.5 ± 0.7	**64.0 ± 0.0**	35.5 ± 0.7	34.0 ± 2.8
	sc-cot	35.5 ± 0.7	33.5 ± 3.5	35.5 ± 0.7	31.5 ± 0.7	**64.0 ± 0.0**	**36.0 ± 0.0**	33.5 ± 3.5
	sc-spp	34.0 ± 2.8	38.0 ± 2.8	36.0 ± 0.0	**32.0 ± 0.0**	**64.0 ± 0.0**	35.5 ± 0.7	35.5 ± 0.7
L3.3-70B	zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	28.0 ± 1.2	54.8 ± 3.6	30.0 ± 1.4	30.8 ± 0.8
	cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.2 ± 1.1	53.4 ± 3.1	30.8 ± 1.1	31.2 ± 1.1
	spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.4 ± 0.5	54.0 ± 2.8	31.0 ± 1.0	31.0 ± 1.0
	sc-zs	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	28.5 ± 0.7	52.5 ± 0.7	31.0 ± 1.4	29.5 ± 3.5
	sc-cot	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.0 ± 0.0	56.0 ± 5.7	30.0 ± 0.0	31.0 ± 1.4
	sc-spp	**64.0 ± 0.0**	**64.0 ± 0.0**	**64.0 ± 0.0**	29.0 ± 1.4	56.0 ± 4.2	31.0 ± 1.4	29.5 ± 0.7
M-L(24.07)	zs	52.6 ± 9.8	50.0 ± 9.9	39.2 ± 9.4	23.8 ± 0.4	**53.8 ± 3.5**	25.6 ± 1.7	26.0 ± 1.4
	cot	58.8 ± 3.7	**59.6 ± 4.4**	55.0 ± 13.1	26.8 ± 2.5	56.6 ± 6.4	28.8 ± 2.0	28.2 ± 3.3
	spp	**57.2 ± 4.9**	56.8 ± 7.3	51.8 ± 11.7	24.4 ± 4.1	54.0 ± 1.2	27.2 ± 1.6	24.8 ± 2.2
	sc-zs	54.0 ± 1.4	**58.0 ± 5.7**	54.0 ± 14.1	24.0 ± 0.0	56.0 ± 0.0	25.0 ± 1.4	26.0 ± 2.8
	sc-cot	58.5 ± 3.5	**63.5 ± 9.2**	43.5 ± 3.5	28.0 ± 5.7	56.5 ± 0.7	31.5 ± 2.1	30.0 ± 0.0
	sc-spp	59.0 ± 0.0	**61.5 ± 3.5**	32.5 ± 12.0	26.0 ± 4.2	51.0 ± 5.7	24.0 ± 0.0	27.0 ± 1.4

Table 6.5: Total Points Averaged Over All Iterations (sh-alt)

other hand, $64 = 4 \cdot 16$, thus LLM players lean more towards cooperation in general. Players from Claude Sonnet 4 (both versions) and DeepSeek-R1 get lower results, since they stick more to the dominant strategy of defection.

Tables 6.3 and 6.4 refer to games played with Stag Hunt style payoff matrices. Both "Stag" and "Hare" are pure strategies, but we observe a lot of values close to $96 = 6 \cdot 16$ meaning most LLM players prefer to go for "Stag". Once again, players from Claude Sonnet 4 (both versions) and DeepSeek-R1 get lower results, as they attempt different strategies.

6.1.2 LLM vs non-LLM

The last 4 columns represent non-LLM players, that follow simple strategies. These players have been described in 5.7. Successful outcomes against them should be a good indication of reasoning abilities of the LLMs, since they imply an agent's ability to analyze their opponent and make informed decisions.

Srep

In Prisoner's Dilemma (6.2 and 6.5) the **srep** player will always choose to defect and a rational player should also always defect in response. Perhaps, a rational player will cooperate a few times - either at the first few rounds or once or twice more later on to give their opponent a chance to "change" their mind, in the case where the rational player is not strictly motivated by playing the dominant strategy -. This expectation of ours is reflected in results of models against **srep** (result values close to 32), with the exception of the *Mistral Large (24.07)* model, which seems to lag behind.

In Stag Hunt (6.3 and 6.4) the **srep** player will follow the Single-Round Mixed Nash Equilibrium strategy in every round. This strategy is given by (5.4):

$$p = \frac{bb - ab}{aa - ab + bb - ba}, \quad q = \frac{aa - ba}{aa - ab + bb - ba}$$

Which, in this specific case, would be:

$$p = \frac{2 - 1}{6 - 1 + 2 - 4}, \quad q = \frac{6 - 4}{6 - 1 + 2 - 4}$$

Or, $p = \frac{1}{3} and q = \frac{2}{3}$. The expected payoff of a single round is $p \cdot aa + q \cdot ab = \frac{1}{3} \cdot 6 + \frac{2}{3} \cdot 1 = 2.66\ldots$

This result is reflected in the resulting total point values, which generally are close to $16 \cdot 2.66\ldots = 42.66\ldots$ across all models.

Pp

The pattern player is a player that plays in cycles:

- defection, cooperation, ... if the game uses Prisoner's Dilemma strategy names
- hare, stag, ... if the game uses Stag Hunt strategy names

Should a player match these patterns exactly, they would get:

- in Prisoner Dilemma payoff matrix games (6.2 and 6.5): opponent defects, so the player would optimally defect getting two points; then opponent cooperates, so the player would optimally defect again getting 6 points. The player should play the dominant strategy of defection (as expected). This would lead them to accumulating $(2 + 6) \cdot 8 = 64$ points after 16 rounds.
- in Stag Hunt payoff matrix games (6.3 and 6.4): opponent plays hare, so the player should respond with hare as well. On the next round the opponent will play stag, so the player should also play stag. This would lead them to accumulating $(2 + 6) \cdot 8 = 64$ points after 16 rounds.

In Prisoner's Dilemma we see results close to 64 across all models. However, this could be a coincidence, since models might just stick to the well-known optimal strategy of defection without much reasoning, i.e., simply regurgitating memorized information. Stag Hunt's pattern player requires slightly better reasoning abilities in order to adapt to. We notice LLMs achieving lower scores when faced with this player. This could be a sign that LLMs need more experience before understanding their opponent's strategy; this hypothesis will be investigated more thoroughly later in this work.

Mf & Tft

Mf is a player that aims to maximize their own rewards off of the opponent's most frequent move, while **tft** aims to maximize their own rewards off of the opponent's most recent move.

In games with Prisoner's Dilemma payoff matrices, both these players will behave similarly and will end up defecting in every round (apart from the first one perhaps); thus, the only appropriate response from a rational player is to also always defect.

In games with Stag Hunt's payoff matrices, these players will reproduce their opponent's move (the most frequent in the case of **mf** or the most recent in the case of **tft**). At this point, an important reminder is that in Stag Hunt both players playing "Stag" or both playing "Hare" are both pure strategy Nash Equilibria for the single-round variant. If the player sticks to "Stag" fairly early in the game, they can get a large cumulative reward at the end. If, however, the player started by favoring "Hare" in the early rounds, although they might be incentivized to keep playing the same move in every round, since their opponent also plays "Hare" in every round and this situation is an equilibrium, a more risky player might hypothesize that their opponent matches their own actions and change their move of choice to "Stag", expecting their opponent to do the same after a while (since the ("Stag", "Stag") combo will eventually yield better results). In this situation, **tft** will respond in kind fairly quickly (after one round of mismatched moves); however, **mf** will be more stubborn since it operates on the basis of frequency; in this way **mf** discourages their opponent, and an LLM agent will revert back to "Hare" having, however, suffered a few losses in the intermediate rounds.

The above theoretical analysis is backed by results for the Prisoner's Dilemma payoff matrix games (6.2 and 6.5) where results are fairly close to $32 = 16 \cdot 2$. Results for Stag Hunt payoff matrix games (6.3 and 6.4) can be interpreted as such: "total points" values close to $6 \cdot 16 = 96$ mean the LLM player stuck to "Stag" for most rounds. Values close to $2 \cdot 16 = 32$ mean the LLM player stuck to "Hare" for most rounds. Values between these could indicate that the LLM started with "Hare" but decided to switch to "Stag" having understood how its opponent works (thus indicating the emergence of reasoning abilities). Finally, smaller values could indicate either that the LLM did not understand how to play with this opponent (thus a lack of reasoning ability) or happened to be in the situation where "Hare" was played for some number of rounds initially and then they attempted to switch to "Stag" but couldn't recover from the loss of strategy change in time to reap the benefits.

6.2 Opponent Comprehension

To disambiguate the above results, shed light on the hypothesized behavior that has been mentioned, and better understand the thinking process of LLM agents, another metric is introduced. Instead of just looking at the "Total Points" accumulated by agents when playing this game, we look at how late a player was at making the most out of their opponent's behavior.

We consider that an AI agent has managed to comprehend their opponent, when the agent systematically responds with actions that use the opponent's moves to their advantage. More formally, suppose players A, B that have played N rounds and the strategies they followed were $(s_A^1, s_B^1), \ldots, (s_A^N, s_B^N)$ - e.g., $(defect, cooperate), \ldots, (defect, defect)$ -. Assume that A is the AI agent, then we call **round of opponent comprehension**, m, the round after which every move that A makes yields a payoff for A that is at least as good as the payoff that B gets.[1]

[1] This move is not necessarily a best response of A to B (described in 3.4.3), since the definition of best response for a single round is a bit more restrictive (e.g., it would not consider mutual cooperation in every round as a best response, but it yields better results in Prisoner's Dilemma)

Furthermore, we have expanded this definition, by including a percentage tp (target percentage) relaxing the requirement of "good" response to *every* move that the opponent makes to the following requirement: in the rounds from m all the way to N, A's moves are at least as good as B's in tp percentage of those rounds.

The following tables are results where $tp = 90\%$ and a lower value is considered better since it indicates that the AI agent understood how to play with its specific opponent earlier in the game.

Finally, a value of 17 means that the agent never *understood* their opponent, since it is out of the range $1-16$ of rounds that were played.

We show the average **round of opponent comprehension** in each game setting for each player. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

6.2.1 LLM vs LLM

LLM agents that play against each other are of the same LLM. These games are represented in the first three columns of the result matrices.

In games with Prisoner's Dilemma style payoff matrices (6.6 and 6.9) LLM agents almost universally immediately - from the start of the game - reach a state of "mutual understanding". In a few cases, agents need to play a few rounds before this happens. Results for *Mistral Large (24.07)* in the *sh-alt* counterfactual setting (table 6.9), which is a strategy counterfactual of Prisoner's Dilemma (moves have different names) indicate that the counterfactual setting confused this LLM and significantly impacted its reasoning abilities.

In games with Stag Hunt style payoff matrices (6.7 and 6.8) LLM agents, also, fairly often manage to reach "mutual understanding" quite quickly in the game. Again *Mistral Large (24.07)* faces issues in the counterfactual setting with Stag Hunt's strategy names.

Lastly, the *Claude Sonnet 4* and *DeepSeek-R1* are worth mentioning. In the resulting tables for total points, it was noted that these models did not achieve maximum total points in the games, and in the tables for **round of opponent comprehension** in this section, we observe (1) mostly low values (typically 1) but (2) also some higher values. The amount of games that these models do not reach mutual agreement very early is much larger than the other LLMs. The combination of low scores and (1) can be attributed to LLMs immediately going for defection in Prisoner's Dilemma (thus favoring the known strategy) or more frequently playing Hare in Stag Hunt. This conclusion coupled with (2) indicates that these two LLMs might attempt creative play-styles and attempt more complicated strategies.

6.2.2 LLM vs non-LLM

The last 4 columns represent non-LLM players, that follow simple strategies. These players have been described in 5.7. Successful outcomes against them should be a good indication of reasoning abilities of LLM, since they imply an agent's ability to analyze their opponent and make informed decisions.

Srep

Since Stag Hunt has a mixed Strategy Nash Equilibrium for the single round variant, there is no possible way for the results in these games to be in any way indicative of anything. We only look at results for Prisoner's Dilemma (tables 6.6 and 6.9). In this game the **srep** player always defects. LLM agents notice this behavior and adapt to it fairly early in the game. The *Llama 3.3 70B Instruct* model seems to struggle a bit more to combat its opponent's strategy, especially in the strategy counterfactual setting (table 6.9), however, it is eventually successful.

The *Mistral* model faces issues again, achieving good results only in the **cot** prompt (both **non-sc** and **sc** variants).

Pp

The pattern player is a player that plays in cycles:

model	prompt	zs	spp	cot	srep	pp	mf	tft
					pd			
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.2 ± 0.4	8.4 ± 7.0	5.6 ± 6.4	2.8 ± 1.3
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	14.0 ± 2.0	3.0 ± 1.2	2.2 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.6 ± 0.5	7.6 ± 6.1	2.0 ± 0.7	3.0 ± 0.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	12.0 ± 0.0	3.0 ± 0.0	4.5 ± 0.7
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	9.0 ± 4.2	3.0 ± 0.0	2.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	2.0 ± 0.0	2.5 ± 0.7	2.5 ± 0.7
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	11.6 ± 5.5	3.6 ± 1.5	3.2 ± 1.6
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	2.0 ± 0.0	3.2 ± 1.8	3.2 ± 2.3
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	9.6 ± 6.2	3.0 ± 1.2	3.6 ± 2.3
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	13.0 ± 1.4	2.5 ± 0.7	2.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	2.0 ± 0.0	3.5 ± 2.1	5.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	8.0 ± 8.5	2.5 ± 0.7	4.0 ± 1.4
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.6 ± 1.3	12.8 ± 4.1	4.0 ± 1.4	2.8 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.4 ± 0.5	10.4 ± 6.1	2.6 ± 1.5	1.8 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.8 ± 0.4	6.8 ± 6.7	3.0 ± 1.2	3.0 ± 1.2
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	13.0 ± 1.4	3.0 ± 0.0	2.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	15.0 ± 1.4	2.5 ± 0.7	1.5 ± 0.7
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	9.0 ± 7.1	3.0 ± 0.0	2.5 ± 0.7
C4S	zs	1.2 ± 0.4	1.4 ± 0.9	3.0 ± 4.5	**1.0 ± 0.0**	12.0 ± 5.5	1.8 ± 1.1	2.2 ± 1.6
	cot	3.0 ± 2.8	10.8 ± 4.1	3.2 ± 2.3	**1.0 ± 0.0**	6.0 ± 6.1	2.2 ± 1.6	1.6 ± 1.3
	spp	5.2 ± 6.9	1.6 ± 1.3	4.6 ± 2.1	**1.0 ± 0.0**	1.8 ± 1.3	4.2 ± 7.2	2.2 ± 1.8
	sc-zs	2.5 ± 2.1	**1.0 ± 0.0**	7.0 ± 8.5	**1.0 ± 0.0**	6.5 ± 7.8	**1.0 ± 0.0**	2.0 ± 1.4
	sc-cot	3.5 ± 3.5	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	4.0 ± 4.2	3.0 ± 2.8	**1.0 ± 0.0**	2.0 ± 1.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
C4S(T)	zs	**1.0 ± 0.0**	2.0 ± 2.2	**1.0 ± 0.0**	**1.0 ± 0.0**	1.2 ± 0.4	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	4.2 ± 7.2	4.0 ± 6.7	**1.0 ± 0.0**	**1.0 ± 0.0**	1.6 ± 1.3	**1.0 ± 0.0**
	spp	1.4 ± 0.9	2.2 ± 2.2	2.4 ± 2.2	**1.0 ± 0.0**	5.2 ± 6.4	1.6 ± 1.3	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	3.0 ± 2.8	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.5 ± 2.1	**1.0 ± 0.0**	2.0 ± 1.4
DS-R1	zs	3.2 ± 2.2	2.0 ± 2.2	8.0 ± 8.3	**1.0 ± 0.0**	4.0 ± 6.7	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.2 ± 2.7	10.6 ± 8.8
	spp	5.2 ± 6.6	7.4 ± 7.6	5.2 ± 6.9	**1.0 ± 0.0**	4.8 ± 5.4	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-zs	6.5 ± 7.8	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	3.5 ± 3.5	**1.0 ± 0.0**	9.0 ± 11.3	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	2.5 ± 2.1	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	7.5 ± 9.2	**1.0 ± 0.0**	**1.0 ± 0.0**
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.8 ± 1.1	14.0 ± 0.0	2.8 ± 0.4	2.8 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.4 ± 0.9	10.0 ± 6.5	2.4 ± 0.5	2.2 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.4 ± 0.9	12.8 ± 5.0	2.4 ± 0.5	2.2 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	4.0 ± 0.0	14.0 ± 0.0	2.5 ± 0.7	3.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	14.0 ± 0.0	3.0 ± 0.0	2.5 ± 0.7
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	3.0 ± 1.4	9.0 ± 7.1	2.0 ± 0.0	3.0 ± 0.0
M-L(24.07)	zs	10.2 ± 8.4	6.8 ± 6.8	**3.8 ± 6.3**	16.4 ± 0.5	15.2 ± 1.1	16.2 ± 0.4	16.0 ± 0.7
	cot	**1.0 ± 0.0**	7.0 ± 8.2	4.2 ± 7.2	4.4 ± 7.1	12.6 ± 6.5	8.6 ± 7.3	16.2 ± 0.8
	spp	**1.0 ± 0.0**	3.2 ± 4.9	4.2 ± 7.2	13.4 ± 6.9	15.6 ± 0.9	17.0 ± 0.0	16.4 ± 0.9
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 0.0	15.0 ± 1.4	16.5 ± 0.7	16.5 ± 0.7
	sc-cot	8.5 ± 10.6	**1.0 ± 0.0**	8.5 ± 10.6	9.0 ± 11.3	8.5 ± 10.6	**1.0 ± 0.0**	10.0 ± 9.9
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	16.0 ± 0.0	17.0 ± 0.0	16.0 ± 0.0

Table 6.6: Round # where the Agent understood the opponent's Strategy (pd)

model	prompt	zs	spp	cot	srep	pp	mf	tft
					pd-alt			
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	14.0 ± 1.9	13.4 ± 2.2	2.0 ± 0.0	10.0 ± 8.2
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.6 ± 0.5	15.4 ± 1.7	1.6 ± 0.5	9.0 ± 8.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.6 ± 1.1	16.6 ± 0.9	1.8 ± 0.4	6.0 ± 7.3
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	11.0 ± 0.0	2.0 ± 0.0	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 1.4	15.0 ± 2.8	2.0 ± 0.0	9.0 ± 11.3
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 1.4	10.5 ± 9.2	1.5 ± 0.7	5.5 ± 2.1
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.6 ± 2.1	15.4 ± 2.2	1.6 ± 0.5	5.4 ± 7.0
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.8 ± 1.1	12.2 ± 3.3	1.4 ± 0.5	8.8 ± 7.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	17.0 ± 0.0	14.6 ± 5.4	1.2 ± 0.4	9.8 ± 7.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.0 ± 2.8	17.0 ± 0.0	1.5 ± 0.7	16.0 ± 1.4
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**11.0 ± 8.5**	11.5 ± 7.8	1.5 ± 0.7	8.5 ± 10.6
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	17.0 ± 0.0	16.0 ± 1.4	1.5 ± 0.7	12.5 ± 3.5
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.4 ± 1.1	14.2 ± 2.7	7.4 ± 7.5	11.4 ± 6.0
	cot	**1.0 ± 0.0**	1.6 ± 1.3	**1.0 ± 0.0**	16.4 ± 0.9	14.0 ± 5.7	**1.0 ± 0.0**	5.8 ± 6.6
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	13.0 ± 4.2	12.4 ± 5.7	1.6 ± 0.5	5.2 ± 6.6
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	16.0 ± 1.4	8.0 ± 8.5	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	**7.5 ± 4.9**	2.0 ± 0.0	9.5 ± 3.5
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	14.0 ± 0.0	16.0 ± 1.4	1.5 ± 0.7	4.0 ± 4.2
C4S	zs	2.4 ± 2.6	**1.2 ± 0.4**	2.0 ± 2.2	17.0 ± 0.0	16.6 ± 0.9	2.6 ± 1.9	7.2 ± 4.5
	cot	**1.2 ± 0.4**	5.8 ± 6.1	8.8 ± 7.4	16.2 ± 0.8	14.0 ± 5.7	3.0 ± 1.4	7.0 ± 6.0
	spp	2.4 ± 2.6	5.8 ± 4.3	4.2 ± 2.6	15.2 ± 2.5	14.8 ± 4.9	**2.2 ± 2.2**	7.2 ± 5.8
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	14.0 ± 1.4	17.0 ± 0.0	2.0 ± 0.0	10.5 ± 7.8
	sc-cot	3.0 ± 2.8	11.5 ± 2.1	10.5 ± 7.8	15.5 ± 2.1	17.0 ± 0.0	**1.0 ± 0.0**	10.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	2.5 ± 0.7	14.5 ± 3.5	17.0 ± 0.0	**1.0 ± 0.0**	9.5 ± 6.4
C4S(T)	zs	**1.0 ± 0.0**	1.2 ± 0.4	2.0 ± 2.2	15.6 ± 1.1	17.0 ± 0.0	2.4 ± 1.5	4.4 ± 5.0
	cot	3.0 ± 4.5	2.2 ± 2.2	2.0 ± 1.7	14.8 ± 2.0	16.6 ± 0.9	**1.6 ± 0.5**	4.6 ± 5.0
	spp	1.2 ± 0.4	1.8 ± 1.3	**1.0 ± 0.0**	12.8 ± 4.8	14.2 ± 6.3	4.2 ± 5.5	1.8 ± 1.8
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	17.0 ± 0.0	14.0 ± 1.4	3.5 ± 3.5	10.0 ± 5.7
	sc-cot	2.5 ± 2.1	**1.0 ± 0.0**	**1.0 ± 0.0**	17.0 ± 0.0	17.0 ± 0.0	**1.0 ± 0.0**	5.0 ± 5.7
	sc-spp	1.5 ± 0.7	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	17.0 ± 0.0	1.5 ± 0.7	3.0 ± 2.8
DS-R1	zs	8.4 ± 4.9	8.2 ± 5.9	11.2 ± 8.0	15.0 ± 1.6	17.0 ± 0.0	**1.6 ± 0.5**	8.2 ± 6.6
	cot	4.2 ± 5.2	5.8 ± 6.6	3.6 ± 5.8	13.6 ± 5.0	17.0 ± 0.0	**3.2 ± 2.7**	3.8 ± 4.1
	spp	8.0 ± 5.8	4.6 ± 3.6	1.6 ± 1.3	14.8 ± 1.6	16.6 ± 0.9	**1.2 ± 0.4**	5.6 ± 6.6
	sc-zs	**1.0 ± 0.0**	8.5 ± 10.6	4.0 ± 4.2	15.5 ± 0.7	17.0 ± 0.0	1.5 ± 0.7	8.5 ± 10.6
	sc-cot	**1.0 ± 0.0**	7.0 ± 8.5	9.0 ± 11.3	16.0 ± 0.0	17.0 ± 0.0	**1.0 ± 0.0**	6.0 ± 7.1
	sc-spp	3.0 ± 2.8	**1.0 ± 0.0**	6.0 ± 1.4	16.0 ± 1.4	17.0 ± 0.0	**1.0 ± 0.0**	5.0 ± 5.7
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.4 ± 0.9	15.0 ± 0.0	3.0 ± 2.7	6.8 ± 3.8
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.4 ± 0.9	15.8 ± 1.1	2.4 ± 2.1	5.2 ± 3.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	13.8 ± 4.5	17.0 ± 0.0	1.6 ± 0.5	2.2 ± 2.7
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	12.0 ± 7.1	15.0 ± 0.0	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	12.5 ± 0.7	17.0 ± 0.0	**1.0 ± 0.0**	3.0 ± 2.8
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.5 ± 2.1	15.0 ± 0.0	3.5 ± 3.5	**1.0 ± 0.0**
M-L(24.07)	zs	**1.0 ± 0.0**	6.4 ± 7.6	**1.0 ± 0.0**	16.6 ± 0.5	17.0 ± 0.0	4.4 ± 6.5	4.0 ± 6.7
	cot	4.2 ± 7.2	4.2 ± 7.2	**1.0 ± 0.0**	16.6 ± 0.5	17.0 ± 0.0	4.4 ± 6.5	7.4 ± 8.8
	spp	**1.0 ± 0.0**	7.6 ± 8.6	1.2 ± 0.4	16.6 ± 0.5	17.0 ± 0.0	4.4 ± 7.1	9.0 ± 8.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	3.0 ± 2.8	16.0 ± 1.4	16.0 ± 1.4	1.5 ± 0.7	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	17.0 ± 0.0	1.5 ± 0.7	9.0 ± 11.3
	sc-spp	**1.0 ± 0.0**	9.0 ± 11.3	**1.0 ± 0.0**	17.0 ± 0.0	17.0 ± 0.0	2.0 ± 0.0	11.0 ± 8.5

Table 6.7: Round # where the Agent understood the opponent's Strategy (pd-alt)

model	prompt	sh zs	spp	cot	srep	pp	mf	tft
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.8 ± 0.8	13.2 ± 5.8	1.6 ± 0.5	4.2 ± 4.1
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 1.2	17.0 ± 0.0	3.4 ± 2.3	6.2 ± 4.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	11.8 ± 5.8	17.0 ± 0.0	2.4 ± 1.5	1.8 ± 1.8
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 0.0	11.5 ± 7.8	1.5 ± 0.7	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.0 ± 2.8	11.5 ± 7.8	4.0 ± 0.0	5.5 ± 6.4
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	17.0 ± 0.0	3.0 ± 2.8	2.5 ± 2.1
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.4 ± 0.9	12.0 ± 5.0	1.4 ± 0.5	6.2 ± 7.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	14.4 ± 2.6	13.2 ± 3.0	1.4 ± 0.5	4.8 ± 6.9
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.0 ± 2.3	11.6 ± 4.0	1.4 ± 0.5	8.6 ± 8.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	17.0 ± 0.0	16.0 ± 1.4	2.0 ± 0.0	4.0 ± 4.2
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	8.5 ± 3.5	1.5 ± 0.7	8.5 ± 4.9
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.5 ± 0.7	10.5 ± 9.2	1.5 ± 0.7	**1.0 ± 0.0**
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.4 ± 0.5	15.0 ± 1.4	2.2 ± 2.2	6.6 ± 6.5
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 1.0	11.2 ± 5.4	1.2 ± 0.4	4.4 ± 4.1
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	14.6 ± 2.4	10.0 ± 4.1	1.6 ± 0.5	7.0 ± 6.6
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 1.4	16.0 ± 1.4	1.5 ± 0.7	10.0 ± 7.1
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	11.5 ± 3.5	7.0 ± 2.8	**1.0 ± 0.0**	4.0 ± 4.2
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**11.0 ± 7.1**	13.0 ± 5.7	**1.0 ± 0.0**	10.0 ± 1.4
C4S	zs	3.8 ± 6.3	7.6 ± 7.2	6.4 ± 7.6	16.8 ± 0.4	13.6 ± 6.5	**1.8 ± 1.3**	8.8 ± 5.9
	cot	**3.0 ± 2.1**	9.6 ± 7.1	4.4 ± 2.7	15.4 ± 1.9	13.4 ± 6.1	10.0 ± 6.6	12.4 ± 6.6
	spp	6.6 ± 5.9	8.4 ± 5.9	6.2 ± 5.4	15.8 ± 2.7	12.0 ± 6.9	4.4 ± 3.8	**2.2 ± 1.1**
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	8.5 ± 9.2	14.5 ± 2.1	9.5 ± 7.8	1.5 ± 0.7	4.0 ± 4.2
	sc-cot	14.5 ± 2.1	**2.0 ± 1.4**	9.5 ± 6.4	14.5 ± 3.5	17.0 ± 0.0	4.0 ± 2.8	14.5 ± 2.1
	sc-spp	**3.5 ± 3.5**	8.5 ± 6.4	6.5 ± 7.8	15.5 ± 2.1	17.0 ± 0.0	4.0 ± 2.8	7.5 ± 7.8
C4S(T)	zs	**1.0 ± 0.0**	4.8 ± 5.9	3.6 ± 3.4	15.0 ± 2.9	15.4 ± 3.6	2.6 ± 1.8	5.4 ± 3.0
	cot	**1.2 ± 0.4**	5.0 ± 6.7	2.6 ± 1.8	16.0 ± 0.0	17.0 ± 0.0	6.0 ± 6.4	10.2 ± 4.8
	spp	3.0 ± 4.5	2.2 ± 1.3	2.2 ± 2.2	16.6 ± 0.5	14.4 ± 5.8	**2.0 ± 1.7**	7.8 ± 6.1
	sc-zs	3.5 ± 3.5	5.0 ± 1.4	1.5 ± 0.7	15.0 ± 1.4	**6.5 ± 3.5**	**1.0 ± 0.0**	9.5 ± 3.5
	sc-cot	6.5 ± 4.9	**1.0 ± 0.0**	1.5 ± 0.7	15.5 ± 2.1	17.0 ± 0.0	4.0 ± 2.8	2.5 ± 2.1
	sc-spp	**2.0 ± 0.0**	**2.0 ± 1.4**	6.0 ± 7.1	17.0 ± 0.0	11.5 ± 7.8	5.0 ± 1.4	7.0 ± 8.5
DS-R1	zs	4.8 ± 3.6	**1.2 ± 0.4**	9.0 ± 5.5	15.8 ± 0.8	17.0 ± 0.0	2.2 ± 1.6	7.2 ± 4.0
	cot	**1.8 ± 1.3**	**1.8 ± 0.8**	5.4 ± 4.9	14.6 ± 1.8	16.6 ± 0.9	3.0 ± 1.6	7.6 ± 5.1
	spp	3.4 ± 4.3	11.2 ± 3.0	3.6 ± 5.8	12.2 ± 3.0	16.6 ± 0.9	**1.6 ± 0.5**	4.2 ± 3.0
	sc-zs	1.5 ± 0.7	4.0 ± 4.2	6.0 ± 7.1	13.5 ± 3.5	17.0 ± 0.0	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	9.0 ± 8.5	10.0 ± 7.1	16.5 ± 0.7	17.0 ± 0.0	3.5 ± 2.1	7.0 ± 5.7
	sc-spp	5.0 ± 5.7	11.0 ± 8.5	7.5 ± 6.4	17.0 ± 0.0	17.0 ± 0.0	**1.0 ± 0.0**	1.5 ± 0.7
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.4 ± 1.8	15.4 ± 0.9	2.2 ± 2.2	11.6 ± 6.8
	cot	4.0 ± 6.7	**1.0 ± 0.0**	**1.0 ± 0.0**	15.6 ± 1.1	16.6 ± 0.9	1.2 ± 0.4	2.2 ± 2.7
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	15.0 ± 1.4	16.2 ± 1.1	5.4 ± 5.2	3.8 ± 3.9
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.0 ± 1.4	15.0 ± 0.0	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	11.5 ± 7.8	17.0 ± 0.0	2.5 ± 2.1	3.0 ± 2.8
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	16.5 ± 0.7	16.0 ± 1.4	**1.0 ± 0.0**	**1.0 ± 0.0**
M-L(24.07)	zs	13.4 ± 7.0	8.2 ± 7.6	12.6 ± 6.5	15.8 ± 1.6	14.4 ± 4.8	**3.4 ± 1.8**	7.4 ± 8.8
	cot	8.0 ± 7.9	**5.0 ± 6.5**	10.0 ± 7.4	16.4 ± 0.5	13.4 ± 7.0	12.8 ± 6.4	12.2 ± 6.9
	spp	7.4 ± 8.3	9.8 ± 8.1	11.2 ± 7.9	16.4 ± 1.3	17.0 ± 0.0	**4.8 ± 6.4**	11.0 ± 8.2
	sc-zs	**1.0 ± 0.0**	8.0 ± 9.9	9.0 ± 11.3	17.0 ± 0.0	17.0 ± 0.0	1.5 ± 0.7	17.0 ± 0.0
	sc-cot	**8.5 ± 10.6**	11.5 ± 7.8	16.0 ± 1.4	17.0 ± 0.0	16.0 ± 1.4	**8.5 ± 10.6**	12.5 ± 4.9
	sc-spp	16.5 ± 0.7	9.0 ± 11.3	15.0 ± 1.4	15.5 ± 2.1	17.0 ± 0.0	**1.0 ± 0.0**	9.0 ± 11.3

Table 6.8: Round # where the Agent understood the opponent's Strategy (sh)

model	prompt	zs	spp	cot	sh-alt srep	pp	mf	tft
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.8 ± 0.4	2.0 ± 0.0	2.6 ± 0.5	2.2 ± 0.8
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.2 ± 0.4	6.0 ± 5.8	1.8 ± 1.1	2.8 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.6 ± 0.5	2.0 ± 0.0	2.8 ± 0.4	1.8 ± 0.8
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	3.0 ± 1.4	2.5 ± 0.7	3.0 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	4.0 ± 2.8	2.0 ± 1.4	3.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	4.0 ± 2.8	2.5 ± 0.7	2.0 ± 1.4
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	11.6 ± 5.7	3.6 ± 1.3	3.4 ± 1.5
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.4 ± 0.5	8.8 ± 4.6	2.2 ± 0.8	4.6 ± 0.9
	spp	**1.0 ± 0.0**	3.6 ± 5.8	**1.0 ± 0.0**	1.8 ± 0.4	4.4 ± 4.3	3.4 ± 1.1	4.0 ± 1.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	8.0 ± 8.5	2.5 ± 0.7	2.0 ± 1.4
	sc-cot	**1.0 ± 0.0**	3.0 ± 2.8	**1.0 ± 0.0**	2.0 ± 0.0	1.5 ± 0.7	3.5 ± 2.1	3.0 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	3.5 ± 3.5	1.5 ± 0.7	3.0 ± 1.4	1.5 ± 0.7	3.0 ± 0.0
C3.7S(T)	zs	**1.0 ± 0.0**	1.2 ± 0.4	**1.0 ± 0.0**	1.6 ± 0.5	12.8 ± 3.9	5.0 ± 1.4	3.2 ± 1.1
	cot	2.2 ± 1.8	**1.0 ± 0.0**	**1.0 ± 0.0**	1.2 ± 0.4	3.2 ± 1.8	2.6 ± 0.9	2.8 ± 1.5
	spp	**1.0 ± 0.0**	1.6 ± 1.3	**1.0 ± 0.0**	1.8 ± 0.4	5.2 ± 5.0	2.0 ± 1.2	2.8 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 0.0	8.0 ± 8.5	3.0 ± 0.0	3.5 ± 2.1
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	1.5 ± 0.7	3.0 ± 2.8	2.0 ± 1.4
	sc-spp	1.5 ± 0.7	**1.0 ± 0.0**	**1.0 ± 0.0**	1.5 ± 0.7	8.0 ± 5.7	3.0 ± 0.0	3.0 ± 0.0
C4S	zs	3.6 ± 4.8	2.2 ± 1.3	1.4 ± 0.9	**1.2 ± 0.4**	7.4 ± 6.3	2.0 ± 1.0	2.2 ± 1.3
	cot	**1.0 ± 0.0**	3.0 ± 4.5	2.0 ± 2.2	1.8 ± 1.1	4.2 ± 5.6	2.2 ± 1.6	1.6 ± 1.3
	spp	2.0 ± 1.4	2.4 ± 1.3	**1.0 ± 0.0**	**1.0 ± 0.0**	5.6 ± 0.9	3.0 ± 1.6	1.6 ± 1.3
	sc-zs	2.0 ± 1.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	2.0 ± 1.4
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	5.5 ± 6.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	3.0 ± 2.8
	sc-spp	3.5 ± 3.5	**1.0 ± 0.0**	3.5 ± 3.5	**1.0 ± 0.0**	3.5 ± 3.5	3.5 ± 3.5	3.5 ± 3.5
C4S(T)	zs	1.2 ± 0.4	**1.0 ± 0.0**	1.2 ± 0.4	**1.0 ± 0.0**	6.2 ± 7.3	1.6 ± 0.9	1.4 ± 0.9
	cot	1.2 ± 0.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	4.0 ± 6.7	**1.0 ± 0.0**	**1.0 ± 0.0**
	spp	4.2 ± 5.2	1.6 ± 1.3	**1.0 ± 0.0**	**1.0 ± 0.0**	5.6 ± 5.1	**1.0 ± 0.0**	1.6 ± 1.3
	sc-zs	2.0 ± 1.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	3.5 ± 3.5	**1.0 ± 0.0**	8.5 ± 10.6	**1.0 ± 0.0**	5.0 ± 5.7
DS-R1	zs	**1.0 ± 0.0**	3.0 ± 2.8	1.4 ± 0.9	1.8 ± 1.8	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	spp	4.2 ± 7.2	2.2 ± 1.8	2.0 ± 2.2	**1.0 ± 0.0**	1.6 ± 1.3	**1.0 ± 0.0**	4.0 ± 6.7
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	6.4 ± 3.3	10.0 ± 5.7	4.2 ± 1.1	3.6 ± 1.8
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	3.8 ± 1.8	12.0 ± 4.5	2.4 ± 0.5	2.6 ± 0.5
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	3.4 ± 1.3	13.2 ± 5.2	3.0 ± 1.2	3.0 ± 1.2
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	5.0 ± 1.4	15.0 ± 1.4	2.5 ± 0.7	9.5 ± 9.2
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	4.0 ± 0.0	8.0 ± 8.5	2.0 ± 0.0	2.5 ± 0.7
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	4.5 ± 3.5	10.0 ± 8.5	2.5 ± 0.7	3.0 ± 1.4
M-L(24.07)	zs	15.6 ± 1.5	16.2 ± 0.8	**10.4 ± 7.7**	16.4 ± 0.5	13.0 ± 6.7	16.6 ± 0.5	17.0 ± 0.0
	cot	9.6 ± 7.9	7.4 ± 7.9	9.4 ± 7.3	11.4 ± 7.2	**5.2 ± 5.4**	13.6 ± 6.0	16.8 ± 0.4
	spp	10.0 ± 7.4	**7.2 ± 8.5**	**7.2 ± 6.5**	14.4 ± 4.2	16.0 ± 0.0	13.8 ± 6.1	16.6 ± 0.5
	sc-zs	16.0 ± 1.4	9.0 ± 9.9	**8.0 ± 9.9**	16.0 ± 0.0	16.0 ± 0.0	16.5 ± 0.7	16.5 ± 0.7
	sc-cot	**8.5 ± 10.6**	10.0 ± 5.7	16.5 ± 0.7	**8.5 ± 10.6**	16.0 ± 0.0	11.0 ± 7.1	11.5 ± 6.4
	sc-spp	16.5 ± 0.7	**9.0 ± 11.3**	16.5 ± 0.7	10.5 ± 9.2	16.5 ± 0.7	16.5 ± 0.7	17.0 ± 0.0

Table 6.9: Round # where the Agent understood the opponent's Strategy (sh-alt)

- defection, cooperation, ... if the game uses Prisoner's Dilemma strategy names
- hare, stag, ... if the game uses Stag Hunt strategy names

The pattern player appears somewhat hard for LLMs to grasp. Analyzing games with Prisoner's Dilemma payoff matrices (tables 6.6 and 6.9); the *Claude 4* model and *DeepSeek* were the only models to consistently understand their opponent's tactics across all prompt styles. In *Claude 3.5 Sonnet v2* **spp** elicits good results, while in *Claude 3.7 Sonnet* **cot** seems to be the better prompting style. On the other hand, games with Stag Hunt style payoff matrices (tables 6.7 and 6.8) indicate that LLMs faced problems in grasping the simple cyclical nature of the opponent's play-style. Values over 10.0 are observed across the board. Only some of the *Claude* models manage to achieve values that are not close to 16 or 17 indicating that they do eventually understand their opponent, but need a lot of experience.

Mf & Tft

Mf is a player that aims to maximize their own rewards off of the opponent's most frequent move, while **tft** aims to maximize their own rewards off of the opponent's most recent move.

In games with Prisoner's Dilemma payoff matrices, both these players will behave similarly and will end up defecting in every round (apart from the first one perhaps); thus, the only appropriate response from a rational player is to also always defect.

The last two columns of all tables of this section inform us that LLMs do manage to work with **mf** and **tft** strategies fairly early, since result values are quite small for all models (excluding *Mistral Large (24.07)* as the odd one out).

We aim to trace hints of reasoning abilities. One such hint - that comes from our analysis on "Total Points" for Stag Hunt - is an LLM that started with "Hare" deciding to switch to "Stag" after realizing that this will prove more profitable in future rounds. Essentially, what we are looking for are "Total Points" values between 32 and 96 in tables 6.3 and 6.4 that coincide with models understand the opponent's strategies in later rounds, i.e., medium values in tables 6.7 and 6.8. Comparing these two tables, we observe that this is indeed the case.

6.3 Cooperation Rates

We show the average cooperation rate for each player averaging results for all opponent types. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

Table 6.14 provides an average of cooperation rates over LLM players when faced only with other LLM players.

6.3.1 LLM vs LLM

LLM agents that play against each other are of the same LLM. These games are represented in the first three columns of the result matrices. Table 6.14 deals specifically with LLM vs LLM scenarios, so observations derived from it are quite impactful to this subsection.

As observed and hypothesized in sections 6.1 and 6.2 agents do typically reach agreement fairly early in the game and opt for cooperation and not defection as a strategy. Exempt from this rule are *Claude Sonnet 4* variants and *DeepSeek-R1*, which have fairly low cooperation rates. This can be attributed to their attempt for more advanced strategic play, which however may backfire, since they do not gather as many points as fully-cooperative players.

6.3.2 LLM vs non-LLM

The last 4 columns represent non-LLM players, that follow simple strategies. These players have been described in 5.7. Successful outcomes against them should be a good indication of reasoning abilities of the LLMs, since they imply an agent's ability to analyze their opponent and make informed decisions.

model	prompt	pd zs	spp	cot	srep	pp	mf	tft
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.4 ± 0.2	0.2 ± 0.1	0.2 ± 0.1
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.1	0.2 ± 0.1	0.1 ± 0.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.2	0.1 ± 0.0	0.2 ± 0.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.4 ± 0.1	0.2 ± 0.0	0.1 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.4 ± 0.2	0.2 ± 0.1	0.2 ± 0.1
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.1 ± 0.0	0.2 ± 0.1	0.2 ± 0.1
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.2	0.2 ± 0.1	0.2 ± 0.1
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.1	0.2 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.3	0.2 ± 0.0	0.2 ± 0.0
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.1	0.2 ± 0.0
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.1	0.1 ± 0.1	0.1 ± 0.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.2	0.2 ± 0.1	0.2 ± 0.1
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.1	0.2 ± 0.0	0.2 ± 0.0
C4S	zs	**0.6 ± 0.4**	0.2 ± 0.4	0.2 ± 0.4	0.0 ± 0.0	0.4 ± 0.2	0.1 ± 0.1	0.1 ± 0.0
	cot	0.1 ± 0.1	0.2 ± 0.0	**0.2 ± 0.2**	0.0 ± 0.0	0.2 ± 0.2	0.1 ± 0.0	0.1 ± 0.0
	spp	**0.5 ± 0.4**	0.4 ± 0.5	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	sc-zs	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.0 ± 0.0	**0.2 ± 0.4**	0.0 ± 0.0	0.1 ± 0.1
	sc-cot	**0.1 ± 0.1**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**0.1 ± 0.0**	0.0 ± 0.0
	sc-spp	**0.1 ± 0.1**	**0.1 ± 0.1**	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
C4S(T)	zs	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	**0.0 ± 0.1**	0.0 ± 0.0	0.0 ± 0.0
	cot	0.0 ± 0.0	**0.2 ± 0.3**	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.0
	spp	0.0 ± 0.1	**0.6 ± 0.4**	0.3 ± 0.3	0.0 ± 0.0	0.2 ± 0.2	0.0 ± 0.1	0.0 ± 0.0
	sc-zs	**0.0 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**0.0 ± 0.0**
	sc-cot	0.0 ± 0.0	**0.1 ± 0.1**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	sc-spp	**0.5 ± 0.7**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.0 ± 0.0	0.1 ± 0.1
DS-R1	zs	0.1 ± 0.0	0.1 ± 0.0	**0.1 ± 0.1**	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.0	0.1 ± 0.0
	cot	0.0 ± 0.0	**0.0 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**0.0 ± 0.1**	0.0 ± 0.0
	spp	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.0	**0.1 ± 0.2**	0.0 ± 0.0	0.0 ± 0.0
	sc-zs	**0.1 ± 0.0**	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
	sc-cot	**0.1 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	sc-spp	0.1 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	**0.2 ± 0.2**	0.0 ± 0.0	0.0 ± 0.0
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.4 ± 0.2	0.1 ± 0.0	0.1 ± 0.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.1	0.1 ± 0.0	0.1 ± 0.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.0	0.4 ± 0.2	0.1 ± 0.0	0.2 ± 0.0
M-L(24.07)	zs	0.8 ± 0.2	0.8 ± 0.2	**0.9 ± 0.3**	0.5 ± 0.0	0.5 ± 0.0	0.5 ± 0.1	0.5 ± 0.1
	cot	**1.0 ± 0.0**	0.6 ± 0.4	0.7 ± 0.4	0.1 ± 0.1	0.3 ± 0.2	0.3 ± 0.2	0.5 ± 0.1
	spp	**1.0 ± 0.0**	1.0 ± 0.1	0.9 ± 0.3	0.4 ± 0.3	**0.5 ± 0.1**	0.4 ± 0.2	0.5 ± 0.1
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**0.5 ± 0.0**	0.5 ± 0.0	0.5 ± 0.0	**0.5 ± 0.0**
	sc-cot	0.9 ± 0.2	**1.0 ± 0.0**	0.7 ± 0.4	0.3 ± 0.3	0.2 ± 0.4	0.0 ± 0.0	0.4 ± 0.3
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**0.5 ± 0.0**	0.5 ± 0.0	**0.8 ± 0.3**	0.5 ± 0.0

Table 6.10: Average Cooperation (Ratio of Cooperative Moves) (pd)

model	prompt	zs	spp	cot	srep	pp	mf	tft
					pd-alt			
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.2	**0.5 ± 0.0**	0.9 ± 0.0	0.6 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.1	0.4 ± 0.1	1.0 ± 0.0	0.6 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.2	0.3 ± 0.2	0.9 ± 0.0	0.7 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.2	0.5 ± 0.0	0.9 ± 0.0	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.0	**0.5 ± 0.0**	0.9 ± 0.0	0.7 ± 0.4
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.3	0.3 ± 0.3	1.0 ± 0.0	0.2 ± 0.0
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.2	**0.5 ± 0.0**	1.0 ± 0.0	0.7 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.1	0.5 ± 0.1	1.0 ± 0.0	0.6 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.1	0.2 ± 0.1	1.0 ± 0.0	0.5 ± 0.3
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.3	0.3 ± 0.2	1.0 ± 0.0	0.3 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.2	**0.5 ± 0.0**	1.0 ± 0.0	0.7 ± 0.5
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.2	0.3 ± 0.3	1.0 ± 0.0	0.3 ± 0.0
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.2	0.4 ± 0.1	0.7 ± 0.4	0.4 ± 0.3
	cot	**1.0 ± 0.0**	0.9 ± 0.1	**1.0 ± 0.0**	0.5 ± 0.1	0.4 ± 0.1	**1.0 ± 0.0**	0.7 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.1	0.4 ± 0.2	1.0 ± 0.0	0.7 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.7 ± 0.1	0.4 ± 0.2	0.6 ± 0.5	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.0	0.5 ± 0.0	0.9 ± 0.0	0.2 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.3	0.3 ± 0.1	1.0 ± 0.0	0.6 ± 0.6
C4S	zs	0.6 ± 0.5	**1.0 ± 0.0**	0.8 ± 0.4	0.4 ± 0.1	0.2 ± 0.2	0.6 ± 0.4	0.4 ± 0.4
	cot	**0.9 ± 0.1**	0.6 ± 0.4	0.5 ± 0.4	0.2 ± 0.1	0.2 ± 0.2	0.3 ± 0.4	0.3 ± 0.4
	spp	**0.8 ± 0.4**	0.3 ± 0.4	0.3 ± 0.4	0.2 ± 0.1	0.3 ± 0.2	0.6 ± 0.5	0.3 ± 0.4
	sc-zs	**1.0 ± 0.0**	0.9 ± 0.1	0.9 ± 0.0	0.2 ± 0.1	0.1 ± 0.0	0.9 ± 0.0	0.2 ± 0.1
	sc-cot	**0.5 ± 0.6**	0.2 ± 0.0	0.2 ± 0.0	0.3 ± 0.1	0.2 ± 0.0	**0.5 ± 0.7**	0.2 ± 0.0
	sc-spp	1.0 ± 0.0	0.9 ± 0.1	0.5 ± 0.5	0.3 ± 0.1	0.2 ± 0.2	**1.0 ± 0.0**	0.5 ± 0.4
C4S(T)	zs	**1.0 ± 0.0**	0.8 ± 0.4	0.8 ± 0.4	0.4 ± 0.1	0.3 ± 0.1	0.8 ± 0.4	0.5 ± 0.5
	cot	0.4 ± 0.5	0.6 ± 0.5	0.7 ± 0.5	0.3 ± 0.2	0.2 ± 0.1	0.4 ± 0.5	**0.7 ± 0.4**
	spp	0.8 ± 0.4	0.6 ± 0.5	0.8 ± 0.4	0.3 ± 0.1	0.3 ± 0.2	**0.9 ± 0.0**	0.8 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	1.0 ± 0.0	**0.7 ± 0.0**	**0.5 ± 0.0**	0.1 ± 0.1	0.2 ± 0.1
	sc-cot	0.5 ± 0.7	0.5 ± 0.7	**1.0 ± 0.0**	0.1 ± 0.1	0.0 ± 0.0	0.5 ± 0.7	0.6 ± 0.6
	sc-spp	**1.0 ± 0.0**	0.5 ± 0.6	0.5 ± 0.7	0.3 ± 0.0	0.2 ± 0.0	0.5 ± 0.7	0.1 ± 0.1
DS-R1	zs	0.3 ± 0.4	0.1 ± 0.0	0.1 ± 0.1	0.3 ± 0.1	0.2 ± 0.2	**0.4 ± 0.5**	0.3 ± 0.4
	cot	0.1 ± 0.1	0.1 ± 0.0	0.2 ± 0.4	**0.3 ± 0.2**	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.4
	spp	0.1 ± 0.1	0.1 ± 0.2	0.3 ± 0.4	0.2 ± 0.1	0.2 ± 0.1	0.2 ± 0.4	**0.3 ± 0.4**
	sc-zs	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.1	0.0 ± 0.0	**0.2 ± 0.2**	0.1 ± 0.0	0.1 ± 0.1
	sc-cot	**0.5 ± 0.7**	0.1 ± 0.1	0.1 ± 0.1	0.3 ± 0.2	0.1 ± 0.0	0.5 ± 0.7	**0.5 ± 0.7**
	sc-spp	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	**0.2 ± 0.3**	0.1 ± 0.1	0.1 ± 0.0	0.1 ± 0.1
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.2	**0.5 ± 0.0**	0.7 ± 0.4	0.3 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.1	0.3 ± 0.2	0.8 ± 0.4	0.3 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.2	0.4 ± 0.2	1.0 ± 0.0	0.8 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	**0.5 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.1	0.1 ± 0.0	**1.0 ± 0.0**	0.6 ± 0.6
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.0	**0.5 ± 0.0**	0.6 ± 0.6	**1.0 ± 0.0**
M-L(24.07)	zs	**1.0 ± 0.0**	0.9 ± 0.2	**1.0 ± 0.0**	0.6 ± 0.1	**0.5 ± 0.0**	0.9 ± 0.2	0.9 ± 0.3
	cot	0.9 ± 0.1	0.9 ± 0.2	**1.0 ± 0.0**	0.5 ± 0.1	0.4 ± 0.2	0.8 ± 0.3	0.8 ± 0.3
	spp	**1.0 ± 0.0**	0.7 ± 0.4	1.0 ± 0.0	0.6 ± 0.1	0.5 ± 0.1	0.9 ± 0.2	0.6 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.1	0.6 ± 0.0	**0.5 ± 0.0**	1.0 ± 0.0	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.2	**0.5 ± 0.0**	1.0 ± 0.0	0.8 ± 0.4
	sc-spp	**1.0 ± 0.0**	0.8 ± 0.3	**1.0 ± 0.0**	0.6 ± 0.1	**0.5 ± 0.0**	0.9 ± 0.0	0.7 ± 0.3

Table 6.11: Average Cooperation (Ratio of Cooperative Moves) (pd-alt)

		sh						
		zs	spp	cot	srep	pp	mf	tft
model	prompt							
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.1	0.3 ± 0.2	1.0 ± 0.0	0.5 ± 0.5
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.3 ± 0.1	0.5 ± 0.5	0.3 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.1	0.8 ± 0.4	0.8 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.3	1.0 ± 0.0	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.3 ± 0.2	0.1 ± 0.0	0.6 ± 0.6
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.0	0.1 ± 0.0	0.6 ± 0.6	0.6 ± 0.6
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.2	0.5 ± 0.1	1.0 ± 0.0	0.7 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.2	0.5 ± 0.0	1.0 ± 0.0	0.7 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.1	0.4 ± 0.2	1.0 ± 0.0	0.6 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.0	0.4 ± 0.2	0.9 ± 0.0	0.6 ± 0.6
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.0	0.5 ± 0.0	1.0 ± 0.0	0.2 ± 0.1
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.0	0.3 ± 0.3	1.0 ± 0.0	**1.0 ± 0.0**
C3.7S(T)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.1	0.4 ± 0.2	0.8 ± 0.4	0.5 ± 0.4
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.1	0.4 ± 0.2	1.0 ± 0.0	0.5 ± 0.5
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.5 ± 0.0	1.0 ± 0.0	0.6 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.2	0.4 ± 0.2	1.0 ± 0.0	0.2 ± 0.1
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.5 ± 0.0	**1.0 ± 0.0**	0.6 ± 0.6
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.3	0.3 ± 0.3	**1.0 ± 0.0**	0.2 ± 0.0
C4S	zs	**0.8 ± 0.3**	0.6 ± 0.4	0.7 ± 0.4	0.2 ± 0.1	0.3 ± 0.2	0.6 ± 0.5	0.5 ± 0.4
	cot	0.2 ± 0.4	0.3 ± 0.4	**0.4 ± 0.5**	0.3 ± 0.1	0.3 ± 0.2	0.2 ± 0.1	0.2 ± 0.2
	spp	0.3 ± 0.4	0.1 ± 0.1	0.6 ± 0.3	0.3 ± 0.1	0.4 ± 0.1	0.3 ± 0.4	**0.9 ± 0.0**
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	0.6 ± 0.5	0.2 ± 0.0	**0.5 ± 0.0**	1.0 ± 0.0	0.6 ± 0.6
	sc-cot	0.2 ± 0.2	0.4 ± 0.6	**0.5 ± 0.4**	0.0 ± 0.0	**0.5 ± 0.0**	0.1 ± 0.1	0.2 ± 0.0
	sc-spp	0.6 ± 0.6	0.1 ± 0.1	**0.6 ± 0.5**	0.4 ± 0.2	0.2 ± 0.1	0.6 ± 0.5	**0.6 ± 0.5**
C4S(T)	zs	0.6 ± 0.5	**0.8 ± 0.3**	0.6 ± 0.4	0.2 ± 0.2	0.3 ± 0.2	0.4 ± 0.5	0.3 ± 0.4
	cot	0.6 ± 0.5	**0.8 ± 0.3**	0.2 ± 0.4	0.4 ± 0.2	0.2 ± 0.1	0.3 ± 0.4	0.5 ± 0.3
	spp	0.8 ± 0.4	**0.9 ± 0.1**	0.7 ± 0.4	0.3 ± 0.1	0.3 ± 0.2	0.8 ± 0.4	0.4 ± 0.4
	sc-zs	0.5 ± 0.6	0.5 ± 0.5	1.0 ± 0.0	0.2 ± 0.1	0.5 ± 0.0	**1.0 ± 0.0**	0.2 ± 0.0
	sc-cot	0.2 ± 0.0	**1.0 ± 0.0**	0.5 ± 0.7	0.2 ± 0.2	0.3 ± 0.0	0.6 ± 0.5	0.5 ± 0.7
	sc-spp	**0.9 ± 0.0**	0.8 ± 0.1	0.6 ± 0.5	0.3 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.1 ± 0.2
DS-R1	zs	0.2 ± 0.4	**0.3 ± 0.4**	0.1 ± 0.0	0.2 ± 0.2	0.1 ± 0.1	0.3 ± 0.4	0.1 ± 0.1
	cot	**0.2 ± 0.4**	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	0.1 ± 0.1
	spp	0.2 ± 0.4	0.1 ± 0.0	**0.4 ± 0.5**	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.4	0.1 ± 0.0
	sc-zs	0.0 ± 0.0	**0.5 ± 0.5**	0.1 ± 0.0	0.2 ± 0.1	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
	sc-cot	0.0 ± 0.0	0.1 ± 0.0	**0.1 ± 0.0**	0.1 ± 0.1	**0.1 ± 0.1**	**0.1 ± 0.0**	**0.1 ± 0.0**
	sc-spp	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	**0.1 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.5 ± 0.1	0.5 ± 0.0	0.8 ± 0.4	0.5 ± 0.3
	cot	0.9 ± 0.2	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.2	0.3 ± 0.1	1.0 ± 0.0	0.8 ± 0.4
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.1	0.5 ± 0.0	0.7 ± 0.4	0.7 ± 0.4
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.1	0.5 ± 0.0	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.3 ± 0.2	0.2 ± 0.0	0.6 ± 0.6	0.6 ± 0.6
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.4 ± 0.4	0.5 ± 0.0	**1.0 ± 0.0**	**1.0 ± 0.0**
M-L(24.07)	zs	0.7 ± 0.2	0.9 ± 0.2	0.7 ± 0.2	0.6 ± 0.1	0.5 ± 0.0	**0.9 ± 0.1**	0.8 ± 0.3
	cot	0.6 ± 0.4	**0.8 ± 0.3**	0.5 ± 0.4	0.5 ± 0.2	0.5 ± 0.0	0.4 ± 0.3	0.5 ± 0.3
	spp	0.7 ± 0.3	0.7 ± 0.4	0.6 ± 0.3	0.5 ± 0.1	0.5 ± 0.1	**0.8 ± 0.2**	0.7 ± 0.3
	sc-zs	**1.0 ± 0.0**	0.8 ± 0.3	0.8 ± 0.2	**0.6 ± 0.0**	0.5 ± 0.0	1.0 ± 0.0	0.4 ± 0.1
	sc-cot	0.7 ± 0.4	0.4 ± 0.5	0.2 ± 0.3	0.4 ± 0.3	0.5 ± 0.0	**0.7 ± 0.4**	0.2 ± 0.2
	sc-spp	0.2 ± 0.4	0.8 ± 0.3	0.5 ± 0.0	0.6 ± 0.1	0.5 ± 0.0	**1.0 ± 0.0**	0.7 ± 0.4

Table 6.12: Average Cooperation (Ratio of Cooperative Moves) (sh)

model	prompt	zs	spp	cot	sh-alt srep	pp	mf	tft
C3.5Sv2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.2	0.2 ± 0.0	0.1 ± 0.1
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.1	0.1 ± 0.1	0.2 ± 0.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.2 ± 0.0
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.1	0.1 ± 0.1	0.2 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.1	0.2 ± 0.0	0.1 ± 0.1
C3.7S	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.4 ± 0.2	0.2 ± 0.1	0.2 ± 0.1
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.4 ± 0.2	0.1 ± 0.1	0.2 ± 0.0
	spp	**1.0 ± 0.0**	0.8 ± 0.3	1.0 ± 0.0	0.1 ± 0.0	0.2 ± 0.1	0.1 ± 0.0	0.2 ± 0.1
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.1	0.1 ± 0.0	0.3 ± 0.3	0.2 ± 0.0	0.1 ± 0.1
	sc-cot	**1.0 ± 0.0**	0.6 ± 0.6	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.3	0.2 ± 0.1	0.2 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	0.6 ± 0.6	0.1 ± 0.0	0.2 ± 0.1	0.1 ± 0.0	0.2 ± 0.0
C3.7S(T)	zs	**1.0 ± 0.0**	0.8 ± 0.4	**1.0 ± 0.0**	0.1 ± 0.0	0.5 ± 0.1	0.2 ± 0.0	0.2 ± 0.0
	cot	0.7 ± 0.5	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.2 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
	spp	**1.0 ± 0.1**	0.8 ± 0.4	**1.0 ± 0.1**	0.1 ± 0.0	0.2 ± 0.1	0.1 ± 0.1	0.2 ± 0.0
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.3	0.2 ± 0.0	0.2 ± 0.1
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.1 ± 0.1	0.2 ± 0.1	0.1 ± 0.1
	sc-spp	0.9 ± 0.1	**1.0 ± 0.0**	**1.0 ± 0.0**	0.1 ± 0.0	0.3 ± 0.2	0.2 ± 0.0	0.2 ± 0.0
C4S	zs	0.3 ± 0.4	0.1 ± 0.1	**0.3 ± 0.4**	0.1 ± 0.1	0.3 ± 0.2	0.1 ± 0.1	0.1 ± 0.1
	cot	0.0 ± 0.0	**0.3 ± 0.4**	0.0 ± 0.1	0.1 ± 0.0	0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	spp	0.1 ± 0.1	0.3 ± 0.3	0.2 ± 0.4	0.0 ± 0.0	**0.3 ± 0.1**	0.1 ± 0.0	0.1 ± 0.1
	sc-zs	**0.1 ± 0.1**	**0.1 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**0.1 ± 0.1**
	sc-cot	0.0 ± 0.0	0.0 ± 0.0	**0.1 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1
	sc-spp	0.1 ± 0.1	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	**0.2 ± 0.1**	0.1 ± 0.1	0.1 ± 0.1
C4S(T)	zs	0.0 ± 0.1	**0.2 ± 0.4**	0.1 ± 0.1	0.0 ± 0.0	0.2 ± 0.2	0.1 ± 0.1	0.1 ± 0.1
	cot	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**0.1 ± 0.2**	0.0 ± 0.0	0.0 ± 0.0
	spp	0.1 ± 0.1	**0.2 ± 0.4**	0.0 ± 0.0	0.1 ± 0.0	0.2 ± 0.2	0.0 ± 0.0	0.1 ± 0.1
	sc-zs	0.1 ± 0.0	0.0 ± 0.0	**0.5 ± 0.7**	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	sc-cot	**0.1 ± 0.1**	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	sc-spp	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.2	0.0 ± 0.0	**0.2 ± 0.4**	0.0 ± 0.0	0.1 ± 0.1
DS-R1	zs	0.1 ± 0.0	**0.1 ± 0.1**	0.1 ± 0.1	0.1 ± 0.0	0.1 ± 0.1	0.1 ± 0.0	0.1 ± 0.0
	cot	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	**0.1 ± 0.1**	0.1 ± 0.0	0.0 ± 0.0
	spp	0.0 ± 0.1	**0.1 ± 0.1**	0.1 ± 0.1	0.1 ± 0.0	0.0 ± 0.1	0.0 ± 0.0	0.1 ± 0.0
	sc-zs	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	0.0 ± 0.0	**0.0 ± 0.0**	0.0 ± 0.0
	sc-cot	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	0.0 ± 0.0	0.0 ± 0.0	**0.0 ± 0.0**
	sc-spp	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**0.0 ± 0.0**	**0.0 ± 0.0**
L3.3-70B	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.4 ± 0.1	0.2 ± 0.0	0.2 ± 0.1
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.5 ± 0.1	0.1 ± 0.0	0.2 ± 0.0
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.0	0.5 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.3 ± 0.2
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.0	0.3 ± 0.2	0.1 ± 0.0	0.2 ± 0.0
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.2 ± 0.1	0.4 ± 0.2	0.2 ± 0.0	0.2 ± 0.0
M-L(24.07)	zs	**0.7 ± 0.2**	0.6 ± 0.2	0.5 ± 0.0	**0.5 ± 0.0**	0.5 ± 0.0	0.5 ± 0.0	**0.5 ± 0.0**
	cot	**0.5 ± 0.2**	0.4 ± 0.2	0.4 ± 0.3	0.3 ± 0.2	0.3 ± 0.2	0.4 ± 0.1	0.3 ± 0.2
	spp	0.7 ± 0.1	0.6 ± 0.3	**0.7 ± 0.2**	0.5 ± 0.3	**0.6 ± 0.0**	0.5 ± 0.2	**0.5 ± 0.0**
	sc-zs	0.7 ± 0.1	**0.9 ± 0.0**	**0.9 ± 0.2**	0.5 ± 0.0	0.5 ± 0.0	**0.5 ± 0.0**	0.5 ± 0.0
	sc-cot	0.4 ± 0.1	0.5 ± 0.0	**0.6 ± 0.0**	0.2 ± 0.4	0.4 ± 0.1	0.3 ± 0.1	0.3 ± 0.0
	sc-spp	0.6 ± 0.1	**0.7 ± 0.3**	0.6 ± 0.1	0.4 ± 0.3	0.5 ± 0.0	0.5 ± 0.0	0.4 ± 0.0

Table 6.13: Average Cooperation (Ratio of Cooperative Moves) (sh-alt)

model	prompt	pd	pd-alt	sh	sh-alt
Claude 3.5 Sonnet v2	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
Claude 3.7 Sonnet	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.2
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.0 ± 0.1
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.3
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.3
Claude 3.7 Sonnet (Thinking)	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.2
	cot	**1.0 ± 0.0**	1.0 ± 0.1	**1.0 ± 0.0**	0.9 ± 0.3
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	0.9 ± 0.2
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	1.0 ± 0.1
Claude Sonnet 4	zs	0.3 ± 0.4	**0.8 ± 0.4**	0.7 ± 0.4	0.2 ± 0.3
	cot	0.1 ± 0.1	**0.7 ± 0.4**	0.3 ± 0.4	0.1 ± 0.2
	spp	0.3 ± 0.4	**0.5 ± 0.4**	0.3 ± 0.3	0.2 ± 0.3
	sc-zs	0.1 ± 0.1	**1.0 ± 0.1**	0.9 ± 0.3	0.1 ± 0.0
	sc-cot	0.0 ± 0.1	0.3 ± 0.3	**0.4 ± 0.4**	0.0 ± 0.1
	sc-spp	0.1 ± 0.1	**0.8 ± 0.3**	0.4 ± 0.4	0.1 ± 0.1
Claude Sonnet 4 (Thinking)	zs	0.0 ± 0.0	**0.9 ± 0.3**	0.7 ± 0.4	0.1 ± 0.2
	cot	0.1 ± 0.2	**0.6 ± 0.5**	0.5 ± 0.4	0.0 ± 0.0
	spp	0.3 ± 0.4	0.7 ± 0.4	**0.8 ± 0.3**	0.1 ± 0.2
	sc-zs	0.0 ± 0.0	**1.0 ± 0.0**	0.7 ± 0.4	0.2 ± 0.4
	sc-cot	0.0 ± 0.1	**0.7 ± 0.5**	0.5 ± 0.5	0.0 ± 0.1
	sc-spp	0.2 ± 0.4	0.6 ± 0.5	**0.8 ± 0.3**	0.0 ± 0.1
DeepSeek-R1	zs	0.1 ± 0.1	0.2 ± 0.2	**0.2 ± 0.3**	0.1 ± 0.0
	cot	0.0 ± 0.0	**0.1 ± 0.3**	0.1 ± 0.3	0.0 ± 0.0
	spp	0.1 ± 0.1	0.2 ± 0.3	**0.2 ± 0.4**	0.1 ± 0.1
	sc-zs	0.1 ± 0.0	0.1 ± 0.1	**0.2 ± 0.3**	0.0 ± 0.0
	sc-cot	0.0 ± 0.1	**0.2 ± 0.4**	0.1 ± 0.0	0.0 ± 0.0
	sc-spp	0.0 ± 0.1	0.1 ± 0.1	**0.1 ± 0.0**	0.0 ± 0.0
Llama 3.3 70B Instruct	zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	cot	**1.0 ± 0.0**	**1.0 ± 0.0**	1.0 ± 0.1	**1.0 ± 0.0**
	spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
Mistral Large (24.07)	zs	0.8 ± 0.2	**1.0 ± 0.1**	0.7 ± 0.2	0.6 ± 0.1
	cot	0.8 ± 0.4	**0.9 ± 0.1**	0.7 ± 0.4	0.5 ± 0.2
	spp	**1.0 ± 0.1**	0.9 ± 0.3	0.7 ± 0.3	0.6 ± 0.2
	sc-zs	**1.0 ± 0.0**	1.0 ± 0.1	0.9 ± 0.2	0.8 ± 0.1
	sc-cot	0.9 ± 0.2	**1.0 ± 0.0**	0.4 ± 0.4	0.5 ± 0.1
	sc-spp	**1.0 ± 0.0**	0.9 ± 0.2	0.5 ± 0.3	0.6 ± 0.1

Table 6.14: Average Cooperation in LLM vs LLM scenarios (Ratio of Cooperative Moves)

In counterfactual settings with typical Prisoner's Dilemma payoff matrices seen in tables 6.10 and 6.13 we see very low cooperation rates. This result is expected, since non-LLM players follow specific strategies, and LLM agents have to understand and counter or work with that strategy. Mutual cooperation is not going to emerge in these scenarios.

In Stag Hunt, however, mutual cooperation can emerge, since it provides pure strategy Nash Equilibria for the players. Players who may take advantage of this are **mf** and **tft** found in the last two columns of tables 6.11 and 6.12, where we do indeed observe high values of cooperation rates (often close to 1.0). This result, once again, indicates the existence of reasoning capabilities of LLMs.

6.4 Efficiency

Performance is often regarded as the most important factor of any technology with efficiency a classic companion. These two concepts are also present in the world of LLMs, where cost is often a function of tokens generated by the AI model. A simple efficiency metric is introduced:

$$efficiency = \frac{points}{tokens}$$

For a more visually pleasing appearance, results are scaled by 1000. In the following table, we represent efficiency as points (gathered by the LLM throughout the whole game) per kilo-tokens.

We show the average efficiency for each player averaging results for all opponent types. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

Across all models the following ranking of prompting styles emerges (from table 6.15):

1. **zs**: zero-shot prompting is the most efficient, followed by
2. **cot**: chain-of-thought, and lastly
3. **spp**: solo-performance prompting is the most inefficient.

Sc (self-consistency) when coupled with any other prompt style is less efficient, and comparing all prompting styles in their **sc** variants, again, the same ranking as before emerges.

The comparison of *Claude* models' **Thinking** and **non-thinking** versions is a point of interest. Claude's implementation of Large Reasoning Models (via their thinking options) is promising, as thinking models often manage similar or better efficiency than their default counterparts.

Lastly, an analysis of table 6.15 from the viewpoint of columns is worthwhile. The first and last columns represent the default Prisoner's Dilemma setting and the strategy counterfactual of that (meaning the counterfactual setting where payoffs are the same but strategies have different names - ("cooperate", "defect") have been mapped to ("stag", "hare") -), while the other two columns represent Stag Hunt variants. Efficiency is higher in the Stag Hunt games, because these allow for better mutual payoffs than Prisoner's Dilemma.

6.5 Failure Rate

Large Language Models (LLMs) are complex structures with various emerging skills, that are not problem-free. Despite our efforts to design an environment for LLMs that handles errors and unexpected or unwanted behavior (described in 4.1.2), LLMs still occasionally face issues with following our instructions. The environment and the error-correcting logic we follow is an expanded version of the gym-like environment used in [28].

All LLMs manage to have almost perfect "validity" rates as seen in table 6.16. When a game is played a validity value is subscribed to each of the 16 rounds that it includes. If any error occurred, this value will be "false". Table 6.16 simply depicts the average amount of valid rounds across all games.

model	prompt	pd	pd-alt	sh	sh-alt
Claude 3.5 Sonnet v2	zs	18.35 ± 10.24	**28.04 ± 12.12**	21.84 ± 12.24	15.56 ± 9.72
	cot	9.00 ± 4.05	**15.88 ± 8.80**	11.25 ± 5.37	7.85 ± 2.92
	spp	7.54 ± 2.98	**11.99 ± 4.73**	9.60 ± 4.14	5.93 ± 2.41
	sc-zs	6.37 ± 3.42	**9.88 ± 4.56**	7.63 ± 3.60	4.59 ± 2.71
	sc-cot	2.96 ± 1.46	**6.06 ± 4.01**	5.50 ± 5.26	2.58 ± 1.00
	sc-spp	2.54 ± 1.09	3.81 ± 1.63	**4.70 ± 4.25**	1.96 ± 0.76
Claude 3.7 Sonnet	zs	14.03 ± 7.88	**24.35 ± 12.64**	24.04 ± 12.10	13.16 ± 7.68
	cot	6.39 ± 2.64	10.14 ± 4.40	**10.68 ± 4.72**	6.17 ± 2.67
	spp	5.48 ± 3.66	**7.89 ± 5.63**	7.22 ± 3.47	4.57 ± 3.10
	sc-zs	4.88 ± 2.49	7.02 ± 4.64	**7.70 ± 4.79**	4.37 ± 2.50
	sc-cot	2.34 ± 1.32	**3.63 ± 1.78**	3.15 ± 1.60	2.01 ± 1.07
	sc-spp	2.36 ± 1.37	**3.21 ± 2.72**	2.67 ± 1.01	1.51 ± 0.69
Claude 3.7 Sonnet (Thinking)	zs	15.67 ± 6.23	22.93 ± 10.25	**30.01 ± 15.17**	19.17 ± 9.07
	cot	8.44 ± 3.39	**13.73 ± 5.86**	13.68 ± 6.63	7.86 ± 3.05
	spp	6.82 ± 2.85	**10.99 ± 4.98**	9.87 ± 4.58	6.10 ± 2.53
	sc-zs	5.35 ± 2.15	9.53 ± 4.24	**9.88 ± 5.87**	6.90 ± 2.89
	sc-cot	2.67 ± 1.19	4.16 ± 2.16	**4.45 ± 2.13**	2.71 ± 1.05
	sc-spp	2.21 ± 0.86	3.45 ± 1.66	**3.49 ± 1.84**	2.08 ± 0.91
Claude Sonnet 4	zs	7.45 ± 3.99	13.85 ± 8.83	**15.12 ± 12.56**	7.92 ± 4.45
	cot	4.15 ± 0.81	**6.34 ± 3.79**	4.53 ± 2.42	4.19 ± 0.82
	spp	3.85 ± 1.26	**4.49 ± 2.59**	4.46 ± 2.23	3.45 ± 0.81
	sc-zs	2.15 ± 0.48	4.86 ± 3.10	**5.45 ± 3.33**	1.99 ± 0.41
	sc-cot	1.36 ± 0.18	1.50 ± 0.94	**1.51 ± 0.47**	1.41 ± 0.25
	sc-spp	1.13 ± 0.35	**2.01 ± 0.99**	1.36 ± 0.81	1.04 ± 0.31
Claude Sonnet 4 (Thinking)	zs	11.59 ± 3.05	**24.90 ± 12.96**	19.88 ± 11.98	11.66 ± 3.63
	cot	5.84 ± 1.51	**9.81 ± 6.49**	8.43 ± 5.02	5.37 ± 1.18
	spp	3.99 ± 1.62	**6.68 ± 2.97**	6.04 ± 2.96	3.50 ± 0.87
	sc-zs	3.99 ± 1.00	**7.91 ± 3.90**	7.51 ± 3.45	4.19 ± 1.89
	sc-cot	1.88 ± 0.38	**3.43 ± 2.19**	3.13 ± 2.20	1.94 ± 0.54
	sc-spp	1.32 ± 0.36	**1.80 ± 0.90**	1.68 ± 0.87	1.07 ± 0.28
DeepSeek-R1	zs	**13.86 ± 5.08**	12.59 ± 5.53	13.47 ± 6.02	12.38 ± 4.01
	cot	8.48 ± 5.19	9.71 ± 4.00	11.77 ± 3.26	**12.48 ± 3.23**
	spp	8.22 ± 3.06	**8.79 ± 5.23**	8.51 ± 4.02	8.68 ± 2.39
	sc-zs	4.37 ± 1.43	3.66 ± 0.95	**4.72 ± 2.32**	4.51 ± 1.36
	sc-cot	2.74 ± 1.31	**4.70 ± 2.15**	4.41 ± 0.99	4.39 ± 1.34
	sc-spp	2.69 ± 0.86	2.29 ± 0.92	2.40 ± 0.37	**3.02 ± 1.29**
Llama 3.3 70B Instruct	zs	25.03 ± 9.53	38.08 ± 17.58	**40.90 ± 17.03**	**25.09 ± 9.76**
	cot	18.62 ± 7.48	28.85 ± 13.32	**30.11 ± 12.13**	17.57 ± 6.97
	spp	16.46 ± 6.41	**27.37 ± 9.98**	25.53 ± 10.79	16.63 ± 6.44
	sc-zs	8.37 ± 3.35	**15.04 ± 5.46**	14.98 ± 5.11	8.50 ± 3.24
	sc-cot	6.22 ± 2.55	**10.28 ± 4.55**	9.55 ± 4.55	5.68 ± 2.40
	sc-spp	5.50 ± 2.14	8.98 ± 3.58	**9.78 ± 3.30**	5.62 ± 2.20
Mistral Large (24.07)	zs	**27.04 ± 11.01**	**47.46 ± 18.77**	39.23 ± 17.11	23.17 ± 8.92
	cot	6.07 ± 3.11	**9.60 ± 4.27**	7.18 ± 3.49	5.57 ± 2.45
	spp	4.51 ± 2.62	**7.15 ± 4.36**	6.01 ± 4.00	4.45 ± 2.33
	sc-zs	9.65 ± 4.33	**16.86 ± 5.59**	13.38 ± 5.80	8.68 ± 3.49
	sc-cot	2.19 ± 0.84	**3.52 ± 1.21**	2.14 ± 1.12	1.72 ± 0.57
	sc-spp	1.84 ± 1.08	**2.71 ± 1.55**	2.10 ± 1.08	1.61 ± 1.01

Table 6.15: Average Efficiency (Points per kilo-token)

model	avg
Claude 3.5 Sonnet v2	100.0 ± 0.0
Claude 3.7 Sonnet	100.0 ± 0.0
Claude 3.7 Sonnet (Thinking)	100.0 ± 0.0
Claude Sonnet 4	100.0 ± 0.0
Claude Sonnet 4 (Thinking)	100.0 ± 0.0
DeepSeek-R1	99.1 ± 6.3
Llama 3.3 70B Instruct	100.0 ± 0.0
Mistral Large (24.07)	99.4 ± 6.5

Table 6.16: Average Valid Rate (% of Valid Outcomes)

At this point, it should be noted that most of the errors faced by the *Mistral Large (24.07)* model were due to its inability to follow our formatting directions. It often used **markdown** style formatting in its output even when specifically asked not to do so (as seen in the hints provided to models in 4.1.2). An example of such a failure is provided in 4.1.4.

6.6 Comparison with Other Work

Firstly, we may compare our results on cooperation rates with that of previous works. [8] found a 65.4% cooperation rate on the single-round variant of Prisoner's Dilemma for *GPT-3.5*. We have broadened the scope of cooperation rates accounting for all rounds of the game in the multi-round variant. Results are shown in tables 6.10, 6.11, 6.12, 6.13, and 6.14. This section focuses on AI agent vs AI agent scenarios so we mainly take into account table 6.14. On classic Prisoner's Dilemma scenario, LLMs achieve 80 to 100% cooperation rates with the exception of *Claude Sonnet 4* and *DeepSeek-R1* models. These models stay in the 10 to 30% cooperation rate range. As has been mentioned, this behavior can be attributed to the attempt of more advanced models to leverage more complex strategies in their gameplan.

Our LLM agents manage to replicate or approach results achieved in other works by non-LLM AI players that were designed and trained specifically for this task. More specifically, LASE agents described in [24] reach cooperation rates upwards of 50% depending on payoff matrix values. It is also shown that through training these agents reach a cooperation probability around 93%. Finally, [7], also, found that their custom agents tend to prefer cooperation "most of the time". These observations in conjunction with our own LLM agent results support the universality and ease of use that is typically associated with LLMs.

Lastly, [7] discusses cooperation rates in Prisoner's Dilemma and Public Goods (an extension of Prisoner's Dilemma). This work deals with LLM agents but they are prompted to assume specific emotional states (such as 'anger' or 'happiness'). They find that negative emotions (e.g., 'anger' and 'fear') lead to higher defection rates, while positive ones (e.g., 'happiness') consistently lead to higher cooperation rates. Such observations can be coupled with our own results on various LLMs to perhaps explain the mentioned discrepancies in cooperation rates.

6.7 Conclusions

- The result of [8] concerning cooperation rates is reproduced (in LLM vs LLM games). As they noted: "Higher cooperation rates could signal more trust and/or more weight on the joint payoff of the two agents, which is highest in the mutually cooperative outcome. In simulations where we did not instruct the model to give a direct answer, we saw it providing reasoning that choosing cooperation was important to maximize joint payoffs or to ensure that both parties were as well off as possible. Thus, in the prisoner's dilemma we see the LLM reaching toward a concern for others, rather than a strictly rational and self-interested player in the game."
- LLMs achieve points close to the expected result when following the mixed-strategy Nash Equilibrium

probability distribution over moves against the **srep** player. This result contradicts findings of [49]. That work highlighted a bias of LLMs in choosing a particular move over others, which is uncharacteristic of rational players. However, they had used single-round variants of Rock-Paper-Scissors for testing. LLMs, because of their acquired knowledge, have developed inherent biases (e.g., "I know that "rock" is a popular first move in Rock-Paper-Scissors) which dissipate in repeated games. We observe that LLMs tend to leave behind such biases as historical information about the previous rounds builds up and players refine their belief about their opponent's play-style.

- More complex LLM models decided to follow strategies with less **cooperation** getting lesser total points and reaching agreement (reflected in the **round of opponent comprehension**) later in the game. An indication of the over-analysis tendency of larger LLMs when facing simple tasks.
- The **pattern** player highlights that simpler LLMs perform worse than more complex ones. Also, in smaller LLMs using a more complex prompting style yields better results.
- The **tft** player was a good benchmark to showcase how LLMs can either maintain a good strategy (if they started with it) or adapt to a better strategy than the one they are currently using. This finding showcases that LLMs can both analyze potential strategies, but make decisions building on such analysis.
- efficiency of AI players drops as the prompt style gets more complex. This result will be quite different in the case of Rock-Paper-Scissors, a more difficult game to play, thus allowing more complex and larger LLMs to better use their capabilities.

Chapter 7

Results - Rock Paper Scissors

Rock-Paper-Scissors is a rather popular game that also attracts researchers and game-theorists [49, 16]. Its nature is slightly more complicated than Prisoner's Dilemma, offering its players three possible choices instead of just two.

We mention once more that agents play 24 rounds in each game and we test 4 scenarios (one base case of Rock-Paper-Scissors and three counterfactuals) with payoff matrices:

	Rock	Paper	Scissors
Rock	(0, 0)	(-1, 1)	(1, -1)
Paper	(1, -1)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(a) eq1

	Rock	Paper	Scissors
Rock	(0, 0)	(-3, 3)	(1, -1)
Paper	(3, -3)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(b) ba3

	Paper	Rock	Scissors
Paper	(0, 0)	(-1, 1)	(1, -1)
Rock	(1, -1)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(c) eq1-alt

	Paper	Rock	Scissors
Paper	(0, 0)	(-3, 3)	(1, -1)
Rock	(3, -3)	(0, 0)	(-1, 1)
Scissors	(-1, 1)	(1, -1)	(0, 0)

(d) ba3-alt

Table 7.1: Payoff matrices for the Rock-Paper-Scissors Counterfactual Settings.

This chapter will showcase results and comment on them.

7.1 Total Points

We show the average total points accumulated in each game setting from each player. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

7.1.1 LLM vs LLM

LLM agents that play against each other are of the same LLM. These games are represented in the first three columns of the result matrices.

When two LLM players face each other in a game like Rock-Paper-Scissors complex strategies can arise. These strategies may depend on the counterfactual settings (e.g., different payoffs) or other thoughts that LLMs have entirely. Rock-Paper-Scissors is a zero-sum game, unlike Prisoner's Dilemma and Stag Hunt. These two games provided players with pure strategy Nash Equilibria they could follow, however, that is

		eq1						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	6.6 ± 13.5	-1.2 ± 9.2	-1.8 ± 6.9	-0.4 ± 3.3	6.0 ± 10.1	11.2 ± 6.4	**16.6 ± 6.9**
	cot	0.8 ± 5.4	2.0 ± 5.7	5.4 ± 4.8	0.2 ± 5.3	3.8 ± 11.9	6.0 ± 3.7	**13.0 ± 6.8**
	spp	8.0 ± 9.9	7.2 ± 11.6	1.6 ± 6.9	-1.4 ± 2.2	9.4 ± 11.2	6.8 ± 1.5	**12.8 ± 7.6**
	sc-zs	-12.5 ± 12.0	-0.5 ± 21.9	0.5 ± 4.9	-0.5 ± 9.2	1.5 ± 2.1	9.0 ± 1.4	**21.5 ± 0.7**
	sc-cot	9.0 ± 0.0	-1.0 ± 2.8	0.0 ± 7.1	1.0 ± 4.2	10.5 ± 14.8	14.0 ± 12.7	**16.5 ± 2.1**
	sc-spp	-2.5 ± 21.9	-9.0 ± 17.0	0.0 ± 9.9	**7.0 ± 1.4**	-1.5 ± 3.5	5.0 ± 5.7	**12.5 ± 14.8**
C3.7S	zs	1.4 ± 9.6	-4.6 ± 11.3	-3.0 ± 11.9	-2.0 ± 5.0	**13.6 ± 12.5**	8.8 ± 2.6	4.8 ± 7.1
	cot	9.6 ± 5.2	2.6 ± 9.8	0.4 ± 5.5	1.4 ± 2.1	19.6 ± 1.3	8.2 ± 1.5	**20.2 ± 3.6**
	spp	3.6 ± 9.6	4.8 ± 6.0	-5.2 ± 8.6	-2.4 ± 2.3	**19.2 ± 1.3**	14.2 ± 7.2	19.0 ± 4.8
	sc-zs	-2.0 ± 28.3	0.0 ± 7.1	-15.5 ± 9.2	-4.0 ± 2.8	**14.0 ± 14.1**	6.5 ± 0.7	0.0 ± 0.0
	sc-cot	**20.5 ± 4.9**	-4.5 ± 4.9	1.5 ± 2.1	-3.5 ± 6.4	21.0 ± 0.0	9.5 ± 0.7	**22.0 ± 2.8**
	sc-spp	9.0 ± 12.7	-0.5 ± 0.7	-1.0 ± 5.7	2.5 ± 10.6	19.5 ± 2.1	12.0 ± 1.4	**21.0 ± 2.8**
C3.7S(T)	zs	0.2 ± 3.6	0.0 ± 5.3	-4.2 ± 7.3	-0.6 ± 3.6	**19.6 ± 1.9**	7.6 ± 2.8	17.6 ± 4.0
	cot	-2.0 ± 6.0	0.2 ± 3.2	-3.8 ± 5.1	-0.4 ± 3.2	**21.0 ± 0.0**	9.8 ± 4.4	15.6 ± 6.7
	spp	-3.2 ± 4.9	2.2 ± 10.9	-1.6 ± 2.9	5.2 ± 4.1	**20.8 ± 0.4**	8.4 ± 1.7	15.8 ± 9.3
	sc-zs	-5.0 ± 2.8	-7.5 ± 0.7	7.5 ± 3.5	-3.5 ± 4.9	**21.0 ± 0.0**	**15.5 ± 9.2**	1.0 ± 1.4
	sc-cot	-2.0 ± 12.7	-8.0 ± 9.9	-9.5 ± 6.4	3.5 ± 4.9	18.5 ± 0.7	14.0 ± 8.5	**22.0 ± 0.0**
	sc-spp	-2.0 ± 11.3	-9.5 ± 12.0	-16.5 ± 4.9	-4.0 ± 4.2	21.0 ± 0.0	9.0 ± 2.8	**22.5 ± 2.1**
C4S	zs	0.2 ± 4.8	-5.4 ± 5.3	-4.2 ± 8.1	2.4 ± 1.9	**18.2 ± 8.0**	10.6 ± 1.7	9.2 ± 8.4
	cot	3.4 ± 17.3	3.2 ± 4.6	4.4 ± 12.5	1.2 ± 3.8	**19.6 ± 2.1**	12.4 ± 6.1	12.2 ± 7.3
	spp	5.0 ± 8.5	-7.6 ± 16.9	4.6 ± 9.4	1.0 ± 6.4	**18.0 ± 3.7**	12.4 ± 3.5	10.4 ± 7.3
	sc-zs	12.5 ± 9.2	-19.5 ± 3.5	-11.5 ± 12.0	0.0 ± 0.0	**24.0 ± 0.0**	10.5 ± 2.1	18.0 ± 8.5
	sc-cot	-5.5 ± 6.4	1.0 ± 18.4	1.5 ± 14.8	3.5 ± 4.9	**21.0 ± 0.0**	9.5 ± 0.7	18.0 ± 8.5
	sc-spp	-0.5 ± 2.1	6.5 ± 9.2	-6.0 ± 4.2	1.0 ± 2.8	**21.0 ± 0.0**	**15.5 ± 9.2**	20.0 ± 1.4
C4S(T)	zs	-0.2 ± 2.3	-2.8 ± 7.3	-0.4 ± 15.9	1.0 ± 6.2	**19.2 ± 4.0**	11.8 ± 0.8	15.0 ± 10.6
	cot	7.6 ± 10.4	0.0 ± 3.4	-2.6 ± 11.5	2.0 ± 4.6	**19.8 ± 1.6**	10.0 ± 7.2	14.6 ± 8.1
	spp	-2.2 ± 5.4	-1.8 ± 10.6	-1.2 ± 5.7	0.4 ± 5.6	**20.8 ± 0.4**	8.4 ± 2.4	15.2 ± 8.3
	sc-zs	-5.0 ± 7.1	0.0 ± 2.8	0.5 ± 9.2	2.5 ± 0.7	21.0 ± 0.0	7.0 ± 5.7	**22.0 ± 2.8**
	sc-cot	11.5 ± 6.4	**8.5 ± 2.1**	**12.0 ± 2.8**	-0.5 ± 3.5	21.0 ± 0.0	11.0 ± 8.5	**23.0 ± 1.4**
	sc-spp	5.0 ± 1.4	-1.0 ± 0.0	1.0 ± 9.9	3.0 ± 1.4	21.0 ± 0.0	10.5 ± 2.1	**23.5 ± 0.7**
DS-R1	zs	1.2 ± 7.8	-1.0 ± 1.6	-3.2 ± 9.7	0.8 ± 3.0	5.8 ± 2.9	6.8 ± 5.0	**14.6 ± 3.8**
	cot	4.0 ± 8.2	5.0 ± 4.3	-4.2 ± 4.5	4.4 ± 4.2	12.4 ± 4.2	9.6 ± 3.0	**17.8 ± 3.4**
	spp	0.8 ± 6.3	-2.4 ± 5.9	-0.4 ± 6.3	-1.4 ± 3.5	6.2 ± 4.8	10.6 ± 2.2	**13.4 ± 5.4**
	sc-zs	5.0 ± 9.9	-3.5 ± 4.9	-0.5 ± 6.4	-3.5 ± 0.7	9.0 ± 8.5	6.5 ± 3.5	**14.5 ± 0.7**
	sc-cot	-0.5 ± 13.4	-3.5 ± 4.9	-3.5 ± 9.2	1.0 ± 1.4	15.0 ± 1.4	**15.5 ± 3.5**	**20.5 ± 2.1**
	sc-spp	-6.0 ± 5.7	-2.5 ± 2.1	-6.0 ± 1.4	0.0 ± 1.4	13.0 ± 2.8	7.0 ± 2.8	**20.0 ± 2.8**
L3.3-70B	zs	-0.8 ± 1.8	-0.4 ± 7.1	1.8 ± 5.4	-2.0 ± 4.7	**14.8 ± 12.6**	6.6 ± 2.1	1.0 ± 1.0
	cot	-0.6 ± 1.3	-1.6 ± 8.8	0.0 ± 2.4	0.6 ± 3.8	**19.0 ± 8.7**	8.2 ± 3.1	10.0 ± 8.9
	spp	1.0 ± 1.7	-1.6 ± 2.4	-1.4 ± 1.8	-0.4 ± 2.9	6.2 ± 7.8	**9.8 ± 6.0**	4.4 ± 4.0
	sc-zs	0.0 ± 0.0	-8.0 ± 11.3	2.0 ± 0.0	-1.0 ± 2.8	**12.5 ± 16.3**	9.0 ± 2.8	0.0 ± 0.0
	sc-cot	-0.5 ± 0.7	2.0 ± 2.8	1.5 ± 2.1	-0.5 ± 2.1	6.0 ± 2.8	7.0 ± 1.4	**16.5 ± 9.2**
	sc-spp	0.0 ± 0.0	-10.5 ± 16.3	-0.5 ± 2.1	5.5 ± 0.7	**19.0 ± 7.1**	12.5 ± 2.1	2.0 ± 2.8
M-L(24.07)	zs	0.0 ± 1.4	-10.6 ± 7.7	-3.8 ± 2.8	0.8 ± 5.2	**20.8 ± 7.2**	5.4 ± 3.6	0.8 ± 0.8
	cot	8.8 ± 5.4	0.0 ± 2.6	-4.2 ± 10.7	0.4 ± 4.2	-1.0 ± 12.9	1.6 ± 4.2	**16.0 ± 9.5**
	spp	9.2 ± 5.4	7.0 ± 13.2	0.6 ± 8.4	0.8 ± 1.5	6.8 ± 8.2	2.2 ± 4.1	**12.6 ± 6.6**
	sc-zs	0.5 ± 0.7	-4.0 ± 2.8	-13.5 ± 0.7	1.0 ± 1.4	**24.0 ± 0.0**	0.5 ± 0.7	0.0 ± 0.0
	sc-cot	11.5 ± 10.6	5.5 ± 3.5	-2.5 ± 30.4	-0.5 ± 0.7	11.5 ± 17.7	-0.5 ± 7.8	**17.5 ± 7.8**
	sc-spp	-0.5 ± 0.7	-0.5 ± 0.7	2.5 ± 17.7	0.0 ± 4.2	6.0 ± 7.1	-4.0 ± 2.8	**6.5 ± 21.9**

Table 7.2: Total Points Averaged Over All Iterations (eq1)

		eq1-alt						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	-1.6 ± 3.5	-3.4 ± 2.3	-0.8 ± 3.2	-0.4 ± 3.4	-4.4 ± 2.5	**9.0 ± 2.6**	5.2 ± 7.1
	cot	0.8 ± 8.0	-2.8 ± 6.7	-0.4 ± 1.7	-5.0 ± 4.7	-9.0 ± 10.9	**4.4 ± 3.9**	-0.8 ± 3.4
	spp	1.6 ± 3.7	-2.6 ± 1.5	1.8 ± 4.6	0.2 ± 5.0	-8.0 ± 5.0	**6.6 ± 3.9**	2.0 ± 4.4
	sc-zs	3.0 ± 5.7	-2.0 ± 4.2	3.5 ± 2.1	3.5 ± 2.1	-12.0 ± 2.8	7.0 ± 1.4	**12.0 ± 1.4**
	sc-cot	-2.5 ± 0.7	-2.5 ± 0.7	-4.5 ± 12.0	0.5 ± 2.1	-20.5 ± 0.7	4.0 ± 0.0	**5.5 ± 4.9**
	sc-spp	-9.0 ± 15.6	5.0 ± 11.3	0.5 ± 6.4	**4.5 ± 2.1**	-17.0 ± 5.7	**9.0 ± 1.4**	3.0 ± 0.0
C3.7S	zs	-4.8 ± 5.5	-2.4 ± 9.2	0.4 ± 8.2	-0.8 ± 5.5	-12.4 ± 11.9	10.2 ± 1.3	**12.0 ± 6.0**
	cot	4.2 ± 6.4	1.4 ± 2.5	-3.4 ± 5.9	-0.2 ± 3.5	**18.4 ± 7.0**	**13.8 ± 5.0**	12.0 ± 12.4
	spp	0.6 ± 5.9	5.6 ± 5.5	-2.2 ± 10.0	-2.4 ± 5.8	**18.8 ± 2.3**	10.6 ± 1.3	17.8 ± 3.3
	sc-zs	-6.0 ± 1.4	-2.5 ± 9.2	0.0 ± 5.7	3.0 ± 1.4	-8.0 ± 22.6	7.0 ± 2.8	**7.5 ± 2.1**
	sc-cot	-1.0 ± 1.4	2.5 ± 0.7	7.0 ± 4.2	3.0 ± 2.8	**18.5 ± 7.8**	10.0 ± 2.8	13.5 ± 12.0
	sc-spp	6.0 ± 7.1	5.0 ± 21.2	4.0 ± 0.0	-3.5 ± 9.2	15.0 ± 0.0	11.5 ± 0.7	**16.5 ± 2.1**
C3.7S(T)	zs	-1.6 ± 5.5	-2.6 ± 8.0	-0.4 ± 3.5	1.0 ± 2.3	2.2 ± 14.8	**6.4 ± 1.9**	4.6 ± 3.4
	cot	0.8 ± 3.6	0.0 ± 4.7	-4.0 ± 2.2	-1.6 ± 4.6	8.8 ± 17.3	9.2 ± 2.2	**13.2 ± 6.1**
	spp	4.2 ± 8.5	-0.2 ± 7.7	-0.8 ± 6.9	0.6 ± 3.4	**18.8 ± 6.1**	9.4 ± 1.8	17.4 ± 5.7
	sc-zs	-8.5 ± 7.8	-11.5 ± 2.1	1.5 ± 13.4	0.5 ± 3.5	18.0 ± 8.5	10.0 ± 1.4	**20.5 ± 2.1**
	sc-cot	1.5 ± 2.1	-3.0 ± 4.2	1.5 ± 3.5	3.5 ± 0.7	14.0 ± 14.1	7.5 ± 2.1	**24.0 ± 0.0**
	sc-spp	**12.5 ± 3.5**	**7.0 ± 2.8**	-4.5 ± 9.2	-4.0 ± 1.4	**21.5 ± 3.5**	9.0 ± 1.4	10.0 ± 12.7
C4S	zs	-2.6 ± 5.6	-1.0 ± 2.1	-2.8 ± 4.8	-0.6 ± 4.2	-2.6 ± 17.9	**7.4 ± 1.9**	3.8 ± 9.1
	cot	1.2 ± 2.7	-2.6 ± 5.9	-4.6 ± 5.9	-1.8 ± 4.6	**15.8 ± 5.4**	9.0 ± 0.7	8.6 ± 4.2
	spp	1.0 ± 11.1	4.4 ± 8.4	2.0 ± 5.8	-2.4 ± 4.4	**16.4 ± 2.6**	1.2 ± 12.8	-1.4 ± 6.8
	sc-zs	-0.5 ± 3.5	-7.5 ± 0.7	-0.5 ± 2.1	0.0 ± 2.8	**17.0 ± 1.4**	7.5 ± 0.7	-2.5 ± 4.9
	sc-cot	1.0 ± 1.4	3.5 ± 7.8	-3.0 ± 2.8	1.0 ± 11.3	**20.0 ± 4.2**	10.0 ± 0.0	3.5 ± 3.5
	sc-spp	1.0 ± 5.7	-2.0 ± 0.0	-2.5 ± 2.1	1.0 ± 2.8	**17.5 ± 3.5**	9.0 ± 1.4	2.5 ± 0.7
C4S(T)	zs	-3.2 ± 1.3	0.6 ± 2.9	-1.2 ± 3.1	-1.6 ± 3.4	-7.2 ± 14.5	**5.6 ± 3.1**	3.2 ± 3.3
	cot	0.0 ± 3.2	0.8 ± 3.3	2.4 ± 4.3	1.6 ± 4.3	**12.8 ± 8.2**	9.4 ± 1.9	4.6 ± 9.8
	spp	0.0 ± 7.5	0.8 ± 6.4	-1.2 ± 2.4	-5.6 ± 4.7	**13.0 ± 8.0**	8.4 ± 2.7	10.0 ± 13.5
	sc-zs	1.0 ± 2.8	-6.0 ± 1.4	-9.0 ± 8.5	-2.0 ± 2.8	-11.5 ± 6.4	**6.5 ± 3.5**	-4.0 ± 0.0
	sc-cot	6.5 ± 3.5	-3.0 ± 4.2	0.0 ± 2.8	-2.0 ± 2.8	**22.5 ± 2.1**	11.0 ± 0.0	11.0 ± 9.9
	sc-spp	4.5 ± 3.5	-4.0 ± 8.5	4.0 ± 7.1	1.5 ± 6.4	**13.5 ± 4.9**	9.0 ± 1.4	6.5 ± 20.5
DS-R1	zs	2.0 ± 3.5	0.4 ± 2.7	2.6 ± 3.2	1.2 ± 2.2	-7.4 ± 2.5	**6.2 ± 2.0**	4.8 ± 2.3
	cot	0.8 ± 3.8	2.8 ± 5.1	0.2 ± 4.8	1.0 ± 4.8	-4.0 ± 4.6	**6.4 ± 2.1**	5.0 ± 4.6
	spp	-3.0 ± 2.5	0.2 ± 1.9	-2.4 ± 5.3	0.0 ± 2.8	-3.8 ± 6.2	**4.6 ± 1.7**	3.2 ± 1.3
	sc-zs	-1.5 ± 2.1	-0.5 ± 3.5	2.5 ± 2.1	-6.0 ± 2.8	-6.5 ± 0.7	**6.0 ± 0.0**	2.5 ± 0.7
	sc-cot	-3.0 ± 1.4	1.0 ± 0.0	4.5 ± 3.5	-2.5 ± 2.1	-0.5 ± 4.9	**9.0 ± 0.0**	6.5 ± 0.7
	sc-spp	2.5 ± 2.1	3.0 ± 4.2	1.0 ± 1.4	0.0 ± 2.8	-8.5 ± 0.7	**6.5 ± 0.7**	6.0 ± 2.8
L3.3-70B	zs	-1.4 ± 2.8	-0.6 ± 1.7	0.6 ± 1.8	0.4 ± 1.7	-7.4 ± 9.4	**7.0 ± 1.6**	4.8 ± 4.5
	cot	-0.6 ± 1.8	0.8 ± 3.2	-0.2 ± 1.6	0.0 ± 1.6	-8.2 ± 2.5	**7.4 ± 1.7**	3.2 ± 2.9
	spp	0.0 ± 1.9	3.2 ± 4.5	2.6 ± 4.0	1.0 ± 2.0	-10.4 ± 8.9	5.0 ± 4.0	**5.4 ± 4.3**
	sc-zs	0.0 ± 0.0	-2.5 ± 0.7	4.0 ± 2.8	3.5 ± 3.5	-14.0 ± 14.1	9.5 ± 2.1	**10.5 ± 2.1**
	sc-cot	0.0 ± 5.7	1.5 ± 6.4	-1.5 ± 3.5	-3.0 ± 1.4	-14.0 ± 14.1	**4.0 ± 1.4**	1.0 ± 2.8
	sc-spp	0.0 ± 0.0	-1.5 ± 2.1	1.0 ± 2.8	1.0 ± 2.8	-6.0 ± 2.8	**8.5 ± 0.7**	4.0 ± 1.4
M-L(24.07)	zs	-0.2 ± 4.2	-3.6 ± 5.2	2.4 ± 5.2	-0.2 ± 3.0	-12.4 ± 9.5	0.0 ± 2.4	**8.4 ± 1.3**
	cot	-3.2 ± 5.4	2.4 ± 5.5	0.4 ± 5.5	-2.4 ± 4.8	-6.2 ± 7.1	3.6 ± 5.9	**5.0 ± 3.7**
	spp	1.4 ± 8.0	-3.0 ± 5.0	0.8 ± 7.2	2.2 ± 4.3	-4.8 ± 5.4	**9.0 ± 10.7**	5.4 ± 6.7
	sc-zs	0.0 ± 0.0	-0.5 ± 0.7	-5.0 ± 12.7	-0.5 ± 4.9	-11.5 ± 17.7	-0.5 ± 0.7	**9.0 ± 4.2**
	sc-cot	-3.0 ± 1.4	2.5 ± 3.5	**8.0 ± 2.8**	0.0 ± 1.4	-14.5 ± 13.4	**12.5 ± 13.4**	11.0 ± 0.0
	sc-spp	-0.5 ± 0.7	0.5 ± 2.1	4.5 ± 4.9	3.0 ± 1.4	-22.0 ± 2.8	7.5 ± 17.7	**19.0 ± 4.2**

Table 7.3: Total Points Averaged Over All Iterations (eq1-alt)

		ba3						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	3.8 ± 16.1	-1.2 ± 25.5	-5.0 ± 12.8	0.2 ± 5.6	-4.0 ± 17.2	9.2 ± 6.2	**29.6 ± 10.7**
	cot	-3.0 ± 21.6	11.6 ± 23.1	7.0 ± 15.1	-0.4 ± 6.3	-7.0 ± 5.1	6.8 ± 2.8	**24.2 ± 14.0**
	spp	**20.2 ± 11.2**	12.4 ± 11.8	7.4 ± 8.5	3.4 ± 3.9	1.6 ± 17.7	6.4 ± 2.1	13.2 ± 12.8
	sc-zs	-14.5 ± 21.9	3.5 ± 3.5	17.0 ± 5.7	2.0 ± 9.9	35.0 ± 0.0	18.0 ± 1.4	**40.0 ± 0.0**
	sc-cot	-0.5 ± 17.7	8.0 ± 15.6	-18.5 ± 26.2	0.0 ± 0.0	16.0 ± 26.9	18.5 ± 9.2	**32.5 ± 4.9**
	sc-spp	4.5 ± 9.2	-24.5 ± 2.1	**33.0 ± 8.5**	-3.0 ± 0.0	34.5 ± 0.7	16.0 ± 18.4	**35.5 ± 0.7**
C3.7S	zs	2.6 ± 20.1	-3.0 ± 13.8	-1.6 ± 11.1	-3.8 ± 6.4	**35.0 ± 5.3**	13.6 ± 12.3	31.4 ± 7.3
	cot	12.2 ± 8.6	-9.2 ± 8.1	-4.4 ± 11.5	0.0 ± 2.9	**34.6 ± 0.9**	16.6 ± 14.9	27.6 ± 11.9
	spp	-6.8 ± 18.5	-4.4 ± 11.8	-5.2 ± 12.7	-3.8 ± 4.6	31.8 ± 4.1	10.6 ± 8.7	**34.6 ± 2.9**
	sc-zs	10.5 ± 12.0	-34.0 ± 7.1	-6.5 ± 13.4	**9.0 ± 8.5**	11.0 ± 2.8	9.5 ± 7.8	**37.0 ± 4.2**
	sc-cot	12.0 ± 2.8	-12.0 ± 15.6	0.5 ± 6.4	-1.0 ± 9.9	35.0 ± 0.0	9.0 ± 1.4	**38.0 ± 2.8**
	sc-spp	7.5 ± 33.2	-5.5 ± 6.4	1.5 ± 4.9	2.0 ± 7.1	34.0 ± 1.4	**31.5 ± 0.7**	**36.5 ± 0.7**
C3.7S(T)	zs	2.4 ± 7.6	7.2 ± 9.0	-3.4 ± 8.3	4.6 ± 5.9	**28.0 ± 4.6**	10.6 ± 0.9	10.8 ± 3.0
	cot	7.0 ± 7.8	-3.2 ± 12.9	-4.4 ± 3.4	-0.6 ± 3.6	**34.8 ± 0.4**	17.2 ± 12.3	30.0 ± 7.4
	spp	-1.2 ± 14.7	-5.6 ± 12.9	-6.6 ± 14.1	2.8 ± 3.0	**34.6 ± 0.5**	28.8 ± 13.4	31.4 ± 5.2
	sc-zs	-8.5 ± 4.9	-4.0 ± 7.1	-6.5 ± 7.8	2.5 ± 4.9	**35.0 ± 0.0**	13.0 ± 2.8	33.0 ± 9.9
	sc-cot	3.5 ± 0.7	7.5 ± 9.2	7.5 ± 0.7	-1.5 ± 2.1	**35.0 ± 0.0**	9.5 ± 6.4	34.5 ± 2.1
	sc-spp	-7.0 ± 1.4	-18.0 ± 1.4	5.0 ± 5.7	4.0 ± 7.1	34.5 ± 0.7	21.5 ± 14.8	**37.0 ± 0.0**
C4S	zs	-3.0 ± 12.7	-1.8 ± 10.5	-4.0 ± 19.1	-1.8 ± 7.4	**34.4 ± 0.5**	11.8 ± 4.0	23.0 ± 14.5
	cot	-2.8 ± 20.2	-9.0 ± 17.4	6.8 ± 7.2	4.0 ± 5.5	**31.8 ± 4.4**	17.4 ± 11.1	23.2 ± 7.8
	spp	13.8 ± 7.6	3.6 ± 10.5	4.8 ± 6.5	3.0 ± 8.0	**28.2 ± 13.5**	13.0 ± 5.4	20.4 ± 15.5
	sc-zs	-13.5 ± 27.6	1.0 ± 7.1	-10.0 ± 2.8	1.0 ± 14.1	**35.0 ± 0.0**	16.5 ± 4.9	11.0 ± 5.7
	sc-cot	7.0 ± 19.8	-3.5 ± 0.7	-7.0 ± 9.9	-1.0 ± 2.8	**36.0 ± 2.8**	17.5 ± 14.8	20.0 ± 28.3
	sc-spp	-3.0 ± 22.6	0.0 ± 5.7	3.5 ± 44.5	4.5 ± 2.1	**35.0 ± 0.0**	12.5 ± 6.4	0.0 ± 0.0
C4S(T)	zs	-8.8 ± 12.6	-9.2 ± 4.5	-6.0 ± 11.0	0.6 ± 2.2	**30.6 ± 6.3**	13.8 ± 3.2	12.0 ± 16.7
	cot	4.0 ± 10.1	6.4 ± 7.6	1.6 ± 10.7	-1.6 ± 5.3	33.6 ± 2.6	9.0 ± 7.2	**33.8 ± 4.5**
	spp	2.8 ± 15.9	3.4 ± 12.0	-0.6 ± 21.2	-3.2 ± 4.0	**34.0 ± 1.4**	12.0 ± 1.6	28.8 ± 9.2
	sc-zs	-5.0 ± 9.9	-9.5 ± 3.5	-10.0 ± 29.7	-5.5 ± 0.7	**32.5 ± 3.5**	17.0 ± 5.7	28.0 ± 12.7
	sc-cot	-5.5 ± 20.5	3.0 ± 1.4	-26.5 ± 12.0	-4.0 ± 14.1	27.5 ± 3.5	12.0 ± 4.2	**37.0 ± 1.4**
	sc-spp	10.5 ± 4.9	-4.5 ± 12.0	3.5 ± 2.1	2.0 ± 7.1	**36.0 ± 1.4**	10.5 ± 4.9	25.0 ± 21.2
DS-R1	zs	-10.4 ± 15.5	1.6 ± 6.9	9.8 ± 10.1	1.0 ± 6.4	13.8 ± 12.8	8.4 ± 4.9	**29.4 ± 4.2**
	cot	9.0 ± 8.4	-8.6 ± 14.6	-0.4 ± 10.6	1.2 ± 7.8	14.8 ± 5.4	11.4 ± 6.3	**20.2 ± 3.8**
	spp	-7.8 ± 9.4	3.0 ± 12.8	2.0 ± 6.8	-1.2 ± 9.2	12.4 ± 10.6	9.8 ± 3.8	**17.0 ± 12.9**
	sc-zs	4.5 ± 10.6	11.5 ± 9.2	5.5 ± 4.9	-2.0 ± 8.5	22.0 ± 11.3	18.5 ± 6.4	**36.5 ± 0.7**
	sc-cot	8.5 ± 0.7	-6.5 ± 19.1	-5.5 ± 4.9	**9.0 ± 8.5**	18.0 ± 1.4	11.5 ± 0.7	**24.5 ± 3.5**
	sc-spp	-14.5 ± 12.0	-8.5 ± 4.9	-2.0 ± 8.5	4.5 ± 3.5	**25.0 ± 8.5**	19.0 ± 4.2	23.5 ± 2.1
L3.3-70B	zs	0.4 ± 17.6	-3.8 ± 16.1	12.2 ± 15.0	2.0 ± 5.3	**15.2 ± 14.5**	11.6 ± 8.2	1.6 ± 3.4
	cot	0.8 ± 7.3	0.4 ± 14.2	2.6 ± 8.5	0.2 ± 1.9	**37.0 ± 6.7**	17.0 ± 0.7	1.6 ± 1.1
	spp	-0.2 ± 1.8	-0.4 ± 10.3	2.6 ± 4.7	0.6 ± 7.1	9.2 ± 5.6	10.4 ± 5.9	**13.2 ± 15.7**
	sc-zs	-10.0 ± 24.0	13.0 ± 18.4	-1.0 ± 1.4	3.0 ± 7.1	**24.5 ± 21.9**	17.5 ± 12.0	-1.0 ± 1.4
	sc-cot	-17.0 ± 25.5	-0.5 ± 3.5	1.5 ± 2.1	2.5 ± 6.4	**20.5 ± 27.6**	13.5 ± 0.7	1.0 ± 1.4
	sc-spp	**27.0 ± 2.8**	3.0 ± 2.8	2.5 ± 0.7	-4.5 ± 9.2	2.5 ± 2.1	13.5 ± 16.3	16.5 ± 26.2
M-L(24.07)	zs	-0.2 ± 2.8	-7.8 ± 10.2	-5.4 ± 4.8	-1.0 ± 2.7	**34.8 ± 11.6**	1.8 ± 13.0	1.8 ± 1.6
	cot	**15.6 ± 11.5**	-3.8 ± 13.7	-1.6 ± 20.9	-0.6 ± 4.2	8.6 ± 7.1	6.8 ± 7.4	14.0 ± 16.5
	spp	9.0 ± 19.0	-13.0 ± 19.1	2.6 ± 13.7	2.2 ± 6.4	-4.8 ± 7.7	-3.0 ± 5.1	**27.8 ± 10.4**
	sc-zs	0.0 ± 0.0	-18.5 ± 17.7	-9.0 ± 5.7	4.5 ± 4.9	**20.5 ± 20.5**	2.5 ± 0.7	0.5 ± 0.7
	sc-cot	11.0 ± 15.6	**17.5 ± 31.8**	3.0 ± 15.6	0.5 ± 2.1	15.5 ± 21.9	13.0 ± 2.8	**22.0 ± 25.5**
	sc-spp	5.0 ± 11.3	-7.5 ± 23.3	-8.5 ± 30.4	3.5 ± 7.8	11.0 ± 17.0	-1.0 ± 21.2	**35.5 ± 3.5**

Table 7.4: Total Points Averaged Over All Iterations (ba3)

		ba3-alt						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	0.6 ± 6.0	2.2 ± 10.3	-1.0 ± 10.9	-3.4 ± 4.7	-8.6 ± 22.2	10.6 ± 5.0	**11.8 ± 12.1**
	cot	2.0 ± 6.8	-5.8 ± 15.9	-4.2 ± 13.0	-0.2 ± 5.4	-8.8 ± 16.5	6.2 ± 11.4	**12.2 ± 7.4**
	spp	-1.4 ± 11.0	5.0 ± 10.2	2.6 ± 9.8	3.6 ± 3.9	-25.8 ± 10.0	**14.0 ± 4.8**	11.8 ± 7.7
	sc-zs	-7.0 ± 8.5	-7.5 ± 4.9	4.5 ± 0.7	0.5 ± 6.4	-17.0 ± 19.8	**8.5 ± 4.9**	7.0 ± 5.7
	sc-cot	8.0 ± 5.7	0.5 ± 9.2	2.0 ± 9.9	-2.5 ± 7.8	-20.5 ± 20.5	**14.0 ± 7.1**	6.5 ± 2.1
	sc-spp	5.0 ± 4.2	-0.5 ± 7.8	2.5 ± 2.1	-4.5 ± 9.2	-14.0 ± 5.7	-11.0 ± 7.1	**18.0 ± 5.7**
C3.7S	zs	1.6 ± 11.9	-2.6 ± 6.9	-10.2 ± 11.0	-1.2 ± 6.5	-13.4 ± 15.8	12.0 ± 7.0	**16.6 ± 16.6**
	cot	1.2 ± 9.9	2.4 ± 9.4	-1.6 ± 5.3	4.2 ± 6.4	**33.4 ± 10.0**	12.4 ± 1.1	20.0 ± 15.4
	spp	6.4 ± 14.7	1.6 ± 11.7	3.0 ± 8.9	-2.6 ± 6.1	27.4 ± 18.1	18.0 ± 10.1	**29.0 ± 11.9**
	sc-zs	10.0 ± 28.3	-3.0 ± 0.0	-19.5 ± 4.9	0.5 ± 4.9	16.0 ± 2.8	**24.0 ± 21.2**	**31.0 ± 8.5**
	sc-cot	-2.0 ± 2.8	-12.0 ± 9.9	-13.0 ± 2.8	1.5 ± 7.8	**31.0 ± 2.8**	19.5 ± 0.7	22.5 ± 24.7
	sc-spp	5.5 ± 14.8	3.5 ± 7.8	2.5 ± 3.5	2.0 ± 8.5	**33.0 ± 2.8**	11.0 ± 4.2	24.5 ± 14.8
C3.7S(T)	zs	4.0 ± 7.0	-6.2 ± 12.2	0.8 ± 14.7	**7.6 ± 6.1**	-8.6 ± 13.7	13.6 ± 4.8	**15.6 ± 12.0**
	cot	6.8 ± 7.3	-1.2 ± 11.5	-1.0 ± 6.0	-0.2 ± 4.3	**18.4 ± 20.2**	15.0 ± 4.4	16.6 ± 18.5
	spp	6.8 ± 5.6	-0.8 ± 10.8	7.8 ± 13.6	-1.6 ± 4.6	**25.8 ± 15.1**	14.0 ± 4.6	19.6 ± 13.6
	sc-zs	-1.5 ± 7.8	-1.0 ± 12.7	-14.0 ± 11.3	4.0 ± 1.4	-14.0 ± 8.5	8.0 ± 0.0	**9.0 ± 5.7**
	sc-cot	2.0 ± 7.1	-13.0 ± 0.0	3.5 ± 6.4	-2.0 ± 1.4	**35.0 ± 7.1**	16.5 ± 2.1	30.5 ± 13.4
	sc-spp	10.0 ± 14.1	5.0 ± 26.9	-3.0 ± 1.4	2.5 ± 6.4	**33.0 ± 9.9**	16.5 ± 0.7	28.5 ± 0.7
C4S	zs	-2.6 ± 12.7	-11.2 ± 8.9	-0.6 ± 4.9	2.2 ± 7.9	7.4 ± 27.3	**15.6 ± 2.3**	10.2 ± 21.1
	cot	7.0 ± 2.5	-2.8 ± 13.1	-4.8 ± 12.6	4.4 ± 9.2	**35.4 ± 4.4**	8.0 ± 11.9	6.8 ± 4.4
	spp	2.8 ± 10.0	-6.8 ± 8.2	2.4 ± 11.9	-3.4 ± 12.8	**24.4 ± 13.8**	15.6 ± 2.1	7.6 ± 12.9
	sc-zs	5.0 ± 7.1	-6.0 ± 5.7	-7.0 ± 4.2	4.0 ± 9.9	**15.0 ± 14.1**	13.5 ± 4.9	-1.5 ± 6.4
	sc-cot	8.0 ± 19.8	-8.5 ± 7.8	0.0 ± 5.7	4.5 ± 12.0	**36.0 ± 5.7**	8.5 ± 0.7	3.0 ± 1.4
	sc-spp	-6.5 ± 10.6	5.5 ± 4.9	-0.5 ± 10.6	-8.0 ± 14.1	27.5 ± 6.4	11.0 ± 1.4	**33.5 ± 9.2**
C4S(T)	zs	-3.2 ± 6.1	-8.6 ± 17.8	4.4 ± 15.1	5.8 ± 4.6	-16.2 ± 11.1	**12.0 ± 2.0**	8.8 ± 16.5
	cot	1.6 ± 5.1	2.0 ± 13.8	-1.2 ± 14.9	5.4 ± 8.6	9.8 ± 27.8	12.0 ± 4.5	**21.6 ± 11.3**
	spp	0.4 ± 8.7	**9.2 ± 9.7**	2.0 ± 4.1	-2.4 ± 7.8	**32.0 ± 4.9**	7.6 ± 7.7	8.8 ± 10.7
	sc-zs	-3.0 ± 4.2	-5.5 ± 3.5	**13.0 ± 4.2**	-3.5 ± 12.0	-0.5 ± 29.0	**16.0 ± 5.7**	-12.5 ± 2.1
	sc-cot	-3.5 ± 16.3	-9.5 ± 9.2	-16.0 ± 7.1	-1.0 ± 9.9	-2.5 ± 0.7	**15.0 ± 1.4**	8.0 ± 7.1
	sc-spp	**10.5 ± 0.7**	-1.5 ± 10.6	-3.0 ± 1.4	-0.5 ± 7.8	**23.5 ± 0.7**	12.5 ± 3.5	14.5 ± 10.6
DS-R1	zs	2.4 ± 7.9	-3.4 ± 8.7	0.2 ± 8.8	-4.0 ± 8.9	-12.6 ± 5.8	**12.4 ± 3.8**	-1.4 ± 7.3
	cot	-4.8 ± 11.6	-0.2 ± 0.8	-4.2 ± 4.3	-3.2 ± 2.8	-7.2 ± 6.9	8.0 ± 5.7	**8.2 ± 4.3**
	spp	-4.8 ± 6.9	-2.8 ± 4.8	-1.8 ± 7.1	1.4 ± 4.5	-14.8 ± 5.3	9.4 ± 7.7	**10.6 ± 9.2**
	sc-zs	-0.5 ± 10.6	2.5 ± 3.5	-1.5 ± 3.5	5.5 ± 2.1	-14.0 ± 0.0	**13.5 ± 0.7**	-0.5 ± 0.7
	sc-cot	6.5 ± 2.1	4.0 ± 4.2	-6.5 ± 2.1	6.5 ± 7.8	-27.5 ± 0.7	**15.5 ± 0.7**	2.0 ± 5.7
	sc-spp	3.5 ± 2.1	2.5 ± 3.5	0.0 ± 1.4	0.0 ± 8.5	-16.5 ± 9.2	**10.5 ± 3.5**	4.0 ± 4.2
L3.3-70B	zs	0.0 ± 3.9	0.0 ± 5.5	-1.2 ± 2.9	3.2 ± 3.3	-16.2 ± 16.5	5.4 ± 5.9	**7.8 ± 2.9**
	cot	-4.8 ± 2.9	2.2 ± 2.7	-0.6 ± 5.6	-0.4 ± 3.2	-8.8 ± 8.6	**7.2 ± 4.4**	5.2 ± 5.5
	spp	-6.6 ± 10.4	1.0 ± 3.3	4.4 ± 3.4	-4.2 ± 7.2	-11.8 ± 9.7	**11.0 ± 3.7**	3.8 ± 2.8
	sc-zs	0.5 ± 0.7	5.0 ± 1.4	2.5 ± 0.7	-1.5 ± 4.9	-20.5 ± 27.6	**6.5 ± 9.2**	**6.5 ± 9.2**
	sc-cot	-4.0 ± 0.0	4.0 ± 2.8	-0.5 ± 3.5	2.5 ± 2.1	-4.5 ± 4.9	**5.0 ± 5.7**	3.5 ± 2.1
	sc-spp	-6.0 ± 1.4	-0.5 ± 3.5	1.5 ± 7.8	-10.0 ± 1.4	-30.5 ± 13.4	**11.5 ± 0.7**	0.5 ± 0.7
M-L(24.07)	zs	0.0 ± 6.4	1.2 ± 8.0	-0.4 ± 10.5	0.0 ± 6.5	-11.6 ± 17.6	-16.4 ± 13.8	**10.2 ± 9.7**
	cot	2.0 ± 12.9	-0.2 ± 6.9	6.8 ± 8.5	-0.6 ± 8.1	-14.6 ± 12.5	-11.2 ± 14.9	**21.8 ± 8.9**
	spp	0.4 ± 10.7	2.8 ± 8.9	0.0 ± 15.0	4.2 ± 5.3	**17.4 ± 18.4**	13.8 ± 13.5	14.6 ± 8.4
	sc-zs	1.0 ± 26.9	2.0 ± 1.4	1.0 ± 5.7	2.5 ± 4.9	-40.0 ± 0.0	-4.5 ± 20.5	**8.5 ± 14.8**
	sc-cot	-3.5 ± 3.5	4.5 ± 4.9	9.5 ± 14.8	6.5 ± 3.5	-16.0 ± 14.1	-6.0 ± 21.2	**19.5 ± 21.9**
	sc-spp	9.5 ± 33.2	-1.0 ± 11.3	-4.5 ± 21.9	1.0 ± 2.8	-32.0 ± 11.3	8.5 ± 16.3	**29.0 ± 8.5**

Table 7.5: Total Points Averaged Over All Iterations (ba3-alt)

not true for Rock-Paper-Scissors. Thus, rational players are not expected to show signs of "agreement" with each other, which will be reflected by mostly low (absolute value wise) results in the "Total Points" tables.

These results provide us with an important observation. More complex prompt styles do better against simpler ones. Maximum values of columns and rows are highlighted in the tables with **bold** values. We can see that **zero-shot (zs)** does not hold any of these values; they all fall in the results of more complex prompt styles.

7.1.2 LLM vs non-LLM

The last 4 columns represent non-LLM players, that follow simple strategies. These players have been described in 5.7. Successful outcomes against them should be a good indication of reasoning abilities of LLM, since they imply an agent's ability to analyze their opponent and make informed decisions.

Srep

In Rock-Paper-Scissors the **srep** player will follow the Single-Round Mixed Nash Equilibrium Strategy in every round. This strategy is given by (5.4):

$$x = \frac{cb}{ba + ac + cb}, \quad y = \frac{ac}{ba + ac + cb}, \quad z = \frac{ba}{ba + ac + cb}$$

Which is:

- **eq1** and **eq1-alt**: $x = y = z = \frac{1}{3}$
- **ba3** and **ba3-alt**: $x = \frac{1}{5}, y = \frac{1}{5}, z = \frac{3}{5}$

The expected payoff of a single round is $-ba \cdot y + ac \cdot z$. Which, of course, in both cases amounts to 0.

This result is reflected in the resulting total point values, which generally are close to $24 \cdot 0 = 0$ across all models.

Pp

The pattern player is a player that plays in cycles: scissors, rock, paper, ... in all game settings.

Should a player match these patterns exactly, they would get:

- in counterfactuals with all payoffs equal to 1 (tables 7.2 and 7.3): $24 \cdot 1 = 24$ points by winning each round.
- in counterfactuals where payoff $ba = 3$ (tables 7.4 and 7.5): $(1 + 3 + 1) \cdot 8 = 40$ points by winning each round.

Even in the default scenario of Rock-Paper-Scissors (table 7.2) most of the simpler models face issues with understanding and taking advantage of their opponent's movement pattern. It seems only the more advanced Claude models, *Claude 3.7 Sonnet* and *Claude Sonnet 4*, can systematically understand the pattern player's tactics. A good runner up is *Llama 3.3 70B Instruct*. Results, also, show a performance boost by enabling **thinking** in *Claude 3.7 Sonnet*. *DeepSeek-R1* shows average performance. Finally, *DeepSeek-R1* manages to achieve somewhat good results when using more advanced prompts (**cot** or **spp**) especially with **self-consistency** enabled).

Moving on to the strategy counterfactual (table 7.3) (where paper beats rock, etc) we see that the *Llama* model mainly achieves negative scores, thus failing to reason on its opponent. The motif for Claude's models continues, but another significant observation can be made: moving from **zero-shot** to more complex prompting types (either **cot** or **spp**) boost LLMs performance and reasoning capabilities (negative values are present in the results from **zero-shot** players).

These results are also reflected in the results for payoff counterfactual settings (tables 7.4 and 7.5). One important observation from the **ba3-alt** is that *Claude Sonnet 4* with **thinking** enabled seems to get confused more easily getting worse scores than its default version.

Ap

Ap is an adaptive player. This player adapts to its opponent's most frequently used move and aims to counter it.

Players (excluding *Mistral* agents) generally manage to accumulate positive values in their final "Total Points". However, these results are mostly not close to the maximums 24 and 40 (depending on game setting). A result, which may be interpreted as LLM agent's needing a bit of experience before they catch onto their opponent's play-style. The results against this player will be further analyzed when we take a look at the **round of opponent comprehension** later.

Tft

Tft player is influenced by the concept of "tit-for-tat". This player counters their opponent's last move.

In the payoff counterfactuals (tables 7.2 and 7.4) (where moves behave the way they typically do in Rock-Paper-Scissors) players perform well and achieve values close the maximum possible rather often. However, in the other counterfactuals, where strategies behave in the opposite way from what is expected, LLMs struggle much more; with more complex models such as *Claude Sonnet 4* showing degraded performance.

7.2 Opponent Comprehension

To disambiguate the above results, shed light on the hypothesized behavior that has been mentioned, and better understand the thinking process of LLM agents, another metric is introduced. Instead of just looking at the "Total Points" accumulated by agents when playing this game, we look at how late a player was at making the most out of their opponent's behavior.

We consider that an AI agent has managed to comprehend their opponent, when the agent systematically responds with actions that use the opponent's moves to their advantage. More formally, suppose players A, B that have played N rounds and the strategies they followed were $(s_A^1, s_B^1), \ldots, (s_A^N, s_B^N)$ - e.g., $(rock, scissors), \ldots, (paper, paper)$ -. Assume that A is the AI agent, then we call **round of opponent comprehension**, m, the round after which every move that A makes yields a payoff for A that is at least as good as the payoff that B gets.

Furthermore, we have expanded this definition, by including a percentage tp (target percentage) relaxing the requirement of "good" response to *every* move that the opponent makes to the following requirement: in the rounds from m all the way to N, A's moves are at least as good as B's in tp percentage of those rounds.

The following tables are results where $tp = 90\%$ and a lower value is considered better since it indicates that the AI agent understood how to play with its specific opponent earlier in the game.

Finally, a value of 25 means that the agent never *understood* their opponent, since it is out of the range $1-24$ of rounds that were played.

We show the average **round of opponent comprehension** in each game setting for each player. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

7.2.1 LLM vs LLM

LLM agents that play against each other are of the same LLM. These games are represented in the first three columns of the result matrices.

Values in these columns are fairly high and closer to 25 meaning that LLMs do not easily counter other LLMs' strategies. A most expected result in the case of Rock-Paper-Scissors.

		eq1						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	10.6 ± 13.1	21.4 ± 4.6	19.6 ± 5.6	21.0 ± 3.5	14.6 ± 12.4	10.4 ± 11.2	**1.0 ± 0.0**
	cot	17.2 ± 7.3	19.6 ± 7.5	20.6 ± 5.6	19.6 ± 7.1	16.0 ± 11.9	22.6 ± 2.1	**5.8 ± 10.7**
	spp	11.2 ± 10.6	11.8 ± 10.9	20.4 ± 5.9	23.0 ± 1.9	11.4 ± 11.3	18.4 ± 9.3	**1.6 ± 0.9**
	sc-zs	25.0 ± 0.0	14.0 ± 15.6	21.5 ± 4.9	13.5 ± 16.3	12.0 ± 15.6	19.5 ± 6.4	**1.0 ± 0.0**
	sc-cot	15.0 ± 14.1	20.5 ± 0.7	22.5 ± 3.5	25.0 ± 0.0	12.5 ± 16.3	7.0 ± 8.5	**1.0 ± 0.0**
	sc-spp	14.5 ± 14.8	15.0 ± 14.1	14.5 ± 14.8	24.0 ± 1.4	16.0 ± 8.5	21.0 ± 0.0	**1.0 ± 0.0**
C3.7S	zs	21.8 ± 5.5	22.2 ± 4.2	20.2 ± 9.7	23.4 ± 2.5	**10.0 ± 12.4**	17.6 ± 9.3	15.0 ± 12.8
	cot	15.4 ± 7.1	16.2 ± 10.0	18.8 ± 4.3	18.0 ± 7.4	**1.0 ± 0.0**	15.6 ± 10.1	**1.0 ± 0.0**
	spp	24.4 ± 1.3	20.0 ± 5.6	22.4 ± 4.3	23.6 ± 1.5	**1.2 ± 0.4**	12.0 ± 10.6	**1.2 ± 0.4**
	sc-zs	13.0 ± 17.0	23.0 ± 0.0	24.5 ± 0.7	22.0 ± 2.8	**11.5 ± 14.8**	15.0 ± 14.1	12.5 ± 16.3
	sc-cot	4.5 ± 4.9	24.5 ± 0.7	11.5 ± 14.8	24.0 ± 1.4	**1.0 ± 0.0**	24.5 ± 0.7	**1.0 ± 0.0**
	sc-spp	9.5 ± 12.0	21.5 ± 3.5	22.0 ± 1.4	**12.5 ± 16.3**	**1.0 ± 0.0**	22.5 ± 2.1	**1.0 ± 0.0**
C3.7S(T)	zs	24.2 ± 1.3	19.4 ± 8.3	22.6 ± 4.8	22.4 ± 2.2	**1.4 ± 0.9**	22.6 ± 1.7	1.6 ± 0.5
	cot	19.2 ± 5.1	21.8 ± 4.9	22.0 ± 2.7	23.8 ± 2.2	**1.0 ± 0.0**	15.6 ± 12.0	4.4 ± 7.6
	spp	23.6 ± 1.5	20.6 ± 4.3	21.2 ± 5.3	15.2 ± 11.3	**1.0 ± 0.0**	17.8 ± 9.7	1.2 ± 0.4
	sc-zs	24.5 ± 0.7	22.5 ± 3.5	13.0 ± 17.0	22.5 ± 3.5	**1.0 ± 0.0**	12.0 ± 15.6	**1.0 ± 0.0**
	sc-cot	24.5 ± 0.7	25.0 ± 0.0	25.0 ± 0.0	19.5 ± 7.8	**1.0 ± 0.0**	10.0 ± 12.7	**1.0 ± 0.0**
	sc-spp	24.0 ± 1.4	25.0 ± 0.0	25.0 ± 0.0	25.0 ± 0.0	**1.0 ± 0.0**	23.0 ± 0.0	**1.0 ± 0.0**
C4S	zs	15.8 ± 8.5	20.0 ± 10.6	23.8 ± 2.2	22.2 ± 0.8	**1.0 ± 0.0**	23.2 ± 1.1	11.8 ± 11.3
	cot	15.2 ± 13.0	15.6 ± 1.9	10.8 ± 11.8	19.4 ± 8.6	**1.0 ± 0.0**	19.0 ± 10.1	3.2 ± 4.9
	spp	16.0 ± 10.5	20.8 ± 9.4	18.6 ± 10.0	19.2 ± 5.5	**3.4 ± 4.8**	14.4 ± 12.3	9.0 ± 8.6
	sc-zs	7.5 ± 9.2	25.0 ± 0.0	25.0 ± 0.0	22.0 ± 4.2	**1.0 ± 0.0**	23.0 ± 1.4	**1.0 ± 0.0**
	sc-cot	21.5 ± 4.9	15.0 ± 14.1	14.5 ± 14.8	21.0 ± 1.4	**1.0 ± 0.0**	24.0 ± 1.4	**1.0 ± 0.0**
	sc-spp	10.0 ± 7.1	11.0 ± 8.5	23.0 ± 2.8	21.5 ± 0.7	**1.0 ± 0.0**	13.0 ± 17.0	**1.0 ± 0.0**
C4S(T)	zs	13.4 ± 12.0	21.0 ± 5.3	17.0 ± 10.0	17.6 ± 8.5	**3.4 ± 5.4**	22.6 ± 1.8	10.6 ± 13.1
	cot	14.4 ± 12.3	21.8 ± 2.7	20.2 ± 10.7	21.6 ± 4.2	**1.0 ± 0.0**	19.6 ± 10.4	3.2 ± 4.9
	spp	24.2 ± 1.1	15.2 ± 9.9	17.8 ± 8.6	18.8 ± 10.1	**1.0 ± 0.0**	19.4 ± 10.4	5.6 ± 6.5
	sc-zs	13.0 ± 17.0	22.0 ± 4.2	24.0 ± 1.4	18.0 ± 5.7	**1.0 ± 0.0**	24.5 ± 0.7	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	7.0 ± 4.2	10.5 ± 13.4	23.5 ± 0.7	**1.0 ± 0.0**	11.0 ± 14.1	**1.0 ± 0.0**
	sc-spp	24.0 ± 1.4	23.0 ± 2.8	24.5 ± 0.7	20.5 ± 2.1	**1.0 ± 0.0**	23.5 ± 0.7	**1.0 ± 0.0**
DS-R1	zs	21.2 ± 5.5	21.2 ± 5.3	19.4 ± 10.3	23.4 ± 1.3	23.2 ± 0.4	16.8 ± 11.1	**7.8 ± 9.7**
	cot	21.2 ± 4.3	20.0 ± 5.5	24.4 ± 0.9	13.6 ± 10.9	12.8 ± 10.3	15.4 ± 12.7	**1.0 ± 0.0**
	spp	18.0 ± 10.4	22.6 ± 2.3	20.6 ± 8.8	23.0 ± 1.4	20.2 ± 4.1	16.8 ± 8.1	**6.2 ± 9.0**
	sc-zs	12.0 ± 15.6	21.0 ± 2.8	24.0 ± 1.4	22.5 ± 3.5	13.5 ± 16.3	22.5 ± 2.1	**1.5 ± 0.7**
	sc-cot	12.0 ± 15.6	23.0 ± 1.4	19.0 ± 7.1	21.5 ± 4.9	3.5 ± 2.1	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-spp	16.0 ± 8.5	24.0 ± 1.4	20.5 ± 3.5	23.0 ± 0.0	13.5 ± 16.3	24.5 ± 0.7	**1.0 ± 0.0**
L3.3-70B	zs	5.8 ± 10.7	16.2 ± 10.4	12.4 ± 11.8	21.8 ± 2.3	**1.0 ± 0.0**	12.8 ± 7.9	**1.0 ± 0.0**
	cot	5.6 ± 10.3	**5.8 ± 10.7**	10.0 ± 12.4	23.0 ± 1.9	2.8 ± 4.0	21.2 ± 4.1	**1.0 ± 0.0**
	spp	**1.0 ± 0.0**	7.8 ± 10.0	21.2 ± 5.8	21.8 ± 5.5	8.6 ± 11.0	15.2 ± 9.4	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	13.0 ± 17.0	20.0 ± 7.1	25.0 ± 0.0	**1.0 ± 0.0**	5.5 ± 4.9	**1.0 ± 0.0**
	sc-cot	10.5 ± 13.4	23.5 ± 2.1	**1.0 ± 0.0**	25.0 ± 0.0	**1.0 ± 0.0**	12.5 ± 16.3	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	13.0 ± 17.0	24.5 ± 0.7	16.0 ± 8.5	2.5 ± 2.1	12.0 ± 15.6	**1.0 ± 0.0**
M-L(24.07)	zs	1.2 ± 0.4	19.2 ± 10.3	23.6 ± 1.1	22.6 ± 4.8	**1.0 ± 0.0**	6.0 ± 10.6	**1.0 ± 0.0**
	cot	18.8 ± 10.4	20.2 ± 6.1	15.8 ± 10.7	21.2 ± 4.3	12.4 ± 12.1	17.0 ± 9.4	**1.0 ± 0.0**
	spp	6.6 ± 10.4	10.6 ± 13.1	17.8 ± 9.6	23.0 ± 0.7	17.2 ± 10.3	19.8 ± 10.5	**1.0 ± 0.0**
	sc-zs	**1.0 ± 0.0**	13.0 ± 17.0	24.5 ± 0.7	23.0 ± 0.0	**1.0 ± 0.0**	24.0 ± 0.0	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	24.0 ± 0.0	13.0 ± 17.0	24.5 ± 0.7	12.5 ± 16.3	24.5 ± 0.7	**1.0 ± 0.0**
	sc-spp	12.5 ± 16.3	11.5 ± 14.8	12.5 ± 16.3	25.0 ± 0.0	11.5 ± 14.8	24.5 ± 0.7	**6.5 ± 7.8**

Table 7.6: Round # where the Agent understood the opponent's Strategy (eq1)

model	prompt	eq1-alt zs	spp	cot	srep	pp	ap	tft
C3.5Sv2	zs	19.0 ± 10.3	22.4 ± 4.2	24.4 ± 1.3	20.4 ± 4.2	21.6 ± 3.8	**11.4 ± 12.5**	14.8 ± 11.4
	cot	21.8 ± 6.1	**19.6 ± 7.3**	19.8 ± 6.8	20.8 ± 5.1	21.6 ± 6.5	24.4 ± 0.9	23.6 ± 1.7
	spp	21.8 ± 2.6	24.0 ± 0.7	**14.6 ± 8.7**	22.6 ± 3.0	18.4 ± 9.0	22.8 ± 2.7	16.4 ± 11.3
	sc-zs	**7.5 ± 9.2**	22.5 ± 0.7	24.0 ± 1.4	**12.0 ± 4.2**	25.0 ± 0.0	19.0 ± 5.7	13.0 ± 17.0
	sc-cot	23.0 ± 1.4	**11.5 ± 3.5**	25.0 ± 0.0	22.5 ± 0.7	25.0 ± 0.0	25.0 ± 0.0	24.0 ± 1.4
	sc-spp	19.0 ± 8.5	12.5 ± 16.3	18.5 ± 9.2	17.0 ± 8.5	25.0 ± 0.0	25.0 ± 0.0	**12.0 ± 15.6**
C3.7S	zs	23.4 ± 1.8	21.0 ± 5.3	22.6 ± 2.3	19.2 ± 7.2	21.2 ± 5.8	22.8 ± 0.8	**7.8 ± 10.0**
	cot	19.8 ± 10.0	19.2 ± 10.2	24.0 ± 1.0	23.2 ± 2.0	3.2 ± 3.9	18.4 ± 9.8	**1.6 ± 1.3**
	spp	18.4 ± 9.5	19.4 ± 8.3	21.6 ± 6.5	22.2 ± 5.2	2.4 ± 1.9	23.6 ± 2.1	**1.0 ± 0.0**
	sc-zs	25.0 ± 0.0	22.0 ± 2.8	20.0 ± 7.1	19.0 ± 5.7	**16.5 ± 12.0**	25.0 ± 0.0	23.0 ± 2.8
	sc-cot	25.0 ± 0.0	21.5 ± 0.7	18.5 ± 4.9	25.0 ± 0.0	**4.0 ± 4.2**	21.5 ± 2.1	13.0 ± 17.0
	sc-spp	12.0 ± 2.8	13.0 ± 17.0	24.0 ± 0.0	25.0 ± 0.0	5.0 ± 0.0	22.5 ± 2.1	**1.0 ± 0.0**
C3.7S(T)	zs	22.6 ± 4.3	24.6 ± 0.9	23.6 ± 1.3	22.4 ± 2.2	15.0 ± 9.7	17.6 ± 10.0	**10.4 ± 12.9**
	cot	19.2 ± 4.4	22.4 ± 2.6	22.0 ± 2.6	23.2 ± 1.9	10.6 ± 13.1	21.8 ± 1.8	**7.2 ± 9.7**
	spp	17.4 ± 10.4	21.6 ± 4.9	23.6 ± 0.5	21.6 ± 3.8	3.4 ± 3.9	24.0 ± 1.0	**1.2 ± 0.4**
	sc-zs	24.5 ± 0.7	22.0 ± 4.2	12.5 ± 16.3	22.5 ± 3.5	3.0 ± 2.8	23.0 ± 0.0	**1.0 ± 0.0**
	sc-cot	19.0 ± 8.5	13.0 ± 17.0	25.0 ± 0.0	22.5 ± 0.7	11.5 ± 14.8	19.0 ± 7.1	**1.0 ± 0.0**
	sc-spp	3.0 ± 1.4	16.0 ± 7.1	24.5 ± 0.7	25.0 ± 0.0	**1.0 ± 0.0**	24.5 ± 0.7	**1.0 ± 0.0**
C4S	zs	19.4 ± 10.3	21.6 ± 7.1	24.0 ± 2.2	22.8 ± 3.0	16.2 ± 10.3	21.6 ± 6.5	**9.4 ± 9.2**
	cot	13.2 ± 10.2	24.2 ± 1.8	23.8 ± 1.3	24.0 ± 1.2	**4.8 ± 2.8**	18.0 ± 9.6	13.6 ± 6.8
	spp	19.6 ± 10.4	17.0 ± 10.1	17.6 ± 10.9	19.6 ± 4.3	**4.2 ± 2.2**	21.2 ± 4.0	21.2 ± 5.4
	sc-zs	25.0 ± 0.0	22.5 ± 3.5	24.5 ± 0.7	24.5 ± 0.7	**2.5 ± 0.7**	25.0 ± 0.0	13.0 ± 17.0
	sc-cot	24.0 ± 0.0	23.5 ± 2.1	24.0 ± 1.4	12.5 ± 16.3	**2.5 ± 2.1**	25.0 ± 0.0	20.5 ± 3.5
	sc-spp	23.5 ± 0.7	24.0 ± 1.4	23.5 ± 0.7	24.0 ± 1.4	**3.0 ± 2.8**	15.0 ± 8.5	24.5 ± 0.7
C4S(T)	zs	22.2 ± 4.7	23.8 ± 0.8	24.2 ± 0.8	22.4 ± 4.2	20.6 ± 9.3	21.4 ± 4.5	**16.4 ± 9.7**
	cot	24.8 ± 0.4	24.6 ± 0.5	22.6 ± 4.3	21.8 ± 4.4	**12.2 ± 10.7**	21.0 ± 5.7	17.6 ± 10.6
	spp	23.6 ± 1.1	21.8 ± 4.9	22.4 ± 2.1	23.4 ± 0.5	**7.8 ± 10.0**	22.0 ± 1.6	14.8 ± 12.7
	sc-zs	25.0 ± 0.0	24.5 ± 0.7	25.0 ± 0.0	23.0 ± 1.4	24.5 ± 0.7	**22.0 ± 1.4**	24.5 ± 0.7
	sc-cot	21.5 ± 4.9	9.0 ± 11.3	24.0 ± 1.4	25.0 ± 0.0	**1.0 ± 0.0**	22.0 ± 1.4	12.0 ± 15.6
	sc-spp	18.5 ± 7.8	24.0 ± 1.4	13.0 ± 17.0	21.5 ± 3.5	**6.5 ± 3.5**	23.0 ± 1.4	13.0 ± 17.0
DS-R1	zs	**14.4 ± 12.3**	23.6 ± 2.1	17.0 ± 6.5	20.6 ± 4.6	23.8 ± 0.8	23.0 ± 1.0	19.8 ± 4.9
	cot	**15.8 ± 10.8**	23.2 ± 1.3	23.6 ± 2.6	24.0 ± 1.7	24.6 ± 0.9	21.2 ± 3.5	16.8 ± 10.5
	spp	23.2 ± 2.2	21.0 ± 3.7	**18.6 ± 8.8**	23.2 ± 0.8	24.6 ± 0.9	22.6 ± 1.1	23.4 ± 0.5
	sc-zs	22.5 ± 3.5	16.5 ± 12.0	**8.0 ± 9.9**	23.0 ± 2.8	25.0 ± 0.0	24.0 ± 0.0	16.5 ± 3.5
	sc-cot	24.0 ± 1.4	22.0 ± 2.8	18.0 ± 9.9	24.5 ± 0.7	25.0 ± 0.0	**13.0 ± 0.0**	15.0 ± 14.1
	sc-spp	21.5 ± 4.9	19.0 ± 5.7	24.0 ± 1.4	24.0 ± 0.0	24.0 ± 0.0	25.0 ± 0.0	**13.0 ± 17.0**
L3.3-70B	zs	16.2 ± 9.8	7.4 ± 9.0	16.4 ± 9.6	23.8 ± 1.6	10.4 ± 9.5	18.6 ± 9.9	**6.8 ± 9.5**
	cot	16.2 ± 9.1	10.2 ± 12.6	12.2 ± 11.1	20.2 ± 5.0	21.0 ± 5.2	19.8 ± 10.0	**1.0 ± 0.0**
	spp	13.0 ± 11.4	**1.4 ± 0.9**	10.2 ± 12.6	23.0 ± 1.2	12.2 ± 9.8	18.8 ± 10.1	11.2 ± 10.7
	sc-zs	5.0 ± 5.7	6.5 ± 3.5	12.5 ± 13.4	21.0 ± 1.4	14.0 ± 15.6	23.5 ± 2.1	**1.0 ± 0.0**
	sc-cot	11.5 ± 14.8	**8.5 ± 6.4**	13.0 ± 17.0	20.5 ± 4.9	14.0 ± 15.6	24.0 ± 1.4	13.0 ± 17.0
	sc-spp	6.0 ± 7.1	6.0 ± 7.1	9.0 ± 11.3	25.0 ± 0.0	7.0 ± 5.7	**1.0 ± 0.0**	**1.0 ± 0.0**
M-L(24.07)	zs	14.2 ± 12.2	19.0 ± 10.2	16.0 ± 12.0	23.8 ± 1.3	19.4 ± 9.1	24.2 ± 0.8	**1.0 ± 0.0**
	cot	20.8 ± 4.2	13.8 ± 9.0	9.0 ± 7.6	23.4 ± 2.1	24.2 ± 0.8	17.2 ± 10.5	**3.2 ± 4.9**
	spp	18.0 ± 7.0	22.2 ± 4.6	20.4 ± 6.4	22.2 ± 2.9	23.8 ± 1.6	12.8 ± 11.3	**5.6 ± 10.3**
	sc-zs	**1.0 ± 0.0**	1.5 ± 0.7	12.0 ± 15.6	21.5 ± 4.9	13.0 ± 17.0	24.5 ± 0.7	7.5 ± 9.2
	sc-cot	17.5 ± 10.6	14.5 ± 13.4	**10.0 ± 7.1**	18.5 ± 4.9	14.5 ± 14.8	13.0 ± 17.0	12.5 ± 16.3
	sc-spp	12.5 ± 16.3	19.0 ± 7.1	23.5 ± 0.7	19.0 ± 7.1	25.0 ± 0.0	12.5 ± 16.3	**1.0 ± 0.0**

Table 7.7: Round # where the Agent understood the opponent's Strategy (eq1-alt)

		ba3						
		zs	spp	cot	srep	pp	ap	tft
model	prompt							
C3.5Sv2	zs	16.2 ± 11.0	17.6 ± 10.3	24.4 ± 0.5	23.4 ± 1.9	19.0 ± 10.4	19.4 ± 10.4	**5.6 ± 10.3**
	cot	20.2 ± 9.1	12.6 ± 11.0	21.0 ± 5.3	17.6 ± 9.8	24.2 ± 1.3	18.4 ± 9.8	**10.0 ± 12.4**
	spp	**10.6 ± 10.1**	13.4 ± 11.5	22.2 ± 4.1	18.4 ± 6.9	15.0 ± 9.1	24.2 ± 0.8	14.2 ± 12.3
	sc-zs	23.5 ± 2.1	24.0 ± 1.4	13.0 ± 14.1	18.0 ± 9.9	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	25.0 ± 0.0	18.5 ± 6.4	25.0 ± 0.0	19.5 ± 2.1	12.0 ± 15.6	11.5 ± 14.8	**1.0 ± 0.0**
	sc-spp	23.0 ± 1.4	25.0 ± 0.0	**2.0 ± 1.4**	24.5 ± 0.7	**1.0 ± 0.0**	12.5 ± 16.3	**1.0 ± 0.0**
C3.7S	zs	20.4 ± 8.6	17.6 ± 10.6	23.2 ± 2.0	22.4 ± 2.5	**1.4 ± 0.9**	18.8 ± 10.1	3.2 ± 3.5
	cot	16.6 ± 9.5	23.0 ± 2.9	23.2 ± 1.6	18.8 ± 6.2	**1.0 ± 0.0**	7.4 ± 8.9	1.2 ± 0.4
	spp	21.0 ± 4.6	24.4 ± 1.3	24.0 ± 1.0	20.8 ± 6.0	3.0 ± 4.5	18.4 ± 9.8	**1.0 ± 0.0**
	sc-zs	16.5 ± 12.0	25.0 ± 0.0	24.5 ± 0.7	17.0 ± 11.3	12.0 ± 15.6	12.0 ± 15.6	**1.0 ± 0.0**
	sc-cot	19.0 ± 7.1	21.0 ± 5.7	21.5 ± 0.7	24.5 ± 0.7	**1.0 ± 0.0**	22.0 ± 1.4	**1.0 ± 0.0**
	sc-spp	13.0 ± 17.0	23.5 ± 2.1	23.5 ± 2.1	24.0 ± 1.4	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
C3.7S(T)	zs	19.2 ± 6.8	20.8 ± 3.5	23.6 ± 2.1	17.8 ± 8.6	5.8 ± 9.1	23.6 ± 1.5	**4.8 ± 4.8**
	cot	18.8 ± 6.6	21.4 ± 4.9	23.6 ± 1.1	21.2 ± 5.8	**1.0 ± 0.0**	13.8 ± 11.9	3.4 ± 5.4
	spp	21.4 ± 5.0	22.4 ± 1.9	24.8 ± 0.4	19.6 ± 8.8	**1.0 ± 0.0**	5.8 ± 10.7	**1.0 ± 0.0**
	sc-zs	25.0 ± 0.0	20.0 ± 7.1	13.0 ± 17.0	24.0 ± 0.0	**1.0 ± 0.0**	19.5 ± 7.8	**1.0 ± 0.0**
	sc-cot	22.5 ± 2.1	24.5 ± 0.7	25.0 ± 0.0	18.5 ± 6.4	**1.0 ± 0.0**	23.5 ± 2.1	**1.0 ± 0.0**
	sc-spp	24.0 ± 1.4	25.0 ± 0.0	24.0 ± 1.4	20.5 ± 0.7	**1.0 ± 0.0**	12.0 ± 15.6	**1.0 ± 0.0**
C4S	zs	23.8 ± 1.6	23.2 ± 2.0	18.0 ± 7.6	24.2 ± 0.8	**1.0 ± 0.0**	23.0 ± 1.9	7.2 ± 8.5
	cot	19.8 ± 10.5	21.8 ± 6.1	19.6 ± 5.6	20.0 ± 4.2	**1.0 ± 0.0**	18.8 ± 10.0	5.6 ± 10.3
	spp	10.2 ± 9.7	19.2 ± 10.3	23.0 ± 1.4	24.0 ± 1.7	**5.8 ± 10.7**	18.6 ± 5.8	15.6 ± 11.6
	sc-zs	20.0 ± 7.1	25.0 ± 0.0	18.0 ± 9.9	25.0 ± 0.0	**1.0 ± 0.0**	23.0 ± 0.0	20.5 ± 3.5
	sc-cot	12.0 ± 15.6	19.0 ± 8.5	13.0 ± 17.0	16.5 ± 9.2	**1.0 ± 0.0**	13.0 ± 17.0	**1.0 ± 0.0**
	sc-spp	25.0 ± 0.0	23.5 ± 2.1	13.0 ± 17.0	13.5 ± 0.7	**1.0 ± 0.0**	24.0 ± 1.4	**1.0 ± 0.0**
C4S(T)	zs	24.0 ± 0.7	24.8 ± 0.4	21.0 ± 4.3	21.4 ± 4.3	**4.2 ± 7.2**	23.4 ± 1.5	15.0 ± 12.3
	cot	22.8 ± 2.3	10.6 ± 10.3	20.2 ± 10.7	17.4 ± 10.1	1.2 ± 0.4	18.2 ± 9.8	**1.0 ± 0.0**
	spp	19.2 ± 6.2	19.4 ± 4.6	12.2 ± 11.5	21.2 ± 4.2	**1.0 ± 0.0**	20.4 ± 9.2	6.0 ± 10.1
	sc-zs	25.0 ± 0.0	24.0 ± 1.4	19.5 ± 7.8	**11.5 ± 0.7**	**1.0 ± 0.0**	23.0 ± 1.4	**1.0 ± 0.0**
	sc-cot	23.5 ± 0.7	23.0 ± 1.4	25.0 ± 0.0	23.0 ± 0.0	3.5 ± 2.1	18.5 ± 7.8	**1.0 ± 0.0**
	sc-spp	22.5 ± 2.1	24.0 ± 1.4	21.5 ± 0.7	24.5 ± 0.7	**1.0 ± 0.0**	23.0 ± 2.8	12.5 ± 16.3
DS-R1	zs	24.8 ± 0.4	19.0 ± 6.7	20.0 ± 10.6	23.0 ± 2.0	18.6 ± 9.3	18.8 ± 8.3	**5.0 ± 8.9**
	cot	19.6 ± 7.4	19.0 ± 8.0	20.8 ± 5.4	17.8 ± 9.1	24.4 ± 0.9	15.2 ± 11.7	**7.0 ± 7.0**
	spp	22.2 ± 3.3	21.2 ± 4.7	19.4 ± 3.0	24.4 ± 1.3	17.8 ± 9.7	14.6 ± 11.6	**9.4 ± 11.5**
	sc-zs	20.0 ± 7.1	20.0 ± 0.0	23.0 ± 1.4	24.0 ± 0.0	11.5 ± 14.8	1.5 ± 0.7	**1.0 ± 0.0**
	sc-cot	17.0 ± 5.7	24.5 ± 0.7	25.0 ± 0.0	18.5 ± 6.4	25.0 ± 0.0	18.5 ± 6.4	**5.5 ± 6.4**
	sc-spp	25.0 ± 0.0	18.0 ± 5.7	24.5 ± 0.7	24.0 ± 0.0	13.0 ± 17.0	**1.5 ± 0.7**	**1.5 ± 0.7**
L3.3-70B	zs	5.8 ± 10.7	15.0 ± 9.7	8.4 ± 10.9	23.8 ± 0.8	7.2 ± 10.4	6.8 ± 9.7	**1.0 ± 0.0**
	cot	9.6 ± 11.9	16.8 ± 10.8	13.4 ± 11.3	21.6 ± 4.9	**1.0 ± 0.0**	16.4 ± 10.5	**1.0 ± 0.0**
	spp	**1.0 ± 0.0**	5.0 ± 4.1	7.6 ± 10.5	24.2 ± 0.8	11.6 ± 11.1	20.2 ± 7.3	1.2 ± 0.4
	sc-zs	13.0 ± 17.0	**1.0 ± 0.0**	11.5 ± 14.8	23.0 ± 0.0	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	13.0 ± 17.0	7.0 ± 8.5	10.5 ± 13.4	16.0 ± 7.1	**1.0 ± 0.0**	22.0 ± 0.0	**1.0 ± 0.0**
	sc-spp	**1.0 ± 0.0**	12.0 ± 15.6	19.0 ± 5.7	18.5 ± 9.2	6.5 ± 7.8	6.0 ± 7.1	**1.0 ± 0.0**
M-L(24.07)	zs	7.6 ± 10.1	15.2 ± 13.0	12.4 ± 10.5	24.8 ± 0.4	**1.0 ± 0.0**	19.8 ± 10.5	**1.0 ± 0.0**
	cot	9.8 ± 11.2	17.0 ± 6.8	19.2 ± 10.2	24.0 ± 1.7	13.8 ± 11.9	15.4 ± 13.1	**4.8 ± 7.9**
	spp	10.0 ± 12.3	20.6 ± 4.6	11.6 ± 9.0	22.0 ± 5.6	20.6 ± 6.6	21.0 ± 6.4	**8.4 ± 9.7**
	sc-zs	**1.0 ± 0.0**	24.5 ± 0.7	11.0 ± 0.0	22.0 ± 2.8	**1.0 ± 0.0**	**1.0 ± 0.0**	**1.0 ± 0.0**
	sc-cot	**1.0 ± 0.0**	11.5 ± 14.8	12.0 ± 15.6	23.5 ± 0.7	13.0 ± 17.0	**1.0 ± 0.0**	1.5 ± 0.7
	sc-spp	24.0 ± 0.0	12.5 ± 16.3	16.5 ± 10.6	17.0 ± 8.5	13.0 ± 17.0	12.5 ± 16.3	**1.0 ± 0.0**

Table 7.8: Round # where the Agent understood the opponent's Strategy (ba3)

model	prompt	zs	spp	cot	srep	pp	ap	tft
					ba3-alt			
C3.5Sv2	zs	23.8 ± 1.3	18.8 ± 7.6	14.6 ± 10.7	24.2 ± 0.8	21.2 ± 7.4	15.0 ± 12.8	**10.2 ± 9.3**
	cot	16.2 ± 9.5	17.6 ± 9.6	19.0 ± 7.4	17.4 ± 10.3	24.4 ± 0.9	18.6 ± 5.6	**14.2 ± 12.1**
	spp	23.4 ± 0.9	24.6 ± 0.5	15.8 ± 9.6	20.0 ± 3.7	25.0 ± 0.0	19.0 ± 8.0	**7.8 ± 10.2**
	sc-zs	24.0 ± 1.4	24.0 ± 0.0	**15.0 ± 0.0**	19.0 ± 8.5	21.0 ± 5.7	23.5 ± 2.1	24.5 ± 0.7
	sc-cot	22.5 ± 2.1	14.0 ± 15.6	12.5 ± 13.4	24.0 ± 1.4	22.5 ± 3.5	24.5 ± 0.7	**10.5 ± 13.4**
	sc-spp	10.5 ± 12.0	12.5 ± 16.3	17.5 ± 10.6	22.0 ± 2.8	24.0 ± 1.4	23.5 ± 2.1	**10.0 ± 12.7**
C3.7S	zs	21.0 ± 4.2	23.2 ± 2.2	23.4 ± 1.8	22.8 ± 2.9	16.6 ± 9.0	19.0 ± 10.2	**12.2 ± 11.8**
	cot	20.0 ± 4.5	20.4 ± 4.6	21.4 ± 4.8	18.0 ± 7.5	**2.6 ± 3.6**	23.2 ± 1.6	9.4 ± 11.7
	spp	17.0 ± 9.7	22.8 ± 2.9	24.6 ± 0.5	20.8 ± 8.8	**4.6 ± 8.0**	17.4 ± 9.9	7.0 ± 10.4
	sc-zs	12.5 ± 16.3	23.5 ± 0.7	25.0 ± 0.0	18.5 ± 9.2	7.5 ± 0.7	**1.0 ± 0.0**	3.0 ± 2.8
	sc-cot	13.0 ± 17.0	23.0 ± 2.8	24.5 ± 0.7	23.0 ± 1.4	2.0 ± 1.4	22.5 ± 2.1	**1.0 ± 0.0**
	sc-spp	19.0 ± 5.7	23.0 ± 1.4	24.5 ± 0.7	23.0 ± 1.4	**1.0 ± 0.0**	25.0 ± 0.0	**1.0 ± 0.0**
C3.7S(T)	zs	22.6 ± 4.3	20.0 ± 4.7	21.6 ± 4.6	24.2 ± 0.4	24.8 ± 0.4	19.0 ± 10.1	**9.4 ± 8.8**
	cot	17.4 ± 6.7	21.6 ± 4.1	22.6 ± 3.2	21.6 ± 5.1	11.8 ± 12.1	21.4 ± 3.9	**2.4 ± 3.1**
	spp	17.0 ± 7.2	19.8 ± 8.4	22.8 ± 1.9	17.4 ± 7.4	**2.4 ± 1.3**	22.0 ± 1.6	6.2 ± 10.0
	sc-zs	24.5 ± 0.7	18.5 ± 0.7	24.5 ± 0.7	21.0 ± 1.4	25.0 ± 0.0	23.0 ± 1.4	**8.0 ± 9.9**
	sc-cot	17.0 ± 5.7	23.5 ± 0.7	25.0 ± 0.0	24.0 ± 0.0	1.5 ± 0.7	23.5 ± 0.7	**1.0 ± 0.0**
	sc-spp	14.0 ± 12.7	18.0 ± 9.9	23.5 ± 2.1	**13.0 ± 8.5**	3.0 ± 2.8	25.0 ± 0.0	**1.0 ± 0.0**
C4S	zs	20.4 ± 6.3	23.4 ± 1.7	19.8 ± 10.5	21.6 ± 3.8	**13.6 ± 9.9**	23.4 ± 0.5	17.4 ± 9.9
	cot	24.0 ± 1.0	20.4 ± 7.6	23.2 ± 2.2	19.2 ± 5.9	**1.4 ± 0.9**	24.2 ± 0.8	18.6 ± 9.9
	spp	23.4 ± 1.3	24.2 ± 1.3	21.4 ± 4.8	19.0 ± 5.6	**6.4 ± 8.8**	22.0 ± 4.0	14.0 ± 9.7
	sc-zs	22.5 ± 2.1	23.0 ± 1.4	24.0 ± 1.4	18.5 ± 6.4	**8.5 ± 4.9**	24.0 ± 1.4	22.0 ± 1.4
	sc-cot	15.0 ± 14.1	25.0 ± 0.0	22.5 ± 3.5	21.0 ± 1.4	**1.0 ± 0.0**	24.0 ± 1.4	24.0 ± 0.0
	sc-spp	23.5 ± 2.1	11.5 ± 2.1	25.0 ± 0.0	24.0 ± 1.4	3.0 ± 2.8	24.5 ± 0.7	**1.0 ± 0.0**
C4S(T)	zs	17.4 ± 10.0	24.8 ± 0.4	19.4 ± 10.3	18.2 ± 9.1	23.4 ± 3.0	24.0 ± 2.2	**14.6 ± 12.4**
	cot	21.0 ± 6.2	20.4 ± 8.6	23.2 ± 1.6	15.2 ± 11.3	**12.2 ± 11.5**	24.6 ± 0.9	13.6 ± 11.8
	spp	23.6 ± 1.3	19.4 ± 7.2	23.6 ± 0.9	16.8 ± 7.0	**2.6 ± 3.0**	22.0 ± 1.9	19.8 ± 9.4
	sc-zs	25.0 ± 0.0	24.0 ± 1.4	23.5 ± 2.1	21.0 ± 2.8	**14.0 ± 12.7**	24.0 ± 1.4	25.0 ± 0.0
	sc-cot	23.5 ± 0.7	24.5 ± 0.7	24.0 ± 1.4	**18.5 ± 9.2**	24.5 ± 0.7	24.5 ± 0.7	24.0 ± 0.0
	sc-spp	14.5 ± 13.4	**8.0 ± 9.9**	23.5 ± 0.7	24.5 ± 0.7	13.5 ± 13.4	23.5 ± 0.7	22.5 ± 2.1
DS-R1	zs	**19.2 ± 7.1**	24.0 ± 1.2	22.6 ± 2.5	21.8 ± 6.6	24.6 ± 0.5	20.8 ± 5.0	22.4 ± 2.1
	cot	23.6 ± 1.1	21.0 ± 5.7	**20.4 ± 8.6**	24.6 ± 0.5	24.4 ± 0.5	23.4 ± 0.9	22.0 ± 2.2
	spp	22.2 ± 4.1	24.2 ± 0.4	19.6 ± 4.9	21.0 ± 5.9	24.8 ± 0.4	20.8 ± 5.3	**16.6 ± 9.8**
	sc-zs	25.0 ± 0.0	**10.5 ± 13.4**	19.0 ± 7.1	22.0 ± 4.2	24.0 ± 1.4	18.5 ± 4.9	12.0 ± 2.8
	sc-cot	**18.0 ± 4.2**	22.0 ± 2.8	25.0 ± 0.0	24.0 ± 1.4	24.5 ± 0.7	19.5 ± 7.8	25.0 ± 0.0
	sc-spp	24.5 ± 0.7	20.0 ± 7.1	**6.0 ± 2.8**	25.0 ± 0.0	25.0 ± 0.0	23.0 ± 0.0	13.0 ± 15.6
L3.3-70B	zs	**10.4 ± 11.6**	14.4 ± 12.3	14.4 ± 10.7	24.2 ± 1.3	11.4 ± 12.4	16.8 ± 11.1	10.6 ± 13.1
	cot	18.2 ± 5.4	6.8 ± 6.3	**5.6 ± 10.3**	22.4 ± 4.2	16.2 ± 8.9	22.2 ± 1.8	10.4 ± 12.9
	spp	20.0 ± 10.6	6.2 ± 10.5	**4.6 ± 5.1**	24.2 ± 1.1	16.0 ± 9.9	14.8 ± 12.7	**1.0 ± 0.0**
	sc-zs	10.0 ± 12.7	**1.0 ± 0.0**	7.5 ± 9.2	23.5 ± 2.1	13.0 ± 17.0	24.0 ± 1.4	**1.0 ± 0.0**
	sc-cot	12.5 ± 10.6	12.0 ± 15.6	20.5 ± 4.9	25.0 ± 0.0	10.5 ± 13.4	24.5 ± 0.7	**1.0 ± 0.0**
	sc-spp	21.0 ± 1.4	13.0 ± 14.1	14.0 ± 15.6	20.0 ± 1.4	19.0 ± 8.5	**1.0 ± 0.0**	**1.0 ± 0.0**
M-L(24.07)	zs	**6.0 ± 6.2**	17.6 ± 10.3	14.6 ± 12.4	23.8 ± 0.8	13.4 ± 11.2	19.2 ± 10.2	7.8 ± 10.5
	cot	11.2 ± 12.2	16.8 ± 5.8	14.0 ± 12.1	22.6 ± 2.1	25.0 ± 0.0	22.0 ± 4.6	**1.6 ± 1.3**
	spp	17.2 ± 10.0	19.6 ± 6.1	17.6 ± 9.7	21.4 ± 4.5	7.0 ± 10.0	12.0 ± 10.4	**5.8 ± 10.7**
	sc-zs	13.0 ± 17.0	20.0 ± 7.1	19.0 ± 8.5	19.0 ± 8.5	25.0 ± 0.0	24.5 ± 0.7	**7.5 ± 9.2**
	sc-cot	24.5 ± 0.7	23.5 ± 0.7	23.5 ± 2.1	24.5 ± 0.7	17.5 ± 10.6	12.0 ± 15.6	**1.5 ± 0.7**
	sc-spp	13.0 ± 17.0	20.5 ± 0.7	23.5 ± 0.7	24.0 ± 1.4	25.0 ± 0.0	24.5 ± 0.7	**1.0 ± 0.0**

Table 7.9: Round # where the Agent understood the opponent's Strategy (ba3-alt)

7.2.2 LLM vs non-LLM

The last 4 columns represent non-LLM players, that follow simple strategies. These players have been described in 5.7. Successful outcomes against them should be a good indication of reasoning abilities of LLM, since they imply an agent's ability to analyze their opponent and make informed decisions.

Srep

Since Stag Hunt has a mixed Strategy Nash Equilibrium for the single round variant, there is no possible way for the results in these games to be in any way indicative of anything. So we decide to skip this opponent.

Pp

The pattern player is a player that plays in cycles: scissors, rock, paper, ... in all game settings.

The pattern player appears somewhat hard for LLMs to grasp. However, LLMs that gathered good results in the "Total Points" section 7.1), are LLMs that comprehended this opponent fairly early in the game and adjusted their strategy to face the pattern player. Thus, observations on **round of opponent comprehension** supporting our earlier observations.

Ap

Ap is an adaptive player. This player adapts to its opponent's most frequently used move and aims to counter it.

We theorized that the mediocre results of LLMs against this player might be explainable by them understanding their opponent's play-style late in the game. However, this would mean that we would observe medium values in the resulting tables, i.e., most values close to 15. Most values are closer to 25 which means that LLMs did - for the most part - not completely understand this opponent.

The adaptive player can be a bit misleading: if one plays the same move against this player fairly often, then for a certain amount of rounds the adaptive player will consistently use the same move (until the opponent's most frequent move changes). This behavior can lead a rational player to misjudge the adaptive player's strategy as something else. Thus, under-performance against the adaptive player is not an immediate indication of a lack of reasoning abilities in LLMs.

Tft

Tft player is influenced by the concept of "tit-for-tat". This player counters their opponent's last move.

Results mentioned in the section 7.1 are further backed by the results on **round of opponent comprehension**. In tables 7.6 and 7.8 that concern payoff counterfactuals, where moves behave in their typical fashion, most values are low and often close to 1. LLMs manage to grasp their opponent's counter-strategy. However, in the strategy counterfactuals (tables 7.7 and 7.9) LLMs understand the **tft** opponent much later in the game (if at all), which explains why they struggled to gather high "total points".

7.3 Efficiency

Performance is often regarded as the most important factor of any technology with efficiency a classic companion. These two concepts are also present in the world of LLMs, where cost is often a function of tokens generated by the AI model. A simple efficiency metric is introduced:

$$efficiency = \frac{points}{tokens}$$

For a more visually pleasing appearance, results are scaled by 1000. In the following table, we represent efficiency as points (gathered by the LLM throughout the whole game) per kilo-tokens.

We show the average efficiency for each player averaging results for all opponent types in table 7.10. For **non-sc** players, the results have been averaged for 5 repetitions of the experiments, while for **sc** players, the results have been averaged for 2 repetitions of the experiments.

Results on efficiency vary from what was observed in the case of Prisoner's Dilemma in section 6.4. There we saw a strong favoritism towards simpler prompting styles. However, Prisoner's Dilemma and Stag Hunt are simpler games, so simpler methods are enough to effectively combat them (as observed by [48]). On the other hand, Rock-Paper-Scissors is a more complex game and in many cases its performance benefits from using **chain-of-thought** or **solo-performance prompting**.

Sc (self-consistency) when coupled with any other prompt style is less efficient, as observed in the case for Prisoner's Dilemma.

The comparison of *Claude* models' **Thinking** and **non-thinking** versions is a point of interest. Claude's implementation of Large Reasoning Models (via their thinking options) is promising, as thinking models often manage similar or better efficiency than their default counterparts.

7.4 Failure Rate

Large Language Models (LLMs) are complex structures with various emerging skills, that are not problem-free. Despite our efforts to design an environment for LLMs that handles errors and unexpected or unwanted behavior (described in 4.1.2), LLMs still occasionally face issues with following our instructions. The environment and the error-correcting logic we follow is an expanded version of the gym-like environment used in [28].

All LLMs manage to have almost perfect "validity" rates as seen in table 7.11. When a game is played, a validity value is subscribed to each of the 24 rounds that it includes. If any error occurred, this value will be "false". Table 7.11 simply depicts the average amount of valid rounds across all games.

At this point, it should be noted that most of the errors faced by the *Mistral Large (24.07)* model were due to its inability to follow our formatting directions. It often used **markdown** style formatting in its output even when specifically asked not to do so (as seen in the hints provided to models in 4.1.2). An example of such a failure is provided in 4.1.4.

7.5 Comparison With Other Work

The researchers in [16] analyzed LLM performance against predefined non-LLM algorithmic agents. The predefined strategies they used influenced the formulation of our own opponents, so some level of comparison can be done between the two works. They only used *GPT* models in their experiments. We cannot directly compare point values, since they use a different point system, however, we can compare outcomes and takeaways from experiments. They make similar observations to ours when it comes to the pattern player (called loop-2 and loop-3 in their work), where they notice that the average payoff of agents rises in subsequent rounds - especially, for more advanced results -. We, also, made similar observations using our "Round of Opponent Comprehension" metric in table 7.6. Furthermore, our experimentation showed promising results against the Tit-for-Tat player (counter player in their work) something which contradicts their results. Although, they played repeated Rock-Paper-Scissors for 10 rounds, the average payoff does not seem to significantly increase over time; this contrasts our results in the mentioned table. Nonetheless, *GPT-4* manages an above random average score, which is also observed in our models (e.g., *Claude 4 Sonnet* or *DeepSeek-R1*). A lot of LLM agents mistook the Tit-for-Tat strategy as a perfect mirroring of their own moves instead of a counter to their previous move, thus opting to keep playing cyclical moves, since they expected their opponent to keep mirroring their strategy.

7.6 Conclusions

- Complex prompt styles (**Solo-Performance**, **Chain-of-Thought**, or **Self-Consistency** yield better results (more points collected) when facing their simpler (**Zero-Shot**) opponents.

model	prompt	eq1	eq1-alt	ba3	ba3-alt
Claude 3.5 Sonnet v2	zs	**1.09 ± 2.37**	0.09 ± 1.34	0.47 ± 2.65	0.25 ± 1.61
	cot	0.44 ± 0.74	-0.23 ± 0.84	**0.49 ± 1.64**	-0.03 ± 1.28
	spp	0.49 ± 0.71	0.01 ± 0.48	**0.66 ± 0.83**	0.05 ± 1.10
	sc-zs	0.20 ± 0.60	0.10 ± 0.48	**0.46 ± 0.56**	-0.03 ± 0.35
	sc-cot	0.14 ± 0.18	-0.08 ± 0.25	**0.15 ± 0.41**	-0.01 ± 0.36
	sc-spp	0.01 ± 0.20	-0.03 ± 0.17	**0.19 ± 0.34**	-0.01 ± 0.17
Claude 3.7 Sonnet	zs	0.58 ± 2.39	0.20 ± 2.12	**2.02 ± 4.02**	0.07 ± 2.02
	cot	0.65 ± 0.75	0.51 ± 0.80	0.63 ± 1.07	**0.71 ± 1.10**
	spp	0.51 ± 0.70	0.41 ± 0.61	0.43 ± 1.13	**0.56 ± 0.76**
	sc-zs	-0.08 ± 0.72	0.00 ± 0.36	0.00 ± 0.93	**0.32 ± 0.71**
	sc-cot	**0.21 ± 0.25**	0.14 ± 0.17	0.17 ± 0.26	0.11 ± 0.28
	sc-spp	0.14 ± 0.18	0.09 ± 0.12	**0.16 ± 0.21**	0.13 ± 0.16
Claude 3.7 Sonnet (Thinking)	zs	1.92 ± 3.30	**0.60 ± 2.55**	**3.26 ± 4.15**	**1.19 ± 4.43**
	cot	0.61 ± 1.04	0.25 ± 1.02	**0.87 ± 1.27**	0.60 ± 1.15
	spp	0.57 ± 0.89	0.60 ± 0.80	**1.06 ± 1.53**	0.70 ± 0.97
	sc-zs	0.23 ± 0.77	0.31 ± 0.98	**0.70 ± 1.44**	-0.08 ± 0.81
	sc-cot	0.11 ± 0.27	0.15 ± 0.21	**0.24 ± 0.28**	0.18 ± 0.31
	sc-spp	0.04 ± 0.26	0.11 ± 0.18	0.17 ± 0.35	**0.18 ± 0.24**
Claude Sonnet 4	zs	0.80 ± 1.82	-0.12 ± 2.25	**0.92 ± 2.96**	0.23 ± 2.49
	cot	0.62 ± 0.89	0.31 ± 0.71	**0.75 ± 1.41**	0.57 ± 1.10
	spp	0.37 ± 0.70	0.19 ± 0.56	**0.73 ± 0.83**	0.30 ± 0.74
	sc-zs	**0.13 ± 0.50**	0.07 ± 0.29	0.12 ± 0.43	0.07 ± 0.21
	sc-cot	0.10 ± 0.18	0.07 ± 0.13	**0.15 ± 0.29**	0.11 ± 0.25
	sc-spp	0.10 ± 0.14	0.05 ± 0.09	0.09 ± 0.20	**0.12 ± 0.21**
Claude Sonnet 4 (Thinking)	zs	**2.07 ± 3.74**	-0.16 ± 2.00	1.51 ± 4.69	0.04 ± 3.76
	cot	1.11 ± 1.68	0.60 ± 0.92	**1.50 ± 2.05**	0.81 ± 1.93
	spp	0.31 ± 0.72	0.28 ± 0.68	**0.69 ± 1.20**	0.57 ± 0.91
	sc-zs	**0.44 ± 0.67**	-0.21 ± 0.39	0.37 ± 1.14	0.03 ± 0.75
	sc-cot	**0.35 ± 0.26**	0.20 ± 0.28	0.16 ± 0.61	-0.04 ± 0.27
	sc-spp	0.13 ± 0.13	0.07 ± 0.12	**0.14 ± 0.18**	0.10 ± 0.13
DeepSeek-R1	zs	0.47 ± 0.73	0.17 ± 0.44	**0.89 ± 1.49**	0.03 ± 0.86
	cot	0.80 ± 0.84	0.23 ± 0.49	**1.04 ± 1.77**	0.02 ± 0.76
	spp	0.41 ± 0.74	0.03 ± 0.31	**0.49 ± 0.96**	0.09 ± 0.85
	sc-zs	0.13 ± 0.22	-0.01 ± 0.11	**0.42 ± 0.48**	0.03 ± 0.20
	sc-cot	0.19 ± 0.28	0.07 ± 0.14	**0.29 ± 0.32**	0.08 ± 0.27
	sc-spp	0.09 ± 0.20	0.05 ± 0.10	**0.22 ± 0.38**	0.05 ± 0.18
Llama 3.3 70B Instruct	zs	1.16 ± 3.16	0.14 ± 2.06	**2.38 ± 4.85**	-0.07 ± 3.64
	cot	1.18 ± 2.20	0.12 ± 1.26	**1.59 ± 2.91**	-0.01 ± 1.46
	spp	0.57 ± 1.41	0.26 ± 1.05	**1.28 ± 2.29**	-0.13 ± 2.19
	sc-zs	0.13 ± 0.63	0.11 ± 0.56	**0.53 ± 1.31**	0.02 ± 0.85
	sc-cot	**0.22 ± 0.28**	-0.10 ± 0.45	0.12 ± 0.73	0.04 ± 0.21
	sc-spp	0.20 ± 0.53	0.04 ± 0.24	**0.46 ± 0.72**	-0.22 ± 0.69
Mistral Large (24.07)	zs	0.78 ± 4.58	-0.35 ± 3.42	**1.44 ± 6.57**	-1.05 ± 5.91
	cot	0.36 ± 1.04	0.02 ± 0.59	**0.51 ± 1.16**	0.13 ± 1.21
	spp	0.50 ± 0.75	0.15 ± 0.79	0.24 ± 1.68	**0.62 ± 1.14**
	sc-zs	**0.11 ± 0.96**	-0.10 ± 0.74	0.01 ± 1.21	-0.37 ± 1.70
	sc-cot	0.15 ± 0.29	0.03 ± 0.22	**0.22 ± 0.31**	0.03 ± 0.24
	sc-spp	0.03 ± 0.18	0.02 ± 0.26	**0.13 ± 0.35**	0.02 ± 0.39

Table 7.10: Average Efficiency (Points per kilo-token)

model	avg
Claude 3.5 Sonnet v2	100.0 ± 0.0
Claude 3.7 Sonnet	100.0 ± 0.0
Claude 3.7 Sonnet (Thinking)	100.0 ± 0.2
Claude Sonnet 4	99.9 ± 0.9
Claude Sonnet 4 (Thinking)	100.0 ± 0.0
DeepSeek-R1	100.0 ± 0.2
Llama 3.3 70B Instruct	100.0 ± 0.0
Mistral Large (24.07)	99.2 ± 6.8

Table 7.11: Average Valid Rate (% of Valid Outcomes)

- LLMs achieve points close to 0 (expected result when following the mixed-strategy Nash Equilibrium probability distribution over moves) against the **srep** player. This result contradicts findings of [49]. That work highlighted a bias of LLMs in choosing a particular move over others, which is uncharacteristic of rational players. However, they had used single-round variants of Rock-Paper-Scissors for testing. LLMs, because of their acquired knowledge, have developed inherent biases (e.g., "I know that "rock" is a popular first move in Rock-Paper-Scissors) which dissipate in repeated games. We observe that LLMs tend to leave behind such biases as historical information about the previous rounds builds up and players refine their belief about their opponent's play-style.
- The **pattern** player highlights that simpler LLMs perform worse than more complex ones. Also, in the same LLM using a more complex prompting style yields better results. Finally, **thinking** counterparts of *Claude*'s LLMs got confused more easily. This result reinforces the findings of [48].
- The **tft** player was a good benchmark to showcase how strategy counterfactuals can degrade LLM reasoning ability. Earlier research [16] has outlined that LLMs face issues in correctly identifying and playing against the **tft** (or **counter**) player. That work, however, was only performed on openAI models available at that time. Our work deals more with the current landscape of LLMs and we find that LLMs can be successful against the **tft** player in game settings where the game is played following its typical rules (non-stategy counterfactual scenarios). Finally, upon inspection of LLM chats, where agents got very low point values, it seems that a lot of LLMs mistook the **tft** player as perfectly mirroring their own strategy instead of countering their previous move; this lead to them consistently playing the same pattern of moves, because they expected their opponent to mirror it. This situation resulted in consecutive ties and low total point scores.
- efficiency of an AI player is boosted by employing a more complex prompt style. **Self-Consistency** does not achieve results that would excuse its performance hit (token consumption). Finally, **thinking** models are promising; they manage similar or better efficiency than their default counterparts.

Chapter 8

Conclusions

8.1 Conclusions

This thesis explored the strategic behavior and reasoning capabilities of large language models (LLMs) and large reasoning models (LRMs) in interactive, game-theoretic environments. By simulating repeated games of Prisoner's Dilemma and Rock-Paper-Scissors against a variety of agent types and prompt styles, we assessed how LLMs navigate cooperation, rationality, and adaptation under different experimental conditions.

Our results demonstrate that LLMs can reproduce cooperative behaviors similar to those reported in previous work [8], particularly in repeated Prisoner's Dilemma settings. When not instructed to answer directly, LLMs often expressed a desire to maximize joint payoffs—indicative of a reasoning pattern that prioritizes mutual benefit over strictly self-interested play. This suggests a degree of social preference embedded in language model reasoning, likely due to their exposure to human norms in training data.

In contrast to prior findings [49] suggesting LLMs exhibit irrational biases in single-round Rock-Paper-Scissors, our experiments reveal that such biases diminish in repeated interactions. LLMs adapt toward equilibrium play when sufficient historical information is available, particularly against a static random opponent. This suggests that LLMs, when allowed to update beliefs about their opponent, can approximate mixed-strategy Nash behavior more effectively than previously assumed.

We observed that more complex models tended to cooperate less and often over-analyzed simple tasks, leading to delayed strategy convergence and lower cumulative rewards. Conversely, smaller models, though limited in capacity, often achieved better performance when paired with more structured prompt styles—especially against pattern-based opponents. More complex reasoning tasks and games were the area where larger LLMs capitalized on their abilities and achieved better results than smaller ones. These findings emphasize the importance of aligning model capacity with task complexity and prompt design.

Against a vindictive opponent (i.e., the **tft** (Tit-for-Tat) agent), LLMs demonstrated the ability to either maintain or gradually adopt effective reciprocal strategies. This behavior supports the claim that LLMs can engage in dynamic strategy refinement over time, rather than committing to a fixed approach. It further highlights the potential of LLMs to perform iterative reasoning and learn from ongoing interaction sequences. Such opponents may, however, confuse the LLM agents in more complex game scenarios, making them misinterpret their strategies and, thus, leading players to improperly adjust their own strategic play.

Prompt style was a decisive factor across both games. More structured prompting strategies - such as Chain-of-Thought, Solo-Performance Prompting, and Self-Consistency - led to higher efficiency when compared to zero-shot baselines, especially in simpler models. However, in more capable LLMs, increased prompt complexity sometimes introduced unnecessary cognitive overhead without proportional performance gains. In particular, Self-Consistency consumed significantly more resources without consistently improving results, raising concerns about its cost-effectiveness.

Interestingly, "thinking" variants of some LLMs showed mixed results. While they occasionally matched or outperformed their default counterparts in reasoning-heavy situations, they were also more prone to confusion. This aligns with recent concerns [48] about the fragility of such models under certain cognitive load conditions.

Counterfactual settings gave promising results. Reasoning ability was hindered only in simpler/older models - like the *Mistral* and *Llama* models we tested - when asked to participate in counterfactual games. Other models were, also, not completely struggle-free, but all the above analysis shows that these struggles do not overshadow LLM reasoning and belief-refinement abilities.

In summary, our findings suggest that LLMs, when appropriately prompted, can demonstrate strategic behavior aligned with rational and cooperative norms. Their performance, however, is sensitive to game structure, opponent behavior, prompt engineering, and model complexity. These insights contribute to the broader understanding of LLM capabilities in reasoning under uncertainty, strategic adaptation, and human-aligned decision-making.

8.2 Future Work

While this thesis has demonstrated that LLMs are capable of adapting to interactive game-theoretic settings and employing strategies resembling rational or cooperative behavior, several important avenues remain open for future investigation.

- **Human vs. LLM Interactions:** While this thesis focused on LLM vs. LLM matchups, future experiments could involve humans playing against LLMs in repeated settings to study alignment, deception, persuasion, and trust-building.
- **Multi-agent and Multi-round Games with Communication:** Introducing explicit communication between agents opens up questions around negotiation, signaling, deception, and emergent coordination. Investigating whether LLMs can learn to use language strategically to influence outcomes—or recognize when others are doing so—could provide insight into their pragmatic competence in game-theoretic settings.
- **Scaling Behavioral Traits with Model Size:** Larger models displayed a tendency toward over-analysis and delayed cooperation. A systematic investigation of how behaviors scale with model size—especially in zero-shot versus reasoning-heavy prompting scenarios—could illuminate when increased capacity helps or hinders strategic decision-making.
- **Longer Games and Memory Integration:** Our experiments involved relatively short repeated games. Extending game lengths or integrating explicit memory mechanisms (e.g., scratchpads, working memory modules, or external memory APIs) may help determine whether LLMs can develop long-term strategies, learn opponent types more efficiently, or emulate sustained behavioral commitments like trust and retaliation.
- **Formalizing Benchmarks for Rationality in LLMs:** Finally, there is a need for standardized evaluation frameworks that go beyond point-based metrics to assess whether LLMs demonstrate rational, cooperative, or adaptive behavior.

In conclusion, while this work establishes that LLMs can participate in and adapt to game-theoretic scenarios in ways that sometimes resemble rational agents, further exploration is needed to determine the limits of their reasoning, the durability of their strategies, and their generalizability across settings. With continued improvements in model architecture, interpretability, and evaluation methodology, LLMs may one day serve not only as tools for simulating strategic behavior, but as agents that participate meaningfully in complex decision-making processes.

Chapter 9

Bibliography

[1] Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., and Schulz, E. "Playing repeated games with large language models". In: *Nature Human Behaviour* (May 2025). ISSN: 2397-3374. DOI: 10.1038/s41562-025-02172-y. URL:

[2] Amazon Web Services. *Inference Parameters – AWS Bedrock*. Accessed: 2025-07-14. 2025. URL:

[3] Argyrou, G., Dimitriou, A., Lymperaiou, M., Filandrianos, G., and Stamou, G. *Automatic Generation of Fashion Images using Prompting in Generative Machine Learning Models*. 2024. arXiv: 2407.14944 [cs.CV]. URL:

[4] Ba, J. L., Kiros, J. R., and Hinton, G. E. *Layer Normalization*. 2016. arXiv: 1607.06450 [stat.ML].

[5] BBC News. *Rock glaciers – the world's fastest-moving ice?* Accessed: 2025-07-10. Mar. 2014. URL:

[6] Bengio, Y., Ducharme, R., Vincent, P., and Janvin, C. "A Neural Probabilistic Language Model". In: *J. Mach. Learn. Res.* 3.null (Mar. 2003), pp. 1137–1155. ISSN: 1532-4435.

[7] Bertrand, Q., Duque, J., Calvano, E., and Gidel, G. *Self-Play Q-learners Can Provably Collude in the Iterated Prisoner's Dilemma*. 2025. arXiv: 2312.08484 [cs.GT]. URL:

[8] Brookins, P. and Debacker, J. "Playing Games With GPT: What Can We Learn About a Large Language Model From Canonical Strategic Games?" In: *SSRN Electronic Journal* (2023). URL:

[9] Brown, T. B. et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL]. URL:

[10] Chaidos, N., Dimitriou, A., Lymperaiou, M., and Stamou, G. *SCENIR: Visual Semantic Clarity through Unsupervised Scene Graph Retrieval*. 2025. arXiv: 2505.15867 [cs.CV]. URL:

[11] Cheng, P., Hu, T., Xu, H., Zhang, Z., Dai, Y., Han, L., and Du, N. "Self-playing Adversarial Language Game Enhances LLM Reasoning". In: *ArXiv* abs/2404.10642 (2024). URL:

[12] DeepSeek-AI et al. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*. 2025. arXiv: 2501.12948 [cs.CL]. URL:

[13] Dimitriou, A., Chaidos, N., Lymperaiou, M., and Stamou, G. "Graph Edits for Counterfactual Explanations: A Comparative Study". In: *Explainable Artificial Intelligence*. Ed. by L. Longo, S. Lapuschkin, and C. Seifert. Cham: Springer Nature Switzerland, 2024, pp. 100–112. ISBN: 978-3-031-63797-1.

[14] Dimitriou, A., Lymperaiou, M., Filandrianos, G., Thomas, K., and Stamou, G. *Structure Your Data: Towards Semantic Graph Counterfactuals*. 2024. arXiv: 2403.06514 [cs.CV]. URL:

[15] Evangelatos, A., Filandrianos, G., Lymperaiou, M., Voulodimos, A., and Stamou, G. "AILS-NTUA at SemEval-2025 Task 8: Language-to-Code prompting and Error Fixing for Tabular Question Answering". In: *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*. Ed. by S. Rosenthal, A. Rosá, D. Ghosh, and M. Zampieri. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 1423–1435. ISBN: 979-8-89176-273-2. URL:

[16] Fan, C., Chen, J., Jin, Y., and He, H. "Can Large Language Models Serve as Rational Players in Game Theory? A Systematic Analysis". In: *ArXiv* abs/2312.05488 (2023). URL:

[17] Filandrianos, G., Dimitriou, A., Lymperaiou, M., Thomas, K., and Stamou, G. *Bias Beware: The Impact of Cognitive Biases on LLM-Driven Product Recommendations*. 2025. arXiv: 2502.01349 [cs.CL]. URL:

[18] Fisher, L. *Rock, Paper, Scissors: Game Theory in Everyday Life*. Hay House, 2008. ISBN: 9781848502024. URL:

[19] Giadikiaroglou, P., Lymperaiou, M., Filandrianos, G., and Stamou, G. "Puzzle Solving using Reasoning of Large Language Models: A Survey". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Y. Al-Onaizan, M. Bansal, and Y.-N. Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 11574–11591. DOI: 10.18653/v1/2024.emnlp-main.646. URL:

[20] Gibbon, D. *Uses of Game Theory in International Relations*. https://tuecontheoryofnetworks.wordpress.com/2013/02/11/uses-of-game-theory-in-international-relations. Accessed: 2025-07-09. Feb. 2013.

[21] He, K., Zhang, X., Ren, S., and Sun, J. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[22] Jia, J., Yuan, Z., Pan, J., McNamara, P. E., and Chen, D. *Large Language Model Strategic Reasoning Evaluation through Behavioral Game Theory*. 2025. arXiv: 2502.20432 [cs.AI]. URL:

[23] Karkani, D., Lymperaiou, M., Filandrianos, G., Spanos, N., Voulodimos, A., and Stamou, G. "AILS-NTUA at SemEval-2025 Task 3: Leveraging Large Language Models and Translation Strategies for Multilingual Hallucination Detection". In: *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*. Ed. by S. Rosenthal, A. Rosá, D. Ghosh, and M. Zampieri. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 1289–1305. ISBN: 979-8-89176-273-2. URL:

[24] Kong, F., Huang, Y., Zhu, S.-C., Qi, S., and Feng, X. *Learning to Balance Altruism and Self-interest Based on Empathy in Mixed-Motive Games*. 2025. arXiv: 2410.07863 [cs.AI]. URL:

[25] Kritharoula, A., Lymperaiou, M., and Stamou, G. *Language Models as Knowledge Bases for Visual Word Sense Disambiguation*. 2023. arXiv: 2310.01960 [cs.CL]. URL:

[26] Kritharoula, A., Lymperaiou, M., and Stamou, G. "Large Language Models and Multimodal Retrieval for Visual Word Sense Disambiguation". In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2023, pp. 13053–13077. DOI: 10.18653/v1/2023.emnlp-main.807. URL:

[27] Li, X. L. and Liang, P. *Prefix-Tuning: Optimizing Continuous Prompts for Generation*. 2021. arXiv: 2101.00190 [cs.CL]. URL:

[28] Liao, A., Tomlin, N., and Klein, D. "Efficacy of Language Model Self-Play in Non-Zero-Sum Games". In: *ArXiv* abs/2406.18872 (2024). URL:

[29] Lin, W., Roberts, J., Yang, Y., Albanie, S., Lu, Z., and Han, K. *GAMEBoT: Transparent Assessment of LLM Reasoning in Games*. 2025. arXiv: 2412.13602 [cs.CL]. URL:

[30] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. *Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing*. 2021. arXiv: 2107.13586 [cs.CL]. URL:

[31] Lymperaiou, M., Filandrianos, G., Dimitriou, A., Voulodimos, A., and Stamou, G. *HalCECE: A Framework for Explainable Hallucination Detection through Conceptual Counterfactuals in Image Captioning*. 2025. arXiv: 2503.00436 [cs.CV]. URL:

[32] Lymperaiou, M., Filandrianos, G., Thomas, K., and Stamou, G. *Counterfactual Edits for Generative Evaluation*. 2023. arXiv: 2303.01555 [cs.CV]. URL:

[33] Lymperaiou, M., Manoliadis, G., Menis Mastromichalakis, O., Dervakos, E. G., and Stamou, G. "Towards Explainable Evaluation of Language Models on the Semantic Similarity of Visual Concepts". In: *Proceedings of the 29th International Conference on Computational Linguistics*. Ed. by N. Calzolari et al. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, Oct. 2022, pp. 3639–3658. URL:

[34] Lymperaiou, M. and Stamou, G. *The Contribution of Knowledge in Visiolinguistic Learning: A Survey on Tasks and Challenges*. 2023. arXiv: 2303.02411 [cs.CL]. URL:

[35] Lymperopoulos, D., Lymperaiou, M., Filandrianos, G., and Stamou, G. "Optimal and efficient text counterfactuals using Graph Neural Networks". In: *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*. Ed. by Y. Belinkov, N. Kim, J. Jumelet, H. Mohebbi, A. Mueller, and H. Chen. Miami, Florida, US: Association for Computational Linguistics, Nov. 2024, pp. 1–14. DOI: 10.18653/v1/2024.blackboxnlp-1.1. URL:

[36] Mozikov, M. et al. “EAI: Emotional Decision-Making of LLMs in Strategic Games and Ethical Dilemmas”. In: *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. 2024. URL:

[37] MYERSON, R. B. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991. ISBN: 9780674341166. URL: (visited on 06/23/2025).

[38] Panagiotopoulos, I., Filandrianos, G., Lymperaiou, M., and Stamou, G. “AILS-NTUA at SemEval-2024 Task 9: Cracking Brain Teasers: Transformer Models for Lateral Thinking Puzzles”. In: *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Ed. by A. K. Ojha, A. S. Doğruöz, H. Tayyar Madabushi, G. Da San Martino, S. Rosenthal, and A. Rosá. Mexico City, Mexico: Association for Computational Linguistics, June 2024, pp. 1733–1746. DOI: 10.18653/v1/2024.semeval-1.248. URL:

[39] Panagiotopoulos, I., Filandrianos, G., Lymperaiou, M., and Stamou, G. “RISCORE: Enhancing In-Context Riddle Solving in Language Models through Context-Reconstructed Example Augmentation”. In: *Proceedings of the 31st International Conference on Computational Linguistics*. Ed. by O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert. Abu Dhabi, UAE: Association for Computational Linguistics, Jan. 2025, pp. 9431–9455. URL:

[40] Papadakis, C., Filandrianos, G., Dimitriou, A., Lymperaiou, M., Thomas, K., and Stamou, G. *StockSim: A Dual-Mode Order-Level Simulator for Evaluating Multi-Agent LLMs in Financial Markets*. 2025. arXiv: 2507.09255 [cs.CE]. URL:

[41] Papadimitriou, C., Filandrianos, G., Lymperaiou, M., and Stamou, G. *Masked Generative Story Transformer with Character Guidance and Caption Augmentation*. 2024. arXiv: 2403.08502 [cs.CV]. URL:

[42] Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., and Miller, A. “Language Models as Knowledge Bases?” In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Ed. by K. Inui, J. Jiang, V. Ng, and X. Wan. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 2463–2473. DOI: 10.18653/v1/D19-1250. URL:

[43] Poundstone, W. *Prisoner’s Dilemma*. Doubleday, 1992. ISBN: 9780385415675. URL:

[44] Premptis, I., Lymperaiou, M., Filandrianos, G., Menis Mastromichalakis, O., Voulodimos, A., and Stamou, G. “AILS-NTUA at SemEval-2025 Task 4: Parameter-Efficient Unlearning for Large Language Models using Data Chunking”. In: *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*. Ed. by S. Rosenthal, A. Rosá, D. Ghosh, and M. Zampieri. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 1383–1405. ISBN: 979-8-89176-273-2. URL:

[45] Raptopoulos, P., Filandrianos, G., Lymperaiou, M., and Stamou, G. *PAKTON: A Multi-Agent Framework for Question Answering in Long Legal Agreements*. 2025. arXiv: 2506.00608 [cs.CL]. URL:

[46] Roewer-Despr’es, F., Feng, J., Zhu, Z., and Rudzicz, F. “ACCORD: Closing the Commonsense Measurability Gap”. In: *ArXiv* abs/2406.02804 (2024). URL:

[47] Shapley, L. S. and Shubik, M. *Game Theory in Economics: Chapter 1, Introduction, The Use of Models*. Santa Monica, CA: RAND Corporation, 1971.

[48] Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S., and Farajtabar, M. *The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity*. 2025. arXiv: 2506.06941 [cs.AI]. URL:

[49] Silva, A. “Large Language Models Playing Mixed Strategy Nash Equilibrium Games”. In: *ArXiv* abs/2406.10574 (2024). URL:

[50] Skyrms, B. “The Stag Hunt”. In: *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, 2003, pp. 1–14.

[51] Stoikou, T., Lymperaiou, M., and Stamou, G. *Knowledge-Based Counterfactual Queries for Visual Question Answering*. 2023. arXiv: 2303.02601 [cs.CL]. URL:

[52] Stringli, E., Lymperaiou, M., Filandrianos, G., Voulodimos, A., and Stamou, G. “Pitfalls of Scale: Investigating the Inverse Task of Redefinition in Large Language Models”. In: *Findings of the Association for Computational Linguistics: ACL 2025*. Ed. by W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 9445–9469. ISBN: 979-8-89176-256-5. DOI: 10.18653/v1/2025.findings-acl.492. URL:

[53] Sun, K., Xu, Y., Zha, H., Liu, Y., and Dong, X. “Head-to-Tail: How Knowledgeable are Large Language Models (LLMs)? A.K.A. Will LLMs Replace Knowledge Graphs?” In: *ArXiv* abs/2308.10168 (2023). URL:

[54] Thomas, K., Filandrianos, G., Lymperaiou, M., Zerva, C., and Stamou, G. *"I Never Said That": A dataset, taxonomy and baselines on response clarity classification*. 2024. arXiv: 2409.13879 [cs.CL]. URL:

[55] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. “Attention is All you Need”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., 2017. URL:

[56] Vlachos, A., Filandrianos, G., Lymperaiou, M., Spanos, N., Mitsouras, I., Karampinis, V., and Voulodimos, A. *Analyze-Prompt-Reason: A Collaborative Agent-Based Framework for Multi-Image Vision-Language Reasoning*. 2025. arXiv: 2508.00356 [cs.CV]. URL:

[57] Wahle, J. P., Ruas, T., Xu, Y., and Gipp, B. *Paraphrase Types Elicit Prompt Engineering Capabilities*. 2025. arXiv: 2406.19898 [cs.CL]. URL:

[58] Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., and Zhou, D. *Self-Consistency Improves Chain of Thought Reasoning in Language Models*. 2023. arXiv: 2203.11171 [cs.CL]. URL:

[59] Wang, Z., Mao, S., Wu, W., Ge, T., Wei, F., and Ji, H. “Unleashing the Emergent Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration”. In: *North American Chapter of the Association for Computational Linguistics*. 2023. URL:

[60] Wang, Z., Mao, S., Wu, W., Ge, T., Wei, F., and Ji, H. *Unleashing the Emergent Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration*. 2024. arXiv: 2307.05300 [cs.AI]. URL:

[61] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. 2023. arXiv: 2201.11903 [cs.CL]. URL:

[62] Wen, Y., Jain, N., Kirchenbauer, J., Goldblum, M., Geiping, J., and Goldstein, T. *Hard Prompts Made Easy: Gradient-Based Discrete Optimization for Prompt Tuning and Discovery*. 2023. arXiv: 2302.03668 [cs.LG]. URL:

[63] Wu, Z., Qiu, L., Ross, A., Akyürek, E., Chen, B., Wang, B., Kim, N., Andreas, J., and Kim, Y. “Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks”. In: *North American Chapter of the Association for Computational Linguistics*. 2023. URL: